diritto & questioni pubbliche

#specialpublication

recognise

AUGUST 2023

LEGAL REASONING AND COGNITIVE SCIENCE
TOPICS AND PERSPECTIVES

DIRITTO & QUESTIONI PUBBLICHE / RECOGNISE
2023/#SPECIAL**PUBLICATION**

DIRITTO & QUESTIONI PUBBLICHE | RECOGNISE
Special Publication / August, 2023

ISSN 1825-0173

Double-Blind Peer Review

Databases: Scopus, Heinonline, Elsevier

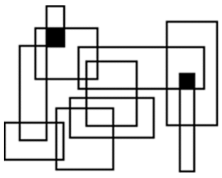
Graphic design rospeinfrantumi

Co-funded by
the European Union

This publication was realized within the frame of the project Recognise-Legal Reasoning and Cognitive Science, co-funded by the Erasmus+ Programme of the European Union under the number 2020-1-IT02-KA203-079834.
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEGAL REASONING AND COGNITIVE SCIENCE
TOPICS AND PERSPECTIVES

edit by
Marco Brigaglia
Corrado Roversi



Diritto & Questioni Pubbliche | Recognise
Special Publication, August 2023

LEGAL REASONING AND COGNITIVE SCIENCE

TOPICS AND PERSPECTIVES

contents

PART IV. Legal reasoning and Cognitive Biases

PART V. Law and Emotions

PART VI. Defeasibility and Legal Cognition

PART VII. Issues on Legal Evidence

PART VIII. Law, Legal Reasoning and Artificial Intelligence

Introduction

This book was conceived and developed within, and with the support of, the Erasmus-KA2 project *RECOGNISE-Legal Reasoning and Cognitive Science*, (https://www.recognise.academy/), carried out by a partnership including the University of Palermo (Coordinator), the University of Bologna, the University of Ljubljana, Maastricht University, the University of Alicante, and the Jagiellonian University in Krakow.

As suggested by its name, the project concerned the contributions that cognitive science—broadly understood as the new waves of empirical or empirically informed inquiries into the mind (both natural and artificial minds) flourished over the last half century—may afford to the understanding, and hopefully to the improvement, of legal reasoning and, more generally, of legal institutions.

Since its inception in the 1950s, cognitive science has developed rapidly and has profoundly impacted many fields of knowledge. So rapid and so profound to be called a "revolution": the "cognitive revolution".

One of the fields where the impact has been more profound is our understanding of human reasoning and decision-making, which has been enriched with many details and probably also deeply transformed. To summarize the dominant trends in the last decades, one could simply resort to four interconnected notions: intuition, emotion, embodiment, and unconscious. To a much more significant extent than granted by common sense and by former mainstream accounts, human "higher order" functions, such as reasoning and decision-making, appear to rely on intuitive rather than logical processes, to be driven by emotions rather than "cool" cognition, to be grounded in the body rather than abstracted from it, and to be pervasively influenced by mechanisms and factors we are not aware of.

In the last twenty years, these kinds of studies and approaches have been applied to *legal* reasoning and decision-making as well, resulting in a series of successful research programs: "Heuristics and biases and the law", "Emotions and the law", "Law and intuition", "Embodied cognition and the law".

Adding to the increasing interest in the sciences of human cognition and interacting with it in various ways is the astonishingly rapid improvement of AI tools and their use and application within legal decision-making processes.

The basic idea of the project was straightforward. All this body of knowledge should become a part of standard legal education. This does not mean that legal education, as it is currently done, should be *replaced* by approaches centered on cognitive science. More modestly, the latter approaches should be somehow *integrated* within traditional legal education: they should be introduced, discussed, and considered.

Given this goal, we built, tested, and implemented a range of didactic activities—a university teaching course, a series of seminars, and two Intensive Study Programs targeted at LLM students, PhD students, legal professionals, and legal scholars. To this, we wanted to add, to the already rich body of knowledge and research, a book including original research papers by experts in the field that were also written in a clear and reader-friendly style and therefore suitable for use as reading materials in higher education (LLM, PhD programs, etc.). This book is the result of this effort, and of the collaboration not only of researchers belonging to the partnership, but also of researchers affiliated to other universities who have generously accepted to take part in the endeavor. The title of the book reflects the variety of perspectives, methods, and topics of the collected contributions.

The book is divided into eight parts.

Part I deals with the naturalization of law and legal theory. The first chapter, by Francesca Poggi and Francesco Ferraro, has a historical character: It introduces currents of thought advocating for empirical approaches in jurisprudence, focusing on American and Scandinavian legal realism. The second chapter, by Bruno Celano, tracks the roots of the anti-psychologistic stance that profoundly influenced 20th-century jurisprudence and advances arguments in favor of its reorientation—at least as far as the theory of norms and norm-based reasoning are concerned—towards a "psychologistic" (or "naturalistic") approach. The third chapter, by Kevin Tobia (the republication of a previously published longer article), provides a comprehensive introduction to the methods and inquiries of a new, exciting field: experimental jurisprudence.

Part II gathers five different cognitive-oriented perspectives on law and legal reasoning. Philip Johnson-Laird and Monica Bucciarelli present their influential application of the theory of mental models to the meaning of causation and its import for the law. Michele Ubertone, Anna Borghi, Caterina Villani and Luisa Lugli apply embodied and grounded cognition perspectives to legal concepts and their use for the legal classification of facts. Marek Jakubiec's chapter investigates the role of metaphorical simulation in legal reasoning. Bartosz Brożek makes the case for an ecological conception of rationality and illustrates some consequences of adopting it to understand the rationality of law. Łukasz Kurek investigates, with an empirically-oriented perspective, mindreading abilities and their role and peculiarities within legal practices.

The topic of Part III is the nature of law and normative phenomena. The first chapter, by Philippe Rochat and Nikita Agarwal, develops a wide-ranging account of the emergence of normative thought and decision-making in human development. The second chapter, by Corrado Roversi, is an inquiry into the cognitive foundations of legal institutions, that is, into the cognitive abilities required to act in light of legal facts one believes exist. In the third chapter, Jaap Hage traces the roots of the disagreement between legal positivists and non-positivists to different views of how our minds shape social reality, and he develops an answer to the latter question based on a particular kind of facts, constructivist facts.

Part IV deals with legal reasoning and cognitive biases. Niek Strohmaier and Sofia de Jong focus on motivated reasoning—the process by which people are unconsciously motivated to reach a particular conclusion while operating under the illusion of objectivity—, and in particular on the role played in motivated reasoning by moral character inferences. The chapter by Przemysław Pałka, Piotr Bystranowski, Bartosz Janik, and Maciej Próchnicki investigates so-called abstract-concrete effects (understood as the general tendency to judge an issue differently depending on the level of abstraction at which it is described), how they affect legal reasoning and what are the consequences for the Rule of Law. Enide Maegherman focuses on confirmation bias, its role during investigation and trial, and methods to help counteract its effects. Finally, Noam Gur's chapter highlights one key function played by legal rules: how they can operate as corrective devices against several systematic biases.

Part V is about law and emotions. The first chapter, by Miha Hafner, deals with the role played by emotions and empathy in criminal procedure. The second chapter, by Kristina Čufar, reconstructs the genealogy of the reason-emotion binary and its implications in Western legal systems from the perspective of feminist critical theory.

Part VI concerns the notion of "defeasibility" and its import in legal cognition. Manuel Atienza focuses on the relations between the notions of "defeasibility" and "balancing" and the problems they deal with—on the one hand, the need to allow for implicit exceptions while remaining within the boundaries of the legal system, and, on the other hand, the need to solve difficult cases argumentatively through the balancing of principles. Rafael Buzón provides a general criticism against legal positivist accounts of defeasibility and develops a post-positivist perspective on the problem. Daniel González Lagier explores a possible application of J. Haidt's social intuitionism to the defeasibility of legal rules and uses it as a point of departure to criticize Haidt's view and, more generally, the normative claims sometimes advanced by cognitive

scientists. Josep Aguiló's chapter deals with legal presuppositions, the defeasibility of presumptive reasoning, and the possible role of the cognitive sciences in the detection of material fallacies.

The topic of Part VII is legal evidence. The first chapter, by Bartłomiej Kucharzyk, provides an overview of psychological factors that may negatively affect the rationality of legal evidence evaluation, providing examples of their impact taken from experimental research. The second chapter, by Barbara Spellman and Adele Quigley-McBride, explains how biases and other conditions may affect the reliability of forensic analyses, reconstructs the (often mistaken) beliefs people have about forensic science and illustrates strategies for evaluating and communicating forensic science results during trials. Finally, Michele Ubertone argues in his chapter that the law often seems to overestimate the human ability to solve problems individually and to underestimate the importance of the division of cognitive labor.

Part VIII turns to Artificial Intelligence. The first chapter, by Antonia Waltermann, explores the challenges posed to legal reasoning by the combined effects of the development of AI and increased knowledge of the cognitive processes (and often biases) involved in legal decision-making. The second chapter, by Susana Navas, provides an overview of the application of AI in the legal field, of different LegalTech tools, and on the probable impact of the European Proposal for a Regulation on Artificial Intelligence.

A final note. Bruno Celano has been one of the creators of the RECOGNISE project and of the book it was meant to produce. Sadly, Bruno died in May 2022. It was an enormous, incalculable loss, both scientifically and emotionally. All our work up to the publication of this book (in which we are happy to have been able to include the English translation of one of his essays) has been accompanied and sustained by the memory of his splendid intelligence, by the care he was able to give us and by the knowledge he passed on to us.

MARCO BRIGAGLIA and CORRADO ROVERSI

PART I.

On the Naturalization of Law and Legal Theory

Ancestors: Early Empirical Approaches to the Analysis of Legal Concepts in Modern Legal Theory

FRANCESCA POGGI, FRANCESCO FERRARO

1. *Introduction*

This chapter provides an introduction to some thinkers who, starting in the 18th century, have advocated an empirical approach to jurisprudence. With regard to the 20th century, the focus will mostly be on Scandinavian Legal Realism and American Legal Realism[1]. The approaches in question can be labelled as "empirical" in one or more of the following senses.

Firstly, they undertake to explain legal concepts—like rights, duties, powers, or obligations—by relating or reducing them to empirical facts. The analysis of legal concepts has always been hotly debated. What does it mean that somebody has a right or that somebody else is under a certain obligation? One of the main problems is that legal concepts are not empirical facts: nobody has ever seen or touched a right or a duty. The existence of rights and duties is usually explained by referring to the legal norms which create them. However, on the one hand, the concept of "legal norm" is troublesome as well, and, on the other, this still does not explain what kinds of entities rights and duties are, or what kind of existence they have.

Secondly, the approaches in question seek to explain the directive function of law in empirical terms. The law has many functions, but especially important among them is that of directing behaviours. And yet, from an empirical point of view, law *prima facie* appears as a set of texts. How is it possible for a set of texts to successfully direct conduct?

Finally, the approaches here examined develop an empirical methodology for the study of legal reasoning in general, and judicial reasoning in particular. They assume that the methods of legal theory must be continuous with those of the natural sciences.

The three previous senses of "empirical approach" are surely interconnected. As we will see, their common trait lies in the translation of legal concepts into empirical facts in order to make their functions amenable to empirical enquiry, both in directing conduct and in legal reasoning. However, those three senses need to be distinguished, since not all of them are expressly adopted by the thinkers under discussion.

2. Early modern empirical approaches to legal reasoning: Hume, Smith, and Bentham.

The first modern attempts to apply an empirical and "scientific" approach to legal reasoning can

[1] We must warn that these are certainly not the only authors in whom it is possible to find traces of an empirical approach to law. Worth mentioning among the authors we have not been able to take into account is at least Petrażycki, the founder of Polish-Russian legal realism. For an accurate comparison between Polish-Russian legal realism and Scandinavian legal realism see FITTIPALDI 2016.

probably be traced back to the 18th century and to the work of David Hume. In the third book of his *Treatise*, Hume propounds his famous theory of justice as an "artificial virtue", namely, a disposition «based on social practices and institutions that arise from conventions» (MORRIS & BROWN 2022). As is well known, Hume followed the Newtonian model, in that he tried to offer an entirely empiricist and naturalistic account of the mind's working and contents. In turn, Hume's explanation of how the very concept of justice arises, and why humans feel moral approbation of justice as a virtue, is based on his theory of the mind. Although it could be objected that Hume's concern with justice is with morality, not law, his account of justice traces morality back to conventional social rules, having «features which jurists often associate with law and not with "pre-legal" social rules or custom» (POSTEMA 2019, 81). That is, justice, for Hume, is an inherently juridical concept, and «law is as much involved in defining of the basic institutions of justice as it is in their interpretation and enforcement by formal governmental institutions» (POSTEMA 2019, 82). So, on the one hand, the *origin* of justice is tied to the development of legal rules, broadly understood as the result of social interaction and the development of conventions; and once these rules are fixed, terms like "obligation" and "right" start to make sense. On the other hand, Hume also provides a theory of obligation which could be seen as applying not only to moral but also to legal obligation. For him, an obligation to perform an action *x* only applies when some *motives* pertaining to the natural and common range of human motivation can prompt us to perform that action. In other words, *ought* implies *can* in the sense that we cannot be under any obligation if no corresponding motive can be found. Moreover, an obligation to do *x* entails that if we fail to do *x*, this is a sign that we lack a quality of character (the aforementioned "natural" motive) and that is a defect or imperfection. The defect or imperfection is such because it provokes a certain kind of displeasure in the observers, as distinguished from the peculiar pleasure we feel when we behold a virtuous action or trait of character. This latter kind of pleasure is generated by the passion of moral approbation. Moreover, when someone lacks the quality of character, ie. the motive needed to do *x*, they usually know it themselves, and they can come to feel shame for this lack and comply with the obligation out of a sense of duty, rather than prompted by a natural motive (HAAKONSSEN 1981, 30-34). This squares with Hume's widespread characterization as a *moral internalist*, namely, as someone who holds that motivation is conceptually connected with the acceptance of, or belief in, some moral standard (see DARWALL 1983, 55).

However, Hume's account does not really enable us to distinguish between purely moral and full-fledged legal obligations. This distinction is one that can instead be found in Adam Smith's jurisprudence. Smith likewise builds on the foundation of justice as a virtue. On his interpretation, however, justice is a purely «negative virtue», one that «only hinders us from hurting our neighbour. [...] We may often fulfil all the rules of justice by sitting still and doing nothing» (SMITH 2002, 95 f.). In contrast to other virtues, the rules of justice are clear and precise and leave room for no unforeseen exceptions—as we would say, they are indefeasible. This, in Smith's view, is due to the fact that justice or, better yet, injustice always elicits the same kind of emotional response by the «impartial spectator». A lack of justice, with the «positive hurt» or *injury* it inflicts, causes resentment in those affected, who will then seek punishment; due to the force of sympathy, the impartial spectator approves of the reaction (punishment) and may even be prompted to cooperate in supporting such a reaction (SMITH 2002, 86–88). The link between Smith's theory of moral sentiments and his «natural jurisprudence» is the concept of rights (HAAKONSSEN 1981, 99). Like "justice", "right" can be defined by appealing to injury; by establishing what actions constitute an injury to us, we can also lay down what our rights are (in the sense of the right *not to be subjected* to such actions). Smith even provides a classification of all subjects of law based on the division of rights, which in turn is determined by a division of the different kinds of injury that someone can suffer.

Smith's empiricist account of rights is based on the impartial spectator's sympathy with those whose rights are violated, that is, who suffer an injury. When it comes to «real rights», that is, rights to *things*, and hence rights held against any other subject—paradigmatically, property rights—the impartial spectator will sympathize with those who are disturbed in their possession by anybody else, and will approve of their reaction in protecting their possession and punishing the trespasser. But why does possession give rise to rights and elicit the spectator's sympathy? Smith's answer lies in the «reasonable expectation» that one can use and dispose of one's own possession: this is what activates the sympathetic mechanism, in that the spectator will sympathize with the expectation in question. However, such "reasonableness" will vary according to the specific circumstances in which the impartial spectator observes the injury being inflicted; hence, different historical and geographical contexts could allow for very different answers in regard to the kinds of things that could be deemed property (HAAKONSSEN 1981, 104-106)—and, one could add, in regard as well to the sheer amount of property that one can rightfully possess. Essentially the same explanation also accounts for «personal rights», i.e., rights to some service from another person. They derive from promises, i.e., statements that can give rise to reasonable expectations. *Obligations*, in turn, are explained in terms of the «expectation and dependance of the promittee that he shall obtain what was promised» (SMITH 1978, 87).

Interestingly, Smith's natural jurisprudence is meant as a *science*, but a *normative* one: it devises a model to be followed by an ideal statesman; hence, it is «the science of a legislator» (SMITH 1976, 39). The idea of a science of legislation was almost commonplace both in British and Continental Enlightenment (BURNS 1984, 7); from Smith himself and from Claude-Adrien Helvétius, this idea was enthusiastically received by Jeremy Bentham (BENTHAM 1968, 99; also HOESCH 2018). In Bentham, we probably find for the first time all three of the senses which we have distinguished with regard to an "empirical approach" to jurisprudence.

However, Bentham's jurisprudence pursues the naturalization of legal concepts as part of a «science of legislation» that is very different from Smith's, in that, as we will see, it draws a sharp distinction between the task of *describing* the law and that of *prescribing* how the law ought to be, as well as between moral and legal concepts.

"Universal expository jurisprudence" (i.e., the descriptive science of law) was, in his view, concerned with defining the basic concepts in law, such as "power", "obligation", "duty", "right", and "liberty", along with many others, among which we find "law" itself (in the sense of *a* law) (BENTHAM 1970, 295). Bentham's empiricist account of legal concepts lies in his theory of fictitious entities. He considered all nouns as being either «names of real entities» or «names of fictitious entities». Names of real entities stood for ideas derived from the perceptions of actually existing objects. By contrast, names of fictitious entities referred to entities to which no real, perceptible existence was ascribed; their existence was only feigned for the sake of discourse (BENTHAM 1997, 84). Conspicuous examples of such names of fictitious entities were nouns like "obligation", "right", and other legal terms: we cannot actually experience, in the sense of perceiving, an obligation or a right. However, the terms in question refer to complex states of affairs, which in turn are reducible to real entities. Hence, we can semantically analyse them by means of the «paraphrastic method», through which whole propositions having as their subject the name of a fictitious entity are translated into equivalent propositions having as their subject a name of a real entity (BENTHAM 1843a, 246 f.). For instance, if we wanted to analyse the term "obligation" (or its synonym "duty"), we would translate the sentence "An obligation to do x is incumbent upon y" into "If y abstains from doing x, she is likely to experience a certain pain (or loss of pleasure)". Obligation and duty are explained in terms of the likely punishment for failing to act accordingly, the punishment deriving from any of three "sanctions", that is, from any of three "causes of pain and pleasure": the religious sanction, in the form of punishments expected at the hand of some divinity; the popular or moral sanction, i.e., punishments deriving from the attitudes of fellow citizens; or the political or legal sanction, i.e. punishments administered by

judges and other legal officials (BENTHAM 1970, 35-37; BENTHAM 1977). Accordingly, we have religious, moral, and legal duties. "Right", in turn, was the name of a fictitious entity of "the second remove", in the sense that propositions about rights had to be translated into propositions about duties (i.e., duties of others towards the right-bearer). Hence, their relation to names of real entities was mediated by other names of fictitious entities (BENTHAM 1997, 164-166).

Thus we see that, for Bentham, the meaning of legal terms must be understood with reference to empirical concepts, like pain and its probability. Also closely tied to this empirically based semantic analysis is his account of the directive function of law. For Bentham, there are no norms understood as mysterious abstract entities. While "duty", "right", and the like are names of fictitious entities, when we speak of "a law" we are referring to a real entity (BENTHAM 2010, 316), namely,

> «an assemblage of signs *declarative* of a *volition* conceived or adopted by the *sovereign* in a state, concerning the conduct to be observed in a certain *case* by a certain person or class of persons [...] subject to his power: such volition trusting for its accomplishment to the expectation of certain events which it is intended such declaration should upon occasion be a means of bringing to pass, and the prospect of which it is intended should act as a motive upon those whose conduct is in question» (BENTHAM 2010, 25 f.).

Hence, it seems that Bentham conceived of normativity as simply the result of a prediction on the law-subjects' part: the anticipation of a certain pain or loss of pleasure, which is sufficient to prompt action or abstention from action[2]. The directive function of law basically consists in its capacity to interfere with the individuals' "motivational set" so as to provide them with new, "artificial" motives to direct their conduct.

Lastly, Bentham endorsed method continuity between jurisprudence and the natural sciences. Despite Bentham's references to the Newtonian conception of science (HALÉVY 1901, 289 f.), a much deeper influence on his expository jurisprudence was probably that of Francis Bacon's and Carl Linnaeus's systems (JACOBS 1990; MIXON 2020). Be that as it may, Bentham conceived of morality as a science mainly based on «sensation and experience» derived from observation, «but partly [...] upon experiment, as much as medicine» (BENTHAM 1988, 26). He saw the science of legislation as one sub-branch of "Deontology", or the science of morality. In his all-comprehensive scheme of all the sciences (an idea he had drawn from Bacon and d'Alembert) (MACK 1962, 97, 105-112), each science was coupled with its respective *art*: on this approach, which has been labelled as «proto-pragmatist», the sciences had value insofar as they could be turned into techniques capable of acting on and changing reality (MACK 1962, 136 f.). Hence, the science of legislation was inextricably connected with the art of legislation, i.e., a technique meant to improve legislation. Another name for this art was "censorial jurisprudence", while "expository jurisprudence" corresponded to the science of legislation (BENTHAM 1970, 293 f.). However, the descriptive and prescriptive aspects of jurisprudence were never to be conflated: studying law *as it is* was different from prescribing how it *ought to be* (BENTHAM 1977, 397 f.). The former task belonged to the "Expositor", who dealt in *facts*; the latter, to the "Censor", who was concerned with *reasons* (BENTHAM 1977, 397 f.). In all cases, even statements about what ought to be were deconstructed by Bentham in purely naturalistic and empirical terms, since they simply expressed a sentiment of approbation on the part of those who uttered them (BENTHAM 1983, 206 f.) — thereby anticipating 20th-century metaethical emotivism.

[2] For simplicity's sake, we are leaving out the possibility (admitted by Bentham) of «praemiary or invitative laws», i.e., laws relying on the anticipation of pleasure (in the form of a reward) rather than pain (in the form of punishment). See BENTHAM 2010, 146.

3. *Scandinavian legal realism*

Scandinavian Legal Realism owes its name to Axel Hägerström's thesis of reality[3]. Against idealistic epistemological subjectivism, Hägerström affirms a subject-object dualism: he claims that, in the cognitive process, the subject comes into contact with a reality independent of her perception. According to Hägerström, in order to identify this reality we must refer to the law of contradiction[4]. This law states that two judgements, one of which denies what the other asserts, cannot both be true. This means that we cannot regard as real a situation in which two contradictory worlds exist. Therefore, according to Hägerström, we can regard as real only objects that are self-identical, that is, determined and, therefore, internally coherent. But in order to conceive of different wholes of objects as unified objects, we need to think of a unifying principle: «a whole in addition to which none other is thinkable» (HÄGERSTRÖM 1964, 53), a *continuum* which is determined, and which therefore enables us to identify other objects as self-identical, determined, and coherent (see HÄGERSTRÖM 1964, 53 f.). This unifying principle, according to Hägerström, cannot be but the world of the senses, that is, the world to which the knowing subject pertains[5]. In other words, in order to explain the possibility of an objective knowledge, Hägerström embraces the thesis which is nowadays labelled "ontological naturalism"[6], according to which there is one (and only one) spatiotemporal framework, and everything that exists is to be found in this framework[7]. Ontological naturalism claims that reality is exhausted by nature, by physical entities, and denies the existence of ideal, "supernatural", or other "spooky" kinds of entities.

Hägerström can be considered the founder of Scandinavian Legal Realism, and his ontological naturalism exerted a strong influence on all exponents of the movement: Anders V. Lundstedt, Karl Olivecrona, and Alf Ross all claim that nothing exists but "mere facts" (See, e.g., LUNDSTEDT 1932, 328 f.; OLIVECRONA 1939, 15, 25 ff.; ROSS 1958, 67). However, law is not a set of physical entities. To be sure, law is a set of texts (statutes, judicial decisions, etc.), but we usually think that such a set creates rights and duties, that it expresses norms directing people's conduct. What sort of entities are duties, rights, and legal norms? Scandinavian legal realists try to explain legal concepts (such duties and rights) as well as the directive function of law in empirical terms.

3.1. *Legal concepts in Scandinavian legal realism*

From Hägerström's thesis of reality it follows that moral and axiological facts do not exist. In fact, Hägerström maintains that moral and normative judgements express just feelings, emotions:

> «If one says, "It is good to possess a barrel of potatoes", this is the same, in so far as "good" actually has a valuational significance, as "How good it is, indeed, to be in possession of a barrel of potatoes"

[3] See HÄGERSTRÖM 1908. However, Hägerström explicitly distances himself from what he calls "realism", that is, the theory according to which objects are "out there" and the subject does not play any role in the cognitive process: see HÄGERSTRÖM 1908, 54 ff.; HÄGERSTRÖM 1964, 74; on this topic see also MINDUS 2009, 48 and 52 ff.

[4] See HÄGERSTRÖM 1964, 42: «The law of contradiction declares, in fact, what reality in itself is».

[5] HÄGERSTRÖM's argument is much more complicated than that just set out. For a more detailed analysis see FRIES 1944; SANDIN 1959; SANDIN 1966; MARC-WOGAU'S 1972; PATTARO 1974, 40 ff.; FARALLI 1982; CASTIGNONE 1995; LYLES 2006; MINDUS 2009, 48 ff.

[6] See BJARUP 2005; SPAAK 2009; FERRARO & POGGI 2014; PAPINEAU 2021. Hägerström himself calls his approach «rational naturalism», so as to stress «the completely logical character of sensible reality» (HÄGERSTRÖM 1964, 37). On this point see BJARUP 2005.

[7] Note that Hägerström does not demonstrate and does not intend to demonstrate that sensible reality truly exists: he argues only that every attempt to demonstrate anything, every judgement, postulates the reality of the world of experience, as proven by the fact that denying that reality gives rise to an inconsistency (see MINDUS 2009, 52).

> or "Oh, if one only had something like that!" Thus it is manifestly an expression of feeling» (HÄGERSTRÖM 1964, 70)[8].

In particular, some normative judgements, like duty-sentences, express «a simultaneous association of a feeling of conative impulse with the idea of an action» (HÄGERSTRÖM 1964, 114). Thus, for example, if the law states, "All citizens ought to pay taxes", what it expresses, and what it aims to create in the citizens, is a «state of consciousness of duty»[9], that is, an association between an unconditional feeling of volition and an idea of action ("paying taxes"). In particular, the "ought to" is perceived and presented as something which belongs objectively to the action of paying taxes, as if there were an «oughtness to be» (HÄGERSTRÖM 1964, 136) as an objective property of the action. Hägerström maintains that this objectification of duties, this feeling of a conative impulse associated with actions, is mainly caused by a process of sociopsychological conditioning, to which a variety of factors contribute (such as education, tradition, habits, the effectiveness of sanctions, and the convergence of different authorities in prohibiting and sanctioning the same conduct). All these factors in combination give rise to «the idea of a system of ways of acting, having the expression of command as an objective property, which carries with it a feeling of conative impulse and leads to a judgement on each particular action» (HÄGERSTRÖM 1964, 131).

In the same manner, Hägerström claims that the concept of a (legal) right conveys the idea of a supernatural power, a mysterious force, which in reality does not exist[10]. The idea of a right can only be explained psychologically, through an account that lays emphasis on the feelings of strength and power associated with the conviction of possessing a right.

It is worth noting that, according to Hägerström, feelings, emotions, and ideas of action are all natural entities: they pertain to the spatiotemporal reality, even if only indirectly, insofar as they belong to a person's mind, through which they therefore exist[11].

Thus legal norms are expressions of commands, i.e., expressions of a conative impulse combined with the idea of another person's action. They do not create duties and rights, but they create the *ideas* of duties and rights as objective properties, and they do so through complex sociopsychological mechanisms, in which an important role is played by the effectiveness of norms themselves and of sanctions[12]. Hägerström also notes that «the expression of a command leads one's thoughts inevitably to a commanding will» (HÄGERSTRÖM 1964, 133). However, according to Hägerström, legal norms are not imperatives: they are not expressions of the will of the State or of other collective entities. This sort of «singularly mystical will» does not exist: it is a fiction[13]. Law is just a system of rules which is actually applied and followed.

[8] As BJARUP (2005, 4) notes, Hägerström's view of morality «is a version of nominalism that holds that there are no moral concepts but only moral words that are used to express various feelings or sensations».

[9] For a more detailed analysis of this idea see PATTARO 2005, 135.

[10] See, e.g., HÄGERSTRÖM 1953, 4 ff. Hägerström also investigates the historical roots of the idea of a right, arguing that the Roman *ius civile* was conceived as a system of rules for acquiring and exercising supernal natural power through magical acts: see HÄGERSTRÖM 1927.

[11] See HÄGERSTRÖM 1908, 76; on this point see also PATTARO 2005, 138; LYLES 2006, 74; MINDUS 2009, 57; FITTIPALDI 2016.

[12] See HÄGERSTRÖM 1964, 167: «It is necessary [...] that fictitious or real commanding authorities should assert themselves effectively and unanimously in a society, in order that the expression of command shall be transformed into a supposed real property of a system of conduct and that the idea of duty shall enter».

[13] According to Hägerström, even if it were plausible to configure the law as an expression of will, in any case, the courts do not ever apply this will. Judges apply a law which does not correspond to the will of all the members of parliament, because historical research into the will of legislators is possible only within certain limits, nor can they be thought to have had a clear appreciation of the full implications of the law they enacted, not even in the case of a single legislator. In fact, a law will sometimes have to be applied to cases which could not have been foreseen at the time of its enactment. See HÄGERSTRÖM 1953, 34. On this point see also MINDUS 2009, 125.

> «When one talks of the "sovereign organs" of a society which is a state, nothing else is meant than that certain rules for the exercise of supreme power come to be applied by persons [...] appointed for that end, in consequence of forces operative within the society» (HÄGERSTRÖM 1953, 37).

Therefore, «[t]he legal order throughout is nothing but a social machine, in which the cogs are men» (HÄGERSTRÖM 1953, 354).

In sum, according to Hägerström, legal vocabulary is «only a matter of using empty words [...] to cause appropriate behaviour» (BJARUP 2005, 7). Hägerström shows how this is possible: he explains how empty legal concepts (or, better yet, sentences containing legal concepts) are connected to "real" psychological events—to false ideas about objective properties which actually do not exist in our world—and how such a connection arises. Moreover, Hägerström enquires into the role of these ideas of objective duties and rights in directing human behaviours[14].

Hägerström's theses strongly influenced other Scandinavian legal realists: Lundstedt and Olivecrona only deepened Hägerström's analysis of legal concepts. The only one to have departed from it to some extent was Alf Ross.

Ross likewise maintains that legal concepts express indefinite ideas about supernatural powers, which in reality do not exist. But Ross thinks that statements about rights can be interpreted in a meaningful way. According to Ross, the word "right" is used as a technical tool of presentation in order to connect a number of conditioning facts with directives for judges[15]. So, for example, to say that "Alf has a property right to the thing T" means that certain facts occurred (e.g., Alf has entered into a valid contract for the purchase of T, or has found T which was *res nullius*) and that, therefore, judges must use coercion is a certain way (e.g., if Karl has stolen T, judges must convict Karl of theft, or, if Alex has damaged T, they must order Axel to pay damages).

Thus Ross develops a logical analysis of legal rights which is very far from Hägerström's orthodoxy[16]. We will return to this point in § 3.3, but before that let us (in the next section) turn to how Olivecrona and Ross developed Hägerström's legacy in explaining the directive function of law.

3.2. *The machinery of law: Scandinavian legal realists and the directive function of law*

As we have seen, Hägerström argues that our false ideas of objective duties and rights play a decisive role in directing human behaviours. The analysis of the role of these false ideas within legal systems was especially deepened by Olivecrona.

Olivecrona maintains that, even if rights are not facts, «[t]he subjective ideas of rights are facts. They cannot be "excised" from the law in the sense of law as fact»[17]. According to Olivecrona, the idea of rights plays a directive function insofar as it exerts a psychological compulsion: it acts as a sign that indicates which actions are permitted and which are not. This psychological effect depends only partly on the threat of sanctions:

> «[the law] would become inefficient if people paid regard to the supposed rights of others only out of the fear of punishment, and if they implemented their legal obligations only under the immediate threat of

[14] Actually, according to Hägerström, the idea of law as a system of objective duties and the idea of law as the product of a superior and compelling force coexist and are both (incorrect and) effective in driving conduct: see HÄGERSTRÖM 1953, 251 ff.

[15] ROSS 1951. For some criticisms see OLIVECRONA 1971, 180 ff.

[16] On this point see BIX 2009.

[17] OLIVECRONA 1971, 185. By contrast, Lundstedt proposes to delete the traditional legal concepts or to use them only in inverted comma. See LUNDSTEDT 1956, 17. Indeed Lundstedt continues to use such traditional concepts in inverted comma.

> sanctions. [...] The directive function of the ideas of rights and duties on the one hand and the working of the machinery of justice on the other are mutually interdependent» (OLIVECRONA 1971, 191 f.).

In this regard, Olivecrona stresses that «[t]o be effective among the general public, legislation on rights presupposes that people [...] attach consequential ideas concerning their own behaviour to the supposed existence of rights» (OLIVECRONA 1971, 198). Also, new rights are always founded on pre-existing ideas of rights, on the «inveterate habit of taking one's own rights as a green light and that of others as a red light» (OLIVECRONA 1971, 199).

In sum, Olivecrona follows Hägerström in claiming that rights and duties are nothing but false ideas generated through a process of sociopsychological conditioning to which the machinery of justice and the fear of sanctions also contribute. However, Olivecrona emphasises, even more strongly than Hägerström, that such false ideas are fundamental cogs in the machinery of law, as well as in the process of psychological conditioning that creates them. It is because people *think* of rights and duties as something real that they act accordingly. It is because everyone *behaves* in this way that we are conditioned to continue to do so. Rights and duties do not exist, but the idea of rights and duties is crucial to explaining the directive function of law.

Alongside and behind the ideas of rights and duties, other factors explain the directive function of law, its ability to direct behaviour. This emerges above all in Olivecrona's analysis of legal rules. According to Olivecrona, legal rules are not commands: they are independent imperatives, that is, imperatives independent of personal relationships. Olivecrona distinguishes two elements of legal rules: an imagined pattern of conduct (which he labels *ideatum*) and «the particular form in which expression is given to it» (OLIVECRONA 1971, 115. See also OLIVECRONA 1939, 29 ff.) Olivecrona labels this second element *imperatum* and claims that it does not consist in «an expression of a wish on the part of the lawgiving authorities»[18]. Instead, it consists in an imperative—an unconditional "shall"—that makes no appeal to any values on the part of the addresses (see OLIVECRONA 1971, 120). In particular, the *imperatum* of legal rules consists in

> «the whole setting in which the enactment takes place: the working constitution, the organization functioning according to its rules, familiar designation of parliamentary bodies and state officials, etc. Once the constitution has been firmly established, the people respond automatically by accepting as binding the texts proclaimed as law through the act of promulgation. Thanks to this attitude among the addressees the *imperatum* becomes effective» (OLIVECRONA 1971, 130).

So, according to Olivecrona as well, there is not an *imperator* behind a legal rule: there is only a set of other rules, a machine functioning according to the rule, and a general, psychological, attitude of obedience.

> «[T]here is no homogeneous source of the rules reckoned as legal. [...] There is no single driving force to the system; the regular application of the rules and their efficacy in governing the life of society depends on a network of psychological and material factors (ideas of rights and duties, habit, belief in authority, fear of sanctions, and so on)» (OLIVECRONA 1971, 77).

What emerges, according to Olivecrona and Hägerström, is that law is a fact, but a very complex one: a network, an interaction, between psychological facts, material facts, ideas, habits, feelings, behaviours that support one another.

[18] OLIVECRONA 1971, 118. While Hägerström claims that legal norms are expressions of commands and not imperatives, Olivecrona claims that they are (impersonal) imperatives and not commands. However, the difference seems merely terminological. For, according to both authors, norms are made up of an idea of action coupled with a normative element that does not express any will.

A partly different analysis of law and legal norms is advanced by Ross. In *On Law and Justice*, Ross claims that the "real content" of a legal rule is always a directive for the judge concerning the use of force: «A national law system is an integrated body of rules, determining the conditions under which physical force shall be exercised against a person» (ROSS 1958, 34). According to Ross, this body of rules is valid if, and only if, judges apply it, and they do so because they feel bound by it. Legal norms are valid if, and only if, they are «effectively followed, and followed because they are experienced and felt to be socially binding» (ROSS 1958, 18). Therefore, to know which legal norms are valid means to know by what norms the courts feel bound: the task of a realistic legal science is to develop descriptions of how judges actually behave, in such a way that predictions can be made about their future conduct.

In the later *Directives and Norms* (1968) this conceptual apparatus becomes more sophisticated but remains essentially unchanged. Here Ross claims that a directive is a linguistic phenomenon, an action-idea conceived as a pattern of behaviour. As far as legal rules are concerned, Ross argues that they are impersonal and heteronomous directives that stand in a relation of correspondence to social facts, meaning that by and large they are followed by the members of a given society. According to Ross, «it is necessary for the establishment of a norm that it be followed not only with external regularity [...], but also with the consciousness of following a rule and being bound to do so» (ROSS 1968, 83). Ross repeats that, from a logical point of view, all legal norms are directed at judges (ROSS 1968, 90 ff., 113 ff.). However, he now admits that the rules addressed to citizens are felt to be independent and therefore «must be recognized as actually existing norms, in so far as they are followed with regularity and experienced as being binding» (ROSS 1968, 92).

In sum, according to Ross, the law is a set of linguistic meanings, but ones that exist in a very peculiar way: legal norms exist if, and only if, they are effective and felt as binding. Their binding force—the "experience of validity", as ROSS (1958, 62) calls it—is a psychological phenomenon, one that is owed to many factors (the same ones analysed by Olivecrona) and which must be accounted for as a fact. So the concepts of (valid) law are reduced to a mix of linguistic, sociological, and psychological facts.

3.3. *Scandinavian methodological naturalism*

Scandinavian Legal Realism contends that law is a set of facts, albeit a very complex one. Conceiving of law as a set of facts is the first step to developing an empirical approach to it. By "empirical approach" we mean the attempt to analyse law and/or legal reasoning in empirical terms. Nowadays, we call *methodological naturalism* the view that the methods of jurisprudence must be continuous with those in science. The continuity can be strictly *methodological* (i.e., philosophical theories must emulate the method of inquiry and styles of explanation employed in the sciences) or it can be *results-based* (i.e., philosophical theories must be supported by scientific results), or both. Whether the Scandinavians can be said to have adopted a methodological naturalism, however, is up for debate.

In this regard, Torben Spaak claims that Olivecrona's analysis of the function of legal rules could—in principle—be empirically tested. However, Spaak concedes that Olivecrona never emphasized the "testability aspect" of his analysis, nor did he devote himself to any sort of empirical analysis (see SPAAK 2009). The issue mainly depends on how we frame empirical analysis, on how we fashion scientific methods and results. Surely, the Scandinavian legal realists never collected statistical data, never expressly formulated predictions, nor did they try to falsify their theses or predictions by comparing them against empirical findings[19]. Even so, the inquiries conducted by

[19] In fact, the thesis that rights and duties are false ideas cannot be tested in this way, since psychological facts are not directly observable. On this point, however, see FITTIPALDI 2016, 309, who underscores how, following Karl Popper and Hans Albert's critical rationalism, entities not amenable to direct observation can be hypothesized in a way that is consistent with the Scandinavian legal realists' analysis.

Olivecrona, Lundstedt, and Hägerström recall scientific investigations in at least four respects.

First, their ontological naturalism places them in the same domain as science: they do not admit the existence of supernatural entities, or of any entities that are not scientifically ascertainable. Second, within that domain, they want to offer a descriptive explanation of legal phenomena. Third, they infer their explanation from empirical data: legal texts, ways of talking, and close historical analysis[20]. And, finally, such an explanation is focused on causal, and hence empirically testable, connections[21].

As far as Ross is concerned, a separate issue needs to be addressed. Ross adheres to logical positivism—which Hägerström instead rejected—and holds that the task of jurisprudence is to develop a logical analysis of the fundamental legal concept of general scope (see ROSS 1958, 25 f.), so as to ensure that the predictions made about law can be empirically verifiable (or falsifiable). But in fact Ross, throughout his career, confined himself to the first part of this task, namely, linguistic analysis. Moreover, his analysis is not merely a description of the current uses (of the various meaning) of these concepts, but rather consists in redefining these concepts, conceiving them anew by defining them in new terms[22]. This is certainly a very different enterprise from Hägerström's. Hägerström and Olivecrona do not want to redefine legal concepts so as to make them meaningful—or to do so according to logical positivism, by relating these concepts to facts—nor does Lundstedt embark on such an enterprise. What they want to do, rather, is to explain how these concepts work and direct behaviours, even if they do not refer to anything. In this respect, their analysis is also far removed from that of the American legal realists[23].

4. *The legacy of early American legal realism: legal concepts and facts*

It is usually accepted that, unlike the Scandinavian legal realists, the early American realists were not concerned with ontological issues. However, they *are* usually regarded as tacitly presupposing some version of ontological naturalism, this in order to make sense of their *semantic* naturalism, namely, the view that meaningful concepts must be directly or indirectly amenable to empirical analysis[24]. This does not necessarily entail subscribing to the view that all existing objects have a factual existence and are thus amenable to empirical proof. For instance, the ontology of Felix S. Cohen explicitly included nonfactual objects, such as relations and processes (COHEN 1960, 62 fn.).

An empirical approach to the analysis of legal concepts is typical of Oliver Wendell Holmes, usually deemed to be the founding father of American Legal Realism. A prominent example of an approach that reduces legal concepts to empirical facts is offered by Holmes's famous reduction of legal science to the point of view of the "bad man", that is, to the prediction of the

[20] See HÄGERSTRÖM 1953, 299: legal science «has become one of the special sciences. Like physics and chemistry, for example, its function is merely to establish the facts within a certain region, to reach general principles by induction, and to make deductive inferences from the inductively established results».

[21] See OLIVECRONA 1971, 84: «the law is a link in the chain of cause and effect». LUNDSTEDT 1956, 126: legal science must be concerned with «social evaluation and other psychological causal connections». On this topic see also BJARUP 2005; FERRARO & POGGI 2014; FITTIPALDI 2016.

[22] This is particularly evident in his analysis of the concept of a legal right (§ 3.1.). Olivecrona criticizes Ross's argument about legal rights, claiming that there is no historical foundation for the assumption that the rules of private law are primarily conceived as rules on the use of the State's force, which rules are then expressed in a more simple and practical way by inserting the word «right». On the contrary, Olivecrona stresses that the rules of private law are generally conceived as rules regulating people's rights and duties: «we cannot take a single step in describing private law without making use of the words "right" and "duty"» (OLIVECRONA 1971, 180).

[23] On this point see, e.g., PATTARO 2005; BIX 2009; FERRARO & POGGI 2012; FITTIPALDI 2016.

[24] LEITER 2007, 35 fn. (Leiter considers this as one of two forms of «substantive» naturalism, the other being an ontological view); SPAAK 2009.

consequences the courts could attach to certain behaviours[25]. For Holmes, a legal duty is «nothing but a prediction that if a man does or omits certain things he will be made to suffer in this or that way by judgment of the court» (HOLMES 1952, 169). Much like Bentham—and unlike Hume and Smith—Holmes's empirical approach allows for (and indeed requires) a clear-cut distinction between moral and legal concepts: the bad man's interest in the law is merely instrumental to avoiding clearly identifiable disagreeable consequences[26]. Legal norms do not seem to belong to Holmes's conceptual toolbox. Moreover, his insistence on the need for a purely predictive approach to legal science explains away the law's directive function as a by-product of its being a set of predictions regarding the disagreeable consequences which the "bad man" would seek to avoid. Of course, this leaves a central question unanswered—namely, how judges (and government agencies) are supposed to be guided by existing law in their decision-making[27]. They would probably also have to think of the undesirable consequences they themselves would incur if they should misapply the law or improperly enforce it.

Almost four decades later, this same empirical approach to the analysis of legal concepts is endorsed by Felix S. Cohen, who develops a "functional approach" meant to analyse the mysterious, "supernatural" concepts of legal language in purely factual terms. Any such concept that would eventually come out as unable to «pay up in the currency of fact, upon demand, is to be declared bankrupt, and we are to have no further dealings with it» (COHEN 1960, 48). Like Holmes's approach, Cohen's is an inquiry into factual, empirically assessable consequences, understood as the real meaning of legal concepts, which are thence brought back to earth from the «ghost-world of supernatural legal entities» (COHEN 1960, 54). Not too distant from Cohen's functional approach is Jerome Frank's, who considers legal concepts as artificial devices with an "operational character"[28]. Frank considers legal concepts to be fictions, understood as ideas used as «means to aid thinking» but with no pretence of real existence[29]. Therefore, both Frank and Cohen hold that legal concepts are not directly amenable to empirical enquiry, but in order to be meaningful they must hold at least an indirect relationship with empirically ascertainable facts. Hence, these concepts are to be seen as shorthand or abbreviations, useful for the sake of discourse and possibly of thought (FRANK 1932, 91).

It is more difficult to find some uniformity in the American realists' treatment of the concept of a legal rule. As noted, Holmes saw legal rules as mere predictions of the courts' decisions. According to Frank, some realists were "rule-sceptics", in the sense that they thought that "paper rules"—rules which could be derived from law books by means of interpretation—concealed the "real rules" that describe judicial behaviour and are useful for predicting future judicial decisions. Other realists, in his view, were "fact-sceptics", in that they thought that the «elusiveness of the facts on which decisions turn» made it impossible, in most (though not all) cases, to predict the outcome, no matter how precise the applicable rules (FRANK 1949, 10 s.). Despite the fact that «Frank's description of the fact-sceptic is basically a description of himself» (DUXBURY 1991, 178), he also acknowledges that «the so-called legal rules have *some* effect» in determining the judge's final decision (FRANK 1931, 43). Frank seems to ascribe to rules a capacity to causally influence judicial decision: they can do so as one of the very many possible factors that lead the judge to

[25] Holmes's "prediction theory", however, was anticipated and probably inspired by Nicholas St. John Green: see HORWITZ 1992, 53 f.

[26] HOLMES 1952, 170-173. However, this by no means makes of Holmes a moral relativist, nor a non-cognitivist, as he thought that ethics could be made the object of a science: see TARELLO 1962, 42 f.

[27] GOLDING 1986, 444. The first to make this point was probably John Dickinson: see DICKINSON 1931, 843.

[28] FRANK 1949, 167, 319 fn. However, Frank himself was dismissive of both Cohen's functionalism and Holmes's predictivism: see DUXBURY 1997, 133.

[29] FRANK 1949, 317. Frank refers to Hans Vaihinger's *Philosophy of As If* and adopts Vaihinger's distinction between dogmas, hypotheses, and fictions properly so called, as well as Pierre de Tourtoulon's distinction among fictions, lies, and myths.

form a "hunch" as to the correct solution to the case at hand[30]. Still, this does not say anything about what rules are and how they can be reduced to empirically ascertainable facts. This also relates to the question of the law's directive function, namely, its capacity to guide conduct and the possibility of explaining that capacity in empirical terms: Frank seems to take it for granted that a knowledge of legal rules can contribute to a sudden intuitive understanding capable of guiding judicial decision-making—as well as it can be helpful to lawyers in persuading courts to decide cases in their clients' favour (FRANK 1932, 761). However, for him, «the rules set forth in a judge's opinion may often be no more the cause of his decision than the cigar he was smoking when he made up his mind» (FRANK 1931, 44); and he provides no account of how rules—or, for that matter, cigars—can be a causal determinant of decision-making.

As much as Karl Llewellyn was reckoned by Frank among the rule-sceptics, his position was probably more nuanced. Llewellyn advocated a jurisprudential approach geared toward the study of dispute settlement and official behaviour. In his view, such an approach would translate traditional legal notions, such as that of legal rule, into purely factual terms. He sought to ascertain how far the "paper rules" of traditional jurisprudence were purely on paper (like "dead-letter" statutes) and how far they were instead "real" (LLEWELLYN 2008, 24). By "real" he famously meant "law in action" and (as reported by Frank) he held that real rules were predictions proper. They belong to the realm of "isness" rather than "oughtness": they purport to describe what *is going to* happen, rather than to prescribe what *should* (or *ought to*) happen. This is the use made of the term "rule" by legal scientists. But Llewellyn finds a place for paper rules as well, because in his view legal rules are in the first place «rules of authoritative ought, addressed *to* officials, telling *officials* what the *officials* ought to do» (LLEWELLYN 2008, 23). The officials may not comply with such authoritative directives, or they may comply only partly. Nonetheless, these directives imply a «tacit statement» that officials are already complying with them and a prediction that they will do so in the future. Llewellyn holds that it is a shared tradition that such implicit statements and predictions are solemn truths, and to a certain extent this belief is self-fulfilling. In any case, this tradition entails that «a good paper justification, in terms of officially accepted paper rules» is necessary for any decision to «be regarded as likely of acceptance» (LLEWELLYN 2008, 24). The "verbal formulation" of rules provides a "stimulus" to which officials react in certain ways (LLEWELLYN 2008, 25). Hence, although he does not explain *how*, Llewellyn holds that some complexes of signs causally and factually determine official behaviour. The "official formulae" of the law are there to be used by lawyers to influence court behaviour (LLEWELLYN 2008, 25).

For Llewellyn, rights are the counterparts to rules; «the right is the shorthand symbol for the rule». When a rule favours someone, that person is considered to have a right. When rights are «ascribed to particular individuals in specific circumstances», they «are deductions which presuppose the rule; the major premise is the general rule on rights; the minor is the proposition hooking up this individual and these circumstances with that general rule». As a legal concept, they help gather all the possible remedies which can be allowed for the violation of certain rules; concrete remedies can be seen, while rights cannot, but they have a «scope independent of the accidents of remedies» and of the concrete multifarious behaviours of the courts. In this sense, they afford a «scientific advance», because they enable us to think more clearly among feigned «ultimate realities» supposedly underlying the endless variety of specific cases (LLEWELLYN 2008, 10).

More generally, Llewellyn acknowledges the role that all legal concepts play in organizing thought by means of classification. The data received through the senses need to be arranged into categories; otherwise, they are useless. Categories and concepts, however, «tend to take on an appearance of solidity, reality and inherent value which has no foundation in experience»

[30] FRANK 1949, 112. Frank borrowed the notion of the judicial hunch from HUTCHESON 1929.

(LLEWELLYN 2008, 28). They are useful organizing tools, but they can also deceive, since they tend to suggest the existence of facts corresponding to their classifications, even when such facts are lacking; moreover, they can distort observation when new data are forced into old categories. Nonetheless, other realists also recognize the undeniable usefulness that «conceptual formulation and a logical arrangement» have for «economy of thought and harmony of structure», as well as for economy and harmony in human minds. That, for example, is the view of Max Radin, who acknowledges the role that legal conceptualism plays in making it possible to handle the whole body of legal material: legal institutions, judicial precedents, and statutes (RADIN 1931, 827).

4.1. *Methodological naturalism in American realist jurisprudence*

American legal realists consistently advocate that the same methods applied in the natural sciences should also be used in jurisprudence and legal science. Walter W. Cook, for instance, propounded «a scientific study of law», by which he meant the application of the 20th-century logical and epistemological developments to jurisprudence. It is time, he said, to discard «[t]he nineteenth century notion of science as the ascertainment of all-embracing laws of nature, holding for all cognizable occasions, which we have seen disappearing from physical science in the twentieth century». He instead argued for the view that

> «[u]nderlying any scientific study of the law [...] will lie one fundamental postulate, viz., that human laws are devices, tools which society uses as one of its methods to regulate human conduct and to promote those types of it which are regarded as desirable. If so, it follows that the worth or value of a given rule of law can be determined only by finding out how it works, that is, by ascertaining, so far as that can be done, whether it promotes or retards the attainment of desired ends» (COOK 1927, 308).

What does this involve, in Cook's view? For him, once the common lawyer's traditional reasoning, based on presumptively fixed principles and rules, is shown to be «as grotesquely inadequate for legal purposes as the childish mechanical notions of the nineteenth century have shown themselves to be in the field of physics»,

> «we discover that the practicing lawyer, as much as, let us say, an engineer or a doctor, is engaged in trying to forecast future events. What he wishes to know is, not how electrons, atoms, or bricks will behave in a given situation, but what a number of more or less elderly men who compose some court of last resort will do when confronted with the facts of his client's case" (COOK 1927, 308).

While Cook also endorsed the view that knowledge of the law is basically about predictions, his theory is not vulnerable to the objection which, as discussed, could be levied against Holmes's application of the bad man's point of view, namely, that it cannot account for the judge's point of view.

> «If we shift our point of view from that of the practicing lawyer to that of the judge who has to decide a new case, the same type of logical problem presents itself. The case is by hypothesis new. This means that there is no compelling reason of pure logic which forces the judge to apply any one of the competing rules urged on him by opposing counsel. His task is not to find the prexisting [*sic*] but previously hidden meaning of the terms in these rules; it is to give them a meaning» (COOK 1927, 308).

In Cook's view, this implies acknowledging that the courts must, in fact, legislate (COOK 1927, 308). A similar approach to method continuity between 20th-century natural sciences and legal science is endorsed by Felix S. Cohen, who claimed that the functional approach he advocated

had already been applied in contemporary mathematics, physics, anthropology, and economics as a «functional attack upon unverifiable concepts» (COHEN 1960, 53).

A looser interpretation of scientific method and its application to legal science is offered by Frank and Llewellyn. Both insisted on an empirical approach to jurisprudence and somehow followed in the footsteps of Roscoe Pound's sociological jurisprudence, who saw the law as a social institution that could be reformed in order to pursue a number of general «social interests» (see, for instance, POUND 1943). However, their endorsement of empirical methods is much more consistent than Pound's.

Frank suggests that realist jurisprudence be renamed "experimental jurisprudence", an approach that, as he describes it, characterizes recent developments in both legal science and economics. Experimentalism is, for him, a complex of attitudes: experimentalists studied existing institutions critically and with an eye to possibilities of reform. «[T]hey repudiate fixed beliefs as to the eternal validity of any particular means for the accomplishment of desired ends» and «are keenly alive to the shifting nature of many of the so-called "facts" upon which all human action is based» (FRANK 1934, 2). Though admittedly vague, Frank's description of such an experimental attitude focuses on the rejection of established principles whose application has led to undesirable consequences. Experimentalists, in economics as in law, «tend to look upon human activities with the eyes of anthropologists» (FRANK 1934, 2). They hold that judges commence their reasonings not with premises, but rather with the conclusions they find desirable, «and work backward to the available premises»; similarly, lawyers seek to justify, if possible, what their client desires (FRANK 1934, 3 f.). It seems, then, that Frank's experimental jurisprudence—which he sought to put to the service of the Roosevelt administration's New Deal—is basically a trial-and-error method for achieving some desirable social and political results by means of legal argumentation.

Llewellyn instead focuses on «behavior analysis» as an alternative to the abstract, deductive thinking of the «devotee of formal logic in the law», who is «concerned with words, with propositions» and symbols. The study of the formal logic of judicial decisions could be very useful, provided it is accompanied by «an equally careful study of the instrumentalism, the pragmatic and socio-psychological decision elements in the same cases» (LLEWELLYN 2008, 20 fn.). Llewellyn seeks to bring jurisprudence closer to the other social sciences, like economics and sociology, and criticizes the kind of analytical approach that sees law as a closed system with its own principles and logic, separated from the other fields of human reality and knowledge. He endorses Roscoe Pound's sociological approach to jurisprudence, but criticizes «Pound's preference for the study of theory, verbalized theory, writer's theory, over study of results, or of how it gets done» (LLEWELLYN 2008, 501). In contrast, Llewellyn favours the consistent application of sociology, anthropology, and ethnography to the study of the actual behaviour of those involved in the "craft" of law: as is well known, he gave an example of this approach himself by studying the law and customs of the Cheyenne together with anthropologist Edward Adamson Hoebel (see LLEWELLYN & ADAMSON HOEBEL 1941).

References

BENTHAM J. 1970. *An Introduction to the Principles of Morals and Legislation*, ed. by J.H. Burns, H.L.A. Hart , The Athlone Press, 1 ff.

BENTHAM J. 1977. *A Comment on the Commentaries and A Fragment on Government*, ed. by J.H. Burns, H.L.A. Hart , The Athlone Press, 576 ff.

BENTHAM J. 1983. *Deontology together with a Table of the Springs of Actions and Article on Utilitarianism*, ed. by A. Goldsworth, Clarendon Press, 394 ff.

BENTHAM J. 1988. *The Correspondence of Jeremy Bentham. Volume 7: January 1802–December 1808*, ed. by J.R. Dinwiddy, Clarendon Press, 394 ff.

BENTHAM J. 1997. *De l'ontologie et autres textes sur les fictions*, Seuil.

BENTHAM J. 2010. *Of the Limits of the Penal Branch of Jurisprudence*, ed. by P. Schofield, Oxford University Press.

BIX B.H. 2009. *The American and Scandinavian Legal Realists on the Nature of Norms*, in DAHLBERG M. (ed.), *De Lege: Uppsala-Minnesota Colloquium: Law, Culture and Values*, Iustus Förlag, 85 ff.

BJARUP J. 2005. *The Philosophy of Scandinavian Legal Realism*, in «Ratio Juris», 18, 1, 1 ff.

CASTIGNONE S. 1995. *Diritto, linguaggio e realtà*, Giappichelli.

COHEN F.S. 1960. *The Legal Conscience: Selected Papers of F.S. Cohen*, in KRAMER COHEN L. (ed.), Yale University Press, 505 ff.

COOK W.W. 1927. *Scientific Method and the Law*, in « American Bar Association Journal», 13, 6, 303 ff.

DARWALL S.L. 1983. *Impartial Reason*, Cornell University Press.

DUXBURY N. 1991. *Jerome Frank and the Legacy of Legal Realism*, «Journal of Law and Society», 18, 175 ff.

DUXBURY N. 1997. *Patterns of American Jurisprudence*, Oxford University Press.

FARALLI C. 1982. *Diritto e magia*, in PATTARO E. (ed.), *Contributi al realismo giuridico*, Giuffrè, 11 ff.

FERRARO F., POGGI F. 2014. *A Commitment to Naturalism: Bentham and the Legal Realists,* in TUSSEAU G. (ed.), *The Legal Philosophy and Influence of Jeremy Bentham*, Routledge, 224 ff.

FITTIPALDI E. 2016. *Introduction: Continental Legal Realism*, in PATTARO E., ROVERSI C. (eds.), *A Treatise of Legal Philosophy and General Jurisprudence*, 297 ff.

FRANK J. 1931. *Are Judges Human? Part One: The Effect on Legal Thinking of the Assumption That Judges Behave Like Human Beings*, in «University of Pennsylvania Law Review», 80, 1, 17 ff.

FRANK J. 1932. *What Courts Do in Fact*, in «Illinois Law Review», 26, 761 ff.

FRANK J. 1934. *Experimental Jurisprudence and the New Deal*, Congressional Record (Bound), 78 . 12, 12412 ff .

FRANK J. 1949. *Law and the Modern Mind*, Stevens & Sons.

FRIES M. 1944. *Verklighetsbegreppet enligt Hägerström*, Harrassowitz.

GOLDING M.P. 1986. *Jurisprudence and Legal Philosophy in Twentieth-Century America—Major Themes and Developments*, in «Journal of Legal Education», 36, 1986, 441 ff.

HAAKONSSEN K. 1981. *The Science of a Legislator: The Natural Jurisprudence of David Hume and Adam Smith,* Cambridge University Press.

HÄGERSTRÖM A. 1908. *Das Prinzip der Wissenschaft. Eine logisch-erkenntnistheorische Untersuchung*, SKHVSU Almqvist & Wiksell.

HÄGERSTRÖM A. 1927. *Der römische Obligationsbegriff im Lichte der allgemeinen römischen Rechtsanschauung. Volume 1*, SKHVSU Almqvist & Wiksell.

HÄGERSTRÖM A. 1953. *Inquiries into the Nature of Law and Morals*, Almqvist & Wiksells boktr.

HÄGERSTRÖM A. 1964. *Philosophy and Religion*, George Allen and Unwin.

HALÉVY E. 1901. *La formation du radicalisme philosophique*, Alcan.

HOESCH M. 2018. *From Theory to Practice: Bentham's Reception of Helvétius*, in «Utilitas», 30, 294 ff.

HOLMES O.W. 1952. *Collected Legal Papers*, Peter Smith.

HORWITZ M.J. 1992. *The Transformation of American Law 1870-1960: The Crisis of Legal Orthodoxy*, Oxford University Press.

HUTCHESON J.C., Jr. 1929. *The Judgement Intuitive: The Function of the 'Hunch' in Judicial Decision,* in «Cornell Law Review», 14, 3, 274 ff.

JACOBS S. 1990. *Bentham, Science and the Construction of Jurisprudence*, in «History of European Ideas», 12, 5, 583 ff.

LEITER B. 2007. *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, Oxford University Press.

LLEWELLYN K., ADAMSON HOEBEL E. 1941. *The Cheyenne Way*, University of Oklahoma Press.

LLEWELLYN K. 2008. *Jurisprudence: Realism in Theory and Practice*, with a new introduction by James J. Chriss, Transaction Publishers.

LUNDSTEDT A.V. 1932. *The Responsibility of Legal Science for the Fate of Man and Nations*, in «New York University Law Quarterly Review», 10, 326ff.

LUNDSTEDT A.V. 1956. *Legal Thinking Revised. My Views on Law*, Almqvist & Wiksell.

LYLES M. 2006. A *Call for Scientific Purity: Axel Hägerström's Critique of Legal Science*, Rönnells.

MARC-WOGAU, K. 1972. *Axel Hägerström's Ontology*, in OLSON R.E., PAREL A.M. (eds.), *Contemporary Philosophy in Scandinavia*, Johns Hopkins Press, 479 ff.

MINDUS P. 2009. *A Real Mind*, Springer.

MIXON R.W. JR. 2020. *Bentham, Science and Utility*, in «Revue d'études benthamiennes», 18. Available on: https://journals.openedition.org/etudes-benthamiennes/8127#toc01112.

MORRIS W.E., BROWN C.R. 2022. *David Hume*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition). Available on: https://plato.stanford.edu/archives/sum2022/entries/hume/.

OLIVECRONA K. 1939. *Law as Fact* (1st Ed.), Einar Munksgaard, Humphrey Milford.

OLIVECRONA K. 1971. *Law as Fact* (2nd Ed.), Stevens & Sons.

PAPINEAU D. 2021. *Naturalism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition). Available on: https://plato.stanford.edu/archives/sum2021/entries/naturalism.

PATTARO E. 1974. *Il realismo giuridico scandinavo I, Axel Hägerström*, Cooperativa libraria universitaria editrice.

PATTARO E. 2005. *The Law and the Right: A Treatise of Legal Philosophy and General Jurisprudence. Volume 1*, Springer.

POSTEMA G.J. 2019. *Bentham and the Common Law Tradition* (2nd Ed.), Oxford University Press.

POUND R. 1943. *A Survey of Social Interests*, «Harvard Law Review», 57, 1 ff.

RADIN M. 1931. *Legal Realism*, «Columbia Law Review», 31, 824 ff.

ROSS A. 1951. *Tû-tû*, in BLOM USSING H. (ed.), *Festskrift till professor, dr. Juris, Henry Ussing*, København, 468 ff.

ROSS A. 1958. *On Law and Justice*, Stevens.

ROSS A. 1968. *Directives and Norms*, Routledge & Kegan Paul.

SANDIN R.T. 1959. *Axel Hägerström's Philosophy of Religion with Special Reference to His Theory of Knowledge*, University of Minnesota.

SANDIN R.T. 1966. *The Concept of Reality and the Elimination of Metaphysics*, in «The Monist», 50, 87 ff.

SMITH A. 1978. *Lectures on Jurisprudence*, ed. by R.L. Meek, D.D. Raphael, P.G. Stein, Oxford University Press.

SMITH A. 2002. *The Theory of Moral Sentiments*, ed. by K. Haakonssen, Cambridge University Press.

SPAAK T. 2009. *Naturalism in Scandinavian and American Realism: Similarities and Differences,* in DAHLBERG M. (ed.), *De Lege: Uppsala-Minnesota Colloquium: Law, Culture and Values*, Iustus Förlag, 33 ff.

TARELLO G. 1962. *Il realismo giuridico americano*, Giuffrè.

Legal Reasoning, Particularism: In Defence of a Psychologistic Approach

BRUNO CELANO

1. *Introduction*

In this contribution, its thrust almost entirely programmatic, I will argue that it is worth carefully exploring the reasons for a drastic reorientation of legal theory, and of the theory of norms and norm-based reasoning more generally, by pivoting toward what, introducing a splashy label savouring of the grandiloquent, I will be calling "psychodeontics".

I will thus be arguing in defence of a psychodeontic approach—not in general terms but in connection with two specific areas: the theory of legal reasoning, and in particular of justificatory judicial reasoning (Sec. 2), and the construction of a two-tier theory of law (we will see in due course what that is) predicated on a particularistic conception of practical reasoning (Sec. 3).

These two areas of investigation are closely related. As we will see, some aspects of the particularistic conception of practical reasoning—those having to do with the relation between rules and exceptions—lend themselves quite naturally to being understood and developed in psychodeontic terms.

Some brief closing remarks will follow (Sec. 4) suggesting that, given the present landscape, a psychologistic approach is our main path toward naturalising jurisprudence.

2. *Psychodeontics*[1]

2.1. *The argument in short*

The argument I will set out in this Section 2 can be summarised as follows.

Two notions of reasoning are usually distinguished: a psychological notion and a logical one. This distinction underlies two further distinctions: explanatory reasons (causes) *vs.* justificatory reasons (reasons in the proper sense), and context of discovery *vs.* context of justification.

It is generally assumed that the theory of legal reasoning, and in particular the theory of judicial reasoning, must have as its object (legal) reasoning understood in a logical rather than a psychological sense.

* This chapter is a translation of an article, originally written in Italian, titled *Ragionamento giuridico, particolarismo. In difesa di un approccio psicologistico* (in «Rivista di filosofia del diritto», 2, 2017, 315 ff.). Its translation was a collective effort. The first draft was written using the DeepL translator. It was then revised by G. Rocchè, M. Taroni, M. Ubertone, A. Zambon, and M. Zgur; copyedited by F. Valente; rechecked and edited by M. Brigaglia and G. Sajeva; and finally approved by G. Todaro. We thank G. Nicolaci and J.J. Moreso for useful discussions.

[1] The ideas set out in this section are deeply indebted to two interlocutors, Marco Brigaglia (see esp. BRIGAGLIA 2016) and Giusi Todaro (see esp. TODARO 2011).

This is a *normative* assumption: A *good* theory of legal reasoning *must* have as its object... Why? Sometimes this is because the assumption, whether implicit or explicit, seems to be the only alternative. Other times the assumption seems to be a matter of epistemological etiquette: In the eyes of a true empiricist, the "things" that are lodged in people's minds (and those quotation marks were meant to convey a sense of restrained indignation) cannot be taken as epistemologically respectable objects. For they are not, in fact, *observable entities*. Linguistic phenomena, on the other hand, are things of the external world and are thus empirically observable. If the theory of legal reasoning is to have any epistemological respectability, then, it must be understood as a theory of (legal) reasoning in the logical sense.

That argument, I will contend, is unacceptable. How can it be claimed that phonemes, morphemes, sentences, propositions (or, more generally, meanings), and logical relations are observable entities any more than are mental events, dispositions, states, acts, and processes?

In truth, and paradoxically, what makes linguistic phenomena and logical relations seem at first sight, from an epistemological standpoint, a more respectable object of investigation than things in the mind is the fact that they are *not*—or not entirely—empirically observable entities. Rather, to use Karl Popper's expression, they are entities belonging to World Three.

My point is this: The assumption that a theory of legal reasoning must deal with reasoning in the logical sense stems from the polemic against psychologism that broke out at the beginning of the twentieth century. It is, indirectly, an expression of twentieth-century anti-psychologism, which (especially through Hans Kelsen) left a deep mark on contemporary legal theory.

Now, it is precisely this basic position, the anti-psychologistic position, that now (at least over the past forty-five years) is in the process of being re-examined. According to some, it is completely discredited. According to many others, the dialectic between psychologism and anti-psychologism is far more complex and problematic than it appeared to the twentieth-century critics of psychologism.

This turning point has come about through the development of cognitive psychology and the cognitive sciences in general, and in particular the development of the model of bounded rationality (seventy years) and of social psychology and the psychology of reasoning (some sixty years), as well as through contemporary research on human inference, heuristics, and biases (some forty-five years) and moral psychology (some twenty years), and then, of course, there is the development of the neurosciences—all very much in vogue, to be sure, but I will do no more than point out the reasons why, in my view, all this research needs to be taken very seriously by anyone intent on engaging with legal reasoning, and with law in general.

2.2. *The illusory solidity of language*

Let us begin, then, with the distinction between psychological and logical notions of reasoning. This distinction underlies—or is reflected in—the distinction between explanatory reasons and justificatory reasons ("good" reasons, or reasons proper), and that between the context of discovery and the context of justification. The three distinctions are not perfectly interchangeable, but they are nevertheless closely related. Addressing the first one is a way of addressing the other two.

In the first, or psychological, sense, reasoning is a particular mental process (a collection of mental states, events, dispositions, and acts). More accurately stated, it is a collection of biochemical processes triggered within a cell tissue, presumably the encephalon: Certain neurons are activated, electrical impulses run along the axons of certain nerve cells, and so on. It is literally a process that takes place "in someone's head".

Reasoning in the logical sense, on the other hand, is a sequence of propositions such that, given some propositions, we obtain from them another proposition (the last proposition in the sequence). It is also said that from the initial propositions, or *premises*, a certain *conclusion* is "inferred," or that

the latter is "derived" or "follows" from the premises, or that the premises "imply" the conclusion[2]. I will say that the premises and the conclusion combined constitute an argument.

In an argument, the premises impart a *certain plausibility* to the conclusion: They lend it a *certain support*, a certain backing, or warrant. An argument is a sequence of propositions such that its conclusion can be said to be (more or less) *justified* or *warranted* in light of the premises assumed.

What does "justification" mean? It means that in an argument, the premises specify or are (more or less good) *reasons* in favour of the conclusion. In general, *p* is a *supporting reason* for *q* if, and only if, were we to hold *p* to be true, we would thereby to some extent be justified in also holding *q* to be true (Watanabe Dauer 1989, 91). The inference from *p* to *q* is correct if the truth of *p* would make the truth of *q* to some extent probable.

So we have an inference when something is taken as a more or less solid reason for something else. From a logical point of view, what is of interest is not reasoning as a mental process, its actual occurrence, but rather the examination of its correctness or incorrectness. The task of logic lies in «the justification and criticism of inference» (Quine 1959, 33), that is, the identification of criteria in light of which the correctness of an argument can be assessed. The logical problem *par excellence* is whether the required relation actually exists between the premises and the conclusion, that is, whether the conclusion actually follows from the premises[3].

As noted (Sec. 2.1), it is commonly held that the theory of legal reasoning must take as its object (legal) reasoning understood in the logical, not in the psychological, sense. More specifically, it is held that the object of the theory of judicial reasoning should be the justificatory reasoning of judges, and *not* a description or explanation of the mental processes that, in point of fact, lead the judge to make a certain decision or the judge's audience to take a certain attitude to it. With some authors, this assumption is explicit. It is often taken for granted.

Why should a theory of legal reasoning assume that its exclusive object of investigation is (legal) reasoning in the logical sense?

This assumption seems to respond to a requirement of epistemological etiquette. The underlying idea, sometimes explicit, seems to be as follows: The things that are in people's minds are not, for an empiricist, epistemologically respectable objects. They are not, in fact, entities whose existence and properties are empirically observable. Behaviourism was right: There can be no scientific knowledge about such things. (A behaviourist runs into another behaviourist: "You look just fine to me! And how am *I* doing?") Language, on the other hand, is a thing of the external world: Linguistic phenomena are empirically observable. So if, as empiricists, we want the theory of legal reasoning to have good epistemological credentials, we will have to understand it as a theory of (legal) reasoning in the logical sense.

That argument lacks any plausibility, and that for two reasons.

First, psychology (of non-behaviourist orientation), and in particular the psychology of reasoning, has been well established as an empirical science for decades. And in contemporary psychology the ban on introspection has in principle been lifted. Not only that: It is now common in the cognitive sciences to refer to mental entities, states, and events that are not introspectively accessible. And, finally, whatever the state of play was, until recently, as regards the possibility of empirically establishing the presence of mental states, describing them, and explaining their nature, the development of neuroscience has, of course, changed all that.

[2] There are most likely to also be inferences that are not reasoning, as in the case of perceptual inferences. Here I will use the term "inference" as a synonym for "reasoning" in the logical sense.

[3] As should be evident from the text, though it is worth clarifying the point explicitly, the term "logical" (reasoning "in the logical sense") is not understood here narrowly as a synonym for "deductive" but is broadly taken to mean "discursive" or "pertaining to discourse", where "discourse" is in turn understood as any set of sentences or propositions. As I am using the term, then, the field of reasoning in the *logical* sense is divided into two sets: that of deductive arguments and that of non-deductive arguments (generalisations, predictions, and estimates of probability, abduction, analogy, counterfactual reasoning, practical inferences of various kinds).

Second, the thesis that linguistic phenomena are more amenable to empirical observation than mental phenomena is untenable. Phonemes, morphemes, and sentences (understood as *types*, not as *tokens*) are, it would seem, abstract entities. A sentence is not a collection of sound waves, or lines of ink on a sheet of paper, or pixels on a computer screen. A sentence is their *form*. Acts of uttering sentences are empirical, spatiotemporally identified phenomena, like earthquakes. What these acts produce are sets of sound waves, lines of ink, and so on. But, as just noted, none of these sets of sounds or lines of ink on paper through which sentences are instantiated are *themselves* sentences, or so it would seem; rather, as just noted, sentences are their form, or, to use a metaphor, the way they appear to the mind's eye, their intelligible body[4]. They are abstract entities that nonetheless somehow guide and control the activity of the bodily systems (the brain-phonatory apparatus, and so on) that produce spatiotemporally identified instances of them.

What, in all this, is there of the purported solidity of the so-called "external" or "observable" phenomena (and those are shudder quotes in the sense remarked on in Sec. 2.1) which are taken to be the object of direct sensible experience, the object which the argument at issue points to?

Not to mention propositions, and in general meanings.

It will be granted that both signifiers (and in particular sentences) and meanings (and in particular propositions) are constituent elements of linguistic phenomena. Why should mental states be considered empirically less respectable objects than propositions, or, in general, than meanings? Propositions, and meanings generally, are entities no more observable than are emotions, desires, or beliefs. Once again: It is true that acts of uttering a sentence produce sound waves, ink on paper, and the like. But the propositions expressed by them—or, in general, the propositions that *can* be so expressed, those that are expressible: the assertibles, or "sayables," to use an old-fashioned term—are, it would seem, something else entirely. And the very possibility of identifying an empirically perceptible event as an act of expressing a proposition depends on the possibility of accessing these phantom objects[5].

In short, language is by no means an empirically respectable object—and it isn't so by the very canons of epistemological respectability which the argument under scrutiny itself assumes as presuppositions. So why rely on such an unreliable object to argue in support of the thesis that only reasoning in the logical sense, and not reasoning in the psychological sense, should be the proper subject of a theory of (legal) reasoning?

The reason for this is probably that language gives a perceptible guise—gives body—to entities that partake of none, or almost none, of the empirical respectability to which the argument appeals. These entities, as we just saw, are phonemes, morphemes, sentences, propositions, and in general meanings, and they include the traditionally paradigmatic case of what cannot be empirically observed, namely, logical relations. The hope of finding refuge under the protective wings of the language is, historically speaking, the offspring of a polemic that has sprung up *against* the possibility, or at any rate the value, of an empirical science that takes reasoning as its object. And this brings us to the polemic against psychologism[6].

2.3. *Anti-psychologism*

As discussed in Section 2.2, reasoning in the psychological sense is a collection of biochemical states (events, acts, processes) in the mind/brain. What kind of collection? Presumably a collection of mental states that takes reasoning in the logical sense as its *content*.

[4] This, of course, is only a metaphor. The issue is a thorny one; see REY 2006.

[5] These, of course, are controversial statements. It is not my intention here to advance any particular conception of linguistic phenomena. The point of interest is this: Whatever position is taken on the matter, there is none of the obviousness ("These are observable entities!") which the argument assumes there to be.

[6] On this transition see DUMMETT 1984.

This way of seeing things presupposes the distinction between a mental state (act, process, etc.) and its content. And accordingly, as we will see shortly, the traditional way of drawing the distinction refers to a particular way of understanding the content of mental states.

Philosophical culture in Central Europe at the turn of the twentieth century centres around a complex set of theses, assumptions, and arguments broadly labelled "anti-psychologism". As is well known, the polemic against psychologism brings together the thought of Gottlob Frege, Husserlian phenomenology and its several offshoots, large swaths of neo-Kantianism, and the beginnings of logical positivism.

In its paradigmatic form, the polemic concerns the status of the laws of logic and the foundations of mathematics. The anti-psychologistic thesis *par excellence* is that the laws of logic cannot be identified with, or explained in terms of, the laws that govern our mental processes, for otherwise it would prove impossible to account for their universality and necessity (their specific objectivity, and it is worth noting here that this has little to do with the illusory solidity of "observable entities"). Mathematical entities, their relations, and logical relations, cannot be identified with representations, nor can they be explained in terms of representations or of relations between representations or of mental activities, for otherwise it would prove impossible to account for their specific objectivity. (As can be appreciated, the point is precisely that the phenomena in question *cannot*, it is believed, be taken as the object of an empirical science.)

In short, logical laws are not «laws of thinking» (*Denkgesetzen*), expressing how, in fact, mental acts or states are produced in our consciousness; they are not «psychological laws», expressing «general features of thinking as a mental occurrence [«das Allgemeine im seelischen Geschehen des Denkens]»; logic and mathematics are not concerned «with the mental process of thinking [*den seelischen Vorgang des Denkens*] and with the psychological laws in accordance with which this takes place [*die psychologischen Gesetze, nach denen es geschieht*]» (FREGE 1984, 351)[7].

The anti-psychologistic position, however, is not confined to the field of logic and the foundations of mathematics. Rather, the polemic against psychologism invests all areas of the theory of knowledge and the theory of judgment as a whole. There are two main directions this anti-psychologistic programme takes.

(1) *Epistemology, or the theory of knowledge*. In this area, the anti-psychologistic position comes down to the thesis that it is one thing to describe, or (causally) explain, the mental processes in virtue of which knowledge, or belief, is in fact formed in our consciousness; it is quite another to clarify its foundations, that is, to justify it. It is one thing to ask what beliefs someone or some group of individuals have, and in what way they have in fact come to form these beliefs; it is quite another to ask whether, and on what basis, these beliefs are justified, or true. We must not confuse an account of how knowledge is acquired as a matter of fact with an account of what *makes* it knowledge (justified true belief)—of that in virtue of which it "counts" as knowledge, of what confers value on it as such[8].

The first type of enquiry is a psychological enquiry, that is, it concerns our representations, the causes of their production, and their relations. The second is not. Thus, for example, it is one thing to investigate the history of a certain discipline, a certain body of beliefs; it is quite another to ask whether that body of beliefs satisfies the necessary and sufficient conditions for it to constitute a body of knowledge. Take, for example, the question, How is physics possible as a science? This is not (a physical or) a psychological question: That in virtue of which the complex of methods and

[7] Cf. FREGE 1980, p. VI: «Never let us take a description of the origin of an idea for a definition, or an account of the mental and physical conditions on which we become conscious of a proposition for a proof of it. A proposition may be thought, and again it may be true: let us never confuse these two things».

[8] «With the psychologistic conception of logic we lose the distinction between the grounds that justify a conviction and the causes that actually produce it. This means that a justification in the proper sense is not possible; what we have in its place is an account of how the conviction was arrived at, from which it is to be inferred that everything has been caused by psychological factors» (FREGE 1979, 147).

beliefs called "physics" has value as knowledge is not a set of (physical or) psychological facts.

In short, epistemology—the clarification of why something counts as knowledge, as justified belief—cannot be understood as falling within the purview of empirical science.

Why? Because, as the anti-psychologist would have it, epistemological enquiry is *normative*: It has to do with what we *must* believe, with what makes a belief, or reasoning, or an inference *correct*. And, the anti-psychologist goes on to say, no collection of facts—not even a collection of facts about the actual production of representations in our consciousness, and the relations that in fact exist between them—makes it possible for us to draw normative conclusions. No description or explanation of psychological processes can account for the truth or correctness of our beliefs or inferences: No such investigation is capable of accounting for the specific normativity of thinking, of knowing (epistemic normativity)[9].

In the lexicon of the anti-psychologists of the early twentieth century, when a certain body of beliefs that claims to be a body of knowledge lives up to that claim, it is said to be "valid". What the anti-psychologist argues is that an investigation into the psychological (as well as the physical and social) processes that explain how knowledge is gained or lost is quite different from an investigation into its conditions of validity as knowledge[10].

(2) *Intentionality*. That first (epistemological) thesis finds support in, and in turn lends support to, a second thesis, stating that (some) mental acts, states, and events are intentional.

Some mental acts and states—e.g., beliefs, desires, hypotheses, hopes, and so on—have the property of "being about", or "being directed toward", objects or states of affairs. The *content* of such acts and states needs to be neatly distinguished from the acts and states themselves. The latter are psychological phenomena (factual data pertaining to internal experience); their contents, on the other hand, enjoy a peculiar form of existence: They exist precisely as meaningful contents. This type of existence—an ideal existence, or "intentional inexistence"—is to be distinguished both from the mode of existence proper to mental acts and states themselves (existing as psychological facts) and from the mode of existence proper to physical objects, states, or processes.

In this second connection (under the rubric of intentionality), the anti-psychologistic polemic is directed against the thesis that all that is given to consciousness is our representations, and consequently that knowledge consists exclusively in comparing our representations, manipulating them, and identifying relations between them. The basic idea is simple: When we have a representation—e.g., when we perceive a physical object or think of a non-existent object—what is given to us is not the representation itself but its object. To see a tree is not to see our representation of a tree: What we see is precisely the tree. Similarly, what we think of when we think of a unicorn is not our representation of the unicorn but the unicorn itself. That in virtue of which a representation we have (e.g., our perception of a tree) is about a physical object (e.g., a tree) is a content, an intentional object (what Husserl called a "noema"), which as such cannot be identified with our representation of the physical object (the tree), nor can it be identified with that object itself[11].

[9] In the contemporary controversy over the naturalisation of epistemology, touched off by QUINE's 1969 essay (cf. Sec. 3.4 below), one of the points—and probably the central point—of disagreement between naturalists and anti-naturalists is precisely over the question of whether or not epistemological enquiry is normative (cf. ENGEL 1996, chs. 1 and 5, and 1998, 375 f.). Also anti-psychologistic are the Wittgensteinian thesis that meaning and understanding cannot be considered as «species of mental acts» (BELL 1992, 402) and the thesis of the normativity of meaning (HALE 1997). And the same goes for the thesis that «the very possession of concepts is a normative matter», and cannot be equated with having a «discriminative ability»: «To have a concept one has to have the idea that one is *justified* in making the relevant discriminations, and such talk of justification is of a piece with talk of rationality and intelligibility—it is a matter of being guided by rules in a fully normative sense» (GUTTENPLAN 1994, 451; italics in the original).

[10] The distinction between an inquiry into the psychological genesis of knowledge (*quid facti*?) and an inquiry into its conditions of validity (*quid juris*?) is clearly of Kantian parentage and is one of the pillars of nineteenth- and twentieth-century neo-Kantianism.

[11] The same argumentative strategy underpins the conception of "propositions" as «non-mental entities expressed

In short, the anti-psychologistic notion of intentionality issues from an attempt to give a non-naturalistic, non-psychological account of the idea that certain mental acts or states are endowed with content. This is the notion of content (the content of mental states) on which rests the traditional way of drawing the distinction between reasoning in the logical sense and reasoning in the psychological sense: Reasoning in the psychological sense is a collection of mental states that takes reasoning in the logical sense as its intentional object (an object endowed with ideal existence, validity, and the like).

2.4. *Anti-psychologism and legal theory: The role of Kelsen*[12]

These different and complementary parts of the polemic against psychologism have exerted a profound influence on contemporary legal theory through Kelsen's theory of law. This is hardly surprising, given that the polemic against psychologism is one of the salient features of the cultural milieu in which Kelsen's philosophical *Bildung* takes place.

According to Kelsen, law is a norm. A norm is a content, or a meaning (*Sinngehalt*, in Kelsen's terms), in the sense specified above:[13] It is the content or meaning of mental acts or states intentionally directed toward certain objects or states of affairs (toward the behaviour of others). As content or meaning, law is neither a psychological phenomenon nor in general a physical phenomenon but something ideal, to be investigated in its specific existence, which Kelsen—no surprise—calls "validity"[14]. In the pure theory of law, the anti-psychologistic thesis that no set of mental or physical facts can account for the truth, or correctness, of our beliefs or inferences is translated into the thesis that what is of interest from the standpoint of a scientific treatment of law is not the law's efficacy but its validity (KELSEN 1945, 30). In the pure theory of law, in other words, there exists between the scientific knowledge of law as such (legal science in the strict sense and the theory of law), on the one hand, and a sociological investigation of the behaviour or mental states determined by law, on the other hand, the same relation that, in the overall framework of the anti-psychologistic polemic, can be found to exist between logic, mathematics, and epistemology, on the one hand, and the sociology or psychology of cognitive processes, on the other.

This is not only a question of historical, or philological, order. The thesis that law, as a norm, is a content or meaning, is key to a particular aspect of the pure theory of law that itself is crucial to that theory: the idea that law is something impersonal, anonymous, "depsychologised", and that therein lies its specific "authority" (KELSEN 1945, 36)[15]. Let me explain.

The pure theory of law offers a particular version of an ancient, and very influential, image of law. According to this image, law enjoys a relative independence, or autonomy (both conceptual and normative), from the preferences, intentions, will, decisions, beliefs—whether actual or possible—of those subject to it. And therein lies its objectivity.

by sentences and forming the objects of propositional attitudes», a conception endorsed, e.g., by Bertrand Russell and George E. Moore. A belief in the existence of such entities—as objects that are neither physical nor psychological—was shared by Bernard Bolzano (*Sätze an sich*), Gottlob Frege (*der Gedanke*), Franz Brentano, Alexius Meinong, and Edmund Husserl (DUMMETT 1991, 250).

[12] In what follows, the basic theses of Kelsen's pure theory of law will be reconstructed as they are in CELANO 1999.

[13] *Translators' note.* In this paper, Celano employs the Italian term "contenuto", literally translated as "content", in the sense specified in the previous section, understood as the (non-psychological) "intentional objects" that mental acts are allegedly directed towards. In talking about Kelsen's views, he shifts to the slightly different expression "contenuto di senso", a literal translation of Kelsen's German expression "Sinngehalt". The latter is usually translated into English as "meaning". In order to make this connection immediately available to the English reader, we have decided to translate "contenuto di senso", on all occurrences, through the phrase "content or meaning".

[14] Kelsen explicitly commits to the anti-psychologistic stance in a couple of key passages in his work. See KELSEN 1960, ix; KELSEN 1966, vii.

[15] For a reading of this Kelsenian passage, see CELANO 1999, secs. 4.3.7, 5.2.3, and 5.2.4.

On this view, law cannot be reduced to any collection of more or less arbitrary preferences, intentions, decisions, or beliefs (or to any collection of actions explained by them). Rather, law is something impersonal, anonymous: It is a set of norms that do not exist as physical or psychological phenomena but are "valid" as such (and yet are not moral truths, either).

From this point of view, the pure theory of law can be understood as the outcome of a particular theoretical operation that consists in transplanting the polemic against psychologism into the terrain of the theory of law, with a view to providing a satisfactory account of the specific objectivity of law. This operation, in other words, consists in using the anti-psychologistic argumentative strategy to claim that law, as such, is independent of the sphere of natural phenomena. How can law, as such, be "valid" independently of human preferences, intentions, decisions, and beliefs? How is it that law cannot be reduced to a set of volitions and beliefs, which are inherently more or less arbitrary? That's simple, Kelsen replies: Law is a norm; a norm is a content or meaning, and—here is the anti-psychologistic thesis—a content or meaning cannot be reduced to, nor can its objectivity be explained in terms of, the mental acts or states (preferences, intentions, volitions, decisions, beliefs) it is the content of.

Stated otherwise, the polemic against psychologism (and we saw this in Sec. 2.3) teaches that the contents of mental acts or states have a particular kind of existence (an ideal existence, an intentional inexistence); unlike psychological phenomena, they enjoy a peculiar objectivity (the objectivity of "thought", or "noema"); and they have an identity independent of human attitudes and beliefs. Law—here is the Kelsenian theoretical operation—is itself a content or meaning. Therefore, just like the laws of logic or of mathematical entities, law is itself something impersonal and anonymous, an entity that is (neither physical nor psychological but) ideal: It is something objective, or "valid", independently of our actual mental processes (whether volitional or cognitive)[16].

That, it seems to me, is the background against which we get the thesis that the object of the theory of legal reasoning must be reasoning understood in a logical, not a psychological, sense. Those who, since Kelsen, have thought they could find in the analysis of language—the language of law—the key to an empirically respectable theory of law have likewise fallen victim to the illusion of the solidity of linguistic entities (Sec. 2.2 above), the trap into which analytic philosophy in general has fallen. In this respect, too, there is a perfect parallelism between the course followed by legal theory and that followed by general philosophy.

In the mid-twentieth century, philosophers came to believe that in language they had an object—a field of phenomena—that gives a perceptible guise, and with it credentials of epistemological respectability, to objects (the laws of logic, and intentional objects in general) which anti-psychologism had carefully distinguished from mental (or generally physical) phenomena. In this way, philosophy of language gained the status of queen of the philosophical disciplines. It would finally no longer have been necessary to engage in disquisitions on the ontological status of "thoughts" or "noemata". It would have been sufficient to carefully examine an object of the external world, namely, language itself. Likewise, dealing with legal

[16] Of course, this theoretical operation raises a swarm of difficulties. Whatever we may think of the laws of logic, of the ontological status of mathematical entities, or in general of the notion of epistemic normativity, we are still left to contend with the idea that law—*positive* law, mind you—can enjoy a form of independence and objectivity comparable to the independence and objectivity the early twentieth-century anti-psychologistic scholars ascribed to logical laws or intentional objects, and law can claim this independent and objective status despite its limitlessly mutable content (a feature of law that Kelsen repeatedly underscored). This is an idea that appears to lack any plausibility. In fact, it is difficult to escape the impression that Kelsen's theoretical operation is exposed to the risk of a twofold error. The first is an unfortunate and consequential confusion between the normativity of law, on the one hand, and the normativity of the laws of logic or of intentional objects, on the other (failing to appreciate the difference between the normativity of law and epistemic normativity). The second is yet another hypostatisation: the transubstantiation of social acts and facts into entities independent of them.

reasoning would have meant, not trying to get into the heads of judges or other legal officials or professionals, but observing and describing observable entities: their discourse.

2.5. *The return of psychologism*

The problem with the language-analysis project as just outlined is that (as we saw in Sec. 2.2) language and discourse are by no means "observable entities", empirically respectable phenomena: By the very canons of this strange and unstable form of empiricism, they are certainly no more so than are mental acts and processes.

Since the 1980s, as we know, the primacy of the philosophy of language has been eroding. The title of queen of philosophical disciplines has passed to the philosophy of mind. We have thus transitioned from linguistic entities to mental phenomena, effectively reversing the course outlined in the previous sections.

But that is not the crux of the matter. The central point is rather that the theses and arguments championed by anti-psychologism have been called into question. The anti-psychologistic consensus has broken down.

Among philosophers, the decisive role in so turning the tide was played by W.V.O. Quine. Working in open and direct antithesis to anti-psychologism, Quine put forward a project to "naturalise" epistemology: Epistemological enquiry is to be understood as «contained in natural science» and specifically «as a chapter of psychology» (QUINE 1969, 83)[17].

In the time since Quine set the agenda, "naturalisation" has become a motto for many philosophers, not only in epistemology but also in the widest range of fields, from the philosophy of mind to metaethics. Naturalism, and with it the rejection of anti-psychologistic theses and arguments[18], is a distinctive feature of much of the contemporary philosophical landscape (see, in general, ENGEL 1996).

The return of psychologism took place not only, as just discussed, along the first of the two paths of development of anti-psychologism, namely, epistemology. It also took place along the second path: intentionality.

The argument, broadly stated, is this. It is a mysterious notion of intentionality which early twentieth-century anti-psychologism relies on (Sec. 2.3 above). Indeed, if the human mind has the capacity to "direct itself toward" objects, if certain mental states or processes have the property of "being about" something (i.e., having content), then it must be the case that we can explain this capacity in terms of natural facts and processes. It must be possible to understand and explain intentionality as a psychological phenomenon.

We have thus come back to the distinction between reasoning in the logical sense and reasoning in the psychological sense. As we have seen (Sec. 2.3 above), the notion of content (the content of mental states) on which rests the traditional way of drawing the distinction is linked to the anti-psychologistic notion of intentionality. As we will see in the next section, the naturalisation of intentionality—the development of a psychological notion of content—cannot but affect the distinction between logical and psychological notions of reasoning.

But today (some fifty years on), we can see that the turning point—the return of psychologism—did not invest the philosophical landscape. Rather, the main outcome is this: In the dialectic between psychologism and anti-psychologism, something new has entered the stage and is playing a decisive role, namely, empirical investigations—investigations in cognitive psychology and, in general, in cognitive science and neuroscience. This, in short, is not an

[17] In setting out his *Epistemology Naturalized*, Quine describes the enterprise as «a surrender of the epistemological burden to psychology» (QUINE 1969, 75). Pascal ENGEL (1998, 391) notes that Quine initially intended to subtitle his essay *Or, the Case for Psychologism* (see also JACQUETTE 2003).

[18] ENGEL (1998, 376) appropriately notes that many contemporary anti-naturalists «believe that the recent naturalistic turn is but a reopening of [the] Pandora's box of psychologism».

internal dispute within philosophy departments. Reasoning has returned to being what it was for the Greeks, prior to the codification of logic, namely, one of the objects of experience.

2.6. *The elusive distinction between reasoning in the logical sense and reasoning in the psychological sense*

How, then, in light of the foregoing considerations, are we to understand the relation between the two notions of reasoning, the logical and the psychological?

From the logical point of view, as we know (Sec. 2.2), the interest is not in reasoning as a mental process but in the examination of its correctness (whether a conclusion really follows from its premises). An argument is «any group of propositions of which one is claimed to follow from the others» (COPI et al. 2014, 6).

Under what conditions can this conjecture be considered justified?

Let us leave aside deductive reasoning, i.e., the relation of logical consequence in the strict sense (where the conclusion *necessarily* follows from the premises). This, along with mathematical relations, is the most resistant case, apparently refractory to treatment from a psychological angle. (As we saw in Sec. 2.3, deduction, or logical consequence, and mathematical relations have been the privileged, though by no means exclusive, domain of twentieth-century anti-psychologism.) Let us instead look at the domain of *non-deductive* inference.

This delimitation of the field of enquiry is justified here by a thesis I am introducing as a postulate (an unproven assumption), which is that in the discursive practice of law, the decisive role is played by non-deductive inferences[19], as well as by practical inferences (that is, inferences relating to the means suitable for achieving given ends or for balancing competing ends). In light of this assumption, we can exclude deductive arguments from our field of enquiry without making this exclusion contrived or gratuitous.

There is, however, a second reason that may justify this narrowing of our field of enquiry.

For decades now, there has been experimental research empirically investigating the way in which human beings actually make inferences, draw conclusions from premises, and make decisions. This line of research has shown that, in reasoning, flesh-and-blood human beings follow heuristics, and that, in decision-making, they do not go in search of *optimal* options (as the standard theory of rational decision-making would have it) but settle for *satisfactory* ones. We are not calculating all the logical consequences of our assumptions or (when making practical inferences) the expected utility of all the possible consequences of all the alternative options in search of an optimum: We take shortcuts; we look for satisfactory options. Thus, for example, probability estimates are based on heuristics (such as representativeness and availability)[20].

The use of heuristics brings with it systematic errors, or *biases*. In forming and evaluating hypotheses, for example, we fall prey to confirmation bias (looking for or overestimating evidence in support of a thesis we already support), and in evaluating options, we fall subject to the framing effect (the same option appears more or less valuable to us depending on how it is framed: A glass half empty seems less valuable than a glass half full).

The human mind, in short, is not, from top to bottom, a computer. It is not so in a very specific sense of that term: The point is not that the human mind is not a machine (the human brain arguably *is* a machine), or that it does not make mistakes (machines sometimes do make mistakes: Think of a system crash in a computer). The idea, rather, is that the human mind is not—not only, and not primarily—a calculator of logical consequences or of the choice that

[19] This assumption is, of course, open to challenge. Here, I will confine myself to taking it as a premise in my argument. That said, its plausibility seems unquestionable to me. For a very clear, solid, and in my view compelling argument to that effect, see DICIOTTI 2007.

[20] These are just a few insights sampled from a very rich and structured body of findings. See, e.g., SIMON 1983, ch. 1; NISBETT & ROSS 1980; KAHNEMAN 2011; GIGERENZER et al. 1999.

maximises expected utility. Human rationality, in general, cannot be represented as a programme for executing deductive inferences or maximising expected utility[21].

We are not, therefore, talking about the undeniable fact that human beings are often confused, uncertain, lost, error-prone, victim to passions, and so on. The thesis, rather, is that the traits just outlined—our recourse to heuristics, and the systematic errors their use leads to, as well as the fact that we are not maximising utility or looking for satisfactory options—are traits proper to human *rationality* (GIGERENZER 2008).

And it is here that the distinction between the two notions of reasoning breaks down.

Under what conditions can a non-deductive, or practical, inference be said to be correct? The question is general, and for a satisfactory answer we should have to distinguish different types of non-deductive argument, or practical inference. But the main point, where we are concerned, is simple: The criteria for the goodness or correctness (or the greater or lesser stringency or plausibility) of non-deductive inferences[22], and of practical inferences, are not independent of the way our mind actually makes those inferences—how could they be? These criteria of correctness are themselves an object of empirical, psychological, investigation[23].

This applies in particular to analogy and concept formation (processes of categorisation), abduction, counterfactual reasoning, and practical inference of various kinds[24]. It is clear that these forms of inference play a central role in legal reasoning, and in particular in judicial reasoning[25].

In what sense, though, are the criteria for the correctness, or plausibility, of inferences of this kind subject to empirical, psychological investigation? (Is this not a blatant *non sequitur*, a blatant violation of the imperative to be careful to distinguish rules from regularities?) The answer is: That certain inferences are more or less good, more or less stringent, is nothing more than a psychological fact—they *are* stringent, good or correct, because they *appear* that way to us. That they are so simply means that they appear that way to us. (That the conclusion follows from the premises is a certain *feeling*, a felt quality.)

In light of these considerations, the distinction between reasoning in the logical sense and reasoning in the psychological sense does not appear so clear-cut. In Section 2.2, this relation was described as follows: Reasoning in the psychological sense is a set of biochemical processes in the brain, whose content is reasoning in the logical sense. But this picture now proves to be inadequate.

[21] See what was said in the previous footnote.

[22] This may even apply to deductive inferences. For an introductory discussion, see JACQUETTE 2003.

[23] It is not that psychologists have an attitude of saying, "We will now explain to you what criteria you should use for determining the correctness of this or that inference". Rather, they mostly confine themselves to investigating the way in which we *in fact* reason. Indeed, the point is precisely that the criteria for the correctness of a certain type of inference cannot be independent of the way the human mind in fact makes that type of inference. As soon as reasoning becomes an object of empirical investigation, the range of inferential possibilities widens out of all proportion. In the maze of these inferential possibilities, how else would it be possible to find one's way around if not by asking oneself how, in fact, the human mind functions? (This aspect of the psychology of reasoning emerges clearly, for example, in the work of Gerd Gigerenzer. See, e.g., GIGERENZER 2008.)

[24] For an introduction, though one that does not address the case of analogy, see GIROTTO 2013, chs. 1, 3, 4, and 5.

[25] In particular, abduction is one of the paths leading to the formulation of unexpressed principles; the construction of intentions that can counterfactually be attributed to the legislator requires, trivially enough, counterfactual reasoning; practical inferences are suited to the function of legislatures and of administrative agencies, as well as to judicial review of the decisions they make. It is a platitude to point out the role that analogy and categorisation play in legal reasoning. Following a widespread practice, even if its justification is uncertain, I will only be concerned here with inferences that seek to answer questions of law (*quaestio iuris*). If, on the other hand, we should also concern ourselves with inferences that seek to give answers to questions of fact (*quaestio facti*), or to problems of proof, we widen the field on which psychological, and generally empirical, investigations have a bearing. Also relevant, in addition to the forms of inference listed in the text, are generalisations, predictions, and probability estimates. These forms of inference, too, are explored in contemporary cognitive psychology: see GIROTTO 2013, chs. 1 and 3.

The content of a set of mental states cannot be understood, in the manner of twentieth-century anti-psychologism, as an intentional object endowed with ideal existence, or as in itself valid independently of the nature of the mind (Sec. 2.3 above). What counts as an argument in the logical sense depends on mental, psychological facts, that is, on the nature of the biochemical processes taking place in the brain—or at least that is the case with non-deductive inferences.

So, then, reasoning in the psychological sense is a collection of mental processes whose content is reasoning in the logical sense, that is, a sequence of propositions whose structure, or form, is determined by psychological facts (events, processes, states, regularities). The two notions of reasoning are like the two sides of a Möbius loop[26].

3. *Rules, exceptions, normality*

3.1. *Particularism*

In a series of writings over the last fifteen years, I have proposed and defended a *particularist* conception of practical reasoning, and specifically of rule-based reasoning (CELANO 2002, 2005, 2006a, 2012, 2016). Before laying out the project for a two-tier theory of law, however, I should clarify what I mean by "particularism".

By this term I mean a conception of practical reasoning revolving around the following thesis (as is customary, I will express myself in terms of "reasons", even if the same thesis can be formulated in terms of "norms", and indeed this latter idiom is the one I will be using in the next section):

(P) Reasons for action are plural[27]. In each case, several reasons apply, most often in conflict with each other[28]. In the event of a conflict, the verdict—that is, the answer (the *right* answer: we are dealing with a

[26] John SEARLE (1983, 1992, 1995) argues that intentionality functions only against a Background of non-intentional abilities, dispositions, and preconditions. I will not elaborate on this thesis or on the arguments Searle advances in support of it. I will merely point out—though this, too, is an idea that would require a separate discussion—that the Background is the natural setting for psychological structures that at the same time play an explanatory and justificatory role. I have elsewhere argued (CELANO 2014) that arguments put forward by very different authors (Pierre Bourdieu, Michel Foucault, David Lewis, Nelson Goodman, and Ludwig Wittgenstein, as well as Searle himself) can be reconstructed as being in agreement in lending some kind of support to the conclusion that in the Background we find such things as psychological phenomena that assume a justificatory role by acting as reasons (and, inextricably, as causes). A clarification, however, is in order. The considerations in this Section 2 take issue with the assumption that the theory of justificatory judicial reasoning must exclusively be concerned with reasoning in the logical sense, and not also with psychological reasoning. But they take issue with this assumption only insofar as the assumption is understood to be justified on the basis of traditional anti-psychologistic arguments or of their revived embodiment in terms of language analysis (Secs. 2.2 and 2.3). In fact, there are excellent ethical-political arguments, ultimately grounded in the rule-of-law ideal, in support of the view that the privileged or exclusive object of consideration should be judges' express statements, even better if set down in writing: Judicial decisions must be reasoned (if, for example, the law is to respect human dignity); the reasoning must be public and clearly identifiable, fixed once and for all, so that it can be subject to scrutiny and evaluation (by the public or by higher courts); the best way to pursue these objectives is to assume that the statements—better yet, the written statements—of judges coincide with their reasoning; and so on. The considerations put forward in this contribution in no way question such ethical-political arguments.

[27] Practical reasons so understood are often incommensurable or indeterminate, but I will not get into that aspect of them here.

[28] I will say that a reason "applies" to a case, be it an individual case or a general one, if the case has the property that, by hypothesis, is a reason for doing or not doing the relevant action. For example, if we assume that the fact that an action is kind is a reason for doing it, the fact that *this* action, or a certain *type* of action, is kind is a reason for doing the action. (The distinction between an individual and a generic case is taken from ALCHOURRÓN & BULYGIN 1971.)

> normative thesis) to the question as to what we have most reasons to do, all things considered, or as to the correctness or otherwise, all things considered, of the conduct being judged—depends on a balancing, or weighing, of the reasons for and against the conduct at issue (there is no preestablished ranking among reasons)[29]. Reasons for action are, hence, *pro tanto* reasons—liable, in each case, to being "defeated" and "overridden" by other reasons, and from time to time subject to balancing. Thus, given a case to which a specific reason or a certain set of reasons applies, which would justify a specific verdict, V1, one cannot exclude in advance the possibility that other reasons also apply to that case, and that the balancing of all the relevant reasons should yield a different and incompatible verdict[30].

Particularism so defined is a conception of the form of practical reasoning, which can be applied both in the moral and in the legal domain. It is not necessarily the case that a particularist in ethics must also be a particularist in legal reasoning, or specifically in the justification of judicial decisions (CELANO 2016). For the sake of simplicity, however, I will not hereafter distinguish between these two domains, even if a comprehensive discussion should.

In Section 2, a position was taken against the anti-psychologistic assumption that the object of the theory of legal reasoning must be reasoning understood in a logical rather than a psychological sense, arguing that this assumption must be called into question and re-examined in light of contemporary empirical investigations, especially psychological ones, into how human beings actually make inferences of various kinds and make decisions. That position—a psychologistic approach to reasoning and decision-making—may at first glance appear unrelated to the position just outlined in this section in favour of particularism. But that appearance is deceptive. Alternatively, some aspects of the particularist conception of practical reasoning naturally lend themselves to a psychological reconstruction, as we will now see.

3.2. *Two-tier theories of law*

Many contemporary legal theorists are in search of a sort of holy grail: a two-tier theory of law (or of practical reasoning in general), hinging on the distinction between first-level norms, or reasons, often referred to as "principles", always subject to balancing, and norms, or reasons, justified on the basis of those principles[31]. At this second level we have "rules": protected reasons for action (RAZ 1979, ch. 1.), entrenched prescriptive generalisations (SCHAUER 1991).

For the purposes of my argument, there is no need—and indeed it is inadvisable—to get mired into the shifting sands of the distinction between rules and principles (debating whether they have different logical forms, whether they differ only in degree, and so on). For our purposes, what matters is only that the proponents of the project for a two-tier theory (I will call them "friends of balancing" or FBs) distinguish two types of norms—let us call them NA and NB—and believe that when two or more NAs conflict, the conflict is solved by balancing the two NAs and formulating, on that basis, an NB[32].

[29] I will not attempt—nor can I attempt—to translate or define the generic and metaphorical notion of *balancing* in terms of a decision-making procedure. For our purposes, it will suffice to rely on an intuitive understanding of that notion, or even on the bare image of a scale: an instrument for measuring the comparative "weight" of the reasons applicable to the case at hand.

[30] This statement of the particularist position differs quite markedly from the currently most representative and influential one developed by Jonathan DANCY (2004). There is no need to discuss this complication here (for an in-depth discussion see CELANO 2005, ch. II), but I should point out straightaway that particularism so understood and defined should not be mistaken for a mysticism of the individual case, of the "concrete" case. It is in virtue of the *properties* of individual cases that one or another set of reasons applies to them. Balancing and its verdict are therefore always concerned with *general* cases (and it is of course through these general cases that individual ones are made to fall under them).

[31] This, of course, does not exclude the possibility that the law of a particular legal system might also include rules that are *not* justified.

[32] By "norm" I mean a conditional that establishes a relation between a general case, acting as an antecedent, and a

For FBs, then, NAs are "unstable" in the following sense: They are norms that do not directly dictate a verdict for the cases to which they are applicable. When a given NA—NA1—applies to a case, it is not certain that the correct normative solution for that case—the correct verdict, all things considered—is the normative solution given by NA1. The balancing could yield a different outcome.

It will be recognised that this interplay is the particularist position as defined Section 3.1 above. The behaviour of NAs is particularist.

According to FBs, then, when a case comes up in which the relevant NAs conflict (these are the NAs applicable to the case, the NAs such that the case falls under their antecedent)[33], they do not *directly* dictate a verdict—how could they, considering that we are dealing precisely with different, incompatible verdicts?—but do so only through the NB that, by hypothesis, constitutes the result of their balancing. The balancing leads to the formulation of a rule, one under which the conflicting principles applicable to the case at hand are ranked, which in turn prescribes a certain verdict—a judgment as to what is to be done, all things considered, or as to whether or not, all things considered, the conduct in question is correct.

To many contemporary legal theorists a two-tier theory appears necessary in order to account for the nature and role of the legal systems of present-day constitutional states governed by the rule of law, marked by a "double level of legality": The acts of public bodies are subordinate to the law, and the law is in turn subordinate to constitutional norms (FERRAJOLI 1993). The basic idea is as follows. First-level norms (principles) provide a justification for the entire system. They are, however, unstable: In each case their application requires balancing them. The role that second-level norms (rules) play in justification should be to spare the decision-maker the burden of balancing the applicable principles on a case-by-case basis:[34] In a case that falls under the rule's antecedent, the correct normative solution will generally be the one indicated by the rule.

Only *generally*, however. The crucial problem for those who share this project is whether it is possible, and if so how, to draw the distinction between the two levels in such a way that the rules really play, in justification, an independent role from that played by principles—in such a way, that is, that the second level does not collapse into the first.

The difficulty is this: Rules, as noted, are meant relieve us of the burden of balancing principles with each new case. We cannot, however, discount the possibility of there being cases which fall under the antecedent of a rule, NB1, but in which the balancing of principles applicable to the case nonetheless points to a normative solution other than the one indicated by NB1, such that to follow NB1 would be to make a mistake. If we want to avoid such errors—if we do not want to fall victim to a kind of "rule fetishism"—we need to leave open the possibility that the rule can sometimes be reconsidered, the possibility that we must sometimes ask whether, even if the case

"normative solution" as a consequent (the idea of a normative solution, roughly understood as the deontic qualification of a given type of behaviour, is taken from ALCHOURRÓN & BULYGIN 1971). I will say that a norm is "applicable" to a case, be it an individual case or a general one, when the case falls under the norm's antecedent (this, to be clear, is the notion of internal applicability; MORESO & NAVARRO 1996). I will say that a norm has been "applied" when, given a case to which it is applicable, the decision-maker in fact adopts—either by forming an intention or by formulating a judgment (I will say, in general, that the decision-maker has thus issued a "verdict")—the normative solution envisaged by the norm (I will also say that, in this case, the decision-maker "applies the normative solution" in question).

[33] If principles are construed not, as many think, as norms with an "open-ended," relatively undefined antecedent (but even then, we will have to be able to distinguish cases that fall under the antecedent from ones that do not), but as norms lacking any antecedent at all, they can still, at any rate, be said to be norms with a tautological antecedent, an antecedent that is always fulfilled, in that the norm is applicable whenever the consequent is applicable (ATIENZA 2014, sec. 1).

[34] This does not, of course, exclude the possibility of rules also being assigned a further role. Thus, for example, they can be used as instruments for allocating and separating decision-making power (SCHAUER 1991). The one specified in the text, however, is the primary role that FBs assign to rules in the justification of decisions. If rules cannot play this role, they cannot play any independent justificatory role at all.

falls under its antecedent, the rule is to be set aside, and the correct normative solution is the one indicated by the balancing of the principles applicable to the case[35].

But this is precisely where the puzzle lies: How can we determine whether or not, in a given case, the rule is to be reconsidered unless we do reconsider it? If in each of the cases falling under the rule's antecedent, we have to determine whether or not the rule is to be applied, and if in order to do so we have to look at the balance of applicable principles (and, in the event of discrepancy, stick to the latter), then the rules do not play the role they are supposed to play. In justification, all the work is done by principles. As compared against principles, rules are completely transparent: In order to determine whether or not a rule is to be applied, we need to look at the balance of principles (and, in the event of discrepancy, follow the latter). Rules, in short, are superfluous relative to principles. In each case, consideration of the principles and their balancing are both necessary and sufficient for the purposes of justification. The second level collapses into the first[36].

NBs, in other words, are likewise unstable. When a given NB—NB1—is applicable to a case, it is not certain that the correct normative solution to that case (the right verdict on it) is the normative solution indicated by NB1. Why? Because it may be that the case also presents other normatively relevant properties (meaning properties apart from those by virtue of which the case falls under the antecedent of NB1)—it may be, that is, that other norms are also applicable to the case at hand—and that, on balance, the correct normative solution is not the one indicated by NB1. Even the behaviour of NB, it would seem, is particularistic.

3.3. *A conjecture*

Is it possible, then, to draw the distinction between the two levels in such a way as to prevent the second level from collapsing into the first, and at the same time to leave open the possibility that the rules may, in some cases, have to be reconsidered?

[35] I say that a rule is "reconsidered" if, and only if, the decision-maker makes its application in a given case (a case to which the rule is applicable) conditional on the answer to the question whether, in that case, the balancing of the applicable principles indicates the normative solution indicated by the rule; if it does not, the decision-maker will have to apply the normative solution indicated by the balancing of these principles.

[36] In CELANO 2006b, I raised this objection (though in a different statement of it) against the theory set out in ATIENZA and RUIZ MANERO 1996 and 2000. Atienza and Ruiz Manero replied as follows: A distinction needs to be drawn between substantive principles and institutional principles, or principles «relating to the following of rules» (concerned with stability, the predictability of decisions, and the allocation and limitation of decision-making power), and «rules are defeated in cases in which, in a proper balance, principles in favour of deviating from the rule carry more weight than ones in favour of following the rule, and only in such cases» (ATIENZA & RUIZ MANERO 2009, sec. 2, my translation, where the Castilian original reads as follows: «las reglas son derrotadas en aquellos casos, pero sólo en aquellos casos, en los que el balance entre los principios que sustentan el apartarse de la regla tiene un peso mayor que el de los principios vinculados al seguimiento de reglas»). (The same reply can be found in RUIZ MANERO (2013), which also distinguishes between the theory that rules are endowed with «*absolute* stability», a theory the author regards a chimerical, and the thesis, which he defends, that their stability is only relative. But the problem, as we will now see, is precisely how to prevent this alleged relative stability from degenerating into absolute instability.) I do not quite understand this reply. I willingly concede that there are first-level norms of two kinds: substantive and institutional. But the second level collapses equally into the first. To determine whether a rule is to be followed we must still, in each case, look at the balance of principles, whether substantive or institutional, that apply to that case. (How can we determine whether or not, in a given case, *the weight of the principles justifying a decision to depart from a rule* is greater than *the weight of applicable institutional principles* if we do not weigh those two sets of principles, that is, if we do not engage in balancing?) The NB remains transparent relative to the NAs. What we have done is only to broaden the list of NAs to be taken into account—to be weighed—for the purpose of reconsidering the rule. ATIENZA and RUIZ MANERO (2009) follow Schauer in distinguishing between cases in which a cursory glance is sufficient to realise that the rule is to be applied, and those in which it is instead necessary to look closely at the balancing of the applicable (substantive) principles in order to establish which normative solution is correct. But that does not solve the problem: How can we distinguish between cases where a cursory glance is sufficient and cases where it is necessary to look closely unless we look closely? (Elsewhere ATIENZA (2014, sec. 2; 2017, sec. 8) outlines a more complex picture that appears to allow one to escape this objection. But I have argued elsewhere (CELANO 2017) that this more complex picture is open to the same objection.)

I will now suggest a way in which this problem can perhaps be solved. I will proceed in two steps. I will first introduce as presuppositions five theses I have defended elsewhere (CELANO 2012, 2016). On the basis of these presuppositions, I will then formulate a conjecture.

First presupposition (in what follows, throughout this section, the term "case" will always be used in the sense referring to a *general* case). The *reasonable* use of rules—their *reasonable* use: this is a normative thesis—, rules that are adequately justified, is based on the possibility of distinguishing between normal and non-normal cases.

Normal cases are those in which it is certainly reasonable to follow the rule (and let me reiterate here that the rules we are talking about are ones understood to be adequately justified). Non-normal cases, on the other hand, are those where it is reasonable to ask whether the course of action indicated by the rule is indeed, under the given circumstances, the course of action to be followed, the right thing to do.

Thus, rules—as such, that is, as protected reasons for action, entrenched prescriptive generalisations (Sec. 3.2 above)—apply to normal cases. In non-normal cases, decision-makers, if reasonable, will not follow the rule. Rather, they will reconsider it: They will ask whether, in a case of that type, the course of action indicated by the rule is the right one, and they will try to answer this question, as best they can, by examining the relevant reasons. (The reasons the rule normally excludes are not, in this case, excluded: The generalisation gets dislodged and is thus no longer entrenched.)

Second presupposition. Normal cases cannot be identified by any exhaustive enumeration of properties. That is, the list of the normatively relevant properties in virtue of which a case is normal is indefinite.

Third presupposition. Correlatively, non-normal cases—and consequently exceptions (meaning *true* exceptions, as per the fourth presupposition below)—cannot be specified in advance except by vacuous clauses ("unless there are decisive reasons to the contrary", "save for adequately justified exceptions", and the like). These cases, too, cannot be specified by any exhaustive enumeration of properties.

Fourth presupposition. In order for a case to properly count as an exception to a rule R, and so as a *true* exception to the rule, it must meet two conditions: (a) the case falls under the antecedent of R, and (b) it would not, in that case, be reasonable to adopt the normative solution provided for by R. True exceptions, in other words, are non-normal cases (cases where it is reasonable to reconsider the rule) in which the course of action indicated by the rule is not the right one: Reconsidering the rule leads the decision-maker (one who is reasonable and adequately informed) to the conclusion that, under the given circumstances, it is reasonable to act in a way other than as the rule prescribes.

Fifth presupposition. So-called "implicit exceptions" are not exceptions proper, unless by this expression one means to refer to the totality of cases covered by the vacuous clauses such as those mentioned under the third presupposition. If, on the other hand, by "implicit" we mean something like "specified in advance by way of an exhaustive, albeit undisclosed, enumeration of normatively relevant properties", then we have a dilemma: Either we concede that the rule was not applicable in the first place (the case was not an exception, after all), or we openly acknowledge that when we make the supposedly implicit condition explicit, what we are in effect doing is we are adopting a new rule, different from the previous one and less general than it. We

have simply changed our mind as to how cases of type T should be regulated (what the right verdict in them is), where T is precisely the antecedent of the rule we have just abandoned[37].

Those are the presuppositions, a set of theses that lay out a specific particularist position, the one I would account to be the most plausible. Now, this conceptual construction can only have any value if we can explain the difference between normal and non-normal cases (and if we assume that the decision-maker is more or less capable of discriminating between them). If we cannot explain where the difference lies—but can no more than say that the former are the cases in which it is reasonable to follow the rule, whereas the latter are the cases in which it is reasonable to ask whether or not the course of action indicated by the rule is the right one—this conceptual construction remains completely idle, uninformative, and gratuitous, even if in itself coherent. That would amount to having said next to nothing, or nothing at all.

Where, then, does the difference lie between normal and non-normal cases? Under what conditions can it reasonably be claimed that the case being decided could be a (true) exception? And, once we concede that the decision-maker has the capacity to discriminate between normal and non-normal cases, how does this discriminating capacity work?

The conjecture can be stated as follows:[38] Whether or not a certain case is normal—recall that on this depends the question of whether or not, in any given case, it is *reasonable* to reconsider the rule (the question, I underscore once again, is a *normative* one)—is something that depends on psychological facts: Certain cases are normal or non-normal if and because they appear that way to us[39], and which cases appear to us to be normal and which do not is a psychological matter of fact. So, then, whether a certain rule, in a certain case, is a reason to act accordingly—whether it is a *justificatory* reason, for that is the crucial point—depends on our psychological makeup.

In conclusion, when the purported reasons are rules, what our reasons are—what we must do, what evaluation we ought to make—depends on a background condition of normality. Given a rule that is adequately justified and applies to a certain case, it is reasonable to follow that rule only under normal conditions. Whether these conditions are fulfilled depends on mental facts. Particularism and psychologism come together.

In other words, as we have seen, the crucial problem for the proponents of a two-tier theory of law—a theory predicated on a distinction between basic reasons (principles) and rules—is whether it is possible to draw that distinction in such a way as to prevent the second level from collapsing into the first, while at the same time leaving open the possibility that, in certain cases, the rules are to be reconsidered. How to determine, in each case, whether the rule is to be reconsidered unless we do reconsider it? The discussion in this section suggests one way in which this question can perhaps be answered. It is not up to us to determine, in each of the cases that fall under their antecedent, whether the rule is to be reconsidered:[40] It is up to our mind.

[37] Cf. BRIGAGLIA & CELANO 2018, sec. 2. Carlos E. ALCHOURRÓN (1996) puts forward a dispositional conception of implicit exceptions that complicates, but does not change, this picture. Implicit exceptions, on this view, are cases which (a) fall under the rule's antecedent, but which (b) the legislators did not have in mind when they enacted the rule (it is worth noting here that we are speaking only of rules attributable to a determinate and identifiable author or group of authors to whom a precise will can be attributed), and which (c)—had they taken these cases into account: Had it only occurred to them that such cases could arise!—they would have regulated differently. Now, when we speak of "implicit exceptions" in these terms, what we are in fact saying is that, if the legislators had only contemplated the case in advance, they would have enacted a different, and less general, rule than the one they in fact enacted.

[38] This conjecture has previously been formulated, or at least suggested, in BRIGAGLIA 2016, sec. 3.3, and is developed in BRIGAGLIA & CELANO 2018, sec. 3.

[39] On the question of who the "us" designates here (or, Who is "we"?), see CELANO 2014, secs. 5 and 6.

[40] Or, *mutatis mutandis*, it is not up to us to determine whether the case in question is one that can be judged simply at a glance or whether it instead requires us to look at it closely (fn. 36).

4. *Conclusion: An invitation to follow the trend*

What I have argued in this essay lines up, or at least is meant to line up, with contemporary research on bounded rationality and human inference, on heuristics and biases, and on moral psychology—all of which is very much the trend now, and mine is indeed an exhortation to pursue that trend before it passes.

There is now a vast, rich, interesting, and rapidly advancing area of psychological, and in general empirical, research that is developing around cognitive science and is concerned with norms and rules, or the way in which these shape human behaviour, and with norm- and rule-based reasoning. These investigations—onto which contemporary developments of neuroscience are grafted—take a more or less consciously psychologistic slant. This whole field of inquiry we might call "psychodeontics" (though this is just a label—a take on Jerry Fodor's "psychosemantics"). It would be inadmissible for these investigations to remain outside the theory of legal reasoning, and the theory of law in general[41]. If they are taken seriously, they are quite likely (and, as I insist, I am not claiming anything more than "likely") to get legal reasoning, and legal theory generally, to fundamentally change course in a more or less drastic way. As I tried to show, there are compelling reasons for exploring the prospect of setting legal theory, and the theory of legal reasoning in particular, on a psychodeontic path.

And that, I believe, is the main path toward a naturalised jurisprudence[42].

[41] By this I do not mean, of course, that there are no legal theorists who have pursued this path. Consider, for instance, the work that, from very different perspectives, has been done by Bartosz Brożek, Dennis Patterson, Giovanni Sartor, and Cass R. Sunstein. There are as well numerous recent contributions to "neurolaw," particularly concerned with teasing out the implications that neuroscience carries for the concepts of guilt and liability. And also a subject of investigation is the psychology of judicial decision-making.

[42] A psychodeontic approach is a way—in fact, it presently strikes me as the *only* way—to accomplish the project of "naturalizing jurisprudence", to go by Brian LEITER's 2007 motto (for Leiter's ideas about how to naturalise normativity, see LEITER 2015).

References

ALCHOURRÓN C.E. 1996. *On Law and Logic*, in «Ratio Juris», 9, 4, 331 ff.

ALCHOURRÓN C.E., BULYGIN E. 1971. *Normative Systems*, Springer.

ATIENZA M. 2014. *Ponderación y sentido común jurídico*, manuscript. Available at: http://lamiradadepeitho.blogspot.it/2014/11/ponderacion-y-sentido-comun-juridico.html.

ATIENZA M. 2017. *Algunas tesis sobre el razonamiento judicial*, in AGUILÓ REGLA J., GRÁNDEZ CASTRO P. (eds.), *Sobre el razonamiento jurídico. Una discusión con Manuel Atienza*, Palestra.

ATIENZA M., RUIZ MANERO J. 1996. *Las piezas del derecho: Teoría de los enunciados jurídicos*, Ariel.

ATIENZA M., RUIZ MANERO J. 2000. *Ilícitos atípicos*, Trotta.

ATIENZA M., RUIZ MANERO J. 2009. *Ancora sugli illeciti atipici: Replica alle critiche italiane*, in «Europa e diritto privato», 203 ff.

BELL D. 1992. *Psychologism*, in DANCY J., SOSA E. (eds.), *A Companion to Epistemology*, Blackwell.

BRIGAGLIA M. 2016. *Rules and Norms: Two Kinds of Normative Behaviour*, in «Revus: Journal for Constitutional Theory and Philosophy of Law», 30.

BRIGAGLIA M., CELANO B. 2018. *Reasons, Rules, Exceptions: Towards a Psychological Account*, in «Analisi e Diritto 2017», 131 ff.

CELANO B. 1999. *La teoria del diritto di Hans Kelsen: Una introduzione critica*, il Mulino.

CELANO B. 2002. *'Defeasibility' e bilanciamento: Sulla possibilità di revisioni stabili*, in «Ragion pratica», 18.

CELANO B. 2005. *Possiamo scegliere fra particolarismo e generalismo?*, in «Ragion pratica», 25, 469 ff.

CELANO B. 2006a. *Pluralismo etico, particolarismo e caratterizzazioni di desiderabilità: Il modello triadico*, in «Ragion pratica», 26, 133 ff.

CELANO B. 2006b. *Principi, regole, autorità: Considerazioni su M. Atienza, J. Ruiz Manero*, Illeciti atipici, in «Europa e diritto privato», 3, 1061 ff.

CELANO B. 2012. *True Exception: Defeasibility and Particularism*, in FERRER BELTRÁN J., RATTI G.B. (eds.), *The Logic of Legal Requirements: Essays on Defeasibility*, Oxford University Press, 268 ff.

CELANO B. 2014. *Pre-convenzioni: Un frammento dello Sfondo*, in «Ragion pratica», 43, 605 ff.

CELANO B. 2016. Rule of Law *e particolarismo etico*, in PINO G., VILLA V. (eds.), Rule of Law: *L'ideale della legalità*, il Mulino.

CELANO B. 2017. *Particularismo, psicodeóntica. A propósito de la teoría de la justificación judicial de Manuel Atienza*, in AGUILÓ REGLA J., GRÁNDEZ CASTRO P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 59 ff.

COPI I.M., COHEN C., MCMAHON K. 2014. *Introduction to Logic* (14th Ed.), Pearson.

DANCY J. 2004. *Ethics Without Principles*, Clarendon Press.

DICIOTTI E. 2007. *Regola di riconoscimento e concezione retorica del diritto*, in «Diritto & Questioni pubbliche», 7, 9 ff.

DUMMETT M. 1984. *Origins of Analytical Philosophy*, Harvard University Press.

DUMMETT M. 1986. *Frege's Myth of the Third Realm*, in ID., *Frege and Other Philosophers*, Clarendon Press.

ENGEL P. 1996. *Philosophie et psychologie*, Gallimard.

ENGEL P. 1998. *The Psychologist's Return*, in «Synthese», 115, 3, 375 ff.

FERRAJOLI L. 1993. *Il diritto come sistema di garanzie*, in «Ragion pratica», I.

FREGE G. 1979. *Logic (1897)*, in ID., *Posthumous Writings*, ed. by H. Hermes, F. Kambartel, F. Kaulbach, trans. By P. Long, R. White, Basil Blackwell, 126 ff. (The German original: *Logik*, 1897.)

FREGE G. 1980. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, trans. By J.L. Austin, Northwestern University Press. (The German original: *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung über den Begriff der Zahl*, 1884.)

FREGE G. 1984. *Logical Investigations. Part. I: Thoughts (1918–19)*, in ID., *Collected Papers on Mathematics, Logic, and Philosophy*, ed. by B. McGuinness, Basil Blackwell, 351 ff. (The German original: *Der Gedanke: Eine logische Untersuchung*, 1918.)

GIGERENZER G. 2008. *Rationality for Mortals: How People Cope with Uncertainty*, Oxford University Press.

GIGERENZER G., TODD P.M., THE ABC RESEARCH GROUP. 1999. *Simple Heuristics That Make Us Smart*, Oxford University Press.

GIROTTO V. (ed.) 2013. *Introduzione alla psicologia del pensiero*, Il Mulino.

GUTTENPLAN S. 1994. *Normative*, in ID. (ed.), *A Companion to the Philosophy of Mind*, Blackwell.

HALE B. 1997. *Rule-Following, Objectivity and Meaning*, in ID., WRIGHT C. (eds.), *A Companion to the Philosophy of Language*, Blackwell.

JACQUETTE D. 2003. *Introduction: Psychologism the Philosophical Shibboleth*, in ID. (ed.), *Philosophy, Psychology, and Psychologism: Critical and Historical Readings on the Psychological Turn in Philosophy*, Kluwer.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Penguin.

KELSEN H. 1960. *Hauptprobleme der Staatsrechtslehre*, 2nd Ed., Scientia Verlag. (Originally published 1911.)

KELSEN H. 1966. *Allgemeine Staatslehre*, Gehlen. (Originally published 1925.)

KELSEN H. 1945. *General Theory of Law and State*, Harvard University Press.

LEITER B. 2007. *Naturalizing Jurisprudence*, Oxford University Press.

LEITER B. 2015. *Normativity for Naturalists*, in «Philosophical Issues», 25, 64 ff.

MORESO J.J., NAVARRO P.E. 1996. *Applicabilità ed efficacia delle norme giuridiche*, in Comanducci P., Guastini R. (eds.), *Struttura e dinamica dei sistemi giuridici*, Giappichelli.

NISBETT R., ROSS L. 1980. *Human inference: Strategies and Shortcomings of Social Judgment*, Prentice Hall.

QUINE W.V.O. 1959. *Methods of Logic*, revised edition, Holt, Rinehart and Winston. (1st ed. 1950.)

Quine W.V.O. 1969. *Epistemology Naturalized*, in ID., *Ontological Relativity and Other Essays*, Columbia University Press, 69 ff.

RAZ J. 1979. *The Authority of Law: Essays on Law and Morality*, Clarendon Press.

REY G. 2006. *The Intentional Inexistence of Language—but Not Cars*, in STAINTON R.J. (ed.), *Contemporary Debates in Cognitive Science*, Blackwell, 237 ff.

RUIZ MANERO J. 2013. *Two Particularistic Approaches to the Balancing of Constitutional Principles*, in «Analisi e diritto», 197 ff.

SCHAUER F. 1991. *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Clarendon Press.

SEARLE J.R. 1983. *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press.

SEARLE J.R. 1992. *The Rediscovery of the Mind*, The MIT Press.

SEARLE J.R. 1995. *The Construction of Social Reality*, Penguin.

SIMON H.A. 1983. *Reason in Human Affairs*, Stanford University Press.

TODARO G. 2011. *Naturalizzazione della dialettica: L'errore nella giustificazione dialogica delle credenze*, Ph.D. dissertation, University of Palermo.

WATANABE DAUER F. 1989. *Critical Thinking: An Introduction to Reasoning*, Oxford University Press.

Experimental Jurisprudence

KEVIN TOBIA

1. *What Is experimental jurisprudence?*

Experimental jurisprudence is scholarship that addresses jurisprudential questions with empirical data, typically data from experiments[1]. This two-part definition is straightforward. But it leads to surprising implications for the nature of jurisprudence and the research that it calls for.

This Article introduces experimental jurisprudence (also known as "XJur"), proposes a framework to understand its contributions[2], and finally explains the central role that XJur should play in two other modern jurisprudential movements: the rise of "ordinary meaning" in legal interpretation and the "New Private Law"[3]. To unpack the two-part definition—"experiments" plus "jurisprudence"—it is helpful to reflect on the meaning of each term. This first Part begins with that background.

The meaning of "jurisprudence" is itself highly controversial[4]. Consider some representative descriptions:

- In the United States, jurisprudence is «mostly synonymous with "philosophy of law" [but there is also] a lingering sense of "jurisprudence" that encompasses high legal theory [...] the elucidation of legal concepts and normative theory from within the discipline of law»[5].
- Jurisprudence is «the most fundamental, general, and theoretical plane of analysis of the social phenomenon called law [...] Problems of jurisprudence include whether and in what sense law is objective [...] the meaning of legal justice [...] and the problematics of interpreting legal texts»[6].
- The «essence of the subject [...] involves the analysis of general legal concepts»[7].

* This article is an adaptation of Tobia K., *Experimental Jurisprudence*, in «University of Chicago Law Review», 89, 2022, 735-802, available on https://lawreview.uchicago.edu/publication/experimental-jurisprudence.

1 The term "experimental jurisprudence" nods to "experimental philosophy," the related experimental approach to questions in philosophy. See KNOBE & NICHOLS 2008, 3; see also STICH & TOBIA 2016, 5 (explaining different versions and goals of experimental philosophy). One of the first modern mention of "experimental jurisprudence" is in SOLUM 2014, 2465 n.5 (2014) (first citing MIKHAIL 2011 and then citing BILZ 2012. Although new, the movement builds on important theoretical work in naturalizing jurisprudence (see generally, e.g., LEITER 2007) and the role of social science in legal philosophy (see generally, e.g., SCHAUER 2020). The term "experimental jurisprudence" had been used fifty years ago, in a very different way. See BEUTEL 1971, 409 (describing an experimental-jurisprudence approach that required «[s]ocial [e]ngineering in [g]overnment»).

2 See below, § 1, 2.

3 See below, §§ 3.1, 3.2 On ordinary meaning, see generally FALLON 2015. On the New Private Law, see generally GOLD et al. 2020.

4 See generally TUR 1978. See also SCHAUER 2020, 95 fn.2: «The word "jurisprudence" is often used these days as a synonym for "philosophy of law". But given the longstanding existence of fields known as historical jurisprudence, sociological jurisprudence, and so on [...] the word [...] remains ambiguous. Nevertheless, it remains important to resist the notion that [jurisprudence] must necessarily be philosophical in method or focus».

5 SOLUM 2018.

6 POSNER 1990.

7 TUR 1978, 152.

These representative descriptions each characterize jurisprudence broadly—and differently. As such, this Article understands jurisprudence inclusively. In the words of legal philosopher Julie Dickson, jurisprudence is a "broad church"[8]. It is concerned with descriptive questions about legal concepts and interpretation as well as normative questions about what law should be[9]. Jurisprudence approaches these questions from a broadly theoretical perspective, but it is not committed to a particular methodology[10].

Despite this inclusivity and breadth, if forced to identify the core of modern jurisprudence, some might point to analytical jurisprudence[11]. A central project of analytical jurisprudence is the examination of legal concepts, including the law itself[12], causation[13], reasonableness[14], punishment[15], and property[16]. That research typically involves "conceptual analysis," in which jurisprudence scholars reflect on legal concepts and attempt to articulate their features. Conceptual analysis also raises questions about which features concepts should have[17]. Despite some skepticism about conceptual analysis, it remains central to jurisprudence. As Alex Langlinais and Professor Brian Leiter put it, «[i]n many areas of philosophy, doubts about [...] conceptual and linguistic analysis [...] have become common [...] but not so in legal philosophy»[18].

Good scholarship has good methods. What are the methods of jurisprudence? Within conceptual analysis, a common method involves reflecting on hypothetical test cases, i.e., thought experiments[19]. One's intuitions about these test cases are taken to provide evidence about whether the proposed analysis is successful[20].

As an example, consider the legal concept of reasonableness. This concept is central to legal determinations such as tort negligence[21], open contract price terms[22], and the line between murder and manslaughter[23]. The term "reasonable" appears in over one-third of modern published judicial decisions[24]. So what are the legal criteria of what is reasonable? Jurisprudential analysis might begin with a proposed criterion before reflecting on test cases to intuitively assess the success of the proposed criterion[25].

As a simple example, consider this criterion: an act is reasonable if—and only if—it is welfare maximizing in expectation. So, in negligence law, the proposed analysis holds that reasonable care is the care that would be expected to lead to the welfare-maximizing result. How might a legal philosopher evaluate the strength of this proposed analysis? They might assess this jurisprudential analysis against the following thought experiment about "life-saving negligence".

A company produces and sells yachts, donating all profits to a high-impact charity. That donation saves five lives per sale. Yacht production also creates pollution, which foreseeably kills

8 DICKSON 2015; see also PRIEL 2019.
9 See generally WEST 2011.
10 DICKSON 2015, 209; SCHAUER 2020, 95 s.
11 TUR 1978, 152.
12 HART 1994.
13 See generally HART & HONORÉ 1985.
14 See generally GARDNER 2015.
15 See FLETCHER 1998.
16 See AUSTIN 1869. See generally COHEN 1954.
17 See, e.g., BIX 1995; see also RAPPAPORT 2014.
18 LANGLINAIS & LEITER 2016, 677.
19 LANGLINAIS & LEITER 2016, 677.
20 LEITER 2003, 43 s. (explaining that jurisprudence «relies on two central argumentative devices—analyses of concepts and appeals to intuition»).
21 *Restatement (Third) of Torts: Liability for Physical and Emotional Harm* § 3 (*Am. L. Inst.* 2005).
22 *U.C.C.* § 2-305 (*Am. L. Inst. & Unif. L. Comm'n* 2020).
23 *Model Penal Code* § 210.3(1)(b) (*Am. L. Inst.* 1980).
24 *Historical Trends*, *Caselaw Access Project*, https://case.law/trends/?q=reasonable.
25 E.g., FLETCHER 1985, 949 s.; see also KEATING 1996, 311.

one person in the nearby town per sale. The company could cheaply install a new production mechanism that would eliminate all pollution and increase production costs. That would eliminate all pollution deaths in the nearby town and decrease profits and thus donations, reducing lives saved to only two per yacht produced. The company does not install the new mechanism, and a number of people die from the pollution, and more are saved by the donations.

This decision appears to be welfare maximizing (five lives saved for each lost—plus the benefits of yachts). But it might seem, intuitively, that the company has not acted with "reasonable care" by failing to install the pollution-eliminating production mechanism[26].

Such a reaction to this thought experiment represents something like a legal-philosophical discovery. The legal philosopher now has some intuitive "data", which might imply that the proposed conceptual analysis needs revision. That intuition (if widely shared) suggests that reasonable care is not simply welfare-maximizing care, and we should refine the analysis, test that revision with more cases, refine the analysis in light of those, and so on.

The life-saving-negligence thought experiment may elicit a shared response (for instance, that the company did not act with reasonable care). This response is an intuition. A jurist describes some scenario (actual or hypothetical) and invites readers to consider some questions about the scenario: Does the care seem reasonable? Which action seems like the cause? Is that rule a legal rule? Thought experiments and corresponding intuitions in legal theory include Justice Oliver Wendell Holmes's Bad Man[27], the criminality of Professor Lon Fuller's Speluncean Explorers[28], excuses for Professor Sandy Kadish's Mr. Fact and Mr. Law[29], and the meaning of rules like Professor Frederick Schauer's "no vehicles in the park"[30] or Fuller's «[i]t shall be a misdemeanor [...] to sleep in any railway station»[31].

Most legal philosophers value this kind of intuitive evidence. Many give intuition great significance. As the philosopher and legal theorist Thomas Nagel puts it: «Given a knockdown argument for an intuitively unacceptable conclusion, one should assume there is probably something wrong with the argument that one cannot detect—though it is also possible that the source of the intuition has been misidentified»[32].

Of course, legal theorists rarely take shared intuitions to settle jurisprudential debate. For one, different cases can give rise to conflicting intuitions. Jurisprudence takes care to understand and resolve those conflicts. In her seminal work on the concept of consent, Professor Heidi Hurd proposes: «What should we do in the face of our conflicting intuitions [...]? One possible solution is to grapple with our intuitions some more in the hope that further thought experiments will help us to determine which set of intuitions misleads us»[33]. Nagel offers another solution: assess whether the «source of the intuition has been misidentified»[34]. These methodological proposals—to assess whether intuitions are shared among philosopher colleagues or other persons, to grapple with

[26] Or maybe not. Perhaps some readers do not share the intuition. The aim here is not to analyze reasonableness but to demonstrate a very familiar method of analysis.
Ignoring unshared intuitions can lead to some problems. There is a danger that the process of conceptual analysis falls victim to groupthink and to information cascades if those «who do not share the intuition are simply not invited to the games» (CUMMINS 1998, 116). The "shared" intuition takes on increasing strength as those with minority views leave the debate.

[27] HOLMES 1897, 459.

[28] See generally FULLER 1949.

[29] PAULSEN & KADISH 1962.

[30] See SCHAUER 2008, 1110 s.

[31] FULLER 1958, 664.

[32] NAGEL 1979, x.

[33] HURD 1996, 143.

[34] NAGEL 1979, x.

further thought experiments, to uncover the "sources" of one's intuitions—are all part of traditional jurisprudence.

Experimental jurisprudence can be seen as providing an empirically grounded method of thought experimentation. As an example, consider the following experimental study about intent[35]. The study, conducted by Professors Markus Kneer and Sacha Bourgeois-Gironde, investigates a jurisprudential question: Does whether a side effect seems to be produced intentionally depend on the severity of the side effect?[36] Traditional jurisprudence might assess that question with thought experimentation, considering two examples of similar actions that lead to differently severe side effects.

Experimental jurisprudence proceeds in a similar way. In this case, the researchers recruited participants and randomly assigned them to evaluate two different scenarios.[37] In the first scenario (the "moderate" one), a mayor decides to build a new highway in order to improve the flow of traffic. However, the highway construction has a foreseeable side effect[38]. It will produce a moderate environmental impact; specifically, it will disturb some animals in the construction zone[39]. The mayor states that he does not "care at all about the environment" and proceeds with the program[40]. In the second scenario ("severe"), another group of participants evaluates a very similar case, except that, in this version, the environmental side effect is severe. It is foreseeable that the impacted animals will die[41]. Again, the mayor makes the same statement and goes ahead with the plan. In both scenarios, participants evaluate the same question: Did the mayor "intentionally" harm the environment?[42]

This study found that, perhaps surprisingly, intuitions about intentionality are sensitive to outcome severity[43]. Even though the mayor expresses the same attitude in both scenarios, participants assess his mental state differently[44]. They more strongly agreed that the harm was produced "intentionally" in the severe case[45].

This result informs conceptual analysis[46]. For example, we can consider two different accounts of intentional action: one in which intentionality is, in fact, sensitive to outcome severity and one in which it is not. As in a traditional jurisprudential analysis, this conceptual analysis makes predictions about how the concept applies. Experimental jurisprudence tests those predictions, supplementing thought experimentation with cognitive-scientific experimentation. Of course, one study does not resolve all debate. In response to this empirical finding, some might argue that the ordinary concept of intentionality is severity sensitive, and future empirical research could seek to test additional predictions of that theory. Others might

[35] See generally KNEER & BOURGEOIS-GIRONDE 2017; KNOBE 2003; below, § 2.
[36] KNEER & BOURGEOIS-GIRONDE 2017, 143 s.
[37] KNEER & BOURGEOIS-GIRONDE 2017,143.
[38] KNEER & BOURGEOIS-GIRONDE 2017,143.
[39] KNEER & BOURGEOIS-GIRONDE 2017,143.
[40] KNEER & BOURGEOIS-GIRONDE 2017,143.
[41] KNEER & BOURGEOIS-GIRONDE 2017, 143.
[42] KNEER & BOURGEOIS-GIRONDE 2017,143.
[43] KNEER, BOURGEOIS-GIRONDE 2017,143.
[44] KNEER & BOURGEOIS-GIRONDE 2017,143.
[45] KNEER & BOURGEOIS-GIRONDE 2017,143.
[46] Or perhaps this method might exceed traditional analyses. For example, Professor Felipe Jiménez has written a generous and insightful critique of experimental jurisprudence, arguing that in legal theory, conceptual analysis should look primarily to the judgments of "legal officials" (JIMÉNEZ 2021). Jiménez's critique is leveled primarily at "folk jurisprudence". Rather than relying on laypeople's judgments, Jiménez argues that conceptual analysis should rely on the judgments of legal officials. But that proposal has a (perhaps) surprising implication: insofar as most legal philosophers are not legal officials, jurisprudence should not generally rely on the judgments of legal-philosophy PhDs. So even some critics of folk jurisprudence may find that experimental jurisprudence has something to offer. For example, in the experimental-jurisprudence study discussed here, the participants were professional judges. See KNEER & BOURGEOIS-GIRONDE 2017, 143.

argue that the experimental participants here exhibit some kind of bias. That latter account might be supported by further empirical research that clarifies that the "source" of the participants' intuitions is in some way inappropriate or untrustworthy[47].

As this example suggests, there are complementarities between traditional and experimental jurisprudence. Traditional jurisprudence often proposes shared intuitions—i.e., claims about a widely shared response to a thought experiment. Experimental jurisprudence can help assess the robustness of that claim by seeking responses from a larger set of persons, including those who have little at stake in the theoretical debate.

Moreover, experimental jurisprudence can help assess questions about intuitions that are hard to address from the armchair. For example, suppose that we tried to test the severity sensitivity of intentionality through thought experiments. Perhaps some can intuitively discern that, all else equal, a very bad outcome seems more intentionally produced than a moderately bad outcome. But it might be hard to feel very confident about those individual intuitions. Other, more subtle patterns of human judgment may be impossible to accurately assess just by thinking hard. The experimental approach, which studies large samples of people and assigns them to consider different versions of thought experiments, can help detect more subtle patterns of judgment, including ones that are not obvious or even introspectively accessible to an individual legal theorist[48].

This example suggests some commonality between traditional jurisprudence and experimental jurisprudence. Both propose theories about legal concepts (e.g., the intentionality of a foreseeable side effect depends on its severity), both test those theories with (thought) experiments, and both revise the conceptual analysis in light of the findings. Perhaps "experimentation" is neither unfamiliar nor unwelcome in jurisprudence.

At the same time, there are differences between the approaches. Traditional jurisprudence occurs in the seminar room—or across the pages of law reviews—among professors and scholars with significant training and expertise. Experimental jurisprudence normally begins online, by surveying laypeople with no special legal training. In analyzing the concepts of legal intent, consent, cause, and reasonableness, why should we think that the views of laypeople with no formal legal training are particularly helpful?

The remainder of this Article answers this question. Part 2 details some examples of recent experimental jurisprudence, proposing a framework to unify these diverse projects. Part 3 argues that experimental jurisprudence is particularly well placed to contribute to two central areas of modern legal theory: the debate about "ordinary meaning" in legal interpretation and the "New Private Law". Together, these Parts aim to clarify and justify the movement of experimental jurisprudence, concluding that it should be understood as a movement at the core of traditional jurisprudence.

2. *Some recent experimental-jurisprudence research*

To understand the significance of the experimental-jurisprudence movement, it is instructive to study examples of work in the area. This Part's brief overview cannot do justice to the enormous and ever-growing number of examples[49]. This Part highlights experimental-jurisprudence studies across several areas: studies of mental states (including knowledge, recklessness, and intent),

[47] In experimental philosophy, the "negative" program has focused on these types of debunking arguments. See, e.g., ALEXANDER et al. 2010, 298 Fn. 2. See generally STICH & TOBIA 2016, 5.

[48] Importantly, experimental jurisprudence does not simply compute answers to legal questions. For example, the key takeaway from this empirical study is not that judges and juries should now hold that foreseeable side effects are more "intentional" as they become more severe. To the contrary, jurisprudential debate about that question continues.

[49] For other introductions to the field of experimental jurisprudence, see generally MAGEN & PROCHOWNIK, https://zrsweb.zrs.rub.de/institut/clbc/legal-x-phi-bibliography; STROHMAIER 2017; PROCHOWNIK 2021; SOMMERS 2021.

consent, causation, and law itself. But experimental jurisprudence has also studied criminal responsibility and punishment[50], blame[51], justice[52], human rights[53], law and morality[54], the internal point of view[55], abstract versus concrete legal principles[56], state paternalism[57], nationality[58], identity and the self[59], free speech[60], custody decisions[61], happiness[62], lying[63], outcome severity[64], attempts[65], harm[66], liability[67], interpretation[68], evidence[69], settlement[70], contract[71], promise[72], ownership[73], disability[74], reasonableness[75], the balancing tests[76], and legal rules[77].

This research has focused largely on lay judgment, but some of it has studied populations with legal training, including law students and judges[78]. And while much of this work involves U.S. participants, today's experimental-jurisprudence movement is the product of international efforts; some of the most impressive examples are conducted by researchers

[50] See generally ALICKE 1992; DARLEY et al. 2000; CARLSMITH et al. 2002; BILZ & DARLEY 2004; DARLEY 2009; CUSHMAN 2008; CUSHMAN 2011; DARLEY 2010; NADELHOFFER et al. 2013; CUSHMAN 2015; BILZ 2016; BREGANT et al. 2016; NAHMIAS & AHARONI 2018; PROCHOWNIK 2017; PROCHOWNIK & UNTERHUBER 2018; KNEER & MACHERY 2019; BAUER & POAMA 2020; DUNLEA & HEIPHETZ 2020; DUNLEA & HEIPHETZ 2021a; DUNLEA & HEIPHETZ 2021b; MARTIN & HEIPHETZ 2021; ROBINSON & DARLEY 1995; ROBINSON 2008; NADELHOFFER 2013.

[51] See generally ALICKE, &DAVIS 1989; ALICKE 2000; SOLAN 2003; PROCHOWNIK et al. 2021; NADLER & MCDONNELL 2012.

[52] See generally DARLEY 2001; DARLEY & PITTMAN 2003; VILLANI et al. 2021.

[53] See generally MIKHAIL 2012.

[54] See generally DONELSON & HANNIKAINEN 2020; FLANAGAN & HANNIKAINEN 2020; HUANG 2019; HUANG 2013; HUANG 2016.

[55] See generally HOEFT 2019; ROVERSI et al. 2022.

[56] See generally BYSTRANOWSKI et al. 2021; STRUCHINER et al. 2020b.

[57] See generally HANNIKAINEN et al. 2017.

[58] See generally HUSSAK & CIMPIAN 2019.

[59] See generally BARTELS & RIPS 2010; EARP et al. 2019; MOLOUKI et al. 2017; MOLOUKI & BARTELS 2017; NEWMAN et al. 2014; STROHMINGER & NICHOLS 2014; TOBIA 2015; TOBIA 2016; STROHMINGER & NICHOLS 2015; MOTT 2018, 243; DUNLEA et al. 2022; SHOEMAKER & TOBIA 2023; DIAMANTIS 2019; DIAMANTIS 2022.

[60] See generally DE KEERSMAECKER et al. 2021.

[61] See generally COSTA et al. 2019.

[62] See generally PHILLIPS 2017; KNEER & HAYBRON (mns); BRONSTEEN et al.(forthcoming).

[63] See generally VIEBAHN et al. 2020; WIEGMANN & MEIBAUER 2019.

[64] See generally KNEER & BOURGEOIS-GIRONDE 2017; RACHLINSKI et al. 2013; KNEER (forthcoming).

[65] See generally NADELHOFFER 2012.

[66] See generally KUGLER 2019.

[67] See generally SANDERS et al. 2014.

[68] See generally SOLAN et al. 2008; GINTHER et al. 2014; BEN-SHAHAR & STRAHILEVITZ 2017; BREGANT et al. 2019; STRUCHINER et al. 2020a; TOBIA 2020; KLAPPER et al. (mns); MACLEOD 2022; TOBIA et al. 2022a; TOBIA, SLOCUM & NOURSE 2022b; NYARKO & SANGA 2022.

[69] See generally BILZ 2012; SOOD 2015; BANDES & SALERNO 2014; GLÖCKNER & ENGEL 2013a.

[70] See generally KOROBKIN & GUTHRIE 1997; BREGANT et al. 2022; BRONSTEEN et al. 2008.

[71] See generally WILKINSON-RYAN 2010; WILKINSON-RYAN & BARON 2009; HOFFMAN & WILKINSON-RYAN 2013; WILKINSON-RYAN & HOFFMAN 2015; KAWASHIMA 1974; WILKINSON-RYAN 2017; FURTH-MATZKIN & SOMMERS 2020.

[72] See generally VANBERG 2008; CHARNESS & DUFWENBERG 2010; WILKINSON-RYAN 2012; EDERER & STREMITZER 2018; EDERER & STREMITZER 2017; MISCHKOWSKI et al. 2019; STONE & STREMITZER 2020.

[73] See generally NANCEKIVELL et al. 2019; NANCEKIVELL et al. 2016; FRIEDMAN et al. 2018; KANNGIESSER & HOOD 2014; FRIEDMAN et. al 2013; NANCEKIVELL et al. 2013; DESCIOLI et al. 2017; ECHELBARGER et al. 2021.

[74] See generally DORFMAN 2021; DORFMAN 2020.

[75] See generally TOBIA 2018a; GROSSMAN et al. 2020; JAEGER 2021; SPRUILL & LEWIS 2022. Experimental work also suggests that ordinary people overestimate the cognitive skills that people possess. See generally RACHLINSKI 2003; KNEER & HAYBRON (mns).

[76] See generally ENGEL & RAHAL (unpublished).

[77] See generally STRUCHINER et al. 2020a; TOBIA et al. 2022a.

[78] TOBIA 2020, 762. See also generally TOBIA et al. 2022a; KNEER & BOURGEOIS-GIRONDE 2017; SPAMANN & KLÖHN 2016; KLERMAN & SPAMANN (unpublished).

outside the United States—for example, by researchers in Brazil, Spain, Lithuania and Germany[79]. Recent studies have also emphasized the importance of cross-cultural samples, employing cross-cultural and cross-linguistic studies[80].

Finally, it is difficult to precisely categorize whether some studies fall neatly into "experimental jurisprudence". The research has important connections to research in behavioral law and economics[81], legal heuristics and biases[82], motivated reasoning[83], experimental bioethics[84], experimental longtermism[85], and research in law and corpus linguistics[86].

The remainder of this Part turns to some recent examples of experimental jurisprudence. Again, most of these examples study ordinary concepts, for example, how laypeople evaluate what is "intentional" or "consensual". Those studies are typically embedded within a particular jurisprudential debate—questions about the nature of intent or causation. But (at the risk of oversimplifying), it may be useful to introduce one broader hypothesis that is relevant to many recent studies.

This hypothesis is the "folk-law thesis"[87]. Broadly speaking, the claim is that ordinary concepts are at the heart of legal concepts. For example, this account would predict that the legal concept of causation reflects features of the ordinary concept of causation and that the legal concept of consent reflects features of the ordinary concept of consent. If this were true, it would provide one general reason for jurisprudential scholars to evaluate empirical research about ordinary concepts. If there are surprising features of ordinary concepts to be discovered, such discoveries might also constitute discoveries of features of legal concepts. For example, a discovery about how people understand ordinary cause (or intent, or consent, or reasonableness) might actually enrich our understanding of the legal notion of cause (or intent, or consent, or reasonableness).

The strong version of this thesis holds that a given legal concept is identical to its ordinary counterpart. The weak version holds that a given legal concept shares some features of the ordinary concept. Even if the strong version of the thesis is false (with respect to a given concept), the weaker folk-law thesis may prove useful in structuring inquiry. For example, learning more about the ordinary concept can help clarify what is, in fact, distinct about the legal concept.

This Part's order of presentation reflects the breadth of experimental jurisprudence. Much of the best-known experimental jurisprudence studies concepts that are specifically referenced in law: knowledge, intent, and consent. But experimental jurisprudence also studies concepts that are not explicitly referenced as outcome-determinative ones, such as the concept of personal identity—or even law itself. Finally, experimental jurisprudence studies broader classes of concepts. For example, in addition to studying the specific concept of causation (which is relevant to the law of tort negligence because it uses the criterion "cause"), it also studies a

[79] See, e.g., STRUCHINER et al. 2020a; DRANSEIKA et al. 2020; KNEER & BOURGEOIS-GIRONDE 2017.
[80] See generally HANNIKAINEN et al. 2021; SPAMANN et al. 2021; LIU et al. 2021.
[81] See generally RACHLINSKI 2011.
[82] See generally GUTHRIE et al. 2002; GUTHRIE et al. 2007; HARLEY 2007; RACHLINSKI et al. 2009; RACHLINSKI et al. 2011; RACHLINSKI 2000; WISTRICH et al. 2015; LIDÉN et al. 2018a; LIDÉN et al. 2018b; WINTER 2020; GLÖCKNER & ENGEL 2013b; MELNIKOFF & STROHMINGER 2020.
[83] See generally KAHAN et al. 2016.
[84] See generally EARP et al. 2020a; EARP et al. 2020b; MIHAILOV et al. 2021.
[85] See generally MARTÍNEZ & WINTER 2022.
[86] See generally MAZZI 2014; SYTSMA et al. 2019; STUBBS 1996; LEE & MOURITSEN 2018.
[87] TOBIA 2021. Professors Steve Guglielmo, Andrew Monroe and Bertram Malle propose that folk psychology may also be at the heart of morality. See GUGLIELMO, et al. 2009.

broader type of causal reasoning (which is relevant to employment-law rules analyzing whether some act was performed "because of" *X* or whether some act "results from" *X*).

This Part concludes with some broader considerations about the nature of experimental jurisprudence, in light of this investigation. Although experimental jurisprudence often focuses on studying terms cited explicitly in law (e.g., "consent"), this is not its primary criterion. Rather, experimental jurisprudence might study any ordinary concept that has a counterpart of legal significance. For example, it studies the ordinary concepts of intent or cause to inform legal theorizing about intent or cause. But it also studies ordinary concepts that do not have counterparts that appear prominently as terms in jury instructions, for example, responsibility and the self. As such, the key question in identifying an area of potential experimental-jurisprudential study is not whether a term is cited explicitly in some legal rule but rather which concepts have jurisprudential significance.

2.1. *Significant Examples*

2.1.1. *Mental states (knowledge, recklessness, intent)*

From criminal law *mens rea* to the distinction between intentional and unintentional torts, mental states have great legal significance. What is knowledge, recklessness, or intent? Experimental jurisprudence has contributed to these legal questions by studying these ordinary concepts[88].

As a first example, consider the legal distinction between knowledge and recklessness. Professor Iris Vilares and her co-authors sought to test whether there is an ordinary distinction between knowledge and recklessness. To do so, they ran a neuroscientific experiment, evaluating whether different brain states were associated with different attributions of knowledge and recklessness[89].

They conducted an fMRI study involving a "contraband scenario"[90]. Participants evaluated different scenarios in which they could carry a suitcase—which might have contraband in it—through a security checkpoint[91]. The probability that the suitcase had contraband varied across different scenarios. In some scenarios, participants were completely sure that the suitcase had contraband (knowledge condition) while in others, there was merely a risk that the suitcase had contraband (recklessness condition)[92].

The study found that participants' evaluations of the two states (knowledge, recklessness) differed and were associated with different brain regions[93]. Moreover, the fMRI data predicted whether participants faced a knowledge or recklessness scenario[94]. The researchers took this as evidence that the legal concepts are, in part, running parallel to an ordinary distinction[95]. This finding suggests that the legal categories of knowledge and recklessness are actually founded on the ordinary notions.

Many other experimental studies have evaluated knowledge[96] and intent[97]. As one example, Markus Kneer and colleagues have found that ordinary people are sometimes sensitive to the

[88] See generally ROBINSON & DARLEY 1995; MALLE & NELSON 2003; NADELHOFFE 2006; NADELHOFFER & NAHMIAS 2011; MUELLER et al. 2012; MACLEOD 2016; KIRFEL & HANNIKAINEN 2022; KNEER & BOURGEOIS-GIRONDE 2017; GINTHER et al. 2014.

[89] VILARES et al. 2017.

[90] VILARES et al. 2017, 3223.

[91] VILARES et al. 2017, 3223.

[92] VILARES et al. 2017, 3223.

[93] VILARES et al. 2017, 3224.

[94] VILARES et al. 2017, 3224.

[95] VILARES et al. 2017,3226.

[96] See generally ROSE et al. 2019; MACHERY 2017.

severity of a side effect when judging whether it was produced intentionally[98]. Recall this Article's earliest hypothetical about life-saving negligence[99]. The company aimed to sell yachts, which would raise money for them and for charity, but that production decision had one bad side effect: creating pollution. Did the company intentionally pollute? Research shows that in these cases, people are more inclined to judge that the pollution is intentionally produced when it is deadly than when it is nonfatal[100]. In other words, the severity of a side-effect affects participants' attributions of intentionality.

These studies about ordinary concepts raise intriguing jurisprudential questions. In an approach similar to that of traditional jurisprudence, experimental-jurisprudence scholars evaluate the source of these judgments: Is severity sensitivity a performance error (a mistaken intuition in response to the thought experiment)? They also ask normative questions: Should the legal criterion of intentional action reflect this "severity sensitivity" feature of the ordinary concept? In response to this latter question, some argue no[101], while others have raised considerations in favor of yes[102]. As in traditional jurisprudence, these debates are not easily resolved. But learning more about the ordinary concept raises new and important questions about how law should understand knowledge, recklessness, intent, and other mental states.

2.1.2. *Consent*

As another example, consider the experimental jurisprudence study of consent. Exciting recent work in this area comes from Professor Roseanna Sommers, who has investigated the ordinary understanding of consent across a range of legal contexts. One important line of her work focuses on the relationship between deception and consent. «Under the canonical view, material deception vitiates consent»[103]. When someone's agreement is gained through deception about a material fact, there is not valid consent. For example, imagine that I offer to sell you a car with "only ten thousand miles", and you agree. In reality, the car has one hundred thousand miles. Your agreement would not be consensual if you relied upon my misrepresentation.

Sommers's experimental jurisprudence of consent has found, however, that ordinary people often attribute "consent" in circumstances in which there has been significant deception[104]. In one of Sommers's experimental hypotheticals, a single woman does not desire to sleep with married men. The woman asks a potential partner about his marital status, and he lies, saying that he is not married[105]. The woman then agrees to sleep with him. In this case, the overwhelming majority of participants judged that the woman did «give consent to sleep with [the man]»[106]. Despite deception regarding a very important fact (the man's marital status), most people attribute consent[107].

[97] See generally MALLE & NELSON, 2003; NADELHOFFER 2005; NADELHOFFER 2006; LEVINE et al. 2018; MIKHAIL 2009; KOBICK & KNOBE 2009; KOBICK, 2010; PROCHOWNIK et al. 2020; SHEN et al. 2011.

[98] See generally KNEER (mns); KNEER et al. (mns).

[99] See above, § 1.

[100] See generally KNEER & BOURGEOIS-GIRONDE 2017. Cf. generally KNOBE 2003.

[101] See, e.g., KNEER & BOURGEOIS-GIRONDE 2017, 140.

[102] See, e.g., TOBIA (mns), 68.

[103] SOMMERS 2020.

[104] SOMMERS 2020. See also generally DEMAREE-COTTON & SOMMERS 2022; SOMMERS & BOHNS 2019; WILKINSON-RYAN 2014; ROBINSON & DARLEY 1995. Empirical work suggests that people are afraid to decline police officers' requests and feel pressure to consent to searches. See generally, e.g., NADLER & TROUT 2012.

[105] SOMMERS 2020, 2252.

[106] SOMMERS 2020.

[107] Note that the same finding arises for various types of deception and various question types: did the woman "let [the man] have sex with her", or did she "give [the man] permission to have sex with her". SOMMERS 2020, 2323.

Given the crucial role that consent plays across tort, criminal, and contract law, these findings raise broad questions[108]. Is the legal notion of consent consistent with the ordinary concept? Of course, this experiment does not settle this complex jurisprudential question. But it does provide the longstanding jurisprudential debate with unique insights. For example, Sommers's further studies suggest that the source of this intuition is something about what seems to be an "essential" part of the agreement[109]. Deception about the essence of the contract or arrangement vitiates consent, but deception about less essential features does not. Here again, this data does not settle the debate about how law should identify the right criteria of legal consent. But it provides important new insight into a jurisprudential analysis; if our intuitions about what seems consensual depend on our view of what is essential to the agreement (rather than what's merely material), does that give us any reason to revise the legal notion of consent?

2.1.3. *Causation*

As a third example, turn to experimental jurisprudence of causation. Jurisprudence has long studied «the plain man's notions of causation»[110]. Experimental jurisprudence makes new progress on that traditional inquiry[111].

When thinking about potential causes, there are several plausible features of significance. One is the potential cause's necessity: Would the outcome have occurred if not for the cause? A second is its sufficiency: Was the cause enough to bring about the outcome?

Studies in cognitive science have shown that the ordinary concept of causation is informed by both of these features[112]. Professor James Macleod has recently conducted important work in this area, designing a study to test whether this feature of the ordinary concept also manifests in people's judgments about cases of legal causation[113]. He considered three legal examples: a scenario asking whether death "result[ed] from" a certain drug, a scenario asking whether an employee was terminated "because of" his age, and a scenario asking whether someone was assaulted "because of" his religion[114].

Participants considered one of four types of cases, each of which varied whether the cause was necessary or sufficient to bring about an outcome:

(i) necessary and sufficient,
(ii) necessary but not sufficient,
(iii) sufficient but not necessary, or
(iv) not sufficient and not necessary[115].

This relates to the discussion, *infra*, of classes of concepts. Many take experimental jurisprudence to focus primarily on concepts cited by law (e.g., consent). But experimental jurisprudence studies a wide range of concepts and classes of concepts that have legal significance.

[108] HURD 1996, 123 («[C]onsent turns a trespass into a dinner party; a battery into a handshake; a theft into a gift; an invasion of privacy into an intimate moment; a commercial appropriation of name and likeness into a biography»).

[109] SOMMERS 2020, 2301.

[110] HART & HONORÉ 1985 1.

[111] See generally SPELLMAN 1997; GREENE & DARLEY 1998; SOLAN & DARLEY 2001; PRENTICE & KOEHLER 2003; LAGNADO & CHANNON 2008; HITCHCOCK & KNOBE 2009; SPELLMAN & HOLLAND 2010; ALICKE et al. 2011; KOMINSKY et al. 2015; HENNE et al. 2017; MACLEOD 2019; KIRFEL & LAGNADO 2021; LAGNADO et al. GERSTENBERG 2017; WILLEMSEN & KIRFEL 2019; GÜVER & KNEER 2022.

[112] See KNOBE & SHAPIRO 2021, 235.

[113] See generally MACLEOD 2019.

[114] MACLEOD 2019, 995.

[115] MACLEOD 2019, 996.

For example, in the drug case, a protagonist buys three different drugs, one from each of three different dealers. The drug that participants were asked about may have been (i) the only drug potent enough to kill by itself, (ii) the only drug potent enough to kill when combined with either of the others, (iii) one of three drugs potent enough to kill by itself, or (iv) one of several drugs potent enough to kill when combined with any other.

That experiment made two striking discoveries. First, people attributed causation in cases in which the cause was not a but-for cause (i.e., (iii) and (iv)). Second, sufficiency had an important effect on ordinary judgments of causation[116]. These findings cohere with recent cognitive-scientific findings that ordinary judgments of causation are influenced by both necessity and sufficiency[117].

2.1.4. *Law*

The preceding examples involve concepts that law cites explicitly: intent, consent, and cause. But experimental jurisprudence also studies some ordinary concepts that have a less explicit legal connection. This next example serves as a proof of this concept.

Consider the concept of law itself. Professors Raff Donelson and Ivar Hannikainen examined how ordinary people understand law. In an important series of experiments, they tested whether ordinary people (and legal experts) endorsed Fuller's conditions of the inner morality of law[118].

In one experiment, they investigated whether people think that law has to be consistent, general, intelligible, public, and stable. That study asked two questions. First, are these conditions of law seen as necessary? Second, do laws in practice observe these principles? The responses—from ordinary people and experts alike—are fascinating. There was strong endorsement of the conditions as principles of law. Yet participants also agreed that there are some laws that are (in fact) not prospective, stable, intelligible, or general[119]. A recent cross-cultural collaboration has replicated the results from this study across eleven different countries[120].

As with the other experimental-jurisprudential studies, the lesson here is not for the law to simply reflect the ordinary notion. (Donelson and Hannikainen do not recommend that law have a self-contradictory nature.) Rather, the experiment adds insight to traditional jurisprudential debates: What features do we believe laws must and do have? And what explains those judgments?

2.2. *A Framework for identifying experimental jurisprudence*

This summary, although not entirely brief, has only scratched the surface. Before turning to the next Part, it is worth elaborating what connects these very diverse projects. One common view is that XJur studies ordinary language and concepts that are invoked explicitly in law. Where the law invokes intent or consent, XJur studies the ordinary concept of intent or consent.

XJur does study those concepts but not because of their explicit legal citation. A more useful criterion is whether the legal object of study has an ordinary-language or ordinary-conceptual counterpart. This difference is clarified in Figure 1 below.

[116] MACLEOD 2019, 999 f.

[117] Recent cognitive science has also highlighted another significant factor in ordinary judgments of causation: (ab)normality. See generally KNOBE & SHAPIRO 2021; SUMMERS 2018; SOLAN & DARLEY 2001.

[118] DONELSON & HANNIKAINEN 2020, 10.

[119] DONELSON & HANNIKAINEN 2020, 18.

[120] HANNIKAINEN et al. 2021, 10.

FIGURE 1: A HEURISTIC FOR IDENTIFYING EXPERIMENTAL JURISPRUDENCE

	Is the object of study referred to explicitly in legal materials (e.g., judicial opinions, statutes, or jury instructions)?		
Does the legally significant object of study have an ordinary counterpart?		Frequently	Rarely
	Yes	Cause; consent; intent; reasonable	Numerical identity; responsibility; the concept of law
	No	Parol evidence rule; stare decisis	Legal positivism; soft law

It may seem that the only objects of XJur's study are those that are invoked explicitly in legal materials. This is partly a result of the (mistaken) view that experimental jurisprudence is principally concerned with predicting how judges or juries will decide cases. Of course, many experimental-jurisprudence studies have focused on those concepts. Unsurprisingly, many important legal concepts also appear with regularity in real legal texts.

However, this criterion—what terms and phrases appear explicitly in legal materials—will not direct us to every useful experimental-jurisprudential project. Many other notions are rarely invoked explicitly in law but are crucial jurisprudential concepts nevertheless. These include concepts of personal identity and the self, (moral) blame and responsibility, and the concept of law itself.

To take just one example, consider the concept of "numerical identity". This is a crucial legal concept[121]. It is implicated as a necessary criterion of most interesting diachronic legal relations. When are you bound by that contract? Only when you are the same person as one of the parties who originally agreed to it. When does he deserve criminal punishment? Only when he is the same person as the one who committed the crime. Yet numerical identity is hardly ever cited explicitly by courts[122].

The same is true of other important legal concepts: although courts cite "law," it is rare for a case to turn on the concept of law or on the concept of soft law. That said, there is not much to learn from studying the ordinary concept of the parole evidence rule, insofar as it has no ordinary-language counterpart[123]. The better criterion for identifying useful experimental-jurisprudential inquiries is whether the legal object has a corresponding ordinary-language counterpart. Because experimental jurisprudence often focuses on ordinary cognition, these ordinary concepts are the most valuable to study.

[121] See generally MATSUMURA 2014; TOOMEY 2022.

[122] A Westlaw search, across all state and federal jurisdictions, returned ten cases.

[123] Of course, if these legal concepts are composed of concepts that have ordinary counterparts, then studying the lay view of those counterparts might prove useful. A good example is the Hand Formula. Although most ordinary people do not speak about or even know the Hand Formula, there may be useful experimental-jurisprudential work in the study of its components (studying how ordinary people weigh burdens of prevention against probability and severity of harm). This reflects an important feature of Figure 1. The top row (reflecting that the legal concept has an ordinary counterpart) is a useful heuristic for finding experimental-jurisprudential projects, but it is not a necessary requirement.

The more useful criterion is whether a counterpart of some ordinary concept plays an important role in the law. This is true of ordinary concepts whose counterparts are cited explicitly (such as the ordinary notion of what is reasonable, consensual, or self-defense) but also ones that are not explicitly cited (such as the ordinary notion of numerical identity or of the concept of law itself).

Before turning to the next Part, there is one final wrinkle. XJur often focuses on specific legally significant concepts, such as ownership. But sometimes it focuses on broader classes of concepts, for example, studying how law treats concepts that admit of a broad range of potential category members[124].

One example is Macleod's important work on causation[125]. His studies examine lay judgments of vignettes that use different phrases—such as "because of" and "result from"—that are taken to reflect ordinary causal reasoning. Another example is Sommers's work on consent[126]. Those experiments study judgments about consent and also whether a person "willingly" acted or gave "permission"[127].

These experiments report similar patterns of judgment across vignettes using these varied terms, and they reveal something more general about ordinary causal or consensual reasoning rather than simply something about some more specific term (e.g., "consent").

3. *Applications*

The preceding Parts have introduced experimental jurisprudence and summarized some of its projects. This Part turns to two more concrete applications, identifying areas in which further XJur work might be particularly useful. The first application concerns the rise of ordinary meaning in legal interpretation, particularly in originalist and textualist legal theory. The second concerns the New Private Law. These serve as a useful pair of illustrations: one associated with public law, one with private law; one focused on interpretation of legal texts, one participating in a broader common law tradition; one debated frequently in the courts, one debated largely in legal-theory circles; but both of tremendous jurisprudential impact and importance. The closer study of these two areas reveals important and diverse ways in which experimental jurisprudence offers unique insights into the most important modern jurisprudential debates.

3.1. *Ordinary meaning*

One of the most significant trends in legal theory—and practice—is the growing importance of ordinary meaning in interpretation.[128] When interpreting legal texts, scholars and courts look to the ordinary meaning or original public meaning of the language[129]. This approach is strongly associated with textualist and originalist theories of statutory[130] and constitutional interpretation[131], but ordinary meaning also plays a significant role within a broad range of other interpretive theories[132] and in the interpretation of other legal texts, including contracts[133] and treaties[134].

[124] See generally TOBIA 2020.
[125] See generally MACLEOD 2019.
[126] See generally, e.g., SOMMERS 2020.
[127] SOMMERS 2020.
[128] See SLOCUM 2015, 4.
[129] See generally SOLUM 2019 (mns); see also NOURSE 2019, 681.
[130] See generally NOURSE 2019.
[131] See generally SOLUM 2019 (mns).
[132] See generally, e.g., SLOCUM 2015; TOBIA 2020.
[133] See generally, e.g., MOURITSEN 2019.

Ordinary meaning's significance varies across legal theories. For example, new textualist theories typically understand a text's communicative content to constrain interpretation and ordinary meaning as a central determinant of communicative content[135]. Pluralistic theories might treat ordinary meaning as one of several interpretive considerations alongside intent, purpose, and even the consequences of a given interpretation.

There are debates about what role ordinary meaning should play in interpretation and what exactly "ordinary meaning" means[136]. But many agree that ordinary meaning is closely connected to empirical facts about how ordinary people understand language[137]. That is, a legal text's ordinary meaning is not necessarily the meaning that its drafters intended it to have or the meaning that would allow the text to achieve the best results. Rather, investigation into a text's ordinary meaning is an investigation into facts about how ordinary people would actually understand the language.

The empirical aspect of ordinary meaning can be traced to the most common justifications for an ordinary-meaning interpretive criterion. There is, of course, great debate about whether and why ordinary meaning should be a legal-interpretive constraint or consideration. But one of the most common justifications invokes rule-of-law values.

These rule-of-law values took center stage in the recent case *Bostock v. Clayton County*[138], a landmark decision protecting gay and transgender persons from employment discrimination under Title VII[139]. Justice Neil Gorsuch's majority opinion was grounded in a textualist analysis of ordinary meaning, holding that Title VII's prohibition against adverse employment actions taken «because of [an individual's] sex" prohibits adverse employment actions taken against persons for being gay or transgender[140]. Consider Justice Gorsuch's discussion of ordinary public meaning:

> «This Court normally interprets a statute in accord with the ordinary public meaning of its terms at the time of its enactment. After all, only the words on the page constitute the law adopted by Congress and approved by the President. If judges could add to, remodel, update, or detract from old statutory terms inspired only by extratextual sources and our own imaginations, we would risk amending statutes outside the legislative process reserved for the people's representatives. And we would deny the people the right to continue relying on the original meaning of the law they have counted on to settle their rights and obligations»[141].

Here, Justice Gorsuch appeals to ordinary meaning as an interpretive criterion that promotes the values of notice and reliance. Interpreting a legal text in line with its ordinary meaning helps ensure that ordinary people can rely on law. Importantly, these rule-of-law justifications reinforce the significance of understanding ordinary meaning empirically. There are facts about what ordinary people would take from statutory language, and interpretation grounded in reliance and fair notice should concern itself with how ordinary people would in fact rely upon or be notified by the legal text.

These rule-of-law justifications are shared among other textualist judges. Consider Justice Brett Kavanaugh's *Bostock* dissent: «Judges adhere to ordinary meaning for two main reasons: rule of law and democratic accountability»[142].

[134] See generally, e.g., SLOCUM & WONG 2021.
[135] See generally, e.g., SOLUM 2019 (mns).
[136] See generally, e.g., FALLON 2015.
[137] See generally, *e.g.*, TOBIA 2020. See also BARNETT 2011, 66: «It cannot be overstressed that the activity of determining semantic meaning at the time of enactment required by the first proposition is *empirical*, not normative» (emphasis in original; citing WHITTINGTON 1999, 6).
[138] 140 S. Ct. 1731 (2020).
[139] See generally *ibid.*
[140] *Ibid.*, 1745.
[141] *Ibid.*, 1738.
[142] *Ibid.*, 1825 (Kavanaugh, J., dissenting).

However, a shared commitment to ordinary meaning does not necessarily resolve all interpretive debate[143]. In *Bostock*, Justices Gorsuch and Kavanaugh both interpreted Title VII in line with what they considered to be its ordinary meaning, but Justice Gorsuch wrote for the majority and Justice Kavanaugh in dissent. That disagreement is the subject of scholarly debate[144]. This Section does not propose a resolution to that debate, but it does use the *Bostock* scholarship as an illustration of what experimental jurisprudence can contribute to jurisprudential study of ordinary meaning.

The key interpretive question in *Bostock* concerned the language of Title VII, which prohibits adverse employment actions taken «because of [an] individual's race, color, religion, sex, or national origin»[145]. One plaintiff employee was fired for being gay, another for being transgender. So *Bostock*'s interpretive question was whether each gay and transgender employee was fired «because of [that individual's] sex»[146].

Justice Gorsuch's majority opinion held that yes, each employee was fired "because of sex"[147]. The reasoning turned on the interpretation of "because of". Justice Gorsuch reasoned that the ordinary meaning of Title VII's "because of" language reflects a but-for test:

> «[T]he ordinary meaning of 'because of' is 'by reason of' or 'on account of.'" In the language of law, this means that Title VII's "because of" test incorporates the simple and "traditional" standard of but-for causation. That form of causation is established whenever a particular outcome would not have happened "but for" the purported cause. In other words, a but-for test directs us to change one thing at a time and see if the outcome changes. If it does, we have found a but-for cause»[148].

According to Justice Gorsuch, an intuitive application of this test suggests that the gay and transgender employees were fired "because of" their sexes. Consider, for example, a transgender woman employee: a person assigned a male sex at birth[149] who identifies as a woman. An antitransgender employer fires her. Now, «change one thing at a time and see if the outcome changes»[150]. Suppose that the employee was instead assigned a female sex at birth and (still) identified as a woman. In this case, the antitransgender employer would not fire her. The antitransgender firing turns entirely on the employee's sex; that is, the employee's sex was a but-for cause of the firing. Thus, firing an individual for being transgender is to fire them "because of" sex[151].

Justices Kavanaugh and Samuel Alito wrote separate dissenting opinions. Each agreed with Justice Gorsuch's starting point—a textualist inquiry into the ordinary meaning of Title VII[152]—but disagreed with details of the analysis. Specifically, both Justices Kavanaugh and Alito suggested that the ordinary meaning of Title VII does not prohibit firing a gay or transgender employee. As Justice Alito put it:

«Suppose that, while Title VII was under consideration in Congress, a group of average Americans decided to read the text of the bill with the aim of writing or calling their representatives in Congress and conveying their approval or disapproval. What would these ordinary citizens have taken

[143] See generally, e.g., GROVE 2020; NOURSE & ESKRIDGE 2021.
[144] See generally, e.g., DEMBROFF et al. 2020; DESEI 2022; EIDELSON 2022; ESKRIDGE et al. 2021; GROVE 2020; KOPPELMAN 2020; MACLEOD 2022; TOBIA & MIKHAIL 2021b.
[145] 42 U.S.C. § 2000e-2(a)(1).
[146] *Bostock*, 140 S. Ct, 1739.
[147] *Ibid.*, 1741.
[148] *Ibid.*, 1739 (citations omitted) (quoting Univ. of Tex. Sw. Med. Ctr. v. Nassar, 570 U.S. 338, 350 (2013)).
[149] The *Bostock* opinions—majority and dissenting—characterize sex as "biological sex". *See ibid.*,1739.
[150] *Ibid.*, 1739.
[151] *Ibid.*, 1741.
[152] See *ibid.*, 1755 (Alito, J., dissenting); *ibid.*, 1824 (Kavanaugh, J., dissenting).

"discrimination because of sex" to mean? Would they have thought that this language prohibited discrimination because of sexual orientation or gender identity?»[153].

Justice Alito assumes that the answer is no. Ordinary citizens would not understand antigay or antitransgender employment actions to be ones taken «because of [the individual employee's] sex»[154].

There are tremendous similarities between the majority and the dissents. All are committed to ordinary meaning, and all justify that approach via rule-of-law values like fair notice and publicity. All are also committed to an empirical conception of ordinary meaning. For each Justice, the key question is what ordinary people would actually understand.

Now, one could proceed by intuition alone, making the best guess about what most ordinary people would say. But those working in experimental jurisprudence have begun to address these empirical questions about ordinary meaning with empirical studies. For example, Macleod has studied how ordinary people today understand language like "because of"[155].

One of Macleod's more recent studies provided participants with the very question in *Bostock*[156]. Participants received a series of questions concerning antigay and antitransgender employers who fired gay or transgender employees. The survey asked: Was the employee fired "because of his sex?"[157] Macleod found that the majority of participants actually agreed in both cases that firing the gay and transgender employee was firing them because of their sex[158].

That experiment tests the empirical question raised by Justices Kavanaugh and Alito. Those dissenting opinions suggested that the answer was no—that ordinary people did not understand the language that way, certainly not in 1964 and probably not today[159]. Justice Kavanaugh's dissent notes that there is likely no difference in the ordinary meaning of Title VII between 1964 and today[160]. Macleod's survey suggests that here, the Justices' individual intuitions may not be a perfect guide to ordinary meaning.

A second experimental study provides further support to rebut the empirical assumptions of Justices Kavanaugh and Alito[161]. That second study presented participants with similar scenarios to those used in Macleod's study. It also included two other hypotheticals with a similar structure: Is someone fired for being in an interracial marriage fired "because of his race"? And is someone fired for being pregnant fired "because of her sex"? The sexual-orientation scenario began:

> «Mike was an employee at an Italian restaurant. Mike had worked there for ten years. Mike was a gay man, who was married to another man. One day, Mike's boss learned that Mike is gay. Two days later, Mike's boss fired him, saying "I'm sorry Mike, I just don't think having gay employees is good for business"»[162].

In one version, participants were presented with a question that mimicked Justices Alito and Kavanaugh's approach. The question for the sexual-orientation case was: "Statement: Mike was

[153] *Bostock*, 140 S. Ct. at 1767 (Alito, J., dissenting). Here, Justice Alito's remarks are ambiguous between referencing the original public meaning and the public's original expected applications. Most new originalists and new textualists are concerned with original public meaning, which is not limited to original expected applications. As such, this Section interprets Justice Alito's opinion in line with the (more common) modern theory focused on public meaning.

[154] *Ibid*., 1767.

[155] See generally MACLEOD 2019.

[156] See generally MACLEOD 2022.

[157] MACLEOD 2022, 19-28.

[158] MACLEOD 2022, 19-28.

[159] The originalist aspect of this interpretation problem raises many more interesting issues, which cannot be addressed here. It is worth noting, however, that some scholarship actually suggests the opposite: "sex" may have had a *broader* meaning in 1964 than today. See ESKRIDGE et al. 2021.

[160] *Bostock*, 140 S. Ct. at 1825 (Kavanaugh, J., dissenting).

[161] See generally TOBIA & MIKHAIL 2021.

[162] TOBIA & MIKHAIL 2021, 476.

fired because of his sex. [Yes or No]"[163]. This question also clarified that "sex" referred only to "biological sex"[164].

The study replicated Macleod's findings. The majority of participants endorsed that antigay and antitransgender firings were because of sex[165]. However, results were more mixed for the other cases. The majority endorsed that anti-interracial marriage and antipregnancy firings were not "because of" race and sex, respectively[166].

That study also presented other participants with a case following Justice Gorsuch's framing. Do ordinary people agree that sex is a but-for cause of antigay and antitransgender firings? The sexual-orientation scenario under that framework concluded instead with this question:

"Imagine that the above scenario were different in exactly one way: Mike was not a man but was instead a woman named 'Michelle,' who is married to a man. Imagine that everything else about the scenario was the same. Would Michelle still have been fired? [Yes or No]".

Here, participants were overall more inclined to find Title VII discrimination. Across all four cases (sexual orientation, transgender, interracial marriage, and pregnancy), the majority of participants chose no, implicitly identifying the Title VII factor (sex or race) as a but-for cause[167].

The details of that experiment to illustrate that one function of experiments is to help evaluate empirical claims implicit in legal theory. The experimental evidence supports Justice Gorsuch's assumption: ordinary people understand sex as a but-for cause of sexual-orientation discrimination. The evidence does not so strongly support Justices Kavanaugh and Alito's assumptions: ordinary people do not agree that antigay and antitransgender firings are not firings "because of [the individual's] sex".

However, the studies offer something beyond this contribution. They help clarify a broader jurisprudential point. For example, the authors of the second experimental study argued that these empirical results support the significance of "a distinction between two types of empirical textualism"[168]. One is "ordinary criteria" textualism, which is reflected in Justice Kavanaugh and Justice Alito's approach to *Bostock*. That view equates ordinary meaning with ordinary understanding. However, as the experiments suggest, the ordinary understanding of Title VII's language may differ from what Justice Kavanaugh and Justice Alito assume. But Justice Gorsuch's textualism in *Bostock* reflects a different theory, "legal criteria textualism". On that view, textualism combines ordinary understanding of statutory terms "with both their previously-established legal meanings and their legal entailments"[169]. This is simply a different type of "empirical textualism". And it is one supported by empirical evidence: Justice Gorsuch's assumption about ordinary people's application of a but-for test is supported by the data.

This distinction—between ordinary- and legal-criteria textualism—is a possibility that could be contemplated and elaborated from the armchair, without empirical evidence. But experimental work helps crystallize the significance of the distinction. Ordinary-criteria textualism and legal-criteria textualism are both theories committed to ordinary meaning in legal interpretation, and both make empirical claims about ordinary people's understanding. It is (in theory) possible that these theories could lead to divergent results. And in the *Bostock* case, the two approaches do

[163] TOBIA & MIKHAIL 2021, 478.
[164] TOBIA & MIKHAIL 2021, 476 fn.74: «Please read the scenario and tell us whether you agree ("yes") or disagree ("no") with the following statement. For the purpose of this question, "sex" should be understood to mean biological sex, per Merriam-Webster's dictionary: "either of the two major forms of individuals that occur in many species and that are distinguished respectively as female or male especially on the basis of their reproductive organs and structures"».
[165] TOBIA & MIKHAIL 2021, 480.
[166] TOBIA & MIKHAIL 2021, 480.
[167] TOBIA & MIKHAIL 2021, 480.
[168] TOBIA & MIKHAIL 2021, 486.
[169] TOBIA & MIKHAIL 2021, 486.

diverge. That divergence is best illuminated by empirical data[170].

This legal-theory distinction, supported by experimental study, is also one that could inform broader jurisprudential debates. Recall that textualists often appeal to fair notice and reliance as justifications for an ordinary-meaning approach to interpretation. By demonstrating the possible divergence of ordinary-criteria and legal-criteria textualism, the experimental study can also be seen as one that raises the question about the concept of publicity and other rule-of-law values.

Specifically, does publicity (as a rule-of-law value) require that law reflect ordinary people's understanding of the language in the text? Or does publicity require that legal criteria are applied consistently with ordinary people's understanding of the application of those criteria? Experimental study itself provides no answer to this question, but it helps articulate it.

Surveys are a very attractive tool in ordinary-meaning debates. As courts and commentators continue to interrogate the ordinary meanings of legal texts, it may become even more tempting to outsource legal interpretation to surveys. But experimental jurisprudence avoids such an incautious use of surveys. Experiments will not tell us in simple terms what the law should be. But they can provide insight into the truth of empirical claims made by legal theories. Moreover, experimental methods can make jurisprudential contributions, calling attention to new theoretical distinctions of practical and jurisprudential significance.

3.2. *The New Private Law*

Like the rise of ordinary meaning, the New Private Law is an influential and impressive movement in modern legal theory[171]. A central theme of the New Private Law is the rejection of reductive, purely instrumental accounts of private law. Professor John Goldberg articulates this vision—a private-law theory that chooses:

> «[T]o stick close to everyday practices and to be wary of concepts, categories, or methods that claim for themselves a certain kind of essential validity or primacy. [This view] supposes that reality is complex and that it will not advance the cause of knowledge to assume that one comes to understand reality by stripping away superstructure to get to base. [It] calls for a patient exploration of the many facets of a phenomenon or problem»[172].

Some might see experimental jurisprudence as a reductive force, one in opposition to this vision of the New Private Law. According to that view, experimentalists simply use surveys to compute answers to private-law-theory questions[173]. Experimental jurisprudence does not typically endorse such reductive analysis. To the contrary, XJur work shares the New Private Law's appreciation for law's complexity. XJur does not take psychology (or legal history or economics or moral philosophy) as the discipline with primacy. And it does not take experimentation (or cost-benefit analysis or moral theory) as the only essentially valid tool.

A second central theme of the New Private Law is that it «takes private law concepts and categories seriously»[174]. It works to appreciate the nuanced «conceptual structure of the law»[175],

[170] Specifically, the empirical data suggest that both approaches favor the plaintiffs in *Bostock*. However, the two approaches are not equivalent. For example, Justice Gorsuch's legal-criteria textualism more strongly supports the gay employee. Both approaches strongly support the transgender employee. *Id.* But see BERMAN & KRISHNAMURTHI 2021.

[171] See generally GOLD et al. 2020.

[172] GOLDBERG 2012, 1650.

[173] See generally, *e.g.*, BEN-SHAHAR & STRAHILEVITZ 2017 (proposing experiments to solve problems of contract interpretation).

[174] GOLD 2020, xvi.

[175] GOLDBERG 2012, 1652.

rejecting the realist critique that law's central concepts are «fictions, nonsense»[176]. This conceptualism also takes seriously people's ordinary concepts. Central figures in the New Private Law even suggest that legal concepts generally should reflect features of ordinary concepts. As Professors Andrew Gold and Henry Smith put it: «The set of legal concepts benefits from its congruence with relatively simple local forms of conventional morality [...] Certainly, contract law can diverge from the morality of promising, just as legislation can go beyond corrective justice. Nevertheless, the ability to draw on simple local morality is an important starting point»[177].

Experimental jurisprudence agrees. Descriptively, many legal concepts share features of the ordinary concept[178]. This supports the "folk-law thesis"[179]. It would be bizarre for law's concepts of good faith, reasonableness, cause, duty, or wrong to be entirely untethered from corresponding ordinary concepts. Some theorists working in experimental jurisprudence also endorse Gold and Smith's normative suggestion that legal concepts benefit from reflecting features of the corresponding ordinary one. Thus, the fact that an ordinary concept has a feature provides a (defeasible) reason that the corresponding legal concept should share that feature[180].

So, there is common ground between experimental jurisprudence and the New Private Law. Experimental jurisprudence can contribute to the New Private Law a richer set of data and questions for jurisprudential debate. As Part 2's experimental studies reveal, ordinary concepts and moral reasoning are not always "simple"[181], intuitive, or obvious. What the seminar room agrees is "ordinarily wrongful" may not reflect what, in fact, all ordinary people understand to be wrongful.

XJur shares the New Private Law's general commitment to understanding ordinary and legal concepts and the belief that such study is truly complex. For example, in assessing the relationship between contract and promise, there is still much to learn about both (legal) contract and (ordinary) promise. Experimental methods have uncovered important insights about lay intuitions of contract[182] and ordinary promising—many of which are not simple or obvious from the armchair[183].

Here again, the New Private Law might be skeptical of empirical approaches to legal scholarship, which are often reductive, inspired by legal realism, and focused on predicting how judges really decide cases—how things really work when "getting down to brass tacks"[184]. Some empirical and psychological studies support criticism of traditional assumptions of reductive and instrumental approaches. For example, legal psychology can illuminate behavioral realities that conflict with standard assumptions of law and economics models[185].

Modern experimental jurisprudence takes a different approach, moving away from the "New Realism"[186], the old experimental jurisprudence (e.g., testing law's effects)[187], and the psychological literature on heuristics and biases. XJur does not primarily study laypeople as potential jurors with choices to model, biases to correct, and decisions to nudge. Instead—like the New Private Law—XJur sees the study of ordinary concepts as central to legal theory. As the New Private Law puts it, laypeople are not merely jurors, but also «norm articulators [...] often

[176] GOLDBERG 2012, 1652.
[177] GOLD & SMITH 2020, 504 f.
[178] See generally TOBIA 2021.
[179] See above, fn. 87 and accompanying text.
[180] See, e.g., above, § 2.
[181] GOLD & SMITH 2020, 505.
[182] See generally, e.g., HOFFMAN & WILKINSON-RYAN 2013.
[183] See generally, e.g., VANBERG 2008; MISCHKOWSKI et al. 2019; EDERER & STREMITZER 2018; STONE & STREMITZER 2020.
[184] GOLDBERG 2012,1642. For an example of empiricism and realism, see generally MILES & SUNSTEIN 2008.
[185] For a helpful example, see WILKINSON-RYAN 2020, 125.
[186] WILKINSON-RYAN 2020, 125.
[187] See generally BEUTEL 1971.

charged with interpreting [what] counts as 'reasonable'»[188]. This observation also supports a reply to Professor Jiménez's suggestion[189] to only count the intuitions of those who contribute to law's content: Ordinary people sometimes contribute to law's content. XJur and the New Private Law both (correctly) understand laypeople not as mere legal objects but as central members of our legal community, poised to contribute meaningfully to law and legal theory[190].

Experimental jurisprudence and the New Private Law agree «that there is often at least a family resemblance between legal and extralegal concepts and norms that bear on questions of personal interaction»[191] and that the nature of that resemblance is worth exploring. Both tend to reject reductive instrumentalism; they agree that legal concepts should not always reflect ordinary ones. Instead, legal theory should grapple with the complex nature of its concepts, and that grappling process should typically include study of the corresponding and constituent ordinary concepts.

Again, neither the New Private Law nor XJur will simply take the ordinary concepts as constitutive of legal ones. Yet both programs recognize the process of studying ordinary concepts as an essential part of jurisprudence. Consider Professors Goldberg and Benjamin Zipursky's description of their task in *Recognizing Wrongs*:

> «We come to the job of explaining the common law somewhat like one trying to explain how the members of a community use their language. The goal is to make explicit the various patterns of thought and conduct that animate this area of the law. If it turns out that many of the concepts and principles utilized in this area have the same character as, or a character very similar to, those which are utilized in non-legal discourse about how one ought (morally) to conduct oneself—indeed, if it turns out that some of the concepts are identical—that is something to be acknowledged, not hidden from view»[192].

This question—how do the concepts in law compare to the concepts in nonlaw—is essential to the New Private Law. Answering that question requires deep knowledge of law and nonlaw. Of course, we all have some knowledge of ordinary concepts such as reasonableness, causation, consent, intent, duty, and wrongfulness. But even those ordinary notions are complex, calling for study from more than one method. Introspection, thought experimentation, and moral theorizing can provide tremendous insight into those ordinary concepts. So too can empirical methods. And as 2 demonstrates, some empirical insights are unique—inaccessible via individual introspection or thought experimentation.

XJur appears complementary to the New Private Law in part because the New Private Law is admirably honest about the project's commitments, the complexity of law, and the multifaceted inquiry that jurisprudence calls for. Consider again Goldberg and Zipursky:

> «Publicity, notice, generality, prospectivity, and the other values that Fuller emphasized seem lacking in a system that relies on judges to articulate rules and principles on a case-by-case basis instead of stating them in canonical form in a code or statute book. [...] There is another way in which tort law achieves a kind of fairness in operation by means apart from Fullerian methods. [...] Part of what it means for tort law to be '"common law" [...] is that the wrongs recognized by tort law are, in their substance, drawn from everyday life rather than constructed de novo by judges in aid of some sort of social engineering project»[193].

[188] GOLDBERG 2012, 1657.
[189] JIMÉNEZ 2021.
[190] GOLDBERG 2012, 1656.
[191] GOLDBERG 2012, 1656.
[192] GOLDBERG & ZIPURSKY 2000, 79.
[193] GOLDBERG & ZIPURSKY 2000, 207 s.

XJur is well-positioned to contribute to debates about publicity, notice, and rule-of-law values[194]. But there is one further (and much more general) justification of XJur and its focus on laypeople—one that Goldberg and Zipursky suggest above: experimental jurisprudence is actually called for by one of the most general and longstanding projects within jurisprudence. One implication of this more general justification is that the experimental-jurisprudence approach is not limited to only those jurisprudential debates premised on publicity or democracy or to situations where there is also a relevant jury instruction referring explicitly to reasonableness or consent.

That longstanding project is the determination of our appropriate legal criteria. For example, what are law's criteria of wrongs, causation, reasonableness, and intent? Countless traditional-jurisprudential projects address this type of question, and most often they address it as the New Private Law does, by considering *everyday* life. The foundation of jurisprudence is not "Judge Hercules", but the ordinary person:

> «Holmes's [...] understandable disdain for the pedantic moralist unfortunately led him to pose a false dilemma between the pedant and the bad man. Missing from his analysis is the ordinary person, the lawyer who counsels this person, and the judge who understands, applies, and crafts the law imagining that her legal community expects her to take this perspective seriously»[195].

Experimental jurisprudence provides unique insight into exactly this question: How does the ordinary person understand what is consensual, causal, reasonable, or intentional?

Although this Section has focused on the New Private Law, XJur's usefulness extends more broadly, to a range of debates in public- and private-law theory. Most legal theorists—not just those in the New Private Law—recognize the crucial connection between law and ordinary people.

As one more example, consider a seminal article in traditional jurisprudence: Professor Gregory Keating's *Reasonableness and Rationality in Negligence Theory*. Keating argues that tort negligence is—and should be—grounded in reasonable (not rational) risk imposition. Moreover, Keating argues, tort reasonableness is better explained by social-contract theory than by law and economics[196].

Perhaps surprisingly, that nonempirical work of jurisprudence begins with a reflection about our ordinary cognition: «Latent in our ordinary moral consciousness, and manifest in philosophical reflection, is a distinction between reasonableness and rationality»[197]. This appeal to ordinary cognition is not merely rhetorical or motivational. Throughout the article, Keating emphasizes the central role that ordinary concepts play in the jurisprudential argument: "[B]ecause negligence law assigns paramount importance to the concept of reasonableness, it receives stronger support from social contract theory than from economics»[198].

Not all of Keating's arguments depend on empirical facts about the ordinary concept of reasonableness, but this first one reflects an important and often overlooked mode of jurisprudence. For legally significant concepts that have an ordinary counterpart, a central question is: What is that ordinary concept? This is a jurisprudential question.

Of course, the legal criteria are not necessarily equivalent to the criteria of the ordinary concept, but traditional jurisprudence understands that there is something critically important in grappling with the features of the corresponding ordinary concept. For example, Keating argues that the

[194] See above, § 3.1.

[195] GOLDBERG & ZIPURSKY 2000, 109; see also GOLDBERG & ZIPURSKY 2000, 364: «It is not only judges and lawyers who interact with the law of torts, directly or indirectly. Everyone does. We all need to know what to expect of the persons, businesses, offices, and organizations around us, and we all need to know what is expected of us. That is why the wrongs recognized by courts as torts cannot be the wrongs that Judge Hercules would endorse for being those whose recognition would make the law the best it can be from the perspective of aspirational political or moral theory».

[196] KEATING 1996, 212 s.

[197] KEATING 1996, 311.

[198] KEATING 1996, 382.

ordinary notion of reasonableness supports social-contract theory: «Social contract theory holds that persons must be held "responsible for their ends". They must, that is, moderate the demands that they make on social institutions so that those demands fit within the constraints of mutually acceptable principles. This is simply an extension of our ordinary idea of reasonableness»[199].

This argument for the social-contract theory of reasonableness depends upon an empirical claim about the ordinary concept. As it happens, recent experimental-jurisprudence research provides some support for Keating's view. Empirical studies confirm that there is an important distinction between the ordinary notions of reasonableness and rationality, which supports Keating's hypotheses[200]. Moreover, as Keating intuited, the ordinary concept of reasonableness reflects what is socially acceptable[201], not necessarily what is economically rational or efficient[202].

In this example, Keating's intuitions about the ordinary concept of reasonableness were impressively accurate. Later experimental-jurisprudence studies lend further support to Keating's (empirical) jurisprudential hypotheses. But it is possible that intuitive, armchair jurisprudence might not capture the whole picture of ordinary cognition in some other cases.

This possible disconnect—between what a legal expert believes about the ordinary concept and what is true of the ordinary concept—could arise for many reasons. One possibility is that the legal expert just makes a mistake. The expert's intuition about the ordinary concept does not actually reflect the features of the ordinary concept. Perhaps the author did not think sufficiently clearly or employed an unconscious bias in favor of some particular theory. Studying the ordinary concept more robustly—with empirical methods—might help strengthen the theorist's conclusions. As Macleod puts it, «Hart and Honoré, after all, had a sample size of two: Hart and Honoré»[203]. Experimental jurisprudence can serve as an empirically grounded method of jurisprudential conceptual analysis.

It could also be that a legal expert, by virtue of all of their expertise and training, has some diminished access to the ordinary concept. When law students encounter a new concept of causation, are their corresponding ordinary concepts entirely unchanged? Or has their concept of causation changed both in and out of law? This is an open and unexplored empirical question. But if legal or philosophical education might sometimes alter one's ordinary concepts, this is another major reason that experimental jurisprudence would play a critical role in linking legal and ordinary concepts.

This process could also interact with the constitution of jurisprudence as a field. If most legal theorists intuit *X*, students who intuit *not X* might (mistakenly) think that they simply don't understand legal theory. As those students eschew legal theory, this preserves the apparent universality of intuition *X* within legal-theory circles. In the words of Professor Robert Cummins: «Those who do not share the intuitions are simply not invited to the games»[204].

Moreover, there are features of a legal concept that laypeople cannot access, but perhaps there also features of the ordinary concept that legal experts cannot perfectly access. Given the classical jurisprudential project of comparing ordinary to legal concepts[205], this possibility could support a jurisprudential division of conceptual labor; with laypeople as the experts of ordinary concepts and cognition.

[199] KEATING 1996, 370 («Reasonable people do not have an extravagant sense of the importance of their own preferences and aspirations in comparison with the aspirations of others. Moreover, reasonable people do not believe that their projects warrant the commitment of a disproportionate share of social wealth, and they do not make demands on others that they would be unwilling to honor themselves»).

[200] See generally GROSSMAN et al. 2020.

[201] KEATING 1996, 383.

[202] See generally JAEGER 2021 (finding that the legal notion of reasonableness is affected more by what is customary than by what is economically rational).

[203] MACLEOD 2019, 1021.

[204] CUMMINS 1998, 116.

[205] See generally, e.g., HONORÉ & GARDNER 2010.

A final possibility is that some features of ordinary concepts are not easily accessible by introspection—by anyone, layperson or expert. For example, consider the "hybrid theory" of reasonableness. On that view, reasonableness judgments reflect a hybrid of statistical and prescriptive considerations. It may not be possible to cleanly test such a subtle feature of the concept with the traditional mode of armchair thought experimentation. No matter how hard one reflects, it might be difficult to identify with certainty whether one's own notion of what is reasonable is a hybrid of considerations of the average and ideal. However, it is possible to begin to assess the predictions of that proposed analysis with cognitive science[206].

A central part of experimental jurisprudence is the study of ordinary language and concepts. And this is precisely because a central part of jurisprudence is such study of ordinary language and concepts. Jurisprudence has long been concerned with "our moral intuitions"[207], the "intuitions of a community"[208]—not the seminar-room community but rather our social and legal community. As Professor Jeremy Waldron explains, «[i]t is not enough that we have considered what Kant said to Fichte»[209]. Intuitions of legal philosophers are to be assessed against what is in fact, «out there, in the world»[210].

4. *Conclusion*

«Whither jurisprudence? Time will tell»[211]. Some offer skeptical and pessimistic prognoses, heralding the «death of jurisprudence»[212].

These reports are greatly exaggerated. Judge Richard Posner described jurisprudence as «the most fundamental, general, and theoretical plane of analysis of the social phenomenon called law»[213]. Scholars will (and should) continue to inquire into longstanding, fundamental, general, and theoretical legal questions.

So why the pessimism? One jurisprudential eulogy concerns the field's dissolution into other disciplines[214]. Perhaps there is nothing distinctively legal about law's notions of causation, knowledge, and reasonableness. The questions of traditional jurisprudence are scattered into questions of law-as-morality[215] or law-as-economics. Experimental jurisprudence might be seen as another destructively instrumental force, proposing law-as-surveys-of-laypeople.

However, this misunderstands the XJur movement. Rather than imagining jurisprudential questions dissolving into nonjurisprudential questions of moral philosophy or social science, experimental jurisprudence reaffirms these questions' fundamentally jurisprudential nature.

Questions about whether and how legal concepts differ from ordinary ones are both longstanding jurisprudential questions and partly empirical ones. Central legal concepts may

[206] In one study, a large number of participants were assigned to separate groups to evaluate average, ideal, and reasonable quantities, and then the mean ratings were statistically analyzed to assess whether average and ideal ratings both predict reasonableness ratings. See generally TOBIA 2018b. This provides some evidence in favor of the proposed account.

[207] E.g., FERZAN 2005, 749 (explaining the role of intuition in the proposed analysis of self-defense).

[208] DRESSLER 2000, 961.

[209] WALDRON 1999, 313. Waldron's point concerns the comparison of our (Western) human-rights intuitions against the intuitions of those from other (non-Western) countries. The point relevant to this Article is that it is also assumed that the relevant intuitions are not just those of expert legal theorists—what matters are the views of «people or whole societies». WALDRON 1999, 306.

[210] WALDRON 1999, 313 (emphasis omitted).

[211] SOLUM 2014, 2497.

[212] See, e.g., BEN-ZVI 2017.

[213] POSNER 1990, xi.

[214] BEN-ZVI 2017, 406 («There is nothing distinctively 'legal' about legal norms»).

[215] BEN-ZVI 2017, 406.

share features with their ordinary counterparts. It is possible that not all features of ordinary concepts are known and that not all features are discoverable by introspection or armchair thought experimentation. Empirical methods contribute unique data about ordinary language and concepts that speak to these central questions of jurisprudence. The XJur program does not assume that law should simply adopt the ordinary concept, but understanding the ordinary features—whether by thought experimentation or modern experimentation—raises important questions and is a central part of the analysis of legal concepts. Experimental jurisprudence thereby reopens a range of fascinating jurisprudential questions about law and its concepts[216]. Moreover, it provides new tools to address these questions. Experiments have revealed new, subtle, and surprising conceptual features, all of which call for further theoretical analysis.

While experimental jurisprudence offers new methods, it also invites analysis from those who do not themselves conduct empirical studies. To participate, one need not run experiments or even collect original data. Empirical data is a prerequisite, but there is already an abundance of data ripe for experimental-jurisprudential analysis. The cognitive science of ordinary language and concepts is full of such data[217]. Those data are "suspended experimental jurisprudence", just waiting to be incorporated into experimental jurisprudence with the addition of thoughtful theoretical analysis. For jurisprudence theorists concerned with our intuitions, empirical work in cognitive science is an essential resource.

This Article has argued that experimental jurisprudence is not a social-scientific replacement of jurisprudence. Rather, it is a form of jurisprudence. Traditional jurisprudence has always contained a central empirical program concerned with law's relation to ordinary people, language, and concepts. Justifications for that traditional project also justify the experimental approach. It is the opposing view—that jurisprudence has nothing to learn from careful study of ordinary language and concepts—that reflects a dramatic break from tradition.

Whither experimental jurisprudence? To the same place as jurisprudence: in search of increasingly sophisticated answers to our fundamental legal questions.

[216] Perhaps legal concepts share many features of the corresponding ordinary ones, or perhaps they are very distinctive. The truth, I would bet, is that legal concepts most often reflect a mixture of features—some drawn from corresponding ordinary concepts and others that are unique. But that remains an open empirical question. And the place to start is with experimental discoveries of the features of these concepts.

[217] See above, fn.49 and accompanying text.

References

ALEXANDER J., MALLON R., WEINBERG J.M. 2010. *Accentuate the Negative*, in «Review of Philosophy and Psychology», 1, 2, 297 ff.

ALICKE M.1992. *Culpable Causation*, in «Journal of Personality and Social Psychology», 63, 3, 368 ff.

ALICKE M. 2000. *Culpable Control and the Psychology of Blame*, in «Psychological Bulletin», 126, 4, 556 ff.

ALICKE M., DAVIS T.L. 1989. *The Role of* a Posteriori *Victim Information in Judgments of Blame and Sanction*, in «Journal of Experimental Social Psychology», 25, 4, 362 ff.

ALICKE M., ROSE D., BLOOM D. 2011. *Causation, Norm Violation and Culpable Control*, in «Journal of Philosophy», 108, 12, 229 ff.

AUSTIN J. 1869. *Lectures on Jurisprudence or the Philosophy of Positive Law*, J. Murray.

BANDES S., SALERNO J.M. 2014. *Emotion, Proof and Prejudice: The Cognitive Science of Gruesome Photos and Victim Impact Statements*, in «Arizona State Law Journal», 46, 1003 ff.

BARNETT R. 2011. *Interpretation and Construction*, in «Harvard Journal of Law and Public Policy», 34, 65 ff.

BARTELS D., RIPS L.J. 2010. *Psychological Connectedness and Intertemporal Choice*, in «Journal of Experimental Psychology», 139, 1, 49 ff.

BAUER P., POAMA A. 2020. *Does Suffering Suffice? An Experimental Assessment of Desert Retributivism*, in «Plos ONE», 15, 4, e0230304. Available on: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230304.

BEN-SHAHAR O., STRAHILEVITZ L. J. 2017. *Interpreting Contracts via Surveys and Experiments*, in «New York University Law Review», 92, 6, 1753 ff.

BEN-ZVI O. 2017. *Zombie Jurisprudence*, in DESAUTELS-STEIN J., CHRISTOPHER T. (eds.), *Searching for Contemporary Legal Thought*, Cambridge University Press, 2017, 406 ff.

BERMAN M., BOSTOCK G. W. 2021. *Was Bogus: Textualism, Pluralism, and Title VII*, in «Notre Dame Law Review», 97, 67 ff.

BEUTEL F. 1971. *The Relationship of Experimental Jurisprudence to Other Schools of Jurisprudence and to Scientific Method*, in «Washington University Law review», 385, 3, 409 ff.

BILZ K. 2012. *Dirty Hands or Deterrence? An Experimental Examination of the Exclusionary Rule*, in «Journal of Empirical Legal Studies», 9, 149 ff.

BILZ K. 2016. *Testing the Expressive Theory of Punishment*, in «Journal of Empirical Legal Studies», 13, 358, ff.

BILZ K., DARLEY J.M. 2004. *What's Wrong with Harmless Theories of Punishment*, in «Chicago-Kent Law Review», 79, 1215 ff.

BIX B. 1995. *Conceptual Questions and Jurisprudence*, in «Legal Theory», 1, 465.

BREGANT J., ROBBENNOLT J.K., WINSHIP V. 2021. *Perceptions of Settlement*, in «Harvard Negotiation Law Review», 27, 93 ff.

BREGANT J., SHAW A., KINZLER K.D. 2016. *Intuitive Jurisprudence: Early Reasoning About the Functions of Punishment*, in «Journal of Empirical Legal Studies», 13, 693 ff.

BREGANT J., WELLBERY I., SHAW A. 2019. *Crime but Not Punishment? Children Are More Lenient Toward Rule-Breaking when the "Spirit of the Law" Is Unbroken*, in «Journal of Experimental Child Psychology», 178, 266 ff.

Bronsteen J., Buccafusco K., Masur J. S. 2008. *Hedonic Adaptation and the Settlement of Civil Lawsuits*, in «Columbia Law Review», 108, 1516 ff.

Bronsteen J., Leiter B., Masur J., Tobia K. 2022. *The Folk Theory of Well-Being*, in in Lombroso T., Knobe J., Nichols S. (eds.), *Oxford Studies in Experimental Philosophy. Volume 5*, Oxford University Press, forthcoming.

Bystranowski P., Janik B., Próchnicki M., Hannikainen I.R., da Franca Couto Fernandes de Almeida G., Struchiner N. 2021. *Do Formalist Judges Abide by Their Abstract Principles? A Two-Country Study in Adjudication*, in «International Journal of the Semiotics of the Law» 35, 1903 ff.

Carlsmith K., Darley J.M., Robinson P.H. 2002. *Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment*, in «Journal of Personality and Social Psychology», 83, 284 ff.

Charness G., Dufwenberg M. 2010. *Bare Promises: An Experiment*, in «Economic Letters», 107, 281 ff.

Cohen F. 1954. *Dialogue on Private Property*, in «Rutgers Law Review», 9, 2, 357 ff.

Costa L., Dillon Esteves A. B., Kreimer R., Struchiner N., Hannikainen I. 2019. *Gender Stereotypes Underlie Child Custody Decisions*, in «Europen Journal of Social Psychology», 49, 3, 548.

Cummins R. 1998. *Reflection on Reflective Equilibrium*, in De Paul M. R., Ramsey W. (eds), *Rethinking Intuition,* Cambridge University Press, 1998, 113 ff.

Cushman F. 2008. *Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment*, in «Cognition», 108, 2, 353 ff.

Cushman F. 2011. *Should the Law Depend on Luck*, in Brockman M. (ed.), *Future Science: Essays from the Cutting Edge*, Oxford University Press,197 ff.

Cushman F. 2015. *Punishment in Humans: From Intuitions to Institutions,* in «Philosophy Compass», 10, 2, 117 ff.

Darley J, Pittman T. S. 2003. *The Psychology of Compensatory and Retributive Justice*, in «Personality and Social Psychology Review», 7, 324 ff.

Darley J. 2001. *Citizens' Sense of Justice and the Legal System*, in «Current Directions Psychological Science», 10, 1, 10 ff.

Darley J. 2009. *Morality in the Law: The Psychological Foundations of Citizens' Desires to Punish Transgressions*, in «Annual Review of Law and Social Science», 5, 1 ff.

Darley J. 2010. *Citizens' Assignments of Punishments for Moral Transgressions: A Case Study in the Psychology of Punishment*, in «Ohio State Journal of Criminal», 8, 101 ff.

Darley J., Carlsmith K.M., Robinson P.H. 2000. *Incapacitation and Just Deserts as Motives for Punishment*, in «Law and Human Behaviuor», 24, 6, 659 ff.

De Keersmaecker J., Bostyn D.H., Van Hiel A., Roets A. 2021. *Disliked but Free to Speak: Cognitive Ability Is Related to Supporting Freedom of Speech for Groups Across the Ideological Spectrum*, in «Social Psychological and Personality Science», 12, 34 ff.

Demaree-Cotton J., Sommers R. 2022. *Autonomy and the Folk Concept of Valid Consent*, in «Cognition», 224, 10 ff.

Dembroff R., Kohler-Hausmann I., Sugarman E. 2020. *What Taylor Swift and Beyoncé Teach Us About Sex and Causes*, in «University of Pennsylvania Law Review Online», 169, 1 ff.

DeScioli P., Karpoff R., De Freitas J. 2017. *Ownership Dilemmas: The Case of Finders Versus Landowners*, in «Cognitive Science», 41, 502 ff.

Desei A. 2022. *Text Is Not Enough*, in «University of Colorado Law Review», 93, 1 ff.

DIAMANTIS M. 2019. *Limiting Identity in Criminal Law*, in «Boston College Law Review», 60, 2011 ff.

DIAMANTIS M. 2022. *Corporate Identity*, in «Experimental Philosophy of Identity and the Self», 34, 14 ff.

DICKSON J. 2015. *Ours Is a Broad Church: Indirectly Evaluative Legal Philosophy as a Facet of Jurisprudential Inquiry*, in «Jurisprudence», 6, 207 ff.

DONELSON R., HANNIKAINEN I.R. 2020. *Fuller and the Folk: The Inner Morality of Law Revisited*, in LOMBROSO T., KNOBE J., NICHOLS S. (eds.), *Oxford Studies in Experimental Philosophy. Volume 3*, Oxford University Press, 6 ff.

DORFMAN D. 2020. *[Un]usual Suspects: Deservingness, Scarcity, and Disability Rights*, in «University of California Law Review», 10, 557 ff.

DORFMAN D. 2021. *Suspicious Species*, in «University of Illinois Law review» 4, 1363 ff.

DRANSEIKA V., DAGYS J., BESNIUNAS R. 2020. *Proper Names, Rigidity, and Empirical Studies on Judgments of Identity Across Transformations*, in «Topoi», 39, 2020, 381.

DRESSLER J. 2000. *Does One* Mens Rea *Fit All?: Thoughts on Alexander's Unified Conception of Criminal Culpability*, in «California Law Review», 88, 955 ff.

DUNLEA J., HEIPHETZ L. 2020. *Children's and Adults' Understanding of Punishment and the Criminal Justice System*, in «Journal of Experimental Social Psychology», 87, 1 ff.

DUNLEA J., HEIPHETZ L. 2021a. *Children's and Adults' Views of Punishment as a Path to Redemption*, in «Child Development», 92. Available on: https://psychology.columbia.edu/sites/default/files/content/Dunlea%20%26%20Heiphetz_Redemption_ChildDev.pdf.

DUNLEA J., HEIPHETZ L. 2021b. *Language Shapes Children's Attitudes: Consequences of Internal, Behavioral, and Societal Information in Punitive and Non-punitive Contexts*, in «Journal of Experimental Psychology: General», 151, 1233 ff.

DUNLEA J., WOLLE R.G., HEIPHETZ L. 2022. *The Essence of an Immigrant Identity: Children's Pro-social Response to Others Based on Perceived Ability and Desire to Change*, in TOBIA K. (ed.), *Experimental Philosophy of Identity and the Self*, Bloomsbury, 6 ff.

EARP B., DEMAREE-COTTON J., DUNN M., DRANSEIKA V. 2020a. *Experimental Philosophical Bioethics*, in «American Journal of Bioethics: Empirical Bioethics», 11, 2020, 30 ff.

EARP B., LATHAM S.R., TOBIA K.P. 2020b. *Personal Transformation and Advance Directives: An Experimental Bioethics Approach*, in «American Journal of Bioethics», 20, 72 ff.

EARP B., SKORBURG J. A., EVERETT J.A.C., SAVULESCU J. 2019. *Addiction, Identity, Morality*, in «American Journal of Bioethics: Empirical Bioethics», 10, 136 ff.

ECHELBARGER M., ROBERTS S.O., GELMAN S.A. 2021. *Children's Concerns for Equity and Ownership in Contexts of Individual-Based and Group-Based Inequality*, in «Journal of Cognition and Development», 23, 3 ff.

EDERER F., STREMITZER A. 2017. *Promises and Expectations*, in «Games and Economic Behaviour»,106, 161 ff.

EDERER F., STREMITZER A. 2018. *Moral Intuitions of Promise Keeping*, in «Principia», 65, 5 ff.

EIDELSON B. 2022. *Dimensional Disparate Treatment*, in «Southern California Law Review», 95, 4, 785 ff.

ENGEL C., RAHAL R.M. 2020. *Justice Is in the Eyes of the Beholder—Eye Tracking Evidence on Balancing Normative Concerns in Torts Cases*, January 2020, unpublished discussion paper (on file with author).

ESKRDIGE W., SLOCUM B.G., GRIES S. T. 2021. *The Meaning of Sex: Dynamic Words, Novel Applications, and Original Public Meaning*, in «Michigan Law Review», 119, 1503 ff.

FALLON R. 2015. *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, in «University of Chicago Law Review», 82, 1235 ff.

FERZAN K. 2005. *Justifying Self-Defense*, in «Law & Philosophy», 24, 6, 711 ff.

FLANAGAN B., HANNIKAINEN I. 2020. *The Folk Concept of Law: Law Is Intrinsically Moral*, in «Australasian Journal of Philosophy», 100, 1, 165 ff.

FLETCHER G. 1985. *The Right and the Reasonable*, in «Harvard Law Review», 98, 949 ff.

FLETCHER G. 1998. *Basic Concepts of Criminal Law*, Oxford University Press.

FRIEDMAN O., PESOWSKI M. L., GOULDING B.W. 2018. *Legal Ownership Is Psychological: Evidence from Young Children*, in PECK K., SHU S.B. (eds), *Psychological Ownership and Consumer Behavior*, Springer, 19 ff.

FRIEDMAN O., VAN DE VONDERVOORT J.W., DEFEYTER M.A., NEARY K.R. 2013. *First Possession, History, and Young Children's Ownership Judgments*, in «Child Development», 84, 1519 ff.

FULLER L. 1949. *The Case of the Speluncean Explorers*, «Harvard Law Review», 62, 616 ff.

FULLER L. 1958. *Positivism and Fidelity to Law: A Reply to Professor Hart*, in «Harvard Law Review» 71, 630 ff.

FURTH-MATZKIN, M., SOMMERS R. 2020. *Consumer Psychology and the Problem of Fine-Print Fraud*, in «Standford Law Review», 72, 503 ff.

GARDNER J. 2015. *The Many Faces of the Reasonable Person*, in «Law Quarterly Review», 131, 563 ff.

GINTHER M., SHEN F. X., BONNIE R. J., HOFFMAN M.B., JONES O. D., MAROIS R., SIMONS K.W., *The Language of* Mens Rea, in «Vanderbilt Law Review», 67, 1327 ff.

GLÖCKNER A., ENGEL C. 2013a. *Can We Trust Intuitive Jurors? Standards of Proof and the Probative Value of Evidence in Coherence-Based Reasoning*, in «Journal of Empirical Legal Studies», 10, 230.

GLÖCKNER A., ENGEL C. 2013b. *Role-Induced Bias in Court: An Experimental Analysis*, in «Behavioural Decision Making», 26, 272.

GOLD A.S. 2020. *Introduction*, in GOLD A.S., GOLDBERG J.C.P., KELLY D.B., SHERWIN E., SMITH H.E. (eds.) 2020., *The Oxford Handbook of the New Private Law*, Oxford University Press, XV ff.

GOLD A.S., GOLDBERG J.C.P., KELLY D.B., SHERWIN E., SMITH H.E. (eds.) 2020. *The Oxford Handbook of the New Private Law*, Oxford University Press.

GOLDBERG J.C.P. 2012. *Introduction: Pragmatism and Private Law*, in «Harvard Law Review», 125, 1640 ff.

GOLDBERG J.C.P., ZIPURSKY B.C. 2020. *Recognizing Wrongs*, Harvard University Press.

GREENE E.J., DARLEY J.M. 1998. *Effects of Necessary, Sufficient, and Indirect Causation on Judgments of Criminal Liability*, in «Law and Human Behavior», 22, 429 ff.

GROSSMAN I., EIBACH R.P., KOYAMA J., SAHI Q.B. 2020. *Folk Standards of Sound Judgment: Rationality Versus Reasonableness*, in «Science Advances» 6 (2), 2020, 103 ff.

GROVE T. 2020. *Which Textualism?*, in «Harvard Law Review», 134, 265 ff.

GUGLIELMO S., MONROE A. E., MALLE B.F. 2009. *At the Heart of Morality Lies Folk Psychology*, in «Inquiry», 52, 5, 449 ff.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A.J. 2007. *Blinking on the Bench: How Judges Decide Cases*, in «Cornell Law Review», 52, 93 ff.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A J. 2002. *Judging by Heuristic: Cognitive Illusions in Judicial Decision Making*, in «Judicature», 86, 44 ff.

GÜVER L., KNEER M. 2022. *Causation and the Silly Norm Effect*, in MAGEN S., PROCHOWNIK K. (eds.), *Advances in Experimental Philosophy of Law*, Bloomsbury Publishing, 133 ff.

HANNIKAINEN I., CABRAL G., MACHERY E., STRUCHINER N. 2017. *A Deterministic Worldview Promotes Approval of State Paternalism*, in «Journal of Experimental Social Psychology», 70, 251 ff.

HANNIKAINEN I., TOBIA K.P., DA FRANCA COUTO FERNANDES DE ALMEIDA G., DONELSON R., DRANSEIKA V., KNEER M., STROHMAIER N., BYSTRANOWSKI P., DOLININA K., JANIK B., KEO S., LAURAITYTĖ E., LIEFGREEN A., PRÓCHNICKI M., ROSAS A., STRUCHINER N. 2021. *Are There Cross-Cultural Legal Principles? Modal Reasoning Uncovers Procedural Constraints on Law*, in «Cognitive Science», 45, 8, e13024. Available on: https://onlinelibrary.wiley.com/doi/10.1111/cogs.13024.

HARLEY E. 2007. *Hindsight Bias in Legal Decision Making*, in «Social Cognition», 25, 48 ff.

HART H.L.A. 1994. *The Concept of Law* (2nd Ed.), Clarendon Press Oxford.

HART H.L.A., HONORÉ A. 1985. *Causation in the Law* (2nd Ed.), Oxford University Press.

HENNE P., PINILLOS A., DE BRIGARD F. 2017. *Cause by Omission and Norm: Not Watering Plants*, in «Australasian Journal of Philosophy», 95, 270 ff.

HITCHCOCK C., KNOBE J. 2009. *Cause and Norm*, in «Journal of Philosophy», 11, 587 ff.

HOEFT L. 2019. *The Force of Norms? The Internal Point of View in Light of Experimental Economics*, in «Ratio Juris», 32, 339.

HOFFMAN D., WILKINSON-RYAN T. 2013. *The Psychology of Contract Precautions*, in «The University of Chicago Law Review», 80, 395.

HOLMES O.W. 1897. *The Path of the Law*, in «Harvard Law Review», 10, 457 ff.

HONORÉ A., GARDNER J. 2010. *Causation in the Law*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Available on: https://plato.stanford.edu/archives/fall2019/entries/causation-law/.

HUANG B. 2013. *Shallow Signals,* in «Harvard Law Review», 126, 2227 ff.

HUANG B. 2016. *Law and Moral Dilemmas*, in «Harvard Law Review», 130, 659 ff.

HUANG B. 2019. *Law's Halo and the Moral Machine*, in «Columbia Law Review», 119, 1811 ff.

HURD H. 1996. *The Moral Magic of Consent*, in «Legal Theory», 2, 121 ff.

HUSSAK L., CIMPIAN A. 2019. *"It Feels Like It's in Your Body": How Children in the United States Think About Nationality*, in «Journal of Experimental Psychology: General», 148, 1153.

JAEGER C. 2021. *The Empirical Reasonable Person*, in «Alabama Law Review», 72, 887 ff.

JIMÉNEZ F. 2021. *Some Doubts About Folk Jurisprudence: The Case of Proximate Cause*, in «University of Chicago Law. Review Online», 66, 1, 27 ff.

KAHAN D., HOFFMAN D., EVANS D., DEVINS N., LUCCI E. 2016. *"Ideology" or "Situation Sense"? An Experimental Investigation of Motivated Reasoning and Professional Judgment*, in «University of Pennsylvania Law Review», 164, 349 ff.

KANNGIESSER P., HOOD B.M. 2014. *Young Children's Understanding of Ownership Rights for Newly Made Objects*, in «Cognitive Development», 29, 30 ff.

KAWASHIMA T. 1974. *The Legal Consciousness of Contract in Japan*, «Law in Japan», 7, 1 ff.

KEATING G. 1996. *Reasonableness and Rationality in Negligence Theory*, in «Stanford Law Review», 48, 311 ff.

Kirfel L., Hannikainen I. R. 2022. *Why Blame the Ostrich? Understanding Culpability for Willful Ignorance*, in Magen S., Prochownik K (eds.), *Advances in Experimental Philosophy of Law*, Bloomsbury Press.

Kirfel L., Lagnado D. 2021. *Causal Judgments About Atypical Actions Are Influenced by Agents' Epistemic States*, in «Cognition», 212, 1 ff.

Klapper S., Schmidt S., Tarantola T. (mns). *Ordinary Meaning from Ordinary People*, unpublished manuscript.

Klerman D., Spamann H. 2024. *Law Matters—Less Than We Thought*, in «Journal of Law, Economics & Organization», 40, forthcoming.

Kneer M. (forthcoming). *Reasonableness on the Clapham Omnibus: Exploring the Outcome-Sensitive Folk Concept of Reasonable*, in «Judicial Decision-Making: Integrating Empirical and Theoretical Perspectives», forthcoming.

Kneer M. (mns). *The Side-Effect Effect as an Instance of a Severity Effect*, unpublished manuscript (on file with author).

Kneer et al. (mns). *The Severity Effect on Intention and Knowledge. A Cross-Cultural Study with Laypeople and Legal Experts*, unpublished manuscript (on file with author).

Kneer M., Bourgeois-Gironde S. 2017. Mens Rea *Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, in «Cognition», 169, 139 ff.

Kneer M., Haybron D.M. (mns). *Happiness and Well-Being: Is It All in Your Head? Evidence from the Folk*, unpublished manuscript (on file with author).

Kneer M., Machery E. 2019 *No Luck for Moral Luck*, in «Cognition», 182, 2019, 331 ff.

Knobe J. 2003. *Intentional Action and Side Effects in Ordinary Language*, in «Analysis» 63, 190 ff.

Knobe J., Nichols S. 2008. *An Experimental Philosophy Manifesto*, in Id., *Experimental Philosophy*, Oxford University Press, 3 ff.

Knobe J., Shapiro S. 2021 *Proximate Cause Explained: An Essay in Experimental Jurisprudence*, in «University of Chicago Law Review», 88, 165 ff.

Kobick J. 2010. *Discriminatory Intent Reconsidered: Folk Concepts of Intentionality and Equal Protection Jurisprudence*, in «Harvard Civil Rights—Civil Liberties Law Review», 45, 517 ff.

Kobick J., Knobe J. 2009. *Interpreting Intent: How Research on Folk Judgments of Intentionality Can Inform Statutory Analysis*, in «Brooklyn Law Review», 75, 409 ff.

Kominsky J., Phillips J., Gerstenberg T., Lagnado D., Knobe J. 2015. *Causal Superseding*, in «Cognition», 137, 196 ff.

Koppelman A. 2020. *Bostock, LGBT Discrimination, and the Subtractive Moves*, in «Minnesota Law Review Headnotes», 105, 1 ff.

Korobkin R., Guthrie C. 1997. *Psychology, Economics, and Settlement: A New Look at the Role of the Lawyer*, in «Texas Law Review», 76, 77 ff.

Kugler M. 2019. *From Identification to Identity Theft: Public Perceptions of Biometric Privacy Harms*, in «University of California Irvine Law Review», 10, 107 ff.

Lagnado D., Channon S. 2008. *Judgments of Cause and Blame: The Effects of Intentionality and Foreseeability*, in «Cognition», 108, 754 ff.

Lagnado D., Gerstenberg T. 2017. *Causation in Legal and Moral Reasoning*, in Waldmann M.R. (ed.), *The Oxford Handbook of Causal Reasoning*, Oxford University Press, 565 ff.

LANGLINAIS A., LEITER B. 2016. *The Methodology of Legal Philosophy*, in CAPPELEN H., SZABÓ GENDLER T., HAWTHORNE J. (eds.), *The Oxford Handbook of Philosophical Methodology*, Oxford University Press, 671 ff.

LEE T., MOURITSEN S.C. 2018. *Judging Ordinary Meaning*, in «Yale Law Journal», 127, 788 ff.

LEITER B. 2003. *Beyond the Hart/Dworkin Debate: The Methodology Problem in Jurisprudence*, in «American Journal of Jurisprudence», 48, 43 ff.

LEITER B. 2007. *Naturalizing Jurisprudence*, Oxford University Press.

LEVINE S., MIKHAIL J., LESLIE A. M. 2018. *Presumed Innocent? How Tacit Assumptions of Intentional Structure Shape Moral Judgment*, in «Journal Experimental Psychology: General», 147, 1728 ff.

LIDÉN M., GRÄNS M., JUSLIN P. 2018a. *'Guilty, No Doubt': Detention Provoking Confirmation Bias in Judges' Guilt Assessments and Debiasing Techniques*, in «Psychology, Crime and Law», 25, 3, 219 ff.

LIDÉN M., GRÄNS M., JUSLIN P. 2018b. *From Devil's Advocate to Crime Fighter: Confirmation Bias and Debiasing Techniques in Prosecutorial Decision-Making*, in «Psychology, Crime and Law», 25, 5, 494 ff.

LIU J., KLÖHN L., SPAMANN H. 2021. *Precedent and Chinese Judges: An Experiment*, in «American Journal of Comparative Law», 69, 93 ff.

MACHERY E., et al. 2017. *The Gettier Intuition from South America to Asia*, in «Journal of Indian Council of Philosophical Research», 34, 517 ff.

MACLEOD J. 2016. *Belief States in Criminal Law*, in «Oklaoma Law Review», 68, 497 ff.

MACLEOD J. 2019. *Ordinary Causation: A Study in Experimental Statutory Interpretation*, in «Indiana Law Journal», 94, 957 ff.

MACLEOD J. 2022. *Finding Original Public Meaning*, in «Georgia Law Review», 1, 56 ff.

MAGEN S., PROCHOWNIK K. M. 2021. *Legal X-Phi Bibliography*, in «The Centre for Law, Behavior and Cognition», 16, 6. Available on: https://zrsweb.zrs.rub.de/institut/clbc/legal-x-phi-bibliography.

MALLE B., NELSON S. E. 2003. *Judging* Mens Rea*: The Tension Between Folk Concepts and Legal Concepts of Intentionality*, in «Behavior Science and Law», 21, 563 ff.

MARTIN J., HEIPHETZ L. 2021. *"Internally Wicked": Investigating How and Why Essentialism Influences Punitiveness and Moral Condemnation*, in «Cognitive Science», 1, 45, 1 ff.

MARTÍNEZ E., WINTER C. 2022. *Experimental Longtermist Jurisprudence*, in MAGAN S., PROCHOWNIK K. (eds.), *Advances in Experimental Philosophy of Law*, Bloomsbury Academic, 2 ff.

MATSUMURA K. 2014. *Binding Future Selves*, in «Louisiana Law Review», 75, 71 ff.

MAZZI D. 2014. *"Our Reading Would Lead to...": Corpus Perspectives on Pragmatic Argumentation in U.S. Supreme Court Judgments*, in «Journal of Argumentation in Context», 3, 103 ff.

MELNIKOFF. D., STROHMINGER N. 2020. *The Automatic Influence of Advocacy on Lawyers and Novices*, in «Nature Human Behavior», 4, 1258 ff.

MIHAILOV E., HANNIKAINEN I. R., EARP B. D. 2021. *Advancing Methods in Empirical Bioethics: Bioxphi Meets Digital Technologies*, in «American Journal of Bioethics», 21, 53 ff.

MIKHAIL J. 2009. *Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge*, in «Psychology of Learning and Motivation», 50, 27 ff.

MIKHAIL J. 2012. *Moral Grammar and Human Rights: Some Reflections on Cognitive Science and Enlightenment Rationalism*, in RYAN GOODMAN, JINKS D., WOODS A.K. (eds.), *Understanding Social Action, Promoting Human Rights*, Oxford University Press,160 ff.

MIKHAIL J. 2021, *Elements of Moral Cognition,* Cambridge University Press.

MILES T., SUNSTEIN C. R. 2008. *The New Legal Realism*, in «The University of Chicago Law Review», 75, 831 ff.

MISCHKOWSKI D., STONE R., STREMITZER A. 2019. *Promises, Expectations, and Social Cooperation*, in «Journal of Law and Economy», 62, 687 ff.

MOLOUKI S., BARTELS D. M. 2017. *Personal Change and the Continuity of the Self*, in «Cognitive Psychology», 93, 1 ff.

MOLOUKI S., BARTELS D. M., URMINSKY O. 2017. *A Longitudinal Study of Difference Between Predicted, Actual, and Remembered Personal Change*, in «Cognitive Science Society», 39, 2748 ff.

MOTT C. 2018. *Statutes of Limitations and Personal Identity*, in LOMBROZO T., KNOBE J., NICHOLS S. (eds.), *Oxford Studies in Experimental Philosophy. Volume 2*, Oxford University Press, 2 ff.

MOURITSEN S., *Contract Interpretation with Corpus Linguistics*, in «Washington Law Review», 94, 1337 ff.

MUELLER P., SOLAN L. M., DARLEY J. M. 2012. *When Does Knowledge Become Intent? Perceiving the Minds of Wrongdoers*, in «Journal of Empirical Legal Studies», 9, 859 ff.

NADELHOFFER T. 2005. *Intentions and Intentional Actions in Ordinary Language and the Criminal Law*, Ph.D. dissertation, Florida State University.

NADELHOFFER T. 2006. *Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality*, in «Philosophical Explorations», 9, 203 ff.

NADELHOFFER T. 2012. *Attempts: In Ordinary Language and the Criminal Law—A Commentary*, in «Jurisprudence», 3, 475 ff.

NADELHOFFER T., HESHMATI S., KAPLAN D., NICHOLS S. 2013. *Folk Retributivism and the Communication Confound*, in «Economics and Philosophy», 29, 235 ff.

NADELHOFFER T., NAHMIAS E. 2011. *Neuroscience, Free Will, Folk Intuitions, and the Criminal Law*, in «Thurgood Marshall Law Review», 36, 157 ff.

NADELHOFFER, T. 2013. *The Future of Punishment*, Oxford University Press.

NADLER J., MCDONNELL M. H. 2012. *Moral Character, Motive, and the Psychology of Blame*, in «Cornell Law Review», 97, 255 ff.

NADLER J., TROUT J. D. 2012. *The Language of Consent in Police Encounters*, in TIERSMA P.M., SOLAN L.M. (eds.), *Oxford Handbook of Language and Law,* Oxford University Press, 326 ff.

NAGEL T. 1979, *Mortal Questions*, Cambridge University Press.

NAHMIAS E., AHARONI E. 2018. *Communicative Theories of Punishment and the Impact of Apology*, in SURPRENANT C. W. (ed.), *Rethinking Punishment in the Era of Mass Incarceration*, Routledge,144 ff.

NANCEKIVELL S., FRIEDMAN O., GELMAN S. A. 2019. *Ownership Matters: People Possess a Naïve Theory of Ownership*, in «Trends in Cognitive Science», 23, 102 ff.

NANCEKIVELL S., MILLAR C. J., SUMMERS P. C., FRIEDMAN O. 2016, *Ownership Rights*, in SYTSMA J., BUCKWALTER W. (eds.), *Companion to Experimental Philosophy*, Blackwell, 247 ff.

NANCEKIVELL S., VAN DE VONDERVOORT J.W., FRIEDMAN O. 2013. *Young Children's Understanding of Ownership*, in «Child Development Perspectives», 7, 243 ff.

NEWMAN G.E., BLOOM P., KNOBE J. 2014. *Value Judgments and the True Self*, in «Personality and Social Psychology Bullutin», 40, 203 ff.

NOURSE V. 2019. *Textualism 3.0: Statutory Interpretation After Justice Scalia*, in «Alabama Law Review», 70, 667 ff.

NOURSE V., ESKRIDGE W. N. 2021. *Textual Gerrymandering; The Eclipse of Republican Government in an Era of Statutory Populism*, in «New York University Law Review », 96, 1718 ff.

NYARKO J., SANGA S. 2022. *A Statistical Test for Legal Interpretation: Theory and Applications*, in «Journal of Law, Economics and Organizations», 38, 2, 539 ff.

PAULSEN M., KADISHM S. 1962. *Criminal Law and its process*, Brown Company.

PHILLIPS J., DE FREITAS J., MOTT C., GRUBER J., KNOBE J. 2017. *True Happiness: The Role of Morality in the Folk Concept of Happiness*, in «Journal of Experimental Psychology: General», 146, 165 ff.

POSNER R. 1990. *The Problems of Jurisprudence*, Harvard University Press.

PRENTICE R., KOEHLER J. J. 2003. *A Normality Bias in Legal Decision Making*, in «Cornell Law Review», 88, 583 ff.

PRIEL D. 2019. *Evidence-Based Jurisprudence: An Essay for Oxford*, in «Analisi e Diritto», 2019, 87 ff.

PROCHOWNIK K. 2017. *Do People with a Legal Background Dually Process? The Role of Causation, Intentionality and Pragmatic Linguistic Considerations in Judgments of Criminal Responsibility*, in BROŻEK B., KUREK Ł., STELMACH J. (eds.), *The Province of Jurisprudence Naturalized*, Wolters Kluwer, 168 ff.

PROCHOWNIK K. 2021. *The Experimental Philosophy of Law: New Ways, Old Questions, and How not to Get Lost*, in «Philosophical Compass», 16. Available on: https://compass.onlinelibrary.wiley.com/doi/10.1111/phc3.12791.

PROCHOWNIK K., KREBS M., WIEGMANN A., HORVATH J. 2020. *Not as Bad as Painted? Legal Expertise, Intentionality Ascription, and Outcome Effects Revisited*, in «Proceedings of the annual meeting of Cognitive Science Society», 42, 1930 ff.

PROCHOWNIK K., UNTERHUBER M. 2018. *Does the Blame Blocking Effect for Assignments of Punishment Generalize to Legal Experts?*, in «Cognitive Science», 43, 2285 ff.

PROCHOWNIK K., WIEGMANN A., HORVATH J. 2021. *Blame Blocking and Expertise Effects Revisited*, in «Proceedings of the annual meeting of Cognitive Science Society», 43, 2323 ff.

RACHLINSKI J. 2000. *Heuristics and Biases in the Courts: Ignorance or Adaptation?*, in «Oregon Law Review», 79, 61 ff.

RACHLINSKI J. 2003. *Misunderstanding Ability, Misallocating Responsibility*, in «Brooklyn Law Review», 68, 1055 ff.

RACHLINSKI J. 2011. *The Psychological Foundations of Behavioral Law and Economics*, «University of Illinois Law Review», 5, 829 ff.

RACHLINSKI J., GUTHRIE C., WISTRICH A. J. 2011. *Probable Cause, Probability, and Hindsight*, in «Journal of Empirical Legal Studies», 8, 72 ff.

RACHLINSKI J., JOHNSON S., WISTRICH A. J., GUTHRIE C. 2009. *Does Unconscious Racial Bias Affect Trial Judges?*, in «Notre Dame Law Review», 84, 1195 ff.

RACHLINSKI J., WISTRICH A. J., GUTHRIE C. 2013. *Altering Attention in Adjudication*, in « University of California Los Angeles Law Review», 60, 1586 ff.

RAPPAPORT A. 2014. *On the Conceptual Confusions of Jurisprudence*, in «Washington University Jurisprudence Review», 7, 77 ff.

ROBINSON P. 2008. *Distributive Principles of Criminal Law: Who Should Be Punished How Much?*, Oxford University Press.

ROBINSON P., DARLEY J. M. 1995. *Justice, Liability, and Blame: Community Views and the Criminal Law*, Westview Press.

ROSE D. et al. 2019. *Nothing at Stake in Knowledge*, in «Noûs», 53, 224 ff.

SANDERS J., KUGLER M. B., SOLAN L. M., DARLEY J. M. 2014. *Must Torts Be Wrongs? An Empirical Perspective*, in «Wake Forest Law Review», 1, 49, 1 ff.

SCHAUER F. 2008. *A Critical Guide to Vehicles in the Park*, in «New York University Law Review», 83, 1109 ff.

SCHAUER F. 2020. *Social Science and the Philosophy of Law*, in TASIOULAS J. (ed.), *Cambridge Companion to the Philosophy of Law*, Cambridge University Press, 95 ff.

SHEN F., HOFFMAN M. B., JONES O. D., GREENE J. D., MAROIS R. 2011. *Sorting Guilty Minds*, in «New York University Law Review», 86, 1306 ff.

SHOEMAKER D., TOBIA K. P. 2022. *Personal Identity*, in DORIS J.M., VARGAS M. (eds.), *Oxford Handbook of Moral Psychology*, Oxford University Press, forthcoming.

SLOCUM B. 2015. *Ordinary Meaning: A Theory of the Most Fundamental Principle of Legal Interpretation*, University of Chicago Press.

SLOCUM B., WONG J. 2021. *The Vienna Convention and the Ordinary Meaning of International Law*, in «Yale Journal of International Law», 46, 191 ff.

SOLAN L. 2003. *Cognitive Foundations of the Impulse to Blame*, in «Brooklyn Law Review», 68, 1003 ff.

SOLAN L., DARLEY J. M. 2001. *Causation, Contribution, and Legal Liability: An Empirical Study*, in «Law and Contemporary Problems», 64, 265 ff.

SOLAN L., ROSENBLATT T., OSHERSON D. 2008. *False Consensus Bias in Contract Interpretation*, in «Columbia Law Review», 108, 1268 ff.

SOLUM L. 2014. *The Positive Foundations of Formalism: False Necessity and American Legal Realism*, in «Harvard Law Review», 127, 2464 ff.

SOLUM L. 2018. *Legal Theory Lexicon 044: Legal Theory, Jurisprudence, and the Philosophy of Law*, in «Legal Theory Lexicon». Available on: https://lsolum.typepad.com/legal_theory_lexicon/2005/05/legal_theory_le.html.

SOLUM L. 2019 (mns). *The Constraint Principle: Original Meaning and Constitutional Practice*, unpublished manuscript.

SOMMERS R. 2020. *Commonsense Consent*, in «Yale Law Journal», 129, 2232 ff.

SOMMERS R. 2021. *Experimental Jurisprudence: Psychologists Probe Lay Understandings of Legal Constructs*, in «Science», 373, 394 ff.

SOMMERS R., BOHNS V.K. 2019. *The Voluntariness of Voluntary Consent: Consent Searches and the Psychology of Compliance*, in «Yale Law Journal», 128, 1962 ff.

SOOD A.M. 2015. *Cognitive Cleansing: Experimental Psychology and the Exclusionary Rule*, in «Georgia Law Journal», 103, 1543 ff.

SPAMANN H., KLÖHN L. 2016. *Justice Is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges*, in «Journal of Legal Studies», 45, 255 ff.

SPAMANN H., KLÖHN L., JAMIN C., KHANNA V., ZHUANG LIU J., MAMIDI P., MORELL A., REIDEL I. 2021. *Judges in the Lab: No Precedent Effects, No Common/Civil Law Differences*, in «Journal of Legal Analysis», 13, 110 ff.

SPELLMAN B. 1997. *Crediting Causality*, in «Journal of Experimental Psychology: General», 126, 323 ff.

SPELLMAN B., HOLLAND C. R. 2010. *When Knowledge Matters to Causation*, in «Annual Conference of Empirical Legal Studies», 5.

SPRUILL M., LEWIS N. A. JR. 2022. *Legal Descriptions of Police Officers Affect How Citizens Judge Them,* in «Journal of Experimental Social Psychology» 101, 104306.

STICH S., TOBIA K. P. 2016. *Experimental Philosophy and the Philosophical Tradition*, in SYTSMA J., BUCKWALTER W. (eds.), *A Companion to Experimental Philosophy*, Blackwell, 5 ff.

STONE R., STREMITZER A. 2020. *Promises, Reliance, and Psychological Lock*, in «Journal of Legal Studies», 49, 33 ff.

STROHMAIER N. 2017. *Introducing Experimental Jurisprudence*, in «Leiden Law Blog», December 1, 2017.

STROHMINGER N., NICHOLS S. 2014. *The Essential Moral Self*, in «Cognition», 131, 159 ff.

STROHMINGER N., NICHOLS S. 2015. *Neurodegeneration and Identity*, in «Psychological Science», 26, 1469 ff.

STRUCHINER N., DA FRANCA COUTO FERNANDES DE ALMEIDA G., HANNIKAINEN I. R. 2020a. *Legal Decision-Making and the Abstract/Concrete Paradox*, in «Cognition», 13, 323 ff.

STRUCHINER N., HANNIKAINEN I. R., DA FRANCA COUTO FERNANDES DE ALMEIDA G. 2020b. *An Experimental Guide to Vehicles in the Park*, in «Judgment & Decision Making», 15, 312 ff.

STUBBS M. 1996. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture,* Blackwell Publishers.

SUMMERS A. 2018. *Common-Sense Causation in the Law*, in «Oxford Journal Legal Studies», 38, 793 ff.

SYTSMA J., BLUHM R., WILLEMSEN P., REUTER K. 2019. *Causal Attributions and Corpus Analysis*, in FISCHER E., CURTIS M. (eds.), *Methodological Advances in Experimental Philosophy*, Bloomsbury, 209 ff.

TOBIA K 2016. *Personal Identity, Direction of Change, and Neuroethics*, in «Neuroethics», 9, 37 ff.

TOBIA K. (mns). *Legal Concepts and Legal Expertise*, unpublished manuscript. Available on: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3536564.

TOBIA K. 2015. *Personal Identity and the Phineas Gage Effect*, in «Analysis», 75, 396 ff.

TOBIA K. 2018a. *How People Judge What is Reasonable*, in «Alabama Law Review», 70, 293 ff.

TOBIA K. 2018b. *Law and the Cognitive Science of Ordinary Concepts*, in BROŻEK B., HAGE J., VINVENT N. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 86 ff.

TOBIA K. 2020. *Testing Ordinary Meaning*, in «Harvard Law Review», 134, 726 ff.

TOBIA K. 2023. *The Cambridge Handbook of Experimental Jurisprudence*, Cambridge University Press.

TOBIA K., MIKHAIL J. 2021. *Two Types of Empirical Textualism*, in «Brooklyn Law Review», 86, 461 ff.

TOBIA K., SLOCUM B. G., NOURSE V. 2022a. *Statutory Interpretation from the Outside*, in «Columbia Law Review» 122, 213 ff.

TOBIA K., SLOCUM B. G., NOURSE V. 2022b. *Progressive Textualism*, in «Georgetown Law Journal», 110, 1437 ff.

TOOMEY J. 2022. *Narrative Capacity*, in «North Carolina Law Review», 100, forthcoming 2022 (on file with author).

TUR R.H.S. 1978. *What Is Jurisprudence?*, in «Philosophical Inquires», 28, 149 ff.

ROVERSI C., UBERTONE M., VILLANI C., D'ASCENZO S., LUGLI L. 2022. *Alice in Wonderland: Experimental Jurisprudence on the Internal Point of View*, in «Jurisprudence», forthcoming.

VANBERG C. 2008. *Why Do People Keep Their Promises? An Experimental Test of Two Explanations*, in «Econometrica», 76, 1467 ff.

VIEBAHN E., WIEGMANN A., ENGELMANN N., WILLEMSEN P. 2020. *Can a Question Be a Lie? An Empirical Investigation*, in «Ergo», 8. Available on: https://journals.publishing.umich.edu/ergo/article/id/1144/.

VILARES I., WESLEY M. J., AHN W.Y., BONNIE R. J., HOFFMAN M., JONES O. D., MORSE S. J., YAFFE G., LOHRENZ T., MONTAGUE P.R. 2017. *Predicting the Knowledge—Recklessness Distinction in the Human Brain,* in «Proceedings of National Academy of Sciences», 114, 3222 ff.

VILLANI C., D'ASCENZO S., BORGHI A., ROVERSI C., BENASSI M., LUGI L. 2021. *Is Justice Grounded? How Expertise Shapes Conceptual Represnetation of Institutional Concepts,* in «Psychological Research», 8, 2434 ff .

WALDRON J. 1999. *How to Argue for a Universal Claim*, in «Columbia Human Rights. Law Review», 30, 305 ff.

WEST R. 2011. *Normative Jurisprudence: An Introduction,* Cambridge University Press.

WHITTINGTON K. 1999. *Constitutional Interpretation: Textual Meaning, Original Intent and Judicial Review,* University Press of Kansas.

WIEGMANN A., MEIBAUER J. 2019. *The Folk Concept of Lying*, in «Philosophical Compass», 14, 8. Available on: https://www.researchgate.net/publication/334806270_The_folk_concept_of_lying.

WILKINSON-RYAN T. 2010. *Do Liquidated Damages Encourage Breach? A Psychological Experiment*, in «Michigan Law Review», 108, 633 ff.

WILKINSON-RYAN T. 2012. *Promise and Psychological Contract*, in «Wake Forest Law Review», 47, 843 ff.

WILKINSON-RYAN T. 2014. *Psychological Account of Consent to Fine Print*, in «Iowa Law Review», 99, 1745 ff.

WILKINSON-RYAN T. 2017. *The Perverse Consequences of Disclosing Standard Terms*, in «Cornell Law Review», 103, 117 ff.

WILKINSON-RYAN T. 2020. *Psychology and the New Private Law*, in GOLD A. S., GOLDBERG J.C.P., KELLY D.B., SHERWIN E., SMITH H.E. (eds.), *Oxford Handbook of the Private Law*, Oxford University Press, 125 ff.

WILKINSON-RYAN T., BARON J. 2009. *Moral Judgment and Moral Heuristics in Breach of Contract*, in «Journal of Empirical Legal Studies», 6, 405 ff.

WILKINSON-RYAN T., HOFFMAN D. A. 2015. *The Common Sense of Contract Formation*, in «Standford Law Review», 67, 1269 ff.

WILLEMSEN P., KIRFEL L. 2019. *Recent Empirical Work on the Relationship Between Causal Judgments and Norms*, in «Philosophy Compass», 14, 1, e12562. Available on: https://compass.onlinelibrary.wiley.com/doi/10.1111/phc3.12562.

WINTER C. 2020. *The Value of Behavioral Economics for EU Judicial Decision-Making*, in «German Law Journal», 21, 240 ff.

WISTRICH A., RACHLINSKI J.J., GUTHRIE C. 2015. *Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings?*, in «Texas Law Review», 93, 855 ff.

PART II.

Cognitive-oriented Perspectives on Law and Legal Reasoning

Causation, the Law, and Mental Models

PHILIP N. JOHNSON-LAIRD, MONICA BUCCIARELLI

> ... the clarification of the structure of ordinary causal statements was and is an indispensable first step towards understanding the use of causal notions in the law
> HART & HONORÉ 1985, p. xxxiv

1. *Introduction*

A long-standing problem for the law is that its notions of causation differ from those of daily life. As jurists admit, these notions are also confusing. Here are three illustrative puzzles. An attacker stabbed an 18-year-old girl and she was taken to hospital. She needed an operation but she was a Jehovah's Witness, and refused a preliminary blood transfusion. She was told that she would die as a result, but she said she did not care. She died the next day. The court accepted that her refusal of the transfusion was the reason she died. Yet, they found her attacker guilty of manslaughter on the grounds that his stabbing caused her death[1]. Why isn't her not having the transfusion the cause of her death?

A man assaulted and raped a woman. She then took poison. He refused to summon medical help and imprisoned her for hours in a hotel room. She died a month later, and the court determined the cause was either the poison or its effects coupled with those of her wounds. It found the defendant guilty of murder. On appeal, the court maintained this verdict[2]. Granted that her death depended on the poison, why didn't her own action cause her death?

A plaintiff fell down an open elevator shaft in the defendant's building and was injured. The court decided that the "proximate" cause of the accident was that the door to the elevator was open, but that the defendant was not guilty of negligence, because no evidence showed that any employee of his had opened it, the plaintiff knew that the elevator sometimes crept up or down, and the open door was a clear danger[3]. How can an open door cause a person to fall?

We return to these cases at the end of this essay to show how it solves their puzzles. Law in the English-speaking world distinguishes between a *factual* cause and, as the previous example illustrates, a *proximate* (or *legal*) cause. Factual causes are supposed to be nothing more than everyday causes, but as we show proximate causes are mystifying. What is most extraordinary is that if you ask for the meaning of an assertion, such as:

The defendant's action caused the outcome

jurists seem unsure. Cognitive scientists sometimes avoid answering this question too, and

[1] *Regina v. Blaue* 61 Cr App R 271 (1975).
[2] *Stephenson v. State*, 179 N.E. 633 (Indiana. 1932).
[3] *Hope v. Longley*, 27 R.I. 579, 65 A. 300 (R.I. 1906).

when they do respond, the disparity in their views is vast. Their theories depend on modal logics of which there is a countable infinity, 'possible worlds', counterfactual conditionals, Bayesian probabilities, mechanisms, forces, powers, principles, and so on and on. The foundational mystery for this essay is therefore: what does the causal assertion above mean? Its meaning, we assume, determines what verifies the assertion, and whether an inference follows of necessity from the assertion. Hence, for example, part of the meaning of *cause* is not that it is an *unusual* event, because this inference does not follow of necessity:

The defendant's action caused the outcome.
Therefore, *the defendant's action was unusual.*

A further mystery is why the answer to our foundational question is not obvious. Our chapter presents solutions to both mysteries. It begins with a review of causation in the law, which demonstrates its self-confessed confusions. It then goes back to Aristotle and examines philosophical analyses of the meaning of causation. None will do. It introduces the theory of mental models—for which it marshals evidence—and it shows how this theory illuminates the concept of causation, leads to proper tests of causal relations, answers our fundamental question, and explains its mysterious difficulty. It concludes with five recommendations for legal practice.

2. *Legal concepts of causation*

Perhaps influenced by John Stuart MILL (1874) whom we'll come to by and by, common law in England and America, and the law in continental Europe, offer a "but for" test (the *sine qua non* test) to answer questions about factual causes, as in:

Is it true that but for *the defendant's action, the victim would not have died?*
If, and only if, it is true, then the defendant's action caused her death.

The test calls for an evaluation of what philosophers nowadays refer to as a "counterfactual" conditional: "If the defendant had not acted as he did, then the victim would not have died". Skeptics argue that it is difficult or impossible to verify counterfactual assertions. But, it is often feasible, as is evident in the many sports with counterfactual rules, e.g., golf, rugby, basketball, and others. In cricket, for instance, a batsman is out if the ball would have hit the wicket had it not hit his body without first hitting his bat. The umpire imagines the path of the ball had the player's body not blocked it, and makes the decision. An automatic TV system, Hawk-Eye, also computes the ball's counterfactual trajectory had it not hit the player's body.

Continental lawyers have proposed other analyses. For instance, Birkmeyer, a nineteenth century jurist, suggested that a cause was whichever condition had the greatest energy and so contributed most to the effect (HART & HONORÉ 1985, Ch. XV). But, continental law also often treats any condition necessary for an event as a cause, just as the *but-for* test does (HART & HONORÉ 1985, Ch. XVI). When we turn from legal principles to actual practice, courts allow for "intervening causes", for assignment of mutual negligence if it is impossible to determine which of several tortfeasors in fact caused damage[4], and sometimes that an action was the probable cause of an effect[5].

[4] *Summers v. Tice* (1948) 33 C2d 80. We thank Stuart Lichten for telling us about this case.
[5] *New York Times co. v. Suillivan* (1964) 376 U.S. 254.

In law, a person who causes harm is not necessarily liable for damages or for punishment. On this point, proximate (aka "legal") cause enters deliberations. It goes back at least to Sir Francis Bacon, the seventeenth century philosopher, and former Lord High Chancellor of England. He assumed that causation peters out over time (F. BACON 2013 [1630]) and so what matters is its immediate harm. MILL (1874, 240) echoes him. But, as the great expert on torts, Prosser, remarked: a proximate cause can be a matter of legal policy and have nothing to do with causation (PROSSER 1964). About proximate cause, he declared: «There is perhaps nothing in the entire field of law which has called forth more disagreement, or upon which the opinions are in such a welter of confusion» (PROSSER 1964, 311), a remark that is primary evidence for our judgment of conceptual turmoil about causation in the law. Yet, we do need to understand the policies and purposes proximate causes are supposed to serve. It aims to curtail the consequences of a person's actions that otherwise might be unforeseeable or continue forever. Perpetrators may also include those who should be protected from accountability. In such cases, the law eliminates liability, and it does so by determining that defendants—even those who caused injury—are not proximate causes, and not liable. At the very least, the concept is misnamed, and sometimes it distorts the concept of causation in order to eliminate liability. It demands spatio-temporal proximity between a cause and its effect. And so only the last in a series of human actors who do harm is its proximate cause (PROSSER 1964, 316). Granted that Brutus was the final conspirator to stab Caesar in Shakespeare's (or Francis Bacon's?) play, he is the proximate cause of Caesar's death. Of course, no causal justification exists to warrant this decision—Casca, who struck first, could have delivered the only mortal blow.

In summary, factual causes in law have an operational definition in the *but-for* test, yet as we will show it yields erroneous diagnoses. Proximate causes do not always concern "cause": their aim is to curtail liability, and so they can also do violence to causation. Likewise, no ground exists for jurists' anxiety that the ramifications of a causal action are boundless, so that proximate causes are needed to curtail them. When you turn off a tap, the water stops flowing, and that can be the end of a causal sequence. Our contention is that the problems of causation in the law have a two-fold solution. Proper policies that have nothing to do with causation should deal with exclusions of liability, and the *but-for* test should be replaced with one based on the correct analysis of factual causation. The first author put forward a skeletal version of this solution over twenty years ago (JOHNSON-LAIRD 1999). We hope that the consequences of the present account will outlast those of turning off a tap.

3. *Theories of causation*

A search for the right analysis of causation starts with Aristotle. He presents a brief typology of four sorts (see his *Metaphysics*, Book V, 1013a and b in ARISTOTLE 1984). Causes, he says, are origins. So, the substance of which something is made is its material cause, e.g., silver. The form in which it is made is its formal cause, e.g., the shape of a saucer. The origin of a change is its efficient cause, e.g., the silversmith who made the saucer. And the end for which something is done is its final cause, e.g., Athenian rituals calling for saucers. Aristotle allows that two things can be causes of one another, though not in the same way, e.g., health causes walking, and walking causes health (*ibidem*, 1013b). As often with Aristotle, the acuteness of his distinctions is striking. Yet, he does not attempt to answer our foundational question about the meaning of a causal assertion.

The first great thinker to try to answer the question was the philosopher David Hume, a pillar of the Scottish enlightenment after the rise of science. He wrote: «We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the

second never had existed» (HUME 1988 [1748], 115). According to his first definition, a general causal assertion, such as:

Stabbing a person causes the person to bleed

has the following construal suitable for analysis in standard logic (see JEFFREY 1981):

If any person is stabbed then the person bleeds.

It refers to a constant conjunction of cause and effect. But, Hume's second definition, which he prefaces with "in other words," is different. It proposes that a singular event such as:

The stabbing of this man caused him to bleed

has a counterfactual interpretation:

If this man had not been stabbed then he would not have bled.

This analysis may be the origin of the *but-for* test (cf. MILL 1874, 237), i.e., *but for* his stabbing this man would not have bled.

After Hume, a deluge. Analyses of causality introduced temporal sequence, spatial contiguity, necessary and sufficient conditions, and scientific laws. As we have seen, some of these ideas entered the law. Yet, Hume anticipated and dismissed many of them. In daily life, however, one constraint is that the onset of a cause does not occur after the onset of its effect. The only violations are in accounts of time travel, and in a physicist's conjecture that there is only one electron that can travel backwards in time. Yet, cause and effect sometimes seem contemporaneous, as when a lead ball causes a dent in a cushion (KANT 1934 [1781], 156).

Spatial contiguity is not strictly necessary for causal relations in everyday life, as in psychological relations, e.g.:

Marcel's jealousy caused him to hate Albertine

or in Einstein's claim about "spooky" action at a distance in quantum physics (GOLDVARG & JOHNSON-LAIRD 2001). Spatio-temporal contiguity, however, can be a cue to a causal interpretation of events.

Other mooted transmitters of causation include a mechanism, power, force, energy, or means of production. KOSLOWSKI (1996, 6) wrote, «a causal mechanism is the process by which a cause brings about an effect». Clocks, engines, and computers, embody hierarchies of causal relations in hardware. Each relation may have its own underlying mechanism, but the "recursion" has to bottom out. So, at least one causal assertion does not invoke mechanisms, or else they exist "all the way down" like the turtles that were once said to support the earth. Our skepticism in no way impugns the role of known mechanisms in the induction of causal relations (e.g., AHN et al. 1995). Likewise, causes can transmit energy that has an effect (HARRÉ & MADDEN 1975, 5). A recent formulation roots causation in people's perceptions of force (WOLFF & BARBEY 2015). When a tennis player smashes a racket on the surface of the court, the energy of the blow is transferred to the racket, bending it out of shape. Other claims are more abstract, and call for scientific principles to lie behind every causal assertion about particular events (HART & HONORÉ 1985, Ch. I., Sec. II). The best evidence for all these accounts is that knowledge of a mechanism, power, force, or principle, can override the covariation of a putative cause and

effect on judgments of causality (e.g., KOSLOWSKI 1996; WHITE 1995). Their presence is a cue to infer a causal relation, but, as we will show, their absence is not enough to infer its absence.

Theorists err when they confuse a philosophical assumption with an element of meaning. MILL (1874, 236) proposed a law of universal causation: «The truth that every fact which has a beginning has a cause». But, this law cannot be part of the meaning of "cause", because if it were, then those who dissent and argue, say, that the big bang had no cause, are not propounding an alternative philosophy, but arguing about the meaning of a word. Lexicography is not metaphysics. The assertion:

Every event has a cause

makes just as good sense as its denial, which is not the self-contradiction that Mill's law implies. The meaning of "causes" embodies neither his universal law nor its quantum mechanical denial. We draw an analogous moral about the crucial case of causation by omission, e.g.:

The signalman's failure to set the signal caused the crash.

MILL (1874, 239) recognized such causes, but argued: «From nothing, from a mere negation, no consequences can proceed». Hence, he claimed that omissions are causes only as a result of their positive effects. Proponents of the theory that causes depend on forces have made an analogous argument (WOLFF et al. 2010). They treat the assertion:

The lack of a jack caused the car to fall to the ground

as an instance of "double prevention". Hitherto, the force of the jack prevented the car from falling, but then another force—say, a mechanic kicking the jack away—prevented this prevention, and so the car fell to the ground. In general terms: A prevented B from preventing C, and so the kick (A) prevented the jack (B) from preventing the fall (C), and so the lack of B caused C. This account works for some assertions, but not for others, such as:

That there's no highest prime number caused Euclid an enormous surprise.

There are no forces or double preventions in this case. The force theory yields an insightful explanation of how interactions among forces yield causal relations. But, as Khemlani and his colleagues have shown, the theory of mental models—to be described below—provides a unified account of omissions as causes, as enabling conditions, and as preventions (KHEMLANI et al. 2018).

In sum, Occam's razor cuts the meanings of causal relations to their rudiments. It excises contiguity, power, force, energy, mechanism, means of production, and scientific principles. Knowledge can add these factors, but causal meanings themselves do not need them. Hume was right.

3.1. *Necessity and counterfactuals*

Is there a necessary connection between cause and effect, i.e., given a cause is its effect bound to occur? Hume dismissed necessity in place of constant connection. But, Kant argued that causation is a notion demanding «that something, A, should be of such a nature, that something else, B, should follow from it necessarily» (KANT 1934 [1781], 90). So, his claim depends on conceptual (or "alethic") necessity. Indeed, no difference exists in the conditions in which the following two assertions are true, or in those in which they are false, i.e., their "truth conditions":

These circumstances cause plants to grow.
Given these circumstances, plants are bound to grow.

Here, "bound to" plays a role akin to certainty or to necessity (JOHNSON-LAIRD & RAGNI 2019).

The second part of Hume's analysis of causation, which we cited earlier, relates it to counterfactual claims, such as:

If the cause hadn't happened then the effect wouldn't have happened.

Various modern philosophers have also advocated a counterfactual semantics for causation (e.g., LEWIS 1973a; MACKIE 1980). So, a particular causal claim, such as:

His lack of vitamin C caused him to develop scurvy

is synonymous with:

If he hadn't lacked vitamin C then he wouldn't have developed scurvy.

Such counterfactual conditionals have well-known semantic accounts based on "possible worlds" (e.g., LEWIS 1973b; STALNAKER 1968). Their essence is that the preceding conditional is true if the claim that he did not develop scurvy is true in the possible world or worlds most like the real world amongst those in which he did not lack vitamin C. Hence, his deficiency in vitamin C leads of necessity to scurvy. The number of possible worlds is too vast to be psychologically plausible, and so the Hawk-Eye TV system is a better model of how humans assess counterfactuals. They envisage how one situation would be without a key component, the lack of vitamin C, and they evaluate the alleged consequence in this modified situation. Yet, the moral remains the same: the truth of the counterfactual above implies the alethic necessity of the effect given the cause.

Counterfactuals are instantiated in PEARL's (2009) use of Bayes nets to explain causation (see also SPIRTES et al. 2000; SLOMAN et al. 2009). Pearl argues that causation is a deterministic notion. A purely probabilistic account, he says, cannot distinguish between *rain causes mud* and *mud causes rain*. Human ignorance, however, demands that causation is treated as probabilistic. The structure of "causal diagrams" is at the core of Pearl's approach. In one of his examples (PEARL 2009, Figure 7.1), a court orders the execution of a prisoner. The captain in charge of the firing squad gives the order to two riflemen. Both of them shoot, and the prisoner dies. Figure 1 presents this scenario in a causal diagram in which each node refers to a proposition that can either be true or false. Pearl's argument begins with an analysis based on standard logic in which each arrow in the diagram denotes a biconditional of the sort, *If and only if A then B*, which we abbreviate as, *Iff A then B*, and so the figure is equivalent to this set of assertions:

Iff the Court orders the execution then the captain gives the order to the riflemen.
Iff the captain gives the order to the riflemen then rifleman 1 fires.
Iff the captain gives the order to the riflemen then rifleman 2 fires.
Iff rifleman 1 fires or rifleman 2 fires, or both fire, then the prisoner dies.

In standard logic, a biconditional, *iff A then B*, is true when *A* & *B* is true or when *not A* & *not B* is true; otherwise, it is false (JEFFREY 1981, 71). It follows that Figure 1 represents only two contingencies: either all its assertions are true, or else all of them are false. So, various conclusions can be drawn from logic alone, such as:

If rifleman 1 didn't fire then the prisoner didn't die.

But, consider the counterfactual event in which rifleman 1 fires without waiting for the captain's order. It violates the logical interpretation of Figure 1. But, it is the sort of event that can occur. PEARL (2013) treats it as demanding an *intervention* that calls for surgery on the diagram in which a "do-operator" cuts the link from the captain's order to rifleman 1 in Figure 1. All else remains the same, and so the following counterfactual conditional is now true:

If rifleman 1 hadn't fired then the prisoner wouldn't have died.

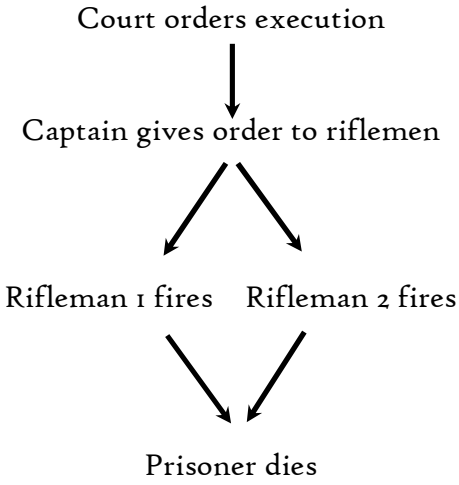


FIGURE 1. A causal diagram of the firing squad example (after PEARL 2009, Figure 7.1)

This treatment amounts to altering a representation of reality, like Hawk-Eye does, by removing one of its components.

Pearl assumes that causation depends on an intervention, not a mere observation. So, whenever an intervention occurs, you remove all arrows from the causal diagram that point to the event (rifleman 1 fires), you fix its value to true, and you use logic to deduce the consequences (the prisoner dies). Pearl's central principle is thus:

Y causes Z if we can change Z by manipulating Y (PEARL 2009, Ch. 11).

He does not intend this principle to define the meaning of "causes". Indeed, it doesn't, because it is not self-contradictory to assert:

The big bang caused the universe to begin, and we cannot manipulate the big bang.

Causal diagrams do not represent conditions that *enable* events to occur, but Sloman and his colleagues have shown how they might do so. The crucial part of *A enables B to happen* is «that A represents a category of events necessary for B, and that an alternative cause of B exists» (SLOMAN et al. 2009, 21). Yet, the lack of a cause does not contradict an enabling assertion, e.g., "The gas enabled an explosion to occur, but luckily no spark occurred to cause it". An enabling condition requires only the possibility of a cause, not its occurrence. As Pearl argues, his analysis supports the *but-for* test in law:

But for the rifleman firing, the prisoner wouldn't have died.

Every arrow in a causal diagram is supposed to signify a potential cause, and the theory presupposes that causation has no simpler analysis. By design, however, it offers no answer to our foundational question about the meaning of "causes".

The example of the firing squad elides an important distinction. The captain ordered the riflemen to fire, but orders are not causes. You cannot disobey a cause. The law reflects the difference: human beings can decide whether or not to obey to an order (HART & HONORÉ 1985, Ch. II). Like causes, words can embody intentions in their meanings, e.g.: chase, emigrate, examine, aim, look for, seek, assassinate (see MILLER & JOHNSON-LAIRD 1976, Sect. 6.3). You can't chase something by accident. Whether an action was intentional can affect a defendant's culpability, but not whether it is a cause.

3.2. *Probabilistic theories*

Uncertainty plagues causality. A philosophical view, originating in the twentieth century, is that the causation is itself probabilistic. REICHENBACH (1956) proposed such an analysis, arguing that *C causes E* if the probability of E given C is greater than the probability of E given not C. He allowed, however, that a putative cause could be "screened off" if there was an earlier probabilistic cause, D, yielding both C and E. Cognitive scientists have also defended probabilistic theories of the meaning of conditionals (e.g., OAKSFORD & CHATER 2007). And Cheng and her colleagues have defended a similar theory of causation based on normalizing the difference between the two conditional probabilities above (CHENG 1997). «Because causal relations are neither observable nor deducible, they must be induced from observable events» (CHENG 1997, 367). In fact, when a display appears to show one billiard ball colliding with another stationary one, it elicits a compelling perception of the first ball causing the second one to move, even though it is, in fact, an illusion (MICHOTTE 1963 [1946]).

The main case for a probabilistic semantics is that people judge that a causal relation holds in instances in which exceptions occur (e.g., CHENG & NOVICK 1991). But, this evidence is hardly decisive. People accept that the coronavirus causes Covid-19, yet they know that not everyone who has the virus develops the disease. Hidden variables affect a person's susceptibility to it. The claim itself is a generic one, akin to *ducks lay eggs*, which people accept, even though they know that only females lay eggs (LESLIE et al. 2011). It is therefore sensible to maintain that an assertion such as:

Smoking caused his death.

has a deterministic meaning, because it differs from:

Smoking probably caused his death.

If causation were probabilistic, the two assertions should share the same truth conditions. Pearl is right: causation is not probabilistic, but the evidence supporting it can be. So, causation and probabilistic considerations need to be kept separate—a division that many theories appear to recognize (see, e.g., WALDMANN 1996; LAGNADO & GERSTENBERG 2017).

3.3. *Causes and conditions*

A spark occurs in a container of combustible gas, and an explosion follows. *But for* the co-occurrence of gas and spark, there would have been no explosion. Hence, MILL (1874, 238) treated both as "conditions", and argued that the choice of one of them to be the cause was capricious. Other theorists have argued that causes are abnormal whereas enabling conditions are normal (HART & HONORÉ 1985, Ch. II), that causes are inconstant whereas enabling

conditions are constant (CHENG & NOVICK 1991), that causes violate a norm whereas enabling conditions occur by default (e.g., EINHORN & HOGARTH 1986), and that causes are relevant to explanations whereas enabling conditions are not (MILL 1874, 238; MACKIE 1980; TURNBULL, SLUGOSKI 1988; HILTON & ERB 1996). What all these accounts presume is that there is no difference in meaning between causes and enabling conditions. Readers might suppose that, say, *abnormal* is part of the meaning of cause. But, as assertion such as: "A spark is not an abnormal cause of an explosion", makes perfect sense, whereas an assertion such as: "A spark occurring after the explosion caused it" is nonsensical. Such tests show that inconstancy, violation of a norm, and relevance to explanation, are also not part of the meaning of "cause". The theory of mental models, to which we now turn, makes the same point.

4. *The mental model theory*

4.1. *The basics*

Mental models are representations of the world, and the development of the "model" theory—as it is referred to—has accumulated five main principles:

(1) Each mental model represents a possibility with many different instances that have in common only what the model represents. Its structure is *iconic* in that it matches the structure of what it represents insofar as possible. So, a kinematic model unfolds in time, just as the sequence of events that it represents does (JOHNSON-LAIRD 1983; KHEMLANI et al. 2013). Models can also contain abstract symbols, which are not iconic, such as negation, and each of them is linked to its semantics.

(2) Individuals cope best with one model at a time (JOHNSON-LAIRD 1983). An *intuitive* model represents only those clauses in the premises that are true; but a *deliberative* model also represents those clauses that are false—using negation to do so. So, humans have an intuitive system of reasoning, and a deliberative system—an idea due to the late Peter Wason, which has since flourished in many forms, though only the model theory appears to have implemented the two systems in a computer program. Intuition focuses on a single model at a time; deliberation can find models that serve as counterexamples to intuitive conclusions (e.g., KHEMLANI & JOHNSON-LAIRD 2021; RAGNI et al. 2018).

(3) Models represent three main sorts of possibility (JOHNSON-LAIRD & RAGNI 2019): what can happen (epistemic possibilities), what is permissible (deontic possibilities), and what is conceptually feasible (alethic possibilities). They can each refer to real or to counterfactual possibilities, i.e., situations that were once possible but that didn't happen (BYRNE 2005; BYRNE & JOHNSON-LAIRD 2020). The manipulation of a model of reality can assess the veracity of a counterfactual (GERSTENBERG et al. 2019; BYRNE et al. 2022).

(4) The interpretation of any assertion is open to *modulation* in which knowledge, meanings, and referents, can eliminate models, add information to them, or assign a truth value to them (e.g., QUELHAS et al. 2019).

(5) An inference is alethically necessary if its conclusion holds only in the models of its premises. Reasoners withdraw a conclusion when they discover that it is false, and revise the premises to try to construct an explanation that resolves the inconsistency (JOHNSON-LAIRD et al. 2004).

Mental models differ from standard logic in several ways. In logic, a contradiction yields valid inferences of any conclusion whatsoever, because there are no cases in which the premises are true and the conclusion is false (JEFFREY 1981, 1). In mental models, contradictory premises yield an

empty model from which it follows only that something is wrong with the premises. Its effects are local. As HINTERECKER and colleagues (2016) showed, people accept inferences of this sort:

Either she died from poison or she died from her wounds.
So, *it is possible that she died from poison.*

In the model theory, the premise yields a conjunction of possibilities, which each holds in default of knowledge to the contrary. One of these possibilities is the conclusion above. Yet it is invalid in all standard modal logics dealing with possibilities.

The model theory applies to all sorts of reasoning, and it explains how reasoners' emotions can improve their reasoning (GANGEMI et al. 2013). It has been implemented in computer programs that take descriptions as their inputs, build models of the situations under description, and use these models both to draw conclusions—just as human reasoners do—and to evaluate given conclusions[6]. We now consider what the theory says about causation.

4.2. *The meanings of general causal and enabling assertions*

Models can represent forces, mechanisms, spatial and temporal contiguity, and much else. But, one of the theory's principal assumptions, going back to GOLDVARG & JOHNSON-LAIRD (2001) and from thence to MILLER & JOHNSON-LAIRD (1976, Sect. 6.3) is:

> «The meanings of causal relations refer solely to sets of possibilities with the temporal constraint that causes do not start after their effects do» (JOHNSON-LAIRD & KHEMLANI 2017, TABLE 10.1).

Any causal assertion is synonymous with a conditional assertion referring to epistemic possibilities (JOHNSON-LAIRD & BYRNE 2002; BYRNE & JOHNSON-LAIRD 2020). For instance, the causal assertion:

Taking substance D causes you to sleep

has the same truth conditions as the conditional assertion:

If you take substance D then you sleep.

It does not follow that all conditionals express causal relations. They can assert other sorts of relation, such as so-called "evidentials", e.g., *If the door is closed, then the elevator is on another floor*. And causal assertions themselves can be elliptical, e.g., *Drivers cause accidents*, implies that the actions or inactions of drivers cause accidents to occur.

The meaning of a general causal assertion refers to a conjunction of default possibilities that have the appropriate temporal constraint. Individuals tend to represent possibilities in *intuitive* models of the world, which the computer program implementing the theory denotes using words. So, the causal claim that taking substance D causes you to sleep has these two intuitive models:

substance D sleep
. . .

The first row in this diagram designates a model of the salient possibility in which you take

[6] These programs are at https://www.modeltheory.org/models/.

substance D and sleep, and the second row, the ellipsis, designates a model with no explicit content but that allows for alternative possibilities in which you do not take substance D. When people think hard, they can enumerate these possibilities in *deliberative* models, where "not" is a symbol for negation:

substance D	sleep
not substance D	not sleep
not substance D	sleep

The first model represents the possibility in which a person takes substance D and sleeps; the second model represents the possibility in which the person does not take substance D and does not sleep; and the third model represents the possibility in which the person does not take substance D and sleeps. The three models establish that substance D causes you to sleep. It suffices, but, as the third model shows, it is not necessary, because other causes of sleep are possible too. What the assertion rules out as impossible is that you take substance D and do not sleep.

Some causes are unique, such as:

Drinking excessive alcohol causes you to get drunk.

Excessive alcohol is the only cause of becoming drunk: it suffices and it is necessary. The assertion refers to only two possibilities, which have these deliberative models:

excessive alcohol	drunk
not excessive alcohol	not drunk

Enabling assertions refer to a different set of possibilities to those for causes. For instance, the assertion:

Taking substance D enables you to sleep

has these intuitive models:

substance D	sleep
. . .	

and so the possibilities seem the same as those for a causal assertion. However, the deliberative models of enabling conditions differ from those above for causes:

substance D	sleep
substance D	not sleep
not substance D	not sleep

Substance D is necessary for you to sleep—without it, you won't sleep, but with it, you may or may not sleep. A weaker sort of enabling condition is one that also allows you to sleep without substance D: its deliberative models represent all four contingencies as possible.

To prevent something is to cause it not to occur, and prevention too can be unique. So, the models for:

Taking substance E prevents you from sleeping

are identical to those for:

Taking substance E causes you not to sleep.

The model theory allows that a non-event can cause an effect (KHEMLANI et al. 2018). So, for example, the assertion:

Not taking substance E causes you to sleep

refers to these possibilities:

not substance E	sleep
substance E	not sleep
substance E	sleep

The second and third possibilities show that on taking substance E, you may or may not sleep. This example could be a double prevention, which we described earlier (WOLFF et al. 2010). But, other examples cannot be:

The idea that Zeus does not exist caused ancient Athenians to laugh.

It is nonsensical to suppose that Zeus's existence prevented Athenians from laughing.

TABLE 1 presents the four main sorts of causal assertion, and their intuitive and deliberative models. Each assertion refers to a conjunction of a set of possibilities that each hold in default of knowledge to the contrary. Causes and preventions can be unique, and therefore have only two models. Enabling conditions can be weak and therefore have all four models. The variables A and B denote events or situations, but they can also denote their non-occurrence or non-existence. The complete set of causal relations therefore consists of seven relations: the four sets of three possibilities in TABLE 1, two sets of two possibilities for strong causes and preventions, and one set of all four possibilities for a weak enabling condition (see JOHNSON-LAIRD & KHEMLANI 2017, for a proof that these exhaust the set of possible causal relations).

4.3. *Empirical corroborations*

Several lines of evidence have corroborated the model theory of causal relations. One line bore out the difference in meaning between causes and enabling conditions. When participants in an experiment had to list what is possible and what is impossible given assertions based on each of the main sorts of causal verb—*causes*, *enables*, and *prevents*, their listings had reliable matches with the predicted possibilities (GOLDVARG & JOHNSON-LAIRD 2001, Experiment 1). A further study showed that individuals distinguish between causes and enablers when they occur together in a ternary relation (GOLDVARG & JOHNSON-LAIRD 2001, Experiment 2). The participants were given brief scenarios, and their task was to identify the cause: "it brings about the event," and the enabling condition: "it makes the event possible". Here is a typical scenario:

Given that there is good sunlight, if a certain new fertilizer is used on poor flowers, then they grow remarkably well. However, if there is not good sunlight, poor flowers do not grow well even if the fertilizer is used on them.

Causal Assertions	Intuitive Models	Deliberative Models
A causes B.	A B . . .	A B not A not B not A B
A enables B.	A B . . .	A B A not B not A not B
A prevents B.	A not B . . .	A not B not A B not A not B
A enables not B.	A not B . . .	A not B A B not A B

TABLE 1. The four principal sorts of causal relation, the intuitive models of their possibilities, and the deliberative models of all their explicit possibilities. The ellipsis denotes an intuitive model with no explicit content, and the variables, A and B, can also have negative instances as their values. Strong interpretations of *causes* and *prevents* refer to only their first two possibilities in the table; weak interpretations of *enables* and *enables not* refer to all four possibilities.

As the participants realized, with sunlight, the fertilizer causes growth; without sunlight, it does not. The ternary relation encapsulating the scenario is:

Sunlight enables the fertilizer to cause the flowers to grow.

The experiment included scenarios embodying a ternary relation switching the roles of the two agents:

The fertilizer enables sunlight to cause the flowers to grow.

It also counterbalanced the order of mention of the causers and enablers, and all the participants saw different contents for each of the four different sorts of scenario. They correctly identified causes and enabling conditions on 85% of occasions. The switch in causal roles between fertilizer and sunlight refutes the role of abnormality, rarity, violation of norms, and other such concepts, as essential to causes. The pattern of possibilities for all eight cases for fertilizer, sunlight, and growth differs between the two assertions above (see GOLDVARG & JOHNSON-LAIRD 2001, 578 f.).

A similar study of the legal implications of causes and enablers used scenarios about harmful effects (FROSCH et al. 2007). A typical scenario was:

Mary threw a lighted cigarette into a bush. Just as the cigarette was going out, Laura deliberately threw petrol on it. The resulting fire burnt down her neighbor's house[7].

[7] Compare: *Watson V. Kentucky & Indiana Bridge & Railway Co.* (1910) 137 KY. 619, 126 S.W. for the doctrine of "intervening cause"—a concept that Stuart Lichten told us about. The initial act of negligence is curtailed by the intervening causal action, but only if it is intentional.

Once again, participants distinguished between causers and enablers. They judged causers as more responsible for the consequences than enablers, and as deserving to serve much longer prison sentences and to pay much greater damages. The causers' responsibility, as a recent study showed, is judged to be even greater when the order of events is described in the same order as they occur in reality than in the opposite order (LIMATA et al. 2022). The natural order of the enabling condition as occurring prior to the cause may make the construction of a kinematic model easier, and so the difference in responsibility between the two agents is clearer. As readers should recall, neither Anglo-American nor continental law recognizes the difference in meaning between a causer and an enabler. Yet, it affects culpability in daily life.

Another line of evidence for the model theory concerns the consequences of causes and enabling conditions. Most participants in a study responded "yes" to the following sort of inference (GOLDVARG & JOHNSON-LAIRD 2001, Experiment 1):

Eating protein will cause her to gain weight.
She will eat protein.

Will she gain weight?
But, as the models of their possibilities in TABLE 1 predict, they responded "no" to the same inference when its initial premise referred to an enabling condition:

Eating protein will enable her to gain weight.

The assertion allows that she may not gain weight.

Still another line of evidence bore out the model theory's deterministic meaning for causal relations (FROSCH & JOHNSON-LAIRD 2011). In various experiments, individuals enumerated the evidence that they thought would refute general causal claims. Most participants called for single pieces of evidence to refute causes, enablers, and preventions. They tended to choose refutations of the form, *A and not B*, for both causes and enablers, but, as predicted, they chose those of the form: *not-A and B* more often for *A enables B* than for *A causes B*.

In sum, contrary to the tradition that MILL (1874) inaugurated, causal and enabling conditions differ in meaning, in inferential consequences, and in evidence that refutes them. Individuals can tell them apart. And the difference does not depend on normal versus abnormal conditions, constant versus inconstant conditions, probable versus improbable conditions, or on what is relevant versus irrelevant to explanations. The contents in the preceding experiments rule out these factors as crucial distinctions. All that matters are the possibilities to which assertions refer and their appropriate temporal constraint.

4.4. *Novel causal relations*

The model theory distinguishes between the possibilities to which various ternary relations refer, such as: *A enables B to cause C*, and *A causes B to cause C*. Their possibilities differ from those for the conjunctions of relations: *A enables B, and B causes C*, and *A causes B, and B causes C* (GOLDVARG & JOHNSON-LAIRD 2001). The theory allows for two-way causal relations—a pertinent example is that people's beliefs about deontic matters affect their emotions, and their emotions about these propositions affect their beliefs in them (BUCCIARELLI & JOHNSON-LAIRD 2019a, 2019b, 2020; BUCCIARELLI et al. 2008). The theory also allows for cyclical causal relations, as in feedback loops in the body and in other systems. It is sensitive to the difference between causes and orders from persons in authority. No other theory, as far as we know, copes with all of these matters.

5. *The legal implications of the model theory*

The principal implications of the model theory for jurisprudence are to refute the *but-for* test, which makes errors of commission and omission, to establish viable replacements for it, and to elucidate hitherto puzzling legal judgments.

The essential problem with the *but-for* test is that it establishes that one event is necessary for the occurrence of another.

But for his failure to wear a seat-belt, he would not hit his head on the dashboard.

Granted that the assertion is true, his failure to wear a seat-belt did not cause him to his head on the dashboard. It enabled the event to happen. An enabling condition is necessary for an event to occur, because without it, the event does not occur (see the possibilities in TABLE 1). Hence, the test works for unique causes, because they are necessary (and sufficient) for the effect to occur. But, it makes an error of commission in treating enabling conditions as causes.

Consider this claim in a situation with a firing squad:

The rifleman's shot caused the prisoner's death.

The three prior possibilities are as follows (see the possibilities in TABLE 1):

rifleman shoots	prisoner dies
not rifleman shoots	not prisoner dies
not rifleman shoots	prisoner dies

The third possibility is that another rifleman shoots and the prisoner dies. The fact described in the assertion above converts the first of the three possibilities into a fact, and the second and third of them into counterfactual possibilities; what remains impossible, as in Pearl's scenario above, is that the rifleman shoots and the prisoner does not die. The assertion above therefore makes a true causal claim, but the but-for test fails to diagnose it:

But for the rifleman shooting, the prisoner would not have died.

The test assertion is false, because another rifleman could have shot the prisoner. So, the *but-for* test also makes errors of omission. It fails to detect certain causes that are sufficient but not necessary. It is remarkable that a great philosopher should have envisaged this flawed test, and that so many eminent scholars have endorsed it. The reason is that unaided thought cannot hold in mind at the same time four different contingencies (JOHNSON-LAIRD 1983). We therefore propose to replace the *but-for* test with two new tests, one for causes, and one for enabling conditions. They derive from the sets of possibilities in TABLE 1.

5.1. *The cause-test and the enable-test*

The *cause*-test determines whether *A caused B* given an observation of their conjunction in an appropriate temporal order. The test does double duty and diagnoses prevention given *A*, and *B* is a non-occurrence. Figure 2 presents the *cause*-test, and we illustrate how it works using an example:

Laura threw petrol on a smouldering cigarette and a fire occurred.

We use common sense and knowledge to test whether Laura's action (A) of throwing petrol on a smouldering cigarette caused the fire (B), and the test allows only a small change to the actual situation in the evaluation of counterfactuals. The answer to the first question in Figure 2 is that Laura's action (A) with no subsequent fire (not B) could not have occurred. The answer to the second question is: Yes, Laura not carrying out her action (not A), with no subsequent fire (not B) could have occurred. The answer to the third question is: No, without her action, no fire could have occurred. The resulting diagnosis is therefore:

Laura's action caused the fire to occur.

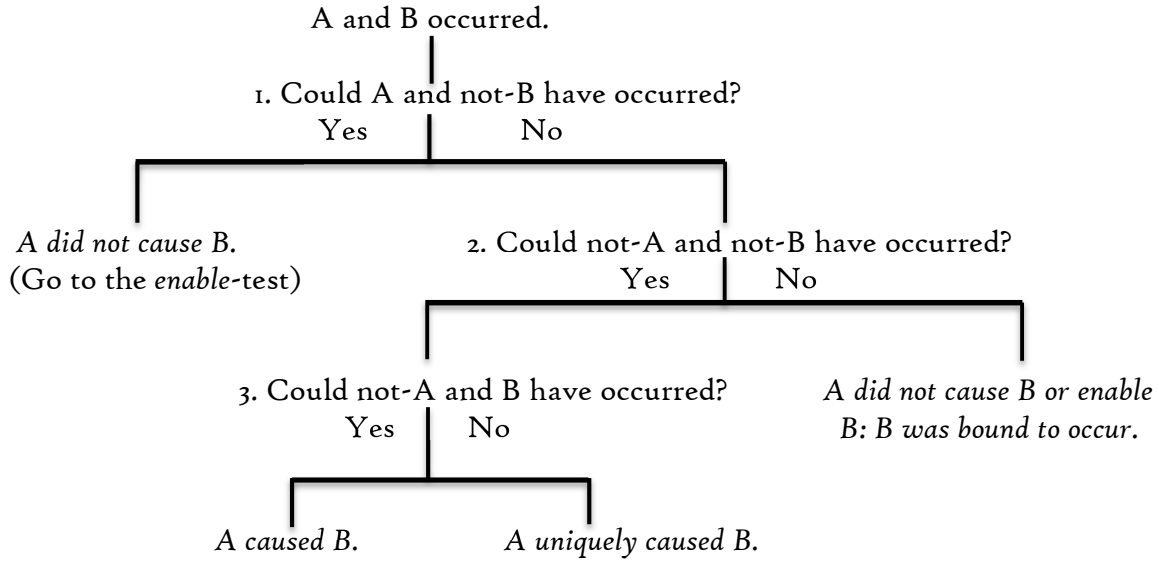


FIGURE 2. The *cause*-test: given the conjunction of A and B in an appropriate temporal order, the answers to three questions diagnose whether *A caused B*. The variables, A and B, each refer either to situations or to their non-occurrences, so when B is a non-occurrence, *causes* is equivalent to *prevents*.

Figure 3 presents the *enable*-test, which determines whether *A enabled B* given their appropriate conjunction. It too depends on the answers to three questions, and we illustrate it using a different example:

Mary threw a cigarette into a bush and a fire occurred.

The answer to the first question in Figure 3 is: Yes, Mary's action (A) could have occurred without the subsequent fire (not B): the smouldering cigarette could have gone out. The answer to the second question is: No, her not carrying out her action (not A) but the fire still occurring (B) couldn't have occurred (without a major change to the situation). The answer to the third question is: Yes, she could have not carried out her action (not A) and the fire not occurred (not B). The resulting diagnosis is:

Mary's action enabled the fire to occur.

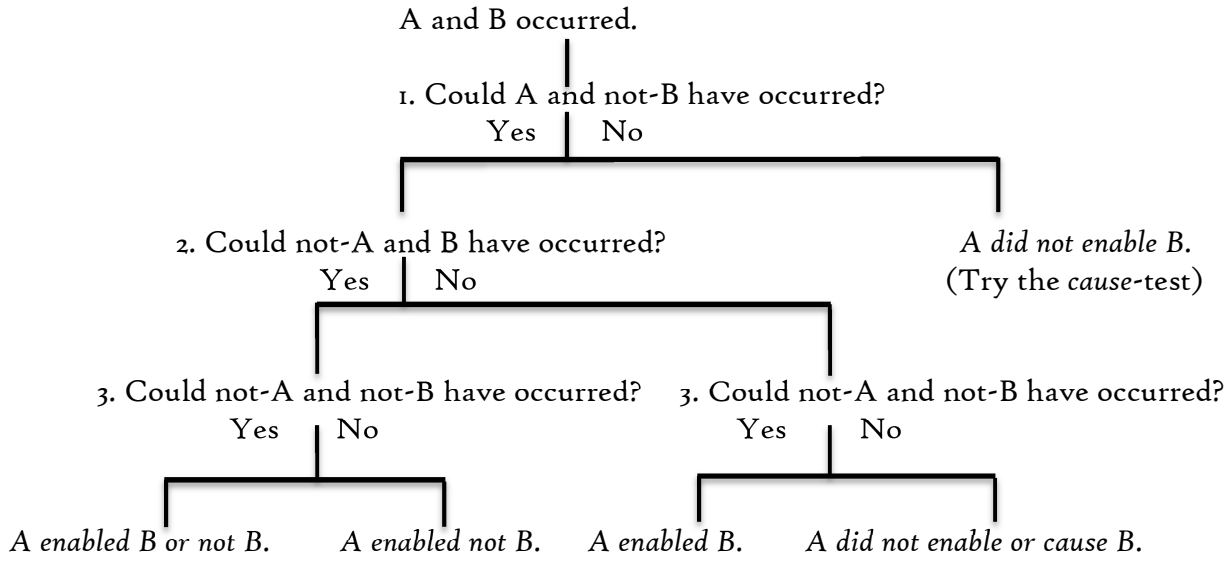


FIGURE 3. The *enable-test*: given the conjunction of A and B in an appropriate temporal order, the answers to three questions about diagnose whether A enabled B. The variables, A and B, each refer either to situations or to their non-occurrences, so when B is a non-occurrence, *enables* is equivalent to *enables B not to occur*.

The tests are more complex than the *but-for* test, but they deliver the right results.

5.2. *An elucidation of the courts' judgements in the three test cases*

The model theory's *cause*-test and *enable*-test elucidate the three court cases, which began this article. The man who stabbed the Jehovah's Witness caused her wounds. Their treatment called for a surgical operation. The court decided that her not having the preliminary blood transfusion for the operation was the "reason" for her death, but assigned its cause to her wounds. The answers to the *cause*-test (where A = No transfusion or operation, and B = death) are:

No	transfusion	death:	Fact
No	transfusion	no death:	Question 1: No: it could not have occurred.
	transfusion	no death:	Question 2: Yes: it could have occurred.
	transfusion	death:	Question 3: Yes: it could have occurred.

So, not having the blood transfusion and operation caused her death. Yet, the court judged her wounds as the cause of her death; they enabled it to occur. The court therefore erred.

The case of the woman who was attacked and raped was different. Her terrible assault was the reason she poisoned herself. As the court decided, the cause of her death was either the poison alone or else its combination with her wounds. The *cause*-test on the poison yields the answers:

poison	died:	Fact
poison	did not die:	Question 1: No: it could hot have happened.
no poison	did not die:	Question 2: Yes: it could have happened.
no poison	died:	Question 3: Yes: it could have happened.

So, her own action caused her death. In contrast, the *cause*-test on her wounds fails the first question: it could have happened that with her wounds she did not die. Hence, the poison with or without her wounds caused her death. Her assailant caused her death only if his attack or his delaying her treatment, or both, had mortal consequences. His actions were the reason for her intentional action of taking the poison.

In the case of the man who fell down an elevator shaft, the court decided that the proximate cause of his accident was the open door. The *cause* and *enable* tests show that the open door did not cause his fall, because he might not have fallen even though the door was open. Instead, it enabled him to fall: had it been closed, he could not have fallen. The court mistook an enabling condition for a cause. They could then judge that the defendant was not guilty of negligence. The *enable*-test yields these results:

Open door	injured:	Fact
Open door	not injured:	Question 1: Yes: it could have happened.
Closed door	injured:	Question 2: No: impossibility
Closed door	not injured:	Question 3: Yes: it could have happened.

So, the open door enabled the man to fall and to injure himself.

We have explored the application of the two tests to complex scenarios characterized by a ternary relationship in which one action enables another to function as a cause. So, when the enabler occurs, the second event functions as a cause; and when the enabler does not occur, the second event cannot function as a cause. This relation is distinct from the conjunction of A enables B, and B causes C (GOLDVARG, JOHNSON-LAIRD 2001), and a typical instance occurs when a catalyst enables a chemical to have a causal effect on another. Future experiments could explore the possibility of developing a test that allows this ternary relationship to be explicitly evaluated, and to be contrasted with the conjunction of the two relations. What makes such tests seem impractical is the difficulty of holding distinct models in mind at the same time.

6. *General Discussion*

The model theory explains the meanings of causal relations, their mental representations, and inferences from them. It postulates only rudimentary meanings for causal relations. With an appropriate temporal order: *A causes B* means that given *A* the occurrence of B is the only possibility, *A enables B* means that given *A* the occurrence of *B* is one possibility, and *A prevents B* means that given *A* the occurrence of not-*B* is the only possibility[8]. These relations are deterministic. If they were probabilistic then *causes* would tolerate exceptions and so it would be indistinguishable from *enables*. Actual legal cases contain many hidden factors, which may be alternative causes, enabling conditions, or preventions, and so diagnosis can be difficult. Alternative theories of causation base their interpretations on forces, powers, mechanisms, and interventions. The model theory accommodates these elements: models can embody them as a result of modulation from knowledge and context. Yet, non-events can be causes, which demonstrate that these factors are not part of the fundamental meanings of causal claims.

What can jurists take away from the model theory's account of causation? In our view, it yields five lessons:

[8] See TABLE 1.

(1) The fundamental assertion:

The defendant's action caused the outcome

has two potential meanings, referring either to one among a set of causes or to a unique cause: the defendant carried out an action with an onset no later than the outcome's onset, and given this action the outcome was bound to occur. Without this action, if the outcome might still have occurred, the cause is just one of several; otherwise, the cause is unique.

(2) Both *A causes B* and *A enables B* have the same intuitive models of a conjunction of default possibilities, one explicit and the other with no explicit content:

A B
. . .

The identity of their intuitive models explains past failures to distinguish between causes and enabling conditions. Only deliberative models, and the *cause*-test and *enable*-test, which make all the possibilities explicit, differ between the two[9].

(3) Contiguity in space and time, force, energy, power, and mechanism, are cues to causation, and modulation can embody them in models. But, they are not part of its essential meaning, because they cannot occur in omissions that are causes, enabling conditions, or preventions.

(4) Subtleties can affect legal judgments: an order from a person in authority can be reason for an action, but not its cause; relations can be ternary so that one event enables another to cause an effect; causal relations can be two way and form feedback loops.

(5) The legal *but-for* test is flawed, and errs both in omitting genuine causes and in treating enabling conditions as causes. The *cause*-test and *enable*-test are better guides to diagnosis in real life.

In conclusion, causation is a matter of possibilities, and possibilities can relate to one another in complex ways. If legal decisions distinguished between the culpability of enabling harm and causing it, replaced the *but-for* test with the *enable*-test and the *cause*-test, and abandoned proximate causes in favor of principles curtailing liability, they would take an indispensable first step towards causation as people conceive it in daily life.

[9] See TABLE 1.

References

ANH W., KALISH C.W., MEDIN D.L., GELMAN S.A 1995. *The Role of Covariation Versus Mechanism Information in Causal Attribution*, in «Cognition», 54, 299 ff.

ARISTOTLE 1984. *Metaphysics*, in ID., *The Complete Works of Aristotle*, Vol. 1 (ed. by J. Barnes), Princeton University Press.

BACON F. 2013. *Maxims of the Law*, Cambridge University Press. (Originally published 1630).

BUCCIARELLI M., JOHNSON-LAIRD P.N. 2019a. *Deontics: Meaning, Reasoning, and Emotion*, in «Materiali per una storia della cultura giuridica», 49, 89 ff.

BUCCIARELLI M., JOHNSON-LAIRD P.N. 2019b. *Emotions and Beliefs about Morality can Change one Another*, in «Acta Psychologica», 198, 102880.

BUCCIARELLI M., JOHNSON-LAIRD P.N. 2020. *Beliefs and Emotions about Social Conventions*, in «Acta Psychologica», 110, 103184.

BUCCIARELLI M., KHEMLANI S., JOHNSON-LAIRD P.N. 2008. *The Psychology of Moral Reasoning*, in «Judgment and Decision Making», 3, 121 ff.

BYRNE R.M.J. 2005. *The Rational Imagination: How People Create Alternatives to Reality*, MIT Press.

BYRNE R.M.J, JOHNSON-LAIRD P.N. 2020. *If and Or: Real and Counterfactual Possibilities in their Truth and Probability*, in «Journal of Experimental Psychology: Learning, Memory, and Cognition», 46, 760 ff.

BYRNE R.M.J., KHEMLANI S., JOHNSON-LAIRD P.N. 2022. *Mental Simulations of Counterfactual Possibilities Qualify Judgments of Truth and Falsity*, in submission.

CHENG P.W. 1997. *From Covariation to Causation: A Causal Power Theory*, in «Psychological Review», 104, 367 ff.

CHENG P.W., NOVICK L.R. 1991. *Causes versus Enabling Conditions*, in «Cognition», 40, 83 ff.

EINHORN H.J., HOGARTH R.M. 1986. *Judging Probable Cause*, in «Psychological Bulletin», 99, 3 ff.

FROSCH C.A., JOHNSON-LAIRD P.N. 2011. *Is Everyday Causation Deterministic or Probabilistic?* In «Acta Psychologica», 137, 280 ff.

FROSCH C.A., JOHNSON-LAIRD P.N., COWLEY M. 2007. *It's not my Fault, your Honor, I'm only the Enabler*, in «Proceedings of the 29th Annual Meeting of the Cognitive Science Society», 1755.

GANGEMI A., MANCINI F., JOHNSON-LAIRD P.N. 2013. *Models and Cognitive Change in Psychopathology*, in «Journal of Cognitive Psychology», 25, 157 ff.

GERSTENBERG T., GOODMAN N.D., LAGNADO D.A., TENENBAUM J.B. 2021. *A Counterfactual Simulation Model of Causal Judgments for Physical Events*, in «Psychological Review», 128, 936.

GOLDVARG Y., JOHNSON-LAIRD P.N. 2001. *Naive Causality: A Mental Model Theory of Causal Meaning and Reasoning*, in «Cognitive Science», 25, 565 ff.

HARRÉ R., MADDEN E.H. 1975. *Causal Powers*, Blackwell.

HART H.L.A., HONORÉ T. 1985. *Causation in the Law* (2nd Ed.), Oxford University Press.

HILTON D.J. ERB H.P. 1996. *Mental Models and Causal Explanation: Judgements of Probable Cause and Explanatory Relevance*, in «Thinking & Reasoning», 2, 273 ff.

HINTERECKER T., KNAUFF M., JOHNSON-LAIRD P.N. 2016. *Modality, Probability, and Mental Models*, in «Journal of Experimental Psychology: Learning, Memory, and Cognition», 42, 1606 ff.

HUME D. 1988. *An Enquiry Concerning Human Understanding*, in FLEW A. (ed.), Open Court. (Originally published 1748).

JEFFREY R. 1981. *Formal Logic: Its Scope and Limits* (2nd Ed.), McGraw-Hill.

JOHNSON-LAIRD P.N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Harvard University Press.

JOHNSON-LAIRD P.N. 1999. *Causation, Mental Models, and the Law*, in «Brooklyn Law Review», 65, 67 ff.

JOHNSON-LAIRD P.N., BYRNE R.M.J. 2002. *Conditionals: A Theory of Meaning, Pragmatics, and Inference*, «Psychological Review», 109, 646 ff.

JOHNSON-LAIRD P.N., GIROTTO V., LEGRENZI P. 2004. *Reasoning from Inconsistency to Consistency*, in «Psychological Review», 111, 640 ff.

JOHNSON-LAIRD P.N., KHEMLANI S.S. 2017. *Mental Models and Causation*, in WALDMANN M.R. (ed.) *The Oxford Handbook of Causal Reasoning*, Oxford University Press, 169 ff.

JOHNSON-LAIRD P.N., RAGNI M. 2019. *Possibilities as the Foundation of Reasoning*, in «Cognition», 193, 130950.

KANT I. 1934. *Critique of Pure Reason* (trans. by J.M.D. Meiklejohn), Dutton. (Originally published 1781).

KHEMLANI S.S., JOHNSON-LAIRD P.N. 2021. *Reasoning about Properties: A Computational Theory*, «Psychological Review», available at: https://doi.org/10.1037/rev0000240.

KHEMLANI S.S., MACKIEWICZ R., BUCCIARELLI M., JOHNSON-LAIRD P.N. 2013. *Kinematic Mental Simulations in Abduction and Deduction*, in «Proceedings of the National Academy of Sciences», 110, 16766 ff.

KHEMLANI S.S., WASYLYSHYN C., BRIGGS G., BELLO P. 2018. *Mental Models and Omissive Causation*, in «Memory & Cognition», 46, 1344 ff.

KOSLOWSKI B. 1996. *Theory and Evidence: The Development of Scientific Reasoning*, MIT Press.

LAGNADO D.A., GERSTENBERG T. 2017. *Causation in Legal and Moral Reasoning*, in WALDMANN M.R. (ed.), *The Oxford Handbook of Causal Reasoning*, Oxford University Press, 565 ff.

LESLIE S.J., KHEMLANI S.S., GLUCKSBERG S. 2011. *Do all Ducks Lay Eggs? The Generic Overgeneralization Effect*, in «Journal of Memory and Language», 65, 15 ff.

LEWIS D. 1973a. *Causation*, in «Journal of Philosophy», 70, 556 ff.

LEWIS D. 1973b. *Counterfactuals*, Harvard University Press.

LIMATA T., IANÌ F., BUCCIARELLI M. 2022. *Story Order in Attribution of Moral Responsibility*, in «Language and Cognition», 14, 228 ff.

MACKIE J.L. 1980. *The Cement of the Universe: A Study in Causation* (2nd Ed.), Oxford University Press.

MICHOTTE A. 1963. *The Perception of Causality*, Methuen. (Originally published 1946)

MILL J.S. 1874. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation* (8th Ed.), Harper.

MILLER G.A. JOHNSON-LAIRD P.N. 1976. *Language and Perception*. Cambridge, MA: Belknap, Harvard University Press.

OAKSFORD M. CHATER N. 2007, *Bayesian Rationality*, Oxford University Press.

PEARL P. 2009. *Causality: Modals, Reasoning, and Inference* (2nd Ed.), Cambridge University Press.

PEARL J. 2013. *Structural Counterfactuals: A Brief Introduction*, in «Cognitive Science», 37, 977 ff.

PROSSER W.L. 1964. *Handbook of the Law of Torts* (3rd Ed.), St. Paul West.

QUELHAS A.C., RASGA C., JOHNSON-LAIRD P.N. 2019. *The Analytic Truth and Falsity of Disjunctions*, in «Cognitive Science», 43, e12739.

RAGNI M., KOLA I. JOHNSON-LAIRD P.N. 2018. *On Selecting Evidence to Test Hypotheses*, in «Psychological Bulletin», 144, 779 ff.

REICHENBACH H. 1956. *The Direction of Time*, University of California Press.

SLOMAN S., BARBEY A.K. HOTALING J.M. 2009. *A Causal Model Theory of the Meaning of Cause, Enable, and Prevent*, in «Cognitive Science», 33, 21 ff.

SPIRITES P., GLYMOUR C., SCHEINES R. 2000. *Causation, Prediction, and Search* (2nd Ed.), MIT Press.

STALNAKER R.C. 1968. *A Theory of Conditionals*, in RESCHER N. (ed.), *Studies in Logical Theory*, Basil Blackwell, 98 ff.

TURNBULL W., SLUGOSKI B.R. 1988. *Conversational and Linguistic Processes in Causal Attribution*, in Hilton D. (ed.), *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality,* Harvester Press, 66 ff.

WALDMANN M.R. 1996. *Knowledge-Based Causal Induction*, in «Psychology of Learning and Motivation», 34, 47 ff.

WHITE P.A. 1995. *Use of Prior Beliefs in the Assignment of Causal Roles: Causal Powers versus Regularity-Based Accounts*, in «Memory & Cognition», 23, 243 ff.

WOLFF P., BARBEY A.K. 2015. *Causal Reasoning with Forces*, in «Frontiers in Human Neuroscience», 9.

WOLFF P., BARBEY A.K., HAUSKNECHT M. 2010. *For Want of a Nail: How Absences Cause Events*, in «Journal of Experimental Psychology: General», 139, 191 ff.

Embodied Cognition and Legal Concepts

MICHELE UBERTONE, ANNA M. BORGHI, CATERINA VILLANI, LUISA LUGLI

1. *Introduction*

When lawyers determine what the law says about a certain case, they claim to do so by way of "classifying" the facts. This means that what the law says about the case is supposed to depend on certain *properties* that the facts display. Another way to put this is that events or objects in the case fall into certain legally relevant concepts in virtue of certain relevant properties they possess. Suppose Giovanni is caught shoplifting. Should Giovanni go to jail? Should he pay a fine? This depends on how we can classify this event. Was it a theft? A robbery? An attempted theft? The solution of the case will depend on the relevant legal concepts the case can be said to fall into. The act of classifying the facts in this way is not equivalent to a simple *description* of them, because it requires the speaker to make a double commitment about the properties of the case. The first commitment is that these properties exist, i.e., that they *de facto* characterise the case in question. The second commitment is that these properties are the ones that should *de iure* be selected as relevant for identifying that particular concept and thus deciding the case. Thus, to classify a fact as a referent of a concept is to take a stance not only on the nature of that fact, but also on the structure of the concept used to present it, and thus on the content of the law and, indirectly, on how the case should be decided.

In this paper, we will try to explain some peculiar aspects of how legal concepts are used to describe and criticise states of affairs, drawing on cognitive science, and in particular the perspective on cognitive science called *embodied and grounded cognition*. In the first section, we will give an initial summary definition of "concept" by briefly explaining the function that concepts, in general, play in our thinking. Since the word "concept" is sometimes used to mean different things, in the first section, we will clear up any ambiguities. At the same time, we will remain as neutral as possible about *theories* of concepts at this stage. In the second section, we will discuss the classical naïve way in which legal concepts are sometimes conceived by lawyers, namely as sets of conditions that are necessary and sufficient to identify whatever they are referring to (e.g., the concept of BACHELOR = criterion 1: unmarried; criterion 2: male; criterion 3: human). We will see that the main problem with this theory in the legal sphere is its inability to explain how subjects using a given legal concept can genuinely disagree about the possibility of applying it to a concrete case of which they have a common representation. In the third section, we will contrast this criterial, naïve view with a more plausible alternative. In the fourth section, we will try to explain how this more plausible view resonates with a strand of contemporary cognitive science: embodied and grounded

* This article is the result of a collaboration between a philosophy of law researcher, Michele Ubertone, and three cognitive psychologists, Anna M. Borghi, Caterina Villani and Luisa Lugli. Although the whole text is the result of a joint effort, the introduction and the sections with a mainly philosophical and jurisprudential content (2, 3 and 4) are mainly attributed to Michele Ubertone, while the following sections with a mainly psychological content are attributed to Anna Borghi, Caterina Villani and Luisa Lugli. The authors would also like to thank Corrado Roversi for his advice and feedback, without which the writing of this article would have been much more difficult.

cognition. We will explain in which sense some of the cognitive science literature talks of concepts in general as being "embodied" and "grounded" and ask what the consequences of these general features may be for the functioning of specifically *legal* concepts. In the fifth section, as an illustration of a possible methodology to meet this challenge, we will present an experiment carried out at the Laboratory of Legal Theory and Cognitive Science of the University of Bologna and explain its significance for legal theory. We believe that this line of research could show that aspects of open problems in legal theory probably cannot be solved with the traditional tools of conceptual analysis, but rest on empirical facts about the functioning of our minds.

This article often refers to the "internal point of view", a phrase used by HLA Hart to identify a particular way of using legal rules and the concepts they contain: using the law to justify and criticise behaviour, rather than merely to describe it. It is therefore appropriate to introduce a caveat at the outset. The purpose of this paper is not to reconstruct the historically most correct interpretation of what Hart meant by his theory of the internal point of view. The purpose of this paper is to explain what we can say about the internal point of view in the light of some lessons from embodied and grounded cognition.

2. *What we mean by "concept"*

Lawyers, philosophers and cognitive psychologists all use the word "concept" in their respective fields of interest, but it is not obvious that in using it they mean exactly the same thing. Different traditions use the word with slightly different meanings and this difference does not always depend on substantive disagreements on mind, language, or logic but merely on different linguistic conventions. For this reason, it is probably a good idea to start with a terminological clarification. In what follows, we will use the word "concept" in the sense provided by Elisabetta Lalumera:

> «Something is a concept by virtue of the function it performs within a cognitive system, and something is the concept of a certain category C (at least partially) by virtue of the further specific function of representing it. It is a further question whether or not the functional kind "concept" is realized by natural kinds[1]. Functional kinds can be individuated and described independently of their realizers. [C]oncepts can perform a double function—namely, abstraction and projection of knowledge—and that are able to be recombined almost freely in order to form more complex concepts and thoughts. [...] Abstraction is the "bottom-up" process of extracting information from a single encounter with an object or property-instance, and generalizing such information to all encounters with that object or property. The experience of tasting rhubarb once and finding it bitter would be of no use if I could not store it as information about rhubarb independently of the specific episode of tasting it, by means of a general representation—a concept. Category induction is the complementary "top-down" process of projecting such knowledge to new encounters. When you tell me you like rhubarb pie, I form the expectation that you will also like other bitter-tasting foods. This is an application of my concept of rhubarb. Thus, the function of a concept is that of a "mental glue," which connects one's past experience with the present. It is because they perform this complex function that concepts are used by default in our higher cognitive capacities, and not vice versa» (LALUMERA 2010, 217 f.)[2].

In adopting this definition, we wish to remain as ecumenical and non-committal as possible. We don't commit ourselves to any theory of what might fulfil the classificatory (*abstraction*)

[1] On natural kinds see fn. 9.

[2] The idea of concepts as "mental glue" originally comes from MURPHY 2004.

and informative (*projection*) functions that characterise concepts. We just want to use the word "concept" to identify whatever cognitive mechanism performs this function. It should be noted that this definition is very different from the one adopted by philosophers such as Gottlob FREGE (1952). For Frege, "concept" is not the name of a cognitive mechanism that performs a particular function, but rather of an abstract entity that exists independently of its comprehension by any human being, an abstract entity which human cognitive faculties merely "grasp" or represent, but don't constitute. Whether such abstract entities exist, and whether it makes sense to consider them as separate and independent from the cognitive mechanisms by which we represent categories of things, is a controversial metaphysical issue that we need not address here. In fact, what we are interested in here are cognitive mechanisms, the existence of which is not disputed, not abstract logical entities.

In order to explain the importance of this cognitive mechanism in our thought it may be useful to refer, as Lalumera does, to a famous short story by Jorge Luis Borges: *Funes el memorioso*. It is the story of an unfortunate individual, Ireneo Funes, who, after falling from a horse and hitting his head, acquires a prodigious memory. From that moment on, Funes remembers everything: every experience of his life, from birth to the present day; the precise facial features of everyone he has ever met; every mental event in his own stream of consciousness, including every past act of recalling memories. But this amazing memory is his curse. For Borges—and this is the most interesting aspect of the story—Funes is not a genius, but an idiot. By unlearning to forget, Funes unlearns to think. Funes' story is useful for illustrating a key principle of concept theory: thinking, in a sense, means forgetting differences.

«Borges suggests that perhaps Funes was incapable of thinking, of really using the mass of data in his possession. Each representation in his mind, we might say, was so detailed that constituted a separate type, it was hopelessly specific. To remember a day, Funes needed another whole day, and that was completely unnecessary» (LALUMERA 2009, 16, our translation).

Thought is realised by classifying into categories the things that exist in the environment with which we interact. Such classification consists in the act of looking at the world, forgetting, so to speak, irrelevant differences. For example, to have the concept APPLE, I need to be able to identify the characteristics common to a particular type of fruit. I need to fix these characteristics in my mind and divert my attention from the contingent details of each individual apple. If, in perceiving an apple, I were not able to select the characteristics it has in common with other apples, I would not be able to think of that apple as an instance of the concept APPLE. The cognitive economy that makes concepts possible and that makes thinking possible would not be taking place.

What criteria of relevance or irrelevance we use to form concepts is highly controversial, and there is much debate in the psychological literature about how the brain comes to determine them. What is certain is that the conceptual network we form through these classifications has the extraordinary property of relating our present interactions with our environment to our past experiences of it. This allows us to recognise in new and unseen objects the characteristics of objects with which we have interacted with in the past. For example, my interaction with a series of apples, pears and bananas since childhood has allowed me to form three distinct concepts: APPLE, PEAR and BANANA. Each of these concepts allows me to classify new and never-before-seen individual fruits and to infer from their superficial characteristics underlying characteristics that are useful to know when interacting with them, but not immediately accessible to those without such concepts. For example, I know that I like pears best, that I like apples only when they are green, and that I like bananas only when they are peeled. If I see a fruit and recognise it as a pear, I can predict that it will taste good; if I recognise it as an apple and see that it is red, I will not eat it; if I recognise it as a banana, I will make sure that I have peeled it before eating it. The fact that I have a concept of APPLE that is distinct from the

concept of PEAR and from the concept of BANANA is closely related to the fact that, given certain sensory inputs, I can infer that a given object is an apple rather than a banana or a pear, and from this qualification I can infer certain interesting information about its nature.

Concepts are often thought of as the meanings of words, but this is not necessarily the case. It is very important, for the understanding of what follows, to admit the possibility that the classificatory function we have designated as distinctive of the notion of concept can be performed by entities that do not coincide with public meanings represented in the same way by various speakers. It is not part of the definition of "concept" that we use here that it should be associated with a word. As we shall see, this is the case in a naive theory of concepts, but in more recent and sophisticated theories the association of concepts with word meanings has been mitigated. Whether or not a subject can create completely idiosyncratic concepts that do not correspond to the shared language of its community is a question that cannot be settled by a stipulative definition of "concept": it is a controversial issue. Ludwig Wittgenstein famously denied that we can follow rules (linguistic or otherwise) that are not shared: If I try to make a rule just for myself, there is no fact in the world that distinguishes the situation in which I think I am following that rule from the fact that I am actually following it. For this reason, according to Wittgenstein, I cannot create a private language and follow a set of rules for identifying references without the help of other subjects in the linguistic community to which we belong (KRIPKE 1982).

Others however have a different opinion to Wittgenstein and think that some concepts are idiosyncratic and others are shared[3]. In many cases, it may seem that the possession of a concept (and the consequent acquisition of the related capacities of abstraction and projection) is independent of the sociolinguistic information that is necessary to name it. That is, it may seem that the name of an object, after all, is not necessary information in order to have the category to which that object belongs as a concept. The fact that an object is called "apple" by other speakers might seem not necessarily relevant to my acquisition and use of *a* concept of APPLE. As Shakespeare's Juliet famously says: "A rose by any other name would smell as sweet" and what applies to roses in this case seems to apply *mutatis mutandis* to apples. If I had never learnt to speak, if, for example, I had been abandoned as a child in the jungle, I would not have given the concepts of APPLE, PEAR and BANANA the names that designate them in the English linguistic community: the names "apple", "pear", "banana". If, however, I had sufficient need to classify apples, pears and bananas in my everyday life, I would arguably have acquired some version of those concepts anyway[4].

In other cases, on the other hand, the possession of a concept seems to depend more crucially on the acquisition of a word[5]. If I believe that Bill has eaten an apple because Anna has told me

[3] Some abilities of categorization that are shown in linguistic behaviour can manifest themselves also in totally non-linguistic tasks. This supports the idea that it is possible to talk of non-lexicalized concepts. The abilities underlying both linguistic and non-linguistic categorization have common, distinctly recognizable properties such as typicality: «It is well-known that typical objects are categorized more quickly and more accurately than atypical objects. (...) [T]hese properties are common to lexicalized concepts and to non-lexicalized concepts. We decide more quickly that a robin is a bird than that a penguin is a bird. Similarly, when subjects learn to classify meaningless, abstract, and non-lexicalized figures into different categories and are then asked to classify new figures into these categories, typical figures are classified more quickly and more accurately than atypical figures» (MACHERY 2009, 59-60).

[4] After all, even animals lacking language are presumably able to classify and recognise different types of fruits and to act according to this classification. By virtue of this acquisition, even without having the ability to recognise APPLES as "apples", I would still have acquired the ability to distinguish them from BANANAS and, for example, the ability to predict that I like red apples statistically less than green ones. The sociolinguistic information that apples are called "apples", which my hypothetical wild alter ego lacks, and which the real me possesses, is simply an additional piece of information about the category and is not necessarily essential for the performance of all the cognitive activities that my concept of APPLE enables me to perform.

[5] This has been evidenced in particular with abstract concepts. See BORGHI 2023, BORGHI et al. 2019, DOVE 2022, CONNELL 2019, HENNINGSEN-SCHOMERS & PULVERMÜLLER 2022, STRIK LIEVERS et al. 2021.

that Bill has eaten an apple, the identity of the concept APPLE used in entertaining this thought will depend essentially on the fact that its extension comprises what the speakers of the English language, and in particular Anna, call "apple". The formation of the concept in this case entails a coordination problem. In yet other cases, not only is the structure of the concept that one subject entertains coordinated with the structure of the concept entertained by others, but the competence in the use of the concept by one depends crucially on the competence of another subject. In other words, subject A could not possess and use concept X without the help of subject B. Imagine you have a pain in one hand. Since this pain persists after a few days, you go to a doctor who, after a series of tests, diagnoses a form of arthritis. Let us assume that you have never heard of arthritis. In this case, your interaction with the doctor will enable you to acquire some concept of ARTHRITIS that will depend crucially on your relationship with the doctor[6]. You will only call "arthritis" that which the doctor, based on criteria only partly known to you, calls that way. The elaboration of concepts through which we interpret and understand the world is often a collaborative activity mediated by language. Almost everything we think is interwoven with beliefs based on the opinions of others, opinions that are conveyed to us through language. The division of cognitive labour goes hand in hand with a division of linguistic labour. Thus, the experts in a certain field, i.e. those in a community who are deputed to perform particular cognitive tasks, are typically also the repositories of particular categorisation criteria.

3. *The naïve theory of concepts and the problem with conceptual disagreements*

The prevailing lay view of the nature of concepts, understood as cognitive mechanisms for representing classes, is that they are something like definitions: a representation of the necessary and sufficient conditions that an object must possess in order to belong to a particular class (MARGOLIS & LAURENCE 2022). For example, according to the naïve view something counts as the referent of the concept BACHELOR if and only if it meets a list of necessary and sufficient conditions (criterion 1: unmarried; criterion 2: male; criterion 3: human). According to this view, linguistic definitions merely make explicit the reference-fixing criteria of the concepts that are conventionally established and shared in a linguistic community. We manage to communicate because we manage to coordinate and match the definitions we give for each term, and thus the structure of the associated concepts.

The idea that communication is a coordination problem and language is fundamentally a complex set of conventional criteria was famously advocated by the American philosopher David LEWIS (1969). The fact of knowing that others use the word "apple" to indicate the concept APPLE gives me an incentive to use the word "apple" in the same way. If others used the word "apple" to mean BANANA I would have a reason to use the word "apple" as meaning BANANA.

	B interprets "apple" as meaning apple	B interprets "apple" as meaning BANANA
A interprets "apple" as meaning APPLE	**Both A and B can communicate**	both obtain nothing
A interprets "apple" as meaning BANANA	both obtain nothing	**Both A and B can communicate**

[6] This example is inspired by a famous thought experiment by BURGE 1979.

It is the same kind of matrix that we can use to explain why different people have reasons to drive on the same side of the road.

	B drives left	B drives right
A drives left	**No car crash**	Car crash
A drives right	Car crash	**No car crash**

It is the kind of situation that game theory (a branch of mathematics concerned with strategy) calls a *pure coordination game*. Pure coordination games can all be solved by establishing conventions[7]. The fact that language is conventional also explains a guiding principle of linguistic interaction: what Paul Grice has named the cooperation principle (GRICE 1989; see also CHRISTIANSEN & CHATER 2022). Two or more people attempting to communicate with each other will try to cooperate to attribute same meanings to same instances of language use[8].

An objection that can be raised against Lewis's theory concerns cases of words standing for *essentially contested concepts*, words like "art" or "science" or "love" (GALLIE 1956). In such cases, it seems that people within the same linguistic community may reasonably disagree on what the true meaning of the words are. We believe these words mean something, but we do not think that the reference of the words can be understood merely by analysing the sociolinguistic conventions associated with them. The same goes with natural kind words[9]. We know that "Water" means H_2O not merely because we know conventions associated with the word, but because experiments were run that *proved* that H_2O was the right meaning[10].

The same problem emerges in the field of law[11]. If legal concepts were to be explained through the naïve theory, this would mean that genuine disagreements about the content of legal concepts are just impossible. Let's go back to Giovanni's shoplifting case. Imagine the staff had noticed his suspicious behaviour since when he entered the supermarket. Someone from the security followed him and waited for him to take a bottle of beer from the shelf and hide it under its coat. At this point the security stopped him and called the police. Was this "theft" or "attempted theft" or maybe neither of the two? If the concept THEFT were merely a set of rigid conventional criteria of application associated with the word "theft", a checklist to determine conclusively whether a given event falls into the category of theft or not, then the disagreement between two people in a case like the one we have imagined would never be a

[7] The profiles in which both players converge on a same course of action are Nash equilibria: the most rational response of each player to a given course of action is choosing the same course of action.

[8] The problem of whether legal concepts are conventionally defined must be kept distinct from the problem of whether legal systems are based on a conventional rule of recognition, a thesis often attributed to Hart. Against the thesis that the rule of recognition is a convention in the sense of Lewis, see CELANO, B. 2023.

[9] A natural kind is a division of reality that is supposed to follow the "natural structure" of reality itself. Supposed examples of natural kind concepts are FOX, ATOM, MALARIA, LIVER. The fact that we have words and concepts to identify foxes, for example, does not seem to be due to arbitrary conventions in the classification of the external world, but to the intention of speakers to accurately reflect in their language and thought the existence of a species whose distinction from other species is assumed to already exist in nature. The very existence of natural kinds, however, is controversial: BIRD,TOBIN 2022.

[10] On the idea that referents of natural kind concepts (like WATER) are identified independently of a mental representation of their properties by speakers see PUTNAM 1975.

[11] Bruno Celano has suggested that legal concepts are defined by engaging in a kind of game, the aim of which is not mere coordination, and which is not simply to establish a convention. Unlike the game described by Lewis, the game lawyers play in defining concepts is non-cooperative and cannot be solved by simply establishing a convention. See: CELANO 2017.

genuine conceptual disagreement. Disagreement about the application of the use of the word "theft" to the case could only be due to one of two circumstances: either the parties disagreeing associate the word to the same concept and then merely disagree because they have different beliefs about the facts of the case or they have a common representation of the fact of the case and then their disagreement must be due to the fact that the associate the word with different criteria of application and then different concepts. In both cases, there would be no identical concept about which the parties are disagreeing.

This resonates with one of Dworkin's critiques to Hart. If legal concepts, and specifically the concept of legal validity, were applied according to a conventional rule of recognition, as Hart says in his *Concept of Law*, then disagreements about legal validity between lawyers would never be genuine. A version of this critique was reformulated by Nicos Stavropoulos in the 1990s with specific reference to the semantic theory that Hart seems to presuppose (STAVROPOULOS 1996). According to Stavropoulos, Hart has an overly internalist, criterial (à la Lewis) conception of legal concepts, a theory that underestimates the interactive relationship that exists between our conceptual classifications of the external world and the actual structure of the external world itself. Stavropoulos thus reformulates Dworkin's theory on the basis of externalist semantics such as those of PUTNAM (1975) and, KRIPKE (1980), theories according to which the meanings of words we use, and the related concepts, are not "in the head" of speakers, but they are something we discover little by little as we interact with the world and with each other. According to these theories, we form the classification criteria associated with concepts on the basis of a pre-existing intuitive and indexical relationship with their referent, not vice versa.

4. *A less naïve approach: Contemporary interpretations of internal point of view statements*

Although the criticism of Hart by Dworkin, Stavropoulos and others assumes that he adopts a naive theory of concepts, his distinction between the "internal" and "external" points of view on law allows a more sophisticated interpretation, one that can explain the phenomenon of disagreement about the content of legal concepts. According to Hart, legal concepts are used differently depending on whether they are intended simply to describe a state of affairs or to criticise (or justify) it. In his language, when one uses law (and thus the concepts it employs) to *describe* a state of affairs, one is looking at law from an external point of view. When, on the other hand, one uses law to *criticise* states of affairs, one looks at law with the internal point of view. While ordinary citizens may or may not be able to adopt the internal point of view, it is necessary for legal officials to be able to do so. It is part of the bread and butter of lawyers, judges and civil servants to use legal concepts in this way.

For example, take the simple statement "This is a contract" and imagine saying it in two different scenarios. First scenario: You walk into the office of a friend of yours and point to a piece of paper on his desk and ask him, "What is this?". "This is a contract," he explains. In this case, the statement has the sole purpose of describing the nature of the indicated object, it does not express any criticism, claim, justification: it simply states a fact. Second scenario: Imagine that the same friend of yours takes the same piece of paper and waves it in the face of his boss during an angry argument in which he claims that the boss is not respecting the contractual salary conditions. In this case, the same claim would have a different meaning. The legal concept CONTRACT would be used not just to give *reasons for belief* by describing a state of affairs, but to give *reasons for action* by expressing a legal claim.

The use of legal concepts with the internal point of view is illustrated by the second scenario. The speaker invokes the legal concepts to justify or criticise a state of affairs not merely to describe it. The speaker, yes, on a literal level merely communicates to his interlocutor that

what he holds in his hands is a contract, but in doing so, on a pragmatic level, he intends to invoke what is written in the contract in order to criticise behaviour.

Hart does not explore much the issue of the non-merely descriptive character of statements made with the internal point of view in *The Concept of Law*, but he does explore this point in an earlier essay, *Definition and Theory in Jurisprudence*.

> «If we take a very simple legal statement like 'Smith has made a contract with Y', we must distinguish the meaning of this conclusion of law from two things: from (1) a statement of the facts required for its truth, e.g. that the parties have signed a written agreement, and also from (2) the statement of the legal consequences of it being true, e.g. that Y is bound to do certain things under the agreement. There is here at first sight something puzzling; it seems as if there is something intermediate between the facts, which make the conclusion of law true, and the legal consequences. But if we refer to the simple case of a game we can see what this is. When 'He is out' is said of a batsman (whether by a player, or by the umpire) this neither makes the factual statement that the ball has struck the wicket nor states that he is bound to leave the wicket; it is an utterance the function of which is to draw a conclusion from a specific rule under which, in circumstances such as these, consequences of this sort arise, and we should obviously neglect something vital in its meaning if, in the attempt to give a paraphrase, we said it meant the facts alone or the consequences alone or even the combination of these two. The combined statement 'The ball has struck the wicket and he must leave the wicket' fails to give the whole meaning of 'He is out' because it does not reproduce the distinctive manner in which the original statement is used to draw a conclusion from a specific but unstated rule under which such a consequence follows on such conditions. And no paraphrase can both elucidate the original and reproduce this feature» (HART 1982, 40).

In this passage, the example taken from the game of baseball is intended to illustrate the fact that in the internal point of view use of a legal concept there is a form of engagement that transcends the mere description of a state of affairs and even the description of a legal situation. The speaker in saying that the ball is out expresses a reflective critical judgement that is not reducible to its factual and normative premises.

In more recent literature, Hart's theory has been characterised as "quasi-expressivist" (FINLAY & PLUNKETT 2018)[12]. According to Finlay and Plunkett, statements made with the internal point of view have a descriptive semantic content, but also and most importantly an expressive pragmatic function. They literally describe, but they do so with the function of prescribing, and thus ultimately of expressing the motivation to act in a certain way or to accept or criticise a certain state of affairs. A statement of the form «"it is the law that L (in X)" semantically expresses the proposition that L is a rule [...] satisfying the criteria of the rule of recognition R of legal system X» (FINLAY & PLUNKETT 2018, 54 f.). At the same time, implicitly relying on the salience of that rule of recognition in that context, it expresses the speaker's motivation to act or have other act in accordance with L.

As already stated in the introduction our aim here is not to reconstruct what Hart really thought about the nature of "internal point of view", but rather to reconstruct This is a charitable interpretation of Hart because it allows his theory to better explain disagreements about the content of legal concepts. According to the naive theory, mastering a concept entails mastering a set of criteria of application. So, if two people both master a concept, a priori they will be able to apply it to specific cases in the same way. In other words, two people who disagree about the application of a concept to a particular case are either talking past each other (because they do not

[12]. The quasi-expressivist interpretation was developed in response to a fully expressivist interpretation proposed in TOH 2005.

understand that they are actually using different criteria of application, and hence different concepts), or have a factual disagreement (they agree about what the criteria of application are, but disagree about whether the object they are describing *de facto* satisfies them or not). But the quasi-expressivist interpretation of Hart's theory allows for a third possibility. This is that two people understand exactly what both mean when they use a word in a particular way, but they disagree about whether the word *should* be used in that way, with that particular practical function. The conceptual disagreement between them is normative. Whether we look at the concepts THEFT or ATTEMPTED THEFT adopting the internal point of view, a problem such as that of whether Giovanni committed one or the other cannot be solved merely based on the linguistic conventions associated to the use of the corresponding words: "theft" or "attempted theft". The speech act of calling this behaviour one thing or another is not merely a description of the events to be used by people who share certain criteria of application of a certain word, but a critical evaluation of those events, and the plea to modulate the extension of the concept in a way that is consistent with this critical evaluation. By calling the event a "theft" or an "attempted theft" a lawyer or a judge *takes a stance* on how certain types of events *should* be classified given the practical normative function that a certain concept or set concepts is understood to have. Classifying the fact as "theft", one expresses his or her views on the appropriateness of a certain type of regulation with respect to a certain type of behaviour.

Another more recent and alternative interpretation of Hart's account of internal point of view statements is an inferentialist one (Ziyu 2023). The inferentialist approach rejects the distinction between semantics and pragmatics and reconstructs the meaning of legal statements solely in terms of commitments, entitlements, and incompatibility. According to inferentialism, the meaning of a speech act is given by the network of inferences that can legitimately be drawn from it. The difference between internal point of view and external point of view statements can then be understood as a difference in meaning because different types of commitments can be inferred from them.

What the quasi expressivist and the inferentialist reading share is the idea that legal statements bear an added layer of meaning which transcends the representation of the referents of the legal concepts employed. Statements employing legal concepts of the type "X falls under legal concept C" can be understood as having different meanings depending of what commitment the speaker is taken to express. The speech act will be ambiguous unless the context clarifies whether the speaker's commitment is either about X or about C. It may not be clear whether the speaker intends to simply say that a certain object has certain properties/a certain fact took place, on the one hand, or whether, on the other hand, their intent is to say that the fact or the object, given its factual properties, should be legally classified in a certain way, and then the relevant concept should be modulated accordingly. For example, in saying "there was a theft at the supermarket", I could simply be informing an unaware person about what happened at the supermarket on the presupposition that she has the same conception of THEFT that I do, or on the other hand I could be telling a person who already knows very well what happened that the fact should be considered a THEFT, rather than, for example, an ATTEMPTED THEFT. In the two cases the meaning is different. In the *de facto*, external point of view, version of the statement I make no claim and express no commitment as to how the concept THEFT should be defined. I merely presuppose that the conception I am presupposing is part of the conventional common ground my interlocutors need to know and accept to be able to infer what I am trying to say. In the *de iure*, internal point of view, version of the statement I make no claim and express no commitment as to what happened. I am presupposing that the fact that a certain event took place is part of the common ground I am sharing with my interlocutors, and based on that common ground I am committing myself to the effect that this event, if it happened, should fall under the concept THEFT (Ubertone 2022).

All of this has consequences on the type of disputes that can arise over the conceptual classification of events, objects or behaviours. It may be that two people who have exactly the same

opinion as to what happened in the supermarket may still be sensibly arguing on whether what happened was an instance of THEFT or not. These types of disputes are conceptual rather than factual and even if they are particularly common in the legal domain, they sometimes also occur in other contexts of everyday life. Consider this example. Anna and Bruce go hiking on the Bologna hills. When the time comes to have lunch, Anna says: "Look at the picnic table at the end of the path! That would do!" Bruce replies: "No, that is not a picnic table... it's just a big stone. It doesn't even have chairs around it". There is something odd about this kind of dispute. How can Bruce consistently recognize that X is the referent of the expression "that picnic table" used by Anna and at the same time deny that X is a picnic table? Bruce *knows* that in Anna's idiolect it is not necessary that an X has chairs around it for X to count as a picnic table. We even can imagine that they both perfectly understand each other's presupposed definitions: they both know that X is a [picnic table]$_{Anna}$ and it is not a [picnic table]$_{Bruce}$. Why then do they still disagree? What Bruce is really saying is not that Anna has wrongly applied *her own* concept of picnic table and is not even necessarily committed to the idea that he has wrongly applied the linguistically correct meaning of the expression "picnic table". What he is really getting at is that [picnic table]$_{Bruce}$ is a *better* concept to use than [picnic table]$_{Anna}$, in the context of the practice in which both Anna and Bruce are involved. This conceptual dispute is actually a dispute over a plan of interaction with the environment. If X is recognized as a picnic table a certain plan is more likely to be performed than if it is not recognized as such. We can easily imagine the emotions that may motivate both parties in this discussion. Maybe Anna thinks sitting on the grass is really part of what is valuable in a picnic experience, and maybe Bruce is afraid of ticks.

The real disagreement between Anna and Bruce is not concerned with empirical facts about X, but about how X-type objects *should* be used. In our example, the dispute is not about picnic tables, or about the world, but about how the signifier "picnic table" and certain objects (candidates as possible referents of the concept PICNIC TABLE) should be used. These are not factual disputes about what is the most adequate description of a particular state of affairs, but normative disputes about what are the best plans of action should a particular factual situation arise in the world. This is a feature of all disputes over the affordances of artefacts as well as all disputes about the legal classification of events. The disputants don't have different opinions about empirical facts, they just feel differently about what is important and have different dispositions as to what should be done. If we accept the inferentialist view, we can say that the real disagreement depends on the difference in commitment between the two disputants, and that while the features of the classified object are part of the common ground between the two speakers, what is really at stake is the criteria of classification that should be employed.

For the purposes of this article, it is irrelevant whether the non-conventionalist and non-representationalist (be it quasi-expressionist or inferentialist) reading of Hart's theory is justified philologically. What is of interest to us is that, as we will see, this type of reading provides a good basis for linking the philosophical idea of an internal point of view on legal concepts with actual experimentally observable mental phenomena.

5. *Embodied and grounded perspectives on legal concepts*

Research on human conceptualisation has long assumed that our ability to form categories and think in terms of them is the result of fixed, abstract, symbolic representations of necessary and sufficient conditions that are somehow encoded in our brains. Concepts have traditionally been thought of as such representations, and have been described as the product of a kind of translation: a translation from sensorimotor language into what we can define as an "a-modal" language, i.e. a language made up of symbols that (much like words written in a non-ideogrammatic alphabet) bear no trace of the sensory modality involved in practical interaction

with the represented object or in the process of perceptual acquisition of the representation itself. The traditional, naïve view conceived concepts as disembodied and detached from the mechanisms that regulate perception and action. According to this view, our concept CHAIR, for example, has nothing to do with the visual image of a chair, the experience of sitting down, or the physical and social environment in which we use the concept of a chair. However, recent research has provided increasing evidence that this traditional view is misguided. Concepts have been shown to be deeply embodied. They are patterns of neural activation that re-enact the experience originally had with their referents[13]. This is consistent with the idea that the possession of concepts may involve not only propositional knowledge, but also certain kinds of know-how and emotional dispositions that may vary according to the degree of expertise of the speakers.

We have seen that the naïve conception has, among others, the shortcoming of not explaining the phenomenon of conceptual disagreements. Important criticisms against Hart were grounded on the assumption that he accepted this view. However, Hart can be read as departing from the naïve view of concepts in at least two important ways that resonate with contemporary cognitive science. First of all, according to Hart, same legal terms are conceptualised differently by different subjects depending on the actual use they make of them. The use of the same word is associated with different practices and experiences, and thus different psychological states and practical dispositions in different people. There is a division of cognitive labour between more and less expert subjects within the same linguistic community. Some users who are legal officials (who take the internal point of view) play an active role in defining the criteria through which a referent is represented while others (who take the external point of view) accept more passively those criteria.

On the other hand, as we saw, Hart seems to depart from the naïve view is that, according to his theory, legal concepts cannot be reduced to mere "representations". Hart famously has a "practice theory of social rules". He believes that rules are defined by practical dispositions to behave in certain ways. Hence, one can legitimately ascribe to Hart the idea that the internal point of view use of concepts defined by legal rules expresses certain practical commitments or dispositions. The fact of subsuming a fact under a legal concept is not necessarily an instance of "representing" reality, but can be also a way of expressing a critical judgment on it, and a practical disposition to act upon that judgment. To master a legal concept is not just to be able to *describe* a given referent presupposing a certain set of properties that conventionally identify it, but also to be able to say, on the assumption that that referent displays certain properties, that a certain set of criteria should be conventionally considered relevant to identify it, and thus that the concept in which the referent is subsumed should be shaped in a certain way. This is a way to pass a practical, albeit not necessarily moral, judgment. This conception of concept-use also explains the way in which concepts are collaboratively shaped and changed. Experts in law take the internal point of view, and take part in a complex conversation that shapes and modulates shared legal concepts, while ordinary citizens often use these concepts to describe the social reality they live in, and thus take the external point of view. Experts in law take active part in a complex conversation that shapes and modulates shared legal concepts, while ordinary citizens often use these concepts merely to describe the social reality they live in.

Both the division of cognitive labour and the non-representational conception of concepts resonate with ways in which contemporary cognitive science looks at concepts. The word "concept" may denote two distinguishable, but related entities. On the one hand, concepts can be described as individual psychological states, or, to put it more crudely, patterns of neurons activated in each individual's brain, encoding plans of interaction to respond to social and physical environmental cues. On the other hand, they can be seen as social objects, tokens of

[13] See BORGHI 2005, BARSALOU 1999, BARSALOU et al. 2008, WELLSBY & PEXMAN 2014, GALLESE & LAKOFF 2005, KIEFER & PULVERMÜLLER 2012.

information, *ways* of shaping patterns of neurons and adapt them to the social and physical environment that can be transmitted to one another. Concepts in this second sense are *memes*: dynamic social tokens that can be transmitted, changed and modulated within a community[14]. These two types of entities coevolve and interact dynamically. Concepts, in the first sense, namely as individual patterns of activations of neurons, are ways in which each of us encodes the complexity of one's own experience to respond to future cues in our environment, heavily influenced by each subject's sensorimotor experiences. Concepts in the second sense are the result of our ability to exchange and compare concepts in the first sense.

If we accept this reconstruction, it is clear how embodied cognition could help us to understand how legal concepts are formed, and that it could prove or disprove the fact that experts and non-experts in law use them differently and contribute to their formation in different ways. Each person using a legal concept will understand it slightly differently, associate it with different experiences and emotions. And, most interestingly for us, depending on their level of legal expertise, experts and non-experts will have a different tendency to associate it with the practical commitments that Hart calls the "internal point of view".

6. *An experimental study on the internal point of view*

The conception of the internal point of view on concepts that we have elaborated in the previous sections, linking it both to practical know-how and to the emotional disposition to act in a certain way, provides valuable insights for empirical investigation. In particular, by testing the use of legal concepts by lawyers and non-lawyers, and in particular data about their perception of their reference as something concrete and artifact-like on the one hand and as emotionally important on the other we can start to gather evidence in favour or against the following two hypotheses.

> *The Division of Cognitive Labor.* Different people conceptualise institutions in a different way. Some of them take the internal point of view on them, some don't.
> *The Internal/External Point of View Divide.* Legal concepts are multifunctional cognitive tools: they are both means to represent the existing institutions as well as means to actively engage in the normative practice which constitutes them.

Researchers at the Legal Theory and Cognitive Science Laboratory of the University of Bologna designed an experiment to show how legal concepts differ from other types of concepts in that they are used differently by people with different legal expertise, and in particular that the difference in use reflects a higher disposition of experts to consider them relevant from an emotional and arguably normative point of view. The data gathered support both hypotheses.

For the experiment, 567 volunteers were recruited from among students and researchers at the University of Bologna and people working in the Bologna area. The participants were divided into two groups: a law group and a control group. The law group was made up of 289 law graduates or professionals, and the control group was made up of 278 graduates or

[14] The word "meme" is used to indicate tokens of social or cultural information that can be passed from a social animal to another one, which get copied, transmitted, modified, evolved in much the way in which genes are. «They are a kind of way of behaving (roughly) that can be copied, transmitted, remembered, taught, shunned, denounced, brandished, ridiculed, parodied, censored, hallowed. [...] [W]e might say that memes are ways: ways of doing something, or making something, but not instincts (which are a different kind of ways of doing something or making something). The difference is that memes are transmitted perceptually, not genetically. They are semantic information, design worth stealing or copying, except when they are misinformation, which, like counterfeit money, is something that is transmitted or saved under the mistaken presumption that it is valuable, useful. [...] Words are the best examples of memes. They are quite salient and well individualized as items in our manifest image» (DENNETT 2017, 206-207).

professionals in fields other than law, such as philosophy, art, communication sciences. The researchers asked them to fill in a Google form and rate 56 Italian words according to various criteria. This allowed to classify the words according to different "dimensions", such as how abstract or concrete the word was considered to be, how easy it was to contextualise its referent, how imaginable its referent was considered to be, and so on. More specifically, both groups were asked to evaluate four types of concepts (i.e. institutional[15], theoretical/scientific, food, artefact) on the following dimensions Abstractness-Concreteness (ABS-CNR); Imaginability (IMG); Contextual Availability (CA); Familiarity (FAM); Age of Acquisition (AoA); Modality of Acquisition (MoA); Social Valence (SOC); Social Metacognition (META); Arousal (ARO); Valence (VAL); Interoception (INT); Metacognition (META); Perceptual Modality Strength in the modalities of vision, hearing, touch, taste and smell (VIS, HEA, TOU, TAS, SME); Body-Object Interaction (BoI); Mouth Involvement (MOUTH) and Hand Involvement (HAND). The 56 words used in the experiment are the following:

- 14 words denoting institutional concepts: contract, state, president, marriage, parliament, process, property, norm, rights, duty, sanction, responsibility, validity, justice;
- 14 words denoting theoretical/scientific abstract concepts: mass, acceleration, subtraction, temperature, sum, energy, litre, metre, gravity, calculation, equation, molecule, electron, multiplication;
- 14 words representing food concepts: banana, carrot, grape, strawberry, mushroom, aubergine, pepper, tomato, pumpkin, basil, apple, orange, chestnut, potato;
- 14 words denoting artefact concepts: hammer, wheel, knife, pot, spoon, tower, umbrella, bed, screwdriver, painting; chair, sculpture, book, computer.

For each dimension, the researchers carried out a Generalised Estimated Equations model (GEE), a statistical analysis that makes it possible to check whether the rating scores obtained with institutional concepts differ significantly from those of other categories of concepts and between the law group and the control group. Specifically, category (institutional, theoretical/scientific, food and artefact) was taken into account as a within-subjects factor and group (legal and control) as a between-subjects factor. Data showed that legal experts rate institutional concepts as being more contextually situated, more familiar, acquired at an earlier age, more "touchable", and associated with a more positive valence. In other words, lawyers understand legal concepts better, they "see" them as concrete things in way in which non-lawyers don't. But what is most interesting is that this epistemic ability is associated with an emotional and arguably practical disposition. Legal experts pair their capacity to concretise more abstract institutional concepts with a special emotional adherence to them. This suggests the existence two related but analytically distinguishable aspects in what Hart calls the internal point of view. This can be taken as a feature of the practical internal point of view, as defined above, and seems to strengthen the idea that a feature of the internal point of view is that of seeing institutions in their social significance. Non-experts, too, perceive broad social notions like validity, justice, rights, and duty as emotionally arousing in a positive way, though their rating is lower than the one given by experts. However, non-experts perceive pure institutional concepts as negatively arousing—as if they amounted to a cold, technical, and often disabling machinery connected with authoritative dictates—, whereas legal experts show a high degree of positive emotional arousal in regard to all institutional notions[16].

[15] Although, in the abstract, the term "institutional concepts" is broader than the term "legal concepts" (there are arguably institutions that are not, strictly speaking, legal institutions), the study we refer to uses the term "institutional concepts" to characterise concepts such as "law", "validity", "judge" or "contract", i.e. concepts that are used in the legal sphere and that shape legal practice and the institutions in which it takes place. So for our purposes the phrases "institutional concepts" and "legal concepts" are interchangeable.

[16] More details about this experiment can be found in VILLANI et al. 2021 and ROVERSI et al. 2023.

7. *Conclusion*

In this article we have attempted to describe a peculiarity of legal concepts. Hart's theory already shows how norms, and thus—we may add—the concepts employed by norms, can be used both to describe states of affairs and to criticise or justify them. In Hart's language, these two ways of using norms are called internal and external point of view. According to Hart, not all people are equally inclined to adopt each of these two points of view. Legal practitioners, professional jurists, and in particular public officials *necessarily* adopt the internal point of view. That is, they use legal concepts to justify particular conclusions of practical reasoning. Ordinary citizens, on the other hand, are sometimes more inclined to use them to describe states of affairs by deferring to a community of more experienced speakers the cataloguing of particular cases within or outside particular concepts for the purpose of justifying practical conclusions. Developing Hart's theory, we can observe that literally identical statements made using the same legal term can be read in alternative ways. To take an example given earlier, the statement "this is a contract" can be understood either as the description of a particular object or event, or as the legal claim that a particular institutional object or event should count as a contract. The first use, the use made with the external point of view, consists in describing a fact (the document or behaviour performed by the contracting parties) assuming the structure of the legal concept (the concept associated with the word "contract") as unproblematic. The second use, the use made with the internal point of view, consists in taking a position on the structure of the legal concept (taking a position on what in general should count as "contract") assuming as unproblematic the accuracy of the fact to which it is to be applied (the nature of the object or fact to which the concept is to be applied).

The idea that the act of applying a concept to a concrete case can be performed in two alternative ways, reflected by two distinct categories of speakers, distances itself from a naïve theory of concepts, according to which concepts are criterial and amodal ways of identifying referents. This idea is closer to more contemporary theories of concepts, and embodied cognition in particular. Various studies carried out within the framework of this theory show how concepts associated with the same terms are modulated differently depending on the background of the individual subject, their experiences, their interaction with the environment. More specifically, the concrete use we make of each concept leaves a perceptible trace in the neuronal pattern to which that concept materially corresponds. The concept for each speaker is not the abstract definition of a term, identical to that of every other speaker in the linguistic community to which they belong, but an idiosyncratic entity that reflects the history of that subject's interaction with a set of concrete situations. Therefore, if, for example, one speaker is accustomed to use the word "contract" solely to describe states of affairs while another speaker is also accustomed to use it to legally justify or criticise states of affairs, this should be somehow reflected in the pattern of neurons that the word activates in the individual speaker. The speaker who uses the word with the internal point of view will presumably describe the concept as eliciting different emotional responses and distinctively practical dispositions. The study of legal concepts therefore lends itself to an interdisciplinary approach that opens up new research perspectives. The empirical results collected by the Laboratory of Law Theory and Cognitive Science at the University of Bologna suggest new ways of specifying the nature of the internal point of view that will be hopefully expanded in the next few years.

References

BARSALOU L.W. 1999. *Perceptual Symbol Systems*, in «Behavioral and Brain Sciences», 22, 4, 577 ff.

BARSALOU L.W, SANTOS A., SIMMONS W.K., WILSON C.D. 2008. *Language and Simulation in Conceptual Processing*, in DE VEGA M., GLENBERG A., GRAESSER A. (eds.), *Symbols, Embodiment, and Meaning*, Oxford University Press, 245 ff.

BIRD A., TOBIN E. 2022. *Natural Kinds, in* ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2023 Edition. Available on: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=natural-kinds.

BORGHI A.M. 2005. *Object, Concepts and Action*, in PECHER D., ZWAAN R (eds.). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, Cambridge University Press, 8 ff.

BORGHI A.M. 2023. *The Freedom of Words: Abstractness and the Power of Language*, Cambridge University Press.

BORGHI A.M., BARCA L., BINKOFSKI F., CASTELFRANCHI C., PEZZULO G., TUMMOLINI L. 2019. *Words as Social Tools: Language, Sociality and Inner Grounding in Abstract Concepts*, in «Physics of life reviews», 29, 120 ff.

BURGE T. 1979. *Individualism and the Mental*, in «Midwest Studies in Philosophy», 4, 73 ff.

CELANO B. 2017. *Due problemi aperti della teoria dell'interpretazione giuridica*, Mucchi.

CELANO B. 2023. *La teoria del diritto di H.L.A. Hart. Una introduzione critica*, Il Mulino.

CHRISTIANSEN M.H., CHATER N. 2022. *The Language Game: How Improvisation Created Language and Changed the World*, Random House.

CONNELL L. 2019. *What Have Labels Ever Done for Us? The Linguistic Shortcut in Conceptual Processing*, in «Language, Cognition and Neuroscience», 34, 10, 1308 ff.

DENNETT D. 2017. *From Bacteria to Bach and Back. The Evolution of Minds*, Norton & Company.

DOVE G. 2022. *Abstract Concepts and the Embodied Mind: Rethinking Grounded Cognition*, Oxford University Press.

FINLAY S., PLUNKETT D., 2018, *Quasi-Expressivism about Statements of Law*, in GARDNER J., LEITER B., GREEN L. (eds.), *Oxford Studies in Philosophy of Law: Volume 3*, Oxford University Press.

FREGE G. 1952. *On Sense and Reference*, in *Translations from the Philosophical Writings of Gottlob Frege* (edited and translated by M. Black, P. Geach), Blackwell, 56 ff. (Originally published in 1892, *Über Sinn und Bedeutung*, in «Zeitschrift fur Philosophie und philosophische Kritik», 100 ff.)

GALLESE V., LAKOFF G. 2005. *The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge*, in «Cognitive neuropsychology», 22, 3-4, 455 ff.

GALLIE W.B. 1956. *Essentially Contested Concepts*, in «Proceedings of the Aristotelian Society», 56, 167 ff.

GRICE H.P. 1989. *Studies in the Ways of Words*, Harvard University Press.

HART H.L.A. 1982. *Definition and Theory in Jurisprudence*, in ID., *Essays in Jurisprudence and Philosophy*, Oxford University Press, 21 ff.

HART H.L.A. 2012. *The Concept of Law*, 3rd Ed., Clarendon Press.

HENNINGSEN-SCHOMERS M.R., PULVERMÜLLER F. 2022. *Modelling Concrete and Abstract Concepts Using Brain-Constrained Deep Neural Networks*, in «Psychological Research», 86, 8, 2533 ff.

KIEFER M., PULVERMÜLLER F. 2012. *Conceptual Representations in Mind and Brain: Theoretical Developments, Current Evidence and Future Directions*, in «Cortex», 48, 7, 805 ff.

KRIPKE S. 1980. *Naming and Necessity*, Harvard University Press.

KRIPKE S. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Harvard University Press.

LALUMERA E. 2009. *Che cosa sono i concetti*, Il Mulino.

LALUMERA E. 2010. *Concepts are a Functional Kind*, in «Behavioral and Brain Sciences», 33, 217 ff.

LEWIS D. 1969. *Convention. A Philosophical Study*, Blackwell.

MACHERY E. 2009. *Doing without Concepts*, Oxford University Press.

MARGOLIS E., LAURENCE S. 2022. *Concepts, in* ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2023 Edition. Available on: https://plato.stanford.edu/entries/concepts/.

MURPHY G. 2004. *The Big Book of Concepts*, MIT Press.

PUTNAM H. 1975. *The Meaning of "Meaning"*, in ID., *Minnesota Studies in the Philosophy of Science*, University of Minnesota Press.

ROVERSI C., UBERTONE M., VILLANI C., D'ASCENZO S., LUGLI L. 2023. *Alice in Wonderland: Experimental Jurisprudence on the Internal Point of View*, in «Jurisprudence», 14, 143.

STAVROPOULOS N. 1996. *Objectivity in Law*, Oxford University Press.

STRIK LIEVERS F., BOLOGNESI M., WINTER, B. 2021. *The Linguistic Dimensions of Concrete and Abstract Concepts: Lexical Category, Morphological Structure, Countability, and Etymology*, in «Cognitive Linguistics», 32, 4, 641 ff.

TOH K. 2005. *Hart's Expressivism and His Benthamite Project*, in «Legal Theory», 11, 2, 75 ff.

UBERTONE M. 2022. *Come non confondere questioni di fatto e questioni di diritto*, in «Ragion Pratica», 1/2022, 201 ff.

VILLANI C., D'ASCENZO S. BORGHI A.M., ROVERSI C., BENASSI M., LUGLI L. 2022. *Is Justice Grounded? How Expertise Shapes Conceptual Representation of Institutional Concepts*, in «Psychological Research» 86, 2434 ff.

WELLSBY M., PEXMAN P. M. 2014. *Developing Embodied Cognition: Insights from Children's Concepts and Language Processing*, in «Frontiers in Psychology», 5, 506 ff.

ZIYU L. 2023. *Hart as an Inferentialist: The Methodological Pragmatist Insight in Hart's Inaugural Lecture*, in «Law and Philosophy», 42, 379 ff.

Metaphorical Simulation and Legal Reasoning: Discovering the Legal Unconscious

MAREK JAKUBIEC

1. *Introduction*

The way lawyers think is the human way of thinking. Legal reasoning, despite its specificity, at a basic level is simply human reasoning. Regardless of the fact that legal reasoning is largely carried out "in the world of abstractness", where abstract concepts are used and norms come into play, the processes that make up legal cognition are firmly grounded in the physical world. In turn, we function as embodied beings in the world, and it is the nature of our bodies that fundamentally affects how we think about the law, how we interpret the law, and make decisions in the context of the law.

The purpose of this text is to present key aspects of the study of embodied conceptual metaphors and their relevance to theories of legal reasoning. First of all, it should be noted that theories of legal reasoning developed as part of legal theory (see MACCORMICK 1994; STELMACH & BROŻEK 2006; SCHAUER 2009) will not be analyzed in details here, as the aim is to highlight unconscious aspects of reasoning that have received relatively little attention in previous research. What is relevant here are the elements of reasoning (or, more broadly, legal thinking or cognition) that are unconscious. These elements are also prevalent in reasoning outside the legal context. This does not mean, however, that legal reasoning is devoid of certain peculiarities. First of all, due to formalization, but also due to the distinctiveness of the reference point for reasoning, which is—on the one hand—a specific state of facts, and on the other—a normative element. Legal reasoning, thus, is constituted by cognitive processes, the final element of which is a certain decision—this can be both a decision in the strict sense (legal decision-making) and, for example, an interpretive decision. Thus, I will understand "decision" in in the broadest possible sense, embracing all types of decisions, without limiting it to administrative decision or judicial decision, taken as a model example of legal decision-making (about legal decisions see e.g. GUTHRIE et al. 2001; BYSTRANOWSKI et al. 2021).

Of particular interest in this article are the unconscious aspects of legal reasoning which can have a very significant impact on the shape of the final decision. To this end, in the first part I will briefly analyse a key aspect of cognitive processes, i.e. conceptual processing that is pointed out—in the light of many research programs of modern cognitive science—as a fundamental element of cognitive processes (MARGOLIS & LAURENCE 2019; BARSALOU 2008; BORGHI & BINKOFSKI 2014). Of course, concepts can be understood in various ways, which will not be analyzed in detail here (see THAGARD 1992; MACHERY 2009). A way of understanding concepts that, for a number of reasons, seems to be the most adequate (although, of course, not without problems in the context of legal concepts) will be briefly described. I will then outline the most important aspects of the theory of mental simulation, which is a crucial element of

* The research was conducted in the framework of the project "The architecture of legal mind", financed by the Polish National Science Center (Grant No 2017/27/B/HS5/01407).

contemporary research on the embodiment of cognition (Bergen 2012; Barsalou 2008; van Dam & Desai 2016; Roversi et al. 2017).

According to this theory, conceptual processing—and, more generally, cognitive processes—are based on unconscious simulation, i.e. the "re-enactment in the mind" of what we have experienced in the past. This means that when we think about objects, actions, etc., the same areas of the brain are reactivated that are active when we see something, hear something (or experience it with other senses), and perform actions such as moving our hands. Importantly, this theory is supported by many studies, including neuroimaging evidence (Barsalou 2008).

In the second part I will indicate the role metaphorical mappings and the related specific type of mental simulation (metaphorical simulation) play in the light of contemporary theories of abstract concepts and reasoning. Metaphorical simulation is a mental simulation that is indirectly—through the mapping process—based on the perceptual experience and bodily action. According to the currently discussed theories of embodied abstract concepts, metaphor—understood not as a linguistic tool but as a cognitive tool—provides a "bridge" between abstract thinking (and abstract concepts) and the bodily experiences we have in the physical world with physical objects (see Jamrozik et al. 2016). Because we do not experience abstract objects with our senses nor interact with them, they cannot be simulated. However, they can be simulated in an indirect way, through the use of metaphorical mappings. If we assume that abstract legal concepts are, at least in part, metaphorical (as research in the cognitive sciences suggests), then unconscious metaphorical simulation plays an important role in legal reasoning.

In the last part it will be shown why the insights of cognitive science should be taken into account in the context of research on legal cognition and—in particular—legal reasoning. Indirectly, I will also hint at some methodological remarks that can be applied to the entire research program (or trend) of cognitive legal studies, in particular to attempts to create theories of legal reasoning that take into account knowledge from cognitive sciences, i.e., theories that are naturalized.

2. *The core of legal unconscious: conceptual processing*

Although the term "unsconciousness" may be associated with psychoanalysis (Lakoff & Jonson 1999; Bargh & Morsella 2008; in legal context see e.g. Bindal & Vashist 2023), unconsciousness of mental simulation (and metaphorical simulation) is simply about remaining outside the realm of our awareness. As Lakoff and Johnson (1999) stated,

«most of our thought is unconscious, not in the Freudian sense of being repressed, but in the sense that it operates beneath the level of cognitive awareness, inaccessible to consciousness and operating too quickly to be focused on».

From the viewpoint of contemporary cognitive science this seems rather uncontroversial. Our cognitive processes are mainly unconscious, including crucial processes that shape our "conscious mind".

Thus, legal unconscious is a group of cognitive processes, relevant to legal cognition, of which we are unaware. This refers not only to negative phenomena noticed in the literature, e.g. implicit bias (see Irwin & Real 2010), but to all mental processes that co-create legal reasoning. Obviously such a formulation is vague, but also enumerating all relevant processes that constitute cognition is, in most cases, impossible.

First, it is worth noting that, according to empirical-grounded theories of cognition, our cognition is overwhelmingly unconscious in nature, and this is not a new idea (see Kihlstrom 1987; Reber 1991, Kahneman 2011; Lakoff & Johnson 1999; Bargh & Morsella 2008; Brożek 2019). In other words, just as we are unaware of most of the physiological processes in our bodies, we are also unaware of the vast majority of cognitive processes. Referring to the metaphor of an iceberg, we see only the top of the iceberg of our cognitive processes, and the

rest remains covered under a sheet of unconsciousness. This observation alone, as long as the legal theorist accepts the findings of cognitive science, is intriguing: even if it seems that lawyers interpret the law in a conscious, formalized way and rely on a (more or less) sophisticated methodology, including also procedural rules, most of the processes that make up legal cognition are unconscious processes. Knowing the importance of the unconscious in cognition allows to take a new perspective on many legal problems that are relevant to both legal theory and legal practice.

As numerous studies indicate, susceptibility to manipulation, reliance on heuristics (KAHNEMAN 2011; GIGERENZER & ENGEL 2006; ENGLICH et al. 2006), nudges (THALER & SUNSTEIN 2008) or the influence of irrelevant environmental stimuli (e.g., the temperature of a drink held in one's hand that influences the attitudes to others—WILLIAMS & BARGH 2008) often turn out to be factors of importance that are, of course, difficult to measure precisely—but impossible to ignore. Also, research on how we use concepts brings a lot of new knowledge in this context. Importantly, the processing of concepts, that is a basic cognitive process, takes place at the unconscious level (BARSALOU 2008; DOVE 2009; LAKOFF & JOHNSON 1999; MARGOLIS & LAURENCE 2019). Thus, as legal concepts constitute a subgroup of concepts, the processing of legal concepts also takes place outside of consciousness. But what are concepts, and legal concepts?

The study of legal concepts belongs to one of the most important strands of the philosophy of law, both at present and in the past centuries (see HAGE & VON DER PFORDTEN 2009).

By "legal concepts" I understand the concepts that, in the form of linguistic expressions (which are symbols), occur in legal language[1]. Legal language is, in a significant part, abstract in nature, as it contains many abstract words. Thus, the conceptual grid of law is also abstract in significant part. Therefore, in the analyses presented here, the relevant concepts are primarily those that constitute the abstract "axis" of the legal system, such as, for example, "intellectual property", "causation", "justice" or, finally, the concept of "law" itself. I thus adopt a definition similar to the one proposed by the Swedish legal theorist Ake Frandberg, according to which legal concepts are those that can usually be found in the catalogue of basic concepts presented to law students (FRANDBERG 2009).

Of course, this definition is imperfect, and it heavily oversimplifies the problem of legal concepts. However, it is not easy to find a good alternative. If we adopt, for example, a broad definition, according to which legal concepts are those concepts that are relevant in a legal context, we must indicate the criteria of relevance. This, in turn, means opening a broad discussion, touching ontological problems. Since legal concepts in a broad sense are relevant for this analysis, the above definition will be harnessed.

So far, the focus has been mainly on the role these concepts play in the legal system in the context of legal language. Legal concepts were presented as inferential links (SARTOR 2009), tools to facilitate the production of normative information (ROSS 1957), "elementary particles of legal discourse" (HOHFELD 1913; CULLISON 1967) or building blocks of "pyramid" that constitutes law (the adherents of *Begriffsjurisprudenz*, see HAFERKAMP 2011). Such an approach is noticeable, for instance, in the famous work *Tu-Tu* by Alf Ross, where the discussion about the meaning of legal concepts takes place in the context of the linguistic analysis of informative functions they perform. Moreover, it is a good example of an analysis in which the distinction between legal concept and legal term is blurred (this is also visible in studies in which Ross's theory is cited, see SARTOR 2009).

[1] This part of the paper is based on a fragment of my book *Metaforyczność prawa* (JAKUBIEC 2022b) in which I analysed the nature of legal concepts.

The meaning of the phrase "legal concept" is unclear not only because of discussions from the field of legal theory, echoing the question of the nature of law (see e.g., RAZ 1983, HART 1961) but first and foremost because the word "concept" occurs in it, to which various meanings are ascribed.

In principle, one can identify two basic ways of understanding concepts—or two perspectives: philosophical and psychological (MARGOLIS 1994). From the philosophical perspective, a concept is the meaning of an expression. The psychological approach, on the other hand, assumes that the relationship between natural language expressions and the world and its features requires the mediation of mental representations—and these are concepts (MARGOLIS 1994). A mental representation is a mental equivalent of an object from the external world. Concepts, therefore, act as a link between the mind and reality—when thinking about the world, we process concepts that represent it.

The idea according to which we possess mental representations, the processing of which is crucial in cognitive processes, can be illustrated as follows: if I think of an object from the outside world, for example, a tree, my mind processes the mental representation of trees. In order to learn about the world, we need something to act as a link between the world (which is outside our brain) and our mental processes—representations.

Within contemporary cognitive science concepts are usually treated as basic mental representations (see CAREY 2009; MARGOLIS 1994; MARGOLIS & LAURENCE 2015; JAKUBIEC 2022a; SUTTON 2004), i.e. basic mental equivalents of objects. This approach, which has so far gone almost unnoticed in legal philosophy, is a transfer of the state of the art from cognitive science to the theory of legal concepts. Cognitive processes are based on the processing of representations (see BARSALOU 2008; FODOR 1975), which allows us to assume that the processing of legal concepts is the basic mechanism of legal cognition.

As a digression, let me mention that there is an intense debate as to the nature of our representations. For example, it is debated whether they are quasi-linguistic and amodal, or rather modal and analogical (see e.g., DOVE 2009; BORGHI et al. 2017).

If legal concepts are to be viewed as mental representations, several problems can be easily pointed out. First, this is a different from looking at legal concepts as linguistic elements—or at least elements analyzed as linguistic elements if not equated with words. Second, while it is easy to understand what a mental representation of a tree or a cat might be, it is more difficult to explain how the mind represents abstract objects.

I understand abstractness (following many representatives of cognitive science) as the absence of reference to material, spatio-temporally existing objects (see e.g. CHATTERJEE 2010; BORGHI, & BINKOFSKI 2014; BORGHI et al. 2017; BORGHI & ZARCONE 2016; CHIAO et al. 2009; COWLING 2017; DESAI et al. 2018), such as law, truth, or justice. *A contrario*, concrete concepts are those concepts that refer to material objects, like cat, table or car. The focus on abstract legal concepts stems from the fact that they constitute the fundamental conceptual basis of law. Law, as an abstract artifact (see the detailed discussion of what this exactly means in BURAZIN et al. 2018), is, after all, a "paradise of abstraction". It is not surprising, then, that in previous theories of legal concepts, abstract concepts have been assigned fundamental importance, such as in Hohfeld's theory. Hohfeld, in building his—highly influential—theory of concepts, analyzed the concepts of "right" and "duty" (HOHFELD 1913).

Of course, as I will try to show in subsequent sections of the text, the theories of embodied abstract concepts, led by the theory of metaphors, come to the rescue here, but this does not change the fact that such an account of legal concepts may be counterintuitive.

Most importantly, there is the question of how to determine the relationship between language and concepts. If concepts are not linguistic creations, then why do we think of them in linguistic terms? Legal concepts can be treated both as mental representations and meanings of words. Language, which allows us to express concepts, gains meaning from concepts (JAKUBIEC 2022a). Abstract language enables us to construct abstract concepts—it is highly likely that

without language we would not be able to create abstract law. However, the fact that language is crucial does not mean that concepts are in any sense reducible to linguistic statements.

While the treatment of concepts as meanings is inherent in traditional thinking about concepts, the combination of the two seems something relatively new in the legal context. This is, of course, particularly relevant to the study of legal concepts, but it may also prove significant for debates within the philosophy of law that are *prima facie* not dependent on a view of concepts. This is the situation we face in the case of legal reasoning. Why?

3. *Mental simulation*

Mental simulation—within the framework of the research program of embodied cognition—is one of the key mechanisms of our cognition and thinking, important for decision-making processes. It can be referred to as the mechanism of embodied concept processing. According to simulation theory, it is crucial for our cognition to "re-experience"—unconsciously—what we have already experienced in some dimension (direct interaction or, for example, by gaining knowledge about something), and the processing of concepts is based on the mechanism of simulation, i.e. re-enactment of the relevant brain areas (BARSALOU 2008; O'SHEA & MORAN 2017). Concepts, therefore, being representations of the external world, constitute a base for unconscious "re-experiencing" of what they represent.

Below I will only outline this concept very briefly, referring to the way it is described by Barsalou, one of the cognitive psychologists who is the main author of simulation theory and its later developments. As he stated, simulation represents a «reactivation of perceptual, motor and introspective states» experienced by the subject (BARSALOU 2008). Thus, during the experience, the brain «integrates different aspects of states, creating a multimodal representation that it stores in memory» (BARSALOU 2008). In other words, as O'SHEA and MORAN (2017) summarize this process:

> «motor simulation can occur in the absence of external input, relying on the re-enactment of previously experienced events which are stored in multi-modal representational format (grounded theories)».

Therefore, to put it simply: when we think of a hand movement, a tree, or a cat, the corresponding areas of the brain (responsible for movement, sight, touch, etc.) are reactivated and we simulate (unconsciously) the events when the brain received information about objects from the external world (BARSALOU 2008). Conceptual processing is thus based on simulating interactions with objects or experiences that these concepts represent. It is not the case that concepts are certain symbols, constituting elements of the language of thought, which only arbitrarily link to the external world (for discussion see BORGHI et al. 2017). The link between concepts and real world is not arbitrary, precisely because of the mechanism of simulation (see JAKUBIEC 2022a, 2022b).

While such an approach may seem counterintuitive—mainly because we are not aware of simulations—it is supported by the results of many experiments, including the ones conducted with functional magnetic resonance imaging (see e.g., BARSALOU 2008; DOVE 2014). Moreover, even given the controversy over the nature of embodied cognition (see MACHERY 2007; MAHON 2015), the mechanism of simulation seems well established in empirical research and its role is difficult to question. As an aside, it is worth noting that here the analysis is limited to unconscious simulation—related to the processing of concepts. From the perspective of legal reasoning, another type of simulation is also relevant—a conscious one, which can be equated with imagination (see BROŻEK 2019).

Of course, this framing of simulation—as re-experiencing something—raises many questions, especially in a legal context. As mentioned earlier, law is largely abstract. Legal language is steeped in abstract words, and the legal conceptual grid is steeped in abstract concepts. These concepts refer to something that is not graspable in the same way as a tree or a cat. None of us, after all, experience—in embodied sense—neither law, obligation nor justice. How, then, can we simulate anything that is represented by such concepts?

4. *Between concreteness and abstractness: metaphorical simulation and law*

Before indicating how this question can be answered, it is worth taking a closer look at the relationship between concrete and abstract concepts—that is, as we assume that concepts are mental representations—between concrete and abstract representations. As mentioned, abstractness may be understood as the absence of reference to material, spatio-temporally existing objects, as in the case of obligation, intellectual property, or justice (see e.g. BORGHI & BINKOFSKI 2014; DESAI et al. 2018). *A contrario*, concrete concepts are concepts that refer to material objects, like a cat, a table or a car. Of course, such a dichotomous presentation has important didactic value, but it is a simplification.

In the context of abstract-concrete distinction, legal concepts constitute a heterogeneous set, but in explaining them one should assume the existence of a certain *continuum*. Let us present a brief typology of legal concepts:

(a) concrete legal concepts (of course not all legal concepts are abstract!)—processed at a basic (cognitive) level through sensorimotor simulations (in line with the results of neuroscientific research; see BARSALOU 2008);
(b) abstract legal concepts with a higher level of detail (e.g. "marriage of Sara and Peter", "contract between Smith and Kowalski")—processed at a basic level through sensorimotor simulations (these may be also simulations of interactions between individuals; on situational simulation see BARSALOU & WIEMER-HASTINGS 2005) and processing of abstract legal language;
(c) abstract legal concepts *sensu stricto* (e.g. "justice", "intellectual property", "rule")—processed at a basic level through metaphorically mediated sensorimotor simulations.

When (b) and (c) are analysed, metaphors—or, more precisely, metaphorical mapping and metaphorical simulation (at the level of mechanism) and conceptual metaphor theory (at the level of explanation)—enter the game.

It should be emphasized, however, that among the theories explaining the mediation mentioned above in (c) not only the theory of conceptual metaphors (LAKOFF & JOHNSON 1999; GIBBS et al. 2004; PECHER et al. 2011; in the context of legal applications e.g.: WOJTCZAK 2017; JAKUBIEC 2022b) is relevant. Other theories that are analysed by legal theorists are, among others: the theory of conceptual blending (FAUCONNIER & TURNER 2002; in the context of legal applications e.g. ROVERSI 2015; ROVERSI et al. 2017), the so-called hybrid theories, e.g. Dove's theory on the role of language in concept processing (DOVE 2009; DOVE 2020) or WAT theory (BORGHI & BINKOFSKI 2014).

The theory of conceptual metaphors, which has its origins in the work of cognitive linguistics, has become a reference point in many discussions especially since the 1980s, when George Lakoff and Mark Johnson began to publish their work (starting with LAKOFF & JOHNSON 1980, that quickly became one of the most cited books devoted to metaphors). Although the proposals they presented were not a complete novelty—one can notice in metaphor theory some elements that literary critic Ivor Armstrong Richards or philosopher Max Black had put forward decades earlier

(Richards 1936; Black 1955), and the seeds of which can be found already in Giambattista Vico (Brożek 2019)—there is no denying that Lakoff and Johnson caused the understanding of metaphor as a cognitive tool to enter the scientific debate. Nowadays, the theory of metaphors arouses a lot of controversy, especially when it comes to its explanatory power and grounding in neuroscientific research, but it is nevertheless pointed out as one of the most important theories of embodied abstract concepts (Jamrozik et al. 2016; Borghi et al. 2017; Roversi et al. 2017). So even if it is not an approach that can be considered as corroborated to the similar extent as simulation theory, it is still an interesting reference point for further research on abstract concepts. This includes abstract legal concepts.

Painting with a broad bush, according to this theory, when we think about the abstract object, we use patterns of thinking about the concrete objects. Somewhat more precisely, metaphoricity means that we understand many aspects of reality by mapping the source domain to the target domain. What do these terms mean? Mapping is an unconscious process in which certain elements of the inferential structure characteristic of a concrete domain (the inference structure of a given conceptual domain) are transferred to the inferential structure associated with an abstract concept (Lakoff & Nunez 2000). The way we think about the concrete is thus reflected in our ways of thinking about the abstract. In other words, the mental simulations involved in processing concrete concepts, described above, are relevant to processing abstract concepts: even though we cannot directly simulate abstract concepts because we have no bodily experience of the objects they represent, we indirectly use these embodied simulations in processing abstract concepts (see e.g. Jamrozik et al. 2016; Borghi et al. 2017; Bergen 2012; Castano & Caroll 2020). Therefore, it is reasonable to conclude—at least on the basis of the theory of metaphorical embodied abstract concept—that our processing of abstract legal concepts is based on the mechanism of metaphorical simulation. Hence, abstract legal concepts can be representations of legal objects—objects that do not exist in the physical sense, constituting only abstract artifacts (see Burazin et al. 2018), and therefore it is difficult to speak of representation in the strict sense that presupposes a relationship between the representation and the represented object.

Such a brief presentation of metaphor theory may seem vague, and therefore it is worth sketching three examples of legal conceptual metaphors. This will also help draw attention to the heterogeneous nature of metaphorical simulation in law.

(1) Law as a material object

We think of law in terms of something material. Of course, law is not something material, but in thinking about it we use the patterns of thinking about concrete objects. That's why we think—and talk—about breaking the law or circumventing it. As Winter states, explaining the significance of cognitive explanation in law,

> «It is the essence of our concept of law that it operates as an external constraint, much like the impenetrable vegetation of the forest. Yet this very conception already places law in the domain of metaphor and imagination, which is to say in the internal realm of the human mind. We cannot even talk about law without metaphorically treating it as an OBJECT: Courts "make" law; criminals "break" the law; vigilantes "take the law into their own hands". The contradiction is devastating. Because the desired constraint turns out to be internal to the mind, the conventional view is defenseless against the various subjectivist critiques that have been leveled against it» (Winter 1995).

(2) Something more important is higher: embodiment of legal hierarchy

This is one example of orientation metaphors, grounded in our most embodied experiences—which take place because of the nature of the human body. In this case, the metaphorization is evident on the linguistic level—we are talking about a higher authority, a supreme court or a

higher-ranking act that overrules a lower one (see empirical results concerning the role of orientational metaphors—SANTANA & DE VEGA 2011; SCHUBERT 2005). The importance of embodied hierarchy in law is evident on many levels. Legal acts are arranged hierarchically—and contain norms of a higher and lower order. The structure of the judiciary is also hierarchical—with many countries referring to the most important court as the "highest court". The same is true of other organs of the state, constituting public administration.

(3) Intellectual property is property (of material objects)
Intellectual property refers to intangible objects—in particular, works. For this reason, it can be seen as a metaphor for the concept of property, which represents a bundle of the owner's rights over tangible things (JOHNSON 2007; LARSSON 2017). This is an interesting example in that the source concept is visible on a linguistic level - unlike the concept of "law", and somewhat similar to the "more is up" metaphor, which is reflected in various aspects of the hierarchical nature of the legal system.

If one seeks to clarify how abstract legal concepts can be embodied, as well as how their representational nature can be explained—which seems essential given the findings of cognitive science—considering the meaning of metaphorical simulation is a necessary step. It is also important from the perspective of understanding the cognitive mechanisms underlying all legal reasoning—or, more broadly, legal cognition.

The importance of metaphor theory, however, goes beyond what I previously highlighted as a possible contribution to the understanding of the legal mind. In the last part, I would like to highlight some practical aspects of the study of the metaphorical nature of law, which show that the importance of this kind of research is not limited to theoretical quandaries and has important implications for the reasoning carried out in the legal context.

First, the analysis of the expressions present in legal language enables us to discover the conceptual metaphors and reconstruct the metaphorical simulations linked to processing of abstract legal concepts. This, in turn, allows us to better understand the aim of certain norms. For instance, in the context of intellectual property law, such analysis allows to discover the primacy that the legislator gives to the protection of creators' rights (LARSSON 2011, 2017; ZALEWSKA 2016). This can be important from the perspective of the interpretation of the law, especially the purposive interpretation, in which we refer to the purpose of legal regulation. Why does it matter? In case of doubts about the meaning of a certain norm, the court may refer to the general concept granting primacy to the rights of the creator over the rights of the beneficiaries. While it is, of course, possible to reach such a conclusion also without taking metaphors into account, their analysis in some cases may prove crucial, for example, when analyzing the court's line of argumentation in drafting an appeal. In a similar vein, Linda Berger, a U.S. legal scholar, notes that:

> «To argue against a dominant metaphor, lawyers must be able to uncover it; to argue for a new metaphor, lawyers must be able to imagine it. Studying the work of cognitive researchers builds such perception and imagination: the more we know about the work of the mind, the use of language, and the means of persuasion, the more critical, insightful, and persuasive we can be.» (BERGER 2004)

Second, as research on the impact of a course of metaphorical mapping on reasoning indicates (such as the experiment on understanding crime—see THIBODEAU & BORODITSKY 2011), specific source concepts shape the processing and understanding of legal concepts in different ways, affecting the assessment of the legal consequences of the events that occurred. This means that seemingly insignificant aspects of how a case is presented are therefore relevant to the decisions made by judges. It also means that by shaping the media message and the implicit

references to concrete objects present in conceptual metaphors, it is relatively easy to manipulate the public's position on the law, which can have a significant impact on legislative action. Also, those who apply the law and perform legal reasoning that results in decisions are not immune to this kind of influence.

In reference to the above first practical "application", it is also worth adding that by analyzing metaphors—or more precisely, by analyzing the language in which a metaphorical simulation manifests itself, it is possible to indicate which metaphor is more common—that is, which metaphorical simulations are more frequently used by people. It can be an argument in favor of a certain interpretation of regulations during doctrinal disputes. Whether such an argument will be justified is a separate issue. After all, the frequent occurrence of a metaphor does not necessarily mean that reasoning based on it is more appropriate than others.

5. *Concluding remarks*

The purpose of this brief paper was to highlight the importance of mental simulation in legal reasoning. Simulation—according to the current state of research within the cognitive sciences—is a key cognitive mechanism. This means that it is necessary to take into account the achievements of cognitive science in the analysis of legal reasoning. As I tried to show, the application of conceptual metaphor theory makes it possible to explain how our mind processes legal concepts, and this, in turn, is crucial for the reasoning carried out in the legal context. Of course, metaphor theory is not the only theory that can be applied in explaining the legal mind, nor is it a theory devoid of controversy. However, this does not change the fact that, being one of the most important theories of embodied abstract concepts (see BORGHI et al. 2017), it can serve as a source of knowledge for naturalization of theories of legal reasoning. There is also no doubt that—even despite questions about its exploratory effectiveness—it makes it possible to draw attention to aspects of legal reasoning that have so far not been the object of research within legal theory. For these reasons, further research on the significance of metaphorical simulation can be a source of knowledge relevant to legal theory.

References

Bargh J.A., Morsella E. 2008. *The Unconscious Mind*, in «Perspectives on Psychological Science» 3, 1, 73 ff.

Barsalou L. 2008. *Grounded Cognition*, in «The Annual Review of Psychology», 59, 617 ff.

Barsalou L., Wiemer-Hastings K. 2005. *Situating Abstract Concepts*, in Pecher D., Zwaan R.A. (eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, Cambridge University Press, 123 ff.

Bergen B. 2012. *Louder than Words: The New Science of How the Mind Makes Meaning*, Basic Books.

Berger L. 2004. *What is the Sound of a Corporation Speaking? How the Cognitive Theory of Metaphor Can Help Lawyers Shape the Law,* in «Scholarly Works», Paper 668, 169 ff. http://scholars.law.unlv.edu/facpub/668.

Bindal A., Vashist L. 2023. *The 'Legal Unconscious': Exploring the Intersection of Law and Psychoanalysis,* in «Asian Journal of Legal Education», 10, 2, 177 ff. Available at: https://doi.org/10.1177/23220058221139.

Black M. 1955. *Metaphor*, in «Proceedings of the Aristotelian Society», New Series, 55, 273 ff.

Borghi A., Binkofski F., Castelfranchi C., Cimatti F., Scorolli C., Tummolini L. 2017. *The Challenge of Abstract Concepts*, in «Psychological Bulletin», 143, 263 ff.

Borghi A., Binkofski F. 2014. *Words as Social Tools: An Embodied View on Abstract Concepts*, Springer.

Borghi A., Zarcone B. 2016. *Grounding Abstractness: Abstract Concepts and the Activation of the Mouth,* in «Frontiers in Psychology», 7, 1 ff.

Brożek B. 2019. *The Legal Mind. A New Introduction to Legal Epistemology,* Cambridge University Press.

Burazin L., Himma K.E., Roversi C. (eds.) 2018. *Law as an Artifact*, Oxford University Press.

Bystranowski P., Janik B., Próchnicki M., Skórska P. 2021. *Anchoring Effect in Legal Decision-Making: A Meta-Analysis*, in «Law and Human Behavior», 45, 1, 1 ff.

Carey S. 2009. *The Origin of Concepts*, Oxford University Press.

Castano E. Caroll G. 2020. *Mental Simulation in the Processing of Literal and Metaphorical Motion Language: An Eye Movement Study*, in «Metaphor and Symbol», 35, 3, 153 ff.

Chatterjee A. 2010. *Disembodying Cognition*, in «Language and Cognition», 2, 1, 79 ff.

Chiao J., Harada T., Oby E.R., Li Z.,Parrish T., Bridge D.J. 2009 *Neural Representations of Social Status Hierarchy in Human Inferior Parietal Cortex*, in «Neuropsychologia», 47, 2, 354 ff.

Cowling S. 2017. *Abstract Entities*, Routledge.

Cullison A.D. 1967. *A Review of Hohfeld's Fundamental Legal Concepts*, in «Cleveland State Law Review», 16, 3, 559 ff.

Desai R. H., Reilly M., Van Dam W. 2018. *The Multifaceted Abstract Brain*, in «Philosophical Transactions of the Royal Society B», 373, 1 ff.

Dove G. 2009. *Beyond Perceptual Symbols: A Call for Representational Pluralism*, in «Cognition», 110, 412 ff.

Dove G. 2011. *On the Need of Embodied and Dis-Embodied Cognition*, in «Frontiers in Psychology», 1, 1 ff.

Dove G. 2014. *Thinking in Words: Language as an Embodied Medium of Thought,* in «Topics in Cognitive Science», 6, 3, 371 ff.

DOVE G. 2020. *More than a Scaffold: Language is a Neuroenhancement*, in «Cognitive Neuropsychology», 37, 288 ff.

ENGLICH B., MUSSWEILER T., STRACK F. 2005. *The Last Word in Court: A Hidden Disadvantage for the Defense*, in «Law and Human Behaviour», 29, 705 ff.

FAUCONNIER G., TURNER M. 2002. *The Way We Think*, New York.

FODOR J. 1975. *The Language of Thought*, Harvard University Press.

FRANDBERG A. 2009. *An Essay on Legal Concept Formation*, in HAGE J., VON DER PFORDTEN D. (eds.), *Concepts in Law*, Springer, 1 ff.

GIBBS R.W., LIMA P.L.C., FRANCOZO E. 2004. *Metaphor is Grounded in Embodied Experience*, in «Journal of Pragmatics», 36, 1189 ff.

GIGERENZER G., ENGEL C. (eds.) 2006. *Heuristics and Law*, The MIT Press.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A.J. 2001. *Inside the Judicial Mind*, in «Cornell Law Review», 86, 777 ff.

HAFERKAMP H-P. 2011. *Begriffsjurisprudenz* (*Jurisprudence of Concepts*), in *Enzyklopaedie zur Rechtsphilosophie* (online: http://www.enzyklopaedie-rechtsphilosophie.net/inhaltsverzeichnis/19-beitraege/105-jurisprudence-of-concepts).

HAGE J., VON DER PFORDTEN 2009. *Concepts in Law*, Springer.

HART H.L.A. 1961. *The Concept of Law*, Oxford University Press.

HOHFELD W. N. 1913. *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, in «The Yale Law Journal», 23, 1, 16 ff.

IRWIN J.F., REAL D.L. 2010. *Unconscious Influences on Judicial Decision-Making: The Illusion of Objectivity*, in «McGeorge Law Review», 43.

JAKUBIEC M. 2022a. *Legal Concepts as Mental Representations*, in «International Journal of Semiotics of Law» ,35, 1837 ff.

JAKUBIEC M. 2022b. *Metaforycznosc prawa*, Copernicus Center Press.

JAMROZIK A., MCQUIRE M., CARDILLO E.R., CHATTERJEE A. 2016. *Metaphor: Bridging Embodiment to Abstraction*, in «Psychonomic Bulletin & Review», 23, 1080 ff.

JOHNSON M. 2007. *Mind, Metaphor, Law*, in «Mercer Law Review», 58, 3, 845 ff.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Penguin.

KIHLSTROM J.F. 1987. *The Cognitive Unconscious*, in «Science», New Series, 237, 4821, 1445 ff.

LAKOFF G., JOHNSON M. 1980. *Metaphors We Live by*, University of Chicago Press.

LAKOFF G., JOHNSON M. 1999. *Philosophy in the Flesh*, Basic Books.

LAKOFF G., NUNEZ R. 2000. *Where Mathematics Comes From*, Basic Books.

LARSSON S. 2011. *Metaphors and Norms. Understanding Copyright Law in a Digital Society*, Lund University.

LARSSON S. 2017. *Conceptions in the Code. How Metaphors Explain Legal Challenges in Digital Times*, Oxford University Press.

MACCORMICK N. 1994. *Legal Reasoning and Legal Theory*, Clarendon Press.

MACHERY E. 2007. *Concept Empiricism: A Methodological Critique*, in «Cognition», 104, 1, 19 ff.

MACHERY E. 2009. *Doing without Concepts*, Oxford.

MAHON B.Z. 2015. *What is Embodied about Cognition?*, in «Language, Cognition and Neuroscience», 30, 4, 420 ff.

MARGOLIS E. 1994. *A Reassessment of the Shift from the Classical Theory of Concepts to Prototype Theory*, in «Cognition» 51: 73–89.

MARGOLIS E., LAURENCE S. 2015. *Conceptual Mind*, MIT Press.

MARGOLIS E., LAURENCE S. 2019. *Concepts*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Available at: https://plato.stanford.edu/entries/concepts/.

O'SHEA H., MORAN A. 2017. *Does Motor Simulation Theory Explain the Cognitive Mechanisms Underlying Motor Imagery? A Critical Review*, in «Frontiers in Human Neuroscience», 11, 1 ff.

PECHER D., BOOT I., VAN DANTZIG S. 2011. *Abstract Concepts: Sensory-Motor Grounding, Metaphors, and Beyond*, in «Psychology of Learning and Motivation», 54, 217 ff.

RAZ J. 1983. *The Problem about the Nature of Law*, in «University of Western Ontario Law Review», 21, 2, 203 ff.

REBER A. 1991. *The Cognitive Unconscious: An Evolutionary Perspective*, in «Consciousness and Cognition», 1, 2, 93 ff.

RICHARDS I.A. 1936. *The Philosophy of Rhetoric*, Oxford University Press.

ROSS A.1957. *Tû-Tû*, in «Harvard Law Review», 70, 5, 812 ff.

ROVERSI C. 2015. *Legal Metaphoric Artifacts*, in The *Emergence of Normative Orders*, STELMACH J., BROZEK B., KUREK L. (eds.), Copernicus Center Press, 215 ff.

ROVERSI C., PASQUI L., BORGHI A.M. 2017. *Institutional Mimesis: An Experimental Study on the Grounding of Legal Concepts*, in «Revus», 32, 73 ff.

SANTANA M., DE VEGA M. 2011. *Metaphors Are Embodied, and So Are Their Literal Counterparts*, in «Frontiers in Psychology», 2, 1 ff.

SARTOR G. 2009. *Understanding and Applying Legal Concepts: An Inquiry on Inferential Meaning*, in HAGE J., VON DER PFORDTEN D. (eds.), *Concepts in Law*, Springer, 35 ff.

SCHAUER F. 2009. *Thinking like a Lawyer*, Harvard University Press.

SCHUBERT T. 2005. *Your Highness: Vertical Positions as Perceptual Symbols of Power*, in «Journal of Personality and Social Psychology», 89, 1, 1 ff.

STELMACH J., BROŻEK B. 2006. *Methods of Legal Reasoning*, Springer.

SUTTON J. 2004. *Are Concepts Mental Representations or Abstracta?*, in «Philosophy and Phenomenological Research», 68, 1, 89 ff.

THAGARD P. 1992. *Conceptual Revolutions*, Princeton.

THALER R. H., SUNSTEIN C. R. 2008. *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press.

THIBODEAU P., BORODITSKY L. 2011. *Metaphors We Think With: The Role of Metaphor in Reasoning*, in «PLOS ONE», 6, 2, e16782.

VAN DAM W.O., DESAI R.H. 2016. *Embodied Simulations Are Modulated by Sentential Perspective*, in «Cognitive Science», 41, 6, 1613.

WILLIAMS L.E., BARGH J.A. 2008. *Experiencing Physical Warmth Promotes Interpersonal Warmth*, in «Science», 322, 5901, 606 f.

WINTER S. 1995. *A Clearing in the Forest*, in «Metaphor and Symbolic Activity», 10, 3, 223 ff.

WOJTCZAK S. 2017. *Metaphorical Engine of Legal Reasoning and Legal Interpretation*, Beck.

ZALEWSKA M. 2016. *Znaczenie metafor pojęciowych na przykładzie prawa autorskiego*, in «Filozofia Publiczna i Edukacja Demokratyczna», 5, 1, 111 ff.

A New Perspective on Law's Rationality. An Experimental Essay

BARTOSZ BROŻEK

1. *The ecological turn* – 2. *The hypothesis* – 3. *The law*

The title of this paper may be promising more than the paper delivers. The perspective on law's rationality offered here may ultimately not be that new. At the very least, similar ideas have been entertained by many legal and moral philosophers, old and contemporary (CF. HUTCHESON 1929; HAIDT 2001; SETMAN 2022). I believe however that I go a step or two further than they have gone, which makes it possible to fully acknowledge and consider the novelty involved.

The subtitle of the paper positions it as an essay. By choosing this form, I want to stress that this is not a typical academic work, both in style and contents, but also to suggest that I do not present a final product, a well-developed theory. I only try to understand law's rationality from a fresh perspective, and even if the attempt results in a failure, my hope is that it will constitute an invitation to rethink opinions and concepts which are taken for granted and may seem unshakable.

I have also dubbed this essay "experimental", as what I present is a kind of thought experiment. I introduce a bold hypothesis and try to consider how our understanding of the law and its rational dimension would look like if the hypothesis were true.

The hypothesis in question says that rationality is *not* in our minds, but in the environment we inhabit. In order to understand the origins of the hypothesis as well as grasp its meaning and significance, it is reasonable to begin with the conception of ecological rationality.

1. *The ecological turn*

«There are three kinds of lies: lies, damn lies, and statistics». This saying was popularised by Mark Twain, who mentioned it in his article *Chapters from My Autobiography* in 1907. However, Twain was not the original author of this phrase. The authorship has been attributed to various well-known figures, such as the British Prime Minister Benjamin Disraeli, Walter Bagehot, Henry Du Pré Labouchère, William Abraham Hewitt, Lord Courtney, or Charles Dilke. Regardless of who came up with this phrase, it is undeniable that it must contain something that resonates with our understanding of the world; otherwise, it would be difficult to explain why this saying is so often and fondly recalled. Yes, it is amusing, but that does not explain the disputes over its authorship, the numerous references in popular culture, the scientific articles dedicated to it, or the Wikipedia entries (cf. WIKIPEDIA CONTRIBUTORS 2023).

This popularity may seem surprising, especially in the times we live in. After all, the 21st century has often been proclaimed the "century of information"; big data and data mining are not only catchy phrases but also thriving businesses, and statistical analyses not only help large corporations better cater to our tastes but also enable more precise medical diagnoses or, as is the case with The Human Brain Project, bring us closer to understanding how the human brain functions. Of course, statistical analyses are challenging, both in execution and interpretation. Mistakes are easily made, and even among specialists, they are all too common. Some lead to

* The first two sections of this essay have been translated from an unpublished Polish text with a kind help of ChatGPT.

catastrophic consequences—for example, the statistical analysis errors contributed to the Challenger space shuttle disaster. But even such tragic mistakes do not explain our aversion to statistics: after all, it is not mathematics that is to blame but the person who misused it.

Perhaps the answer to our question is provided by a simple thought expressed in the previous paragraph: statistics is challenging, and our brains are not adapted to using it. No one has demonstrated this more vividly than Daniel Kahneman and Amos Tversky. Participants in one of the experiments designed by them were asked to rank future scenarios for a young woman named Linda, who is thirty-one years old, unmarried, talkative, and very intelligent. She majored in philosophy and was highly involved in fighting discrimination, promoting social justice, and participating in anti-nuclear energy demonstrations. The scenarios that needed to be evaluated included the following versions of her future: (a) Linda is a feminist activist, (b) Linda works in social care and helps people with mental disorders, (c) Linda is a member of the Women's Choice League, (d) Linda is a bank teller, (e) Linda is an insurance agent, (f) Linda is a bank teller and a feminist activist. Kahneman and Tversky were shocked when it turned out that all participants in the experiment considered the probability of Linda being a bank teller who is also a feminist activist to be greater than the probability of her being just a bank teller (cf. KAHNEMAN 2011, 211). This is an obvious error because the probability of the joint occurrence of two independent events is lower than the probability of the occurrence of only one of them. We are not dealing here, of course, with complicated statistical analyses but with a completely elementary property of probability theory. Let's think about how much more difficult it must be for us to understand complex statistical constructs!

How can we explain all this? Kahneman and Tversky argue that the error we make in cases like Linda's story stems from relying on the unconsciously used representativeness heuristic. Linda's characteristics are representative of a feminist activist, not a bank teller, which is why the scenario of her future life in which she is a feminist activist exerts such a strong force on us that we ignore elementary principles of probability theory. There are, in fact, more of these heuristics—specific "thinking shortcuts"—in the arsenal of our minds.

In their famous work from 1974, *Judgment under Uncertainty: Heuristics and Biases*, Kahneman and Tversky also described the availability heuristic and the anchoring and adjustment heuristic (TVERSKY & KAHNEMAN 1974). The availability heuristic leads us to base our response to an encountered problem on information that is easiest to recall from memory. For example, when asked whether there are more words that start with the letter "r" or have the letter "r" in the third position, English language users typically—incorrectly—indicate the first possibility, probably because it is easier to remember words in which "r" is the first letter (see TVERSKY & KAHNEMAN 1974).

On the other hand, the anchoring and adjustment heuristic comes into play when an "anchor" appears in the "environment" surrounding the problem being considered. This anchor, which is unrelated to the problem, influences the adopted solution. In their classic studies, Kahneman and Tversky asked participants in the experiment to estimate the percentage of UN member countries that were African countries. However, before doing so, the participants had to determine whether this percentage was higher or lower than the number that appeared on a "wheel of fortune". This device was designed in such a way that either the number 10 or 65 would appear. It turned out that participants who "spun" the wheel and obtained the number 10, on average, believed that around 25% of the UN member countries were African countries. Meanwhile, the participants who "spun" the wheel and obtained the number 65 had a significantly higher average estimate—around 45% (see TVERSKY & KAHNEMAN 1974).

In the context of these illustrations, there is no doubt that basing judgments and decisions on unconscious heuristics can lead to embarrassing errors. Heuristics are the source of great foolishness that does not reflect our species in the best light. It seems urgent to develop mechanisms that weaken the destructive effects of these described mechanisms. Perhaps, instead of burdening

children with multiplication tables, we should start by raising their awareness of the dangers associated with excessive reliance on intuitive judgments. Maybe—alongside some Very Important Social Policy—it is worth investing in a broad educational campaign that explains to society the detrimental consequences of heuristic thinking. Why wouldn't the example with Linda replace Anny Apple, Firefighter Fred and Dippy Duck in elementary English textbooks, and why wouldn't words with the letter "r" in the third position appear on cigarette packages? It would be an EGEDOR: everyone's grand effort in defence of rationality.

Such demands made in the name of fighting stupidity would demonstrate a complete misunderstanding of what heuristics are. They would, ironically, succumb to the influence of the availability heuristic. Let us note that the tendency to consider heuristics as the source of all foolishness stems from the fact that when we describe them, we almost exclusively mention the errors that result from their use: the case of Linda, peculiar games with the wheel of fortune combined with questions about the UN, the trivial question about the occurrence of the letter "r". It is not surprising, therefore, that "heuristics" is for us a source of errors and pyramid-like foolishness. Meanwhile, heuristics are very clever and useful products of evolution.

Human beings live in an incredibly complex environment, facing hundreds of more or less significant decisions every day. There would not be enough time or energy for rational deliberation over each of them. Heuristics, by acting automatically, quickly, at an unconscious level, and effortlessly, usually provide us with good or at least acceptable solutions to encountered problems.

Let's consider, for example, what the representativeness heuristic truly gives us. It offers us suggestions based on certain patterns or prototypes. If a situation is similar enough to typical circumstances in which we are in danger, our unconsciousness will prompt us to exercise caution. If someone possesses characteristics typical of friendly and sympathetic individuals—being smiling, helpful, and polite—we intuitively assess that there is no danger coming from them. Certainly, we can make mistakes in both cases. The "dangerous" situation may turn out to be a joke played by our friends (exploiting our unconscious reactions), and the friendly stranger may be a psychopathic killer. The point is that such mistakes will occur relatively rarely; frequent and systematic errors will only arise when dealing with problems that are themselves atypical, at least from the perspective of the evolution of the human mind, such as dealing with statistical analysis or other tasks unrelated to the problems we encounter in our daily lives.

Similar remarks can be made about the availability heuristic and the anchoring and adjustment heuristic. For an organism, events that leave a distinct memory trace are particularly important—they evoke strong emotional reactions or occur relatively frequently. It is good, therefore, that we have quick access to this knowledge, and it is also good—in the overwhelming majority of cases—that we base our decisions on it. The availability heuristic is not primarily a source of troublesome errors but an incredibly helpful mechanism that facilitates navigation in a complex physical and social reality. The anchoring and adjustment heuristic, on the other hand, makes us sensitive to the context in which the problem we are solving appears. Usually, all elements of that context will be relevant, while some random piece of information (such as an unrelated number) will rarely appear there, most likely introduced by a malicious and inventive experimental psychologist.

Therefore, heuristics are not responsible for our stupidity; they rather serve as a useful toolbox that helps us adapt better to the environment in which we live (cf. GIGERENZER 2001). The belief that we can make decisions differently stems from far idealised, and perhaps even counterfactual, standards of rationality that we have internalised. Let's take a look at Kant's philosophy, for example. Kant claimed that only those moral and legal decisions that are in line with the categorical imperative are well justified (rational): «Act only according to that maxim whereby you can at the same time will that it should become a universal law» (KANT 2002, 37). The categorical imperative may seem innocent. It is just one short rule that we intuitively want

to agree with. It does not impose any specific material obligations on us; it only determines a way of acting that is supposed to enable rational decision-making. However, this innocence is deceptive. Let us pause for a moment and consider what Kant expects from us—who a Good Kantian truly is.

Firstly, a Good Kantian can effortlessly handle abstract conceptual constructions. Since the essence of practical rationality according to Kant is universality—the ability to demonstrate that a particular decision is in accordance with a rule of conduct that could be a universally valid rule—a Good Kantian does not focus on specifics but contemplates highly abstract norms of behaviour. In order to effectively do this, the language they use must be abstract, but at the same time, it must not possess the shortcomings that abstraction brings, particularly vagueness and open texture. Therefore, a Good Kantian fulfils Leibniz's dream: they possess an abstract tool for thinking about morality and law that is as perfect as the language of mathematics.

Secondly, a Good Kantian is compelled to think systematically. It may seem that the categorical imperative finds straightforward application to individual isolated problems. For example, when a Good Kantian ponders how to interpret the rule "Vehicles are prohibited from entering the park", they only need to determine which of the possible interpretations would be a good "universal law". The trouble is that in order to do this, one must consider all possible situations in which the interpreted rule could find application in one way or another. A Good Kantian must consider not only whether cars, motorcycles, or bicycles can enter the park. Their universal rule must also encompass toy cars, wartime souvenir transport trucks, electric motor-powered wheelchairs, passenger airplanes, refurbished Spitfires that served in the Battle of Britain, ambulances transporting critically ill individuals, and so on (cf. SCHAUER 2008). Moreover, a Good Kantian must also consider how the adopted interpretation relates to other norms of conduct. If the result of their deliberation is a "universal law", they cannot allow another—equally universal—norm of conduct to lead to a different conclusion. Therefore, a Good Kantian is a true intellectual Hercules: when resolving any problem, they take into account the entire moral or legal system, ensuring that the decision made does not violate the coherence of that system.

Lastly, a Good Kantian is fully autonomous in their decisions. They base them solely on the operations of reason. They are deaf to the voices of other people, capable of ignoring their own intuitions, and dissociating themselves from the strongest emotions. They are, in fact, a perfect inferential machine that reliably draws conclusions based on the premises at hand. (It is worth noting, incidentally, that Kant was not such a naive thinker as to overlook the role of emotions in ethical and legal thinking. The point is that he described this role in a very atypical way. On the one hand, he emphasised that emotions usually considered important for morality, such as pleasure and displeasure, are obstacles on the path to truly moral decisions. On the other hand, he noted that moral conduct is associated with a feeling of respect or esteem (*Achtung*), which, however, does not have a motivating role and rather resembles an epistemic or aesthetic emotion (cf. KANT 2002, 17).)

A Good Kantian is, therefore, a kind of superhero, and their cognitive abilities far surpass what even an exceptional individual can do. Humans do not possess a perfect, unambiguous language; we cannot meet the Herculean task of creating a complete and coherent system of norms ready to uniformly resolve any moral or legal problem. We also lack the self-control required by Kantian autonomy. This can only be summed up in one way: a Good Kantian is like a unicorn—it does not exist in nature.

Similar remarks can be made about other conceptions of rationality. For example, humans are not capable of meeting the requirement of utility maximisation, which sets the standard for economic efficiency. We are also not inferential machines: we not only make logical errors but often do so systematically, even with seemingly simple reasoning patterns such as *modus tollendo tollens* (if p, then q, and it is not the case that q, therefore it is not the case that p). Does this

mean that we are doomed to irrationality? That only machines have a chance at full rationality—devoid of emotional pressure, maximising utility, or reasoning in accordance with logically valid inference patterns? Must we be more or less foolish?

The matter, it seems, is more complicated. In the literature on economics and psychology, a distinction between two types of rationality has gained significance: constructive and ecological rationality (cf. SMITH 2008). Constructive rationality encompasses concepts well-known to us from textbooks: Kantian moral philosophy or rational choice theory, which forms the basis of contemporary economics. Within this approach, we construct standards for decision-making and action, which are then used to evaluate actual decisions and actions. Ecological rationality is something different: we say that the decisions and behaviours of individuals are ecologically rational to the extent that they are adapted to the structure of the environment. Whether our decisions are rational in this sense, therefore, is not determined by our method of reasoning or the quality of the premises we rely on. It also does not matter whether our thinking conforms to a particular abstract rule, such as the categorical imperative or the maximisation of utility. On the contrary, ecologically rational decisions do not have to be the result of reasoning at all. They usually serve as prompts from our unconscious intuitive heuristics, which we simply accept without subjecting them to conscious control. The quality of these decisions is not indicated by the way they are made but by the extent to which they are adapted to the structure of the environment.

Which of these two visions of rationality should we follow, then? Is it more important to meet the abstract and often difficult-to-apply constructive standards or to have a good fit with the environment in which we live? Who is the fool here—the lofty Constructivist who enjoys ignoring facts or the somewhat unambitious Ecologist seeking intuitive solutions that easily fit into the given situation? Arguably, this is a poorly posed question. We need both constructive and ecological rationality. The former, as Vernon L. Smith claims, generates diversity—it allows for the creation of many, often different ideas for solving problems. «Selection, however, occurs during ecological processes of adaptation» (SMITH 2008, 38): some of our theories turn out to be valid, and the standards expressed in them become part of our intuitive understanding of the world, while others fail and fade into oblivion.

2. *The hypothesis*

In the previous section, I discussed two types of rationality: constructive and ecological. It turned out that the high standards of rational thinking and action described by philosophers in thick volumes, taught in schools and university lectures, are quite troublesome. On the one hand, we often lack the resources required to adhere to the rules of rationality. On the other hand, the world we live in, both physical and social, is so complex that even precise guidelines for dealing with problems encountered do not guarantee success. Our sophisticated intellectual constructions crumble in the face of complex reality. Therefore, we simply cannot be as rational as we would like to be.

However, we are rational in another sense—ecological. Evolution has equipped us with decision-making mechanisms that operate at an unconscious level and are based on simplified rules—heuristics. These mechanisms generally work well, allowing us to make decisions that usually turn out to be good, even though they are not preceded by long deliberation and do not consume much time and energy.

It should be noted that both types of rationality—constructive and ecological—are important and necessary. The former is responsible for variation: it involves creating different concepts of thinking and action that provide an opportunity to break free from established but not necessarily most effective behavioural patterns. The latter has selective power: it determines what is truly effective.

In this context, it is worth revisiting the question of what it means for a human being to be a rational creature. Is our rationality, as Aristotle believed, something completely distinct from the emotional mechanisms we share with other animals? Or should we think about rationality differently, allowing the possibility that it is connected to unconscious decision-making processes based on emotional reactions? Let us take a closer look at this matter.

In the famous essay *The Emotional Dog and Its Rational Tail*, American psychologist Jonathan Haidt formulated a controversial thesis (cf. HAIDT 2001). Based on the analysis of numerous studies in experimental psychology and neurobiology, he claimed that human moral decisions hardly have a rational aspect. They are almost entirely based on emotional reactions. In the process of upbringing, we learn—instructed by adults but also simply by observing them—unconscious, semi-automatic reactions to social situations. Yes, sometimes we use rational arguments, but not to make decisions—those are the domain of unconscious emotional mechanisms—but rather to justify them; to convince others that we acted rightly.

It doesn't take long deliberations to notice that Haidt's view is provocative, at least in the context of European thought history. After all, all—well, almost all—the greatest philosophers have always emphasised that the use of reason is what determines our humanity and truly moral stance. This is beautifully illustrated by the example of Socrates, who showed how rational argumentation can expose the falsehood of moral views based on emotions, passions, and prejudices.

It is not surprising, therefore, that Haidt's essay sparked significant criticism (cf. PIZARRO & BLOOM 2003; LEVY 2006; FINE 2006). Philosophers and psychologists launched a vehement attack on his theses, emphasising that humans are not as irrational as Haidt presents them. The popularity of Haidt's essay and the strength of the reactions to it are incredibly instructive. It seems that he touched a sensitive chord—deeply ingrained beliefs that cannot be undermined without causing general outrage. This reaction also indicates, however, that "there is something to it" because if Haidt's concept were entirely baseless, it would be difficult to explain such widespread and sharp criticism. And even though Haidt limited his analyses to the sphere of morality, the conclusions he formulated can be generalised. Isn't it true that all our decisions—whether in morality, mathematics, economy, or quantum chemistry—are made based on unconscious emotional mechanisms, and then only justified with rational arguments if necessary? In other words, is rationality truly crucial to being human, or is it merely a set of tools that allow us to justify *ex post factum* decisions made based on emotions?

Perhaps the greatest advocates of reason and rationality in history were Descartes and Kant. The former believed that humans constitute a kind of "union" of two substances: thinking and extended. In other words, Descartes proclaimed the dualism of the mind and body. Importantly, the mind—the "thinking thing"—was considered by Descartes to be purely rational. This is evident particularly when the author of the *Discourse on the Method* ponders the place of emotions in human cognitive experience. Descartes claims that emotions originate from the body, not the mind. One could say they "contaminate" rational thought processes. The mind itself has nothing to do with joy, anger, or shame. It operates more like a perfect geometer, though its focus extends beyond geometric figures to encompass all ideas. These ideas are "seen" by the mind's eye, compared to one another, combined into larger wholes, or conversely, broken down into elementary factors, thus leading to rational decisions. To be fair to Descartes, it should be added that he acknowledged the significant role of emotions in human life and made efforts to adequately describe this role. However, this does not change the fact that the rational mind—as such—had nothing to do with emotional life for him.

Kant, as we have stressed above, held a similar view. When contemplating the foundations of moral and legal judgments, he argued that the greatest obstacle to meeting rational standards in these domains is succumbing to individual preferences, basing our actions on what we like or dislike. These preferences are, of course, based on emotional reactions. Therefore, to be a

rational moralist, we must learn to ignore any suggestions coming from the emotional layer of our minds. An impartial and thus rational moral actor is only someone who acts based on principles accepted through reason, detached from individual needs and emotions.

There were also great philosophers who viewed emotions differently, even claiming that it is not reason but emotions that shape societies. They constitute a much more powerful force than the most refined intellectual speculations. In this context, David Hume said that «reason is a slave to the passions». However, it should be noted that even Hume—though he placed emotions at the center of human mental life—clearly distinguished them from reason. They are something entirely different from rational thinking.

It is therefore difficult to escape the conclusion that the traditional conception of the mind places the ability to reason at its center—or at least in a distinguished position—and treats emotions as something perhaps important but certainly not prominent. It is not emotions that determine who we are, but our capacity for rational thinking.

Psychological and neurobiological research conducted in recent decades has led to a somewhat different view of the mind appearing more frequently in scientific and popular discourse. Perhaps the most well-known embodiment of this view is the thesis put forth by Daniel Kahneman: the mind consists of two systems, 1 and 2. System 1 is responsible for unconscious decision-making: it operates quickly, without significant energy expenditure, and beyond conscious control. Its functioning is based, if not entirely, then to a large extent on emotions. On the other hand, System 2 enables conscious decision-making, but it is time-consuming and requires effort. Important note: The majority of decisions we make on a daily basis come from the unconscious System 1. The resource-consuming System 2 is not suitable for frequent use. It is useful in exceptional situations, especially when System 1 suggests conflicting solutions or when we encounter an unusual problem (cf. KAHNEMAN 2011).

A variation of this vision of the mind is the above-described conception of Jonathan Haidt. Let us recall that he compares the human mind to an emotional dog wagging a rational tail. By this, he means that our minds are fundamentally emotional mechanisms, which evolution has only complemented with the ability to think rationally. However, rational thinking does not replace emotions in any way; it merely supports them. Haidt himself claims that this support is relatively small: we use rational arguments to *ex post factum* justify decisions made unconsciously through emotional reactions learned through years of experience. Not everyone goes as far as Haidt; some philosophers and psychologists emphasise that rational thinking does sometimes have a real impact on the decisions we make.

Nevertheless, the fact remains that—at least in cognitive science and philosophy inspired by it—the way we think and talk about the mind and rationality has fundamentally changed. The mind is no longer seen as a rational analytical machine that derives unshakable conclusions from available premises, sometimes struggling—or even losing—against irrational emotions. Emotional mechanisms have become a crucial and inseparable layer of the mind. And we don't have to consider them irrational: even if they do not meet the rigorous standards formulated by philosophers, more often than not, they prove to be ecologically rational.

But what if we go even further? What if the human mind is entirely emotional? Some contemporary philosophers suggest this direction of research (cf. HURLEY et al. 2011). Perhaps we should acknowledge that the mind does not operate based on any algorithms but rather constitutes a space in which different emotional states compete or cooperate with each other, with the task of regulating behaviour. In this view, "reason" is merely a derivative of emotional mechanisms, and its task is to organise more fundamental processes based on emotions. Emotions have the first and final say.

Let's indulge our imagination even further. Why not consider that everything rational in the traditional sense is outside the mind? As organisms, we learn to navigate the complex physical and social world from birth. We learn that we can only leave a room by opening the door, that

it is impossible to jump onto a balcony on the third floor, that it is better not to say certain words in the presence of adults, as it will result in punishment, that hot water can burn us, and that uttering certain phrases will lead to our behaviour being accepted or disapproved by a particular audience. We learn all of this thanks to the existence of emotional mechanisms: the conclusion that one cannot jump to the third floor is not the result of any reasoning, just as the feeling that a certain argument will serve as a good defence against disapproval is not the product of any conscious thought process.

From this perspective, logically valid schemes of reasoning—such as *modus ponens* (if *p*, then *q*; *p*; therefore *q*)—do not exist within our minds but are external to them, similar to doors, balconies on the third floor, or reactions to the words we utter. We reason according to *modus ponens* not because we *apply* that inference scheme, but because we encounter certain obstacles in our ecological niche—for example, the disapproval of a logic teacher. We are emotional beings from beginning to end, but ones that inhabit a rational ecosystem. Rationality is not within us but in the world in which we find ourselves living.

This vision may seem very extravagant if not senseless. It is certainly inconsistent with what we are accustomed to thinking about rationality. However, it is worth considering it—not necessarily to accept it, but to better understand what it means to be rational.

3. *The law*

Let us pose the question now how this unorthodox approach to rationality may influence our understanding of the law. The view experimentally embraced here is that the human mind is not rational in any traditional sense of the word. When we make decisions—unconsciously or consciously—we do it always in an intuitive or instinctive way. The fact that some of those decisions *seem* rational is made possible by the fact that we inhabit an environment—physical and social—which is so structured that if favours "rational" decisions.

This is a very strong claim. Imagine that one is trying to solve a geometrical problem, or considers a complex case in tax law. One is fully aware what they are doing and puts much conscious effort in finding a solution. For example, one is drawing certain figures to visualise the geometrical problem at hand, or works back and forth with the geometry axioms to produce a rigorous proof of the claim which constitutes a solution to the problem. Or—in the context of the tax law case—one is perusing past cases and familiarises themselves with doctrinal theories in order to test various ways of interpreting a set of tax law provisions. The claim advanced here is that even with such conscious effort one is not *applying* the rules of logic or other methodological precepts. One is only generating circumstances under which one's instinctive or intuitive faulty has a reasonable chance at stumbling upon a solution to the problem which is coherent with the *external* requirements of rationality.

This is a crucial point. No one ever applies *modus ponens* (or any other rule of rational thinking for that matter); they only *think* they do. What really happens is that a solution comes to one's conscious mind of which they *feel* that it is the right solution. And the emergence of the feeling is conditioned by the training one has gone through in the process of inculturation. The social institutions (understood broadly, as embracing conceptual structures which favour particular kinds of decisions in typical situations) have been simply embodied in one's mental maps responsible for intuitive judgment.

It would be impossible to consider in detail how this dramatic shift in understanding rationality may alter our understanding of legal thinking and legal institutions. However, since the goal of the present paper is only to put forward a bold hypothesis and initiate rather than summarise a discussion, I will limit myself to illustrating some potential consequences of the view of rationality embraced here for legal philosophy by considering three different issues.

Legal philosophers have sometimes expressed the view that legal thinking has little to do with conscious deliberation and the careful application of rules, but rather is the domain of intuition and insight. The *locus classicus* of this minority approach is a lecture of J.C. Hutcheson *The Judgement Intuitive: The Role of the "Hunch" in Judicial Decision* (HUTCHESON 1929). There, Hutcheson advances a conception that the mind of a judge is not a «cold logic engine», which algorithmically deals with the law understood as a «system of rules and precedents, of categories and concepts» (HUTCHESON 1929, 274); it takes advantage—especially in complex cases considered by brilliant judges—of imagination and intuition:

> «While when the case is difficult or involved, and turns upon a hairsbreadth of law or of fact [...], I, after canvassing all the available material at my command, and duly cogitating upon it, give my imagination play, and brooding over the cause, wait for the feeling, the hunch—that intuitive flash of understanding which makes the jump-spark connection between question and decision, and at the point where the path is darkest for the judicial feet, sheds its light along the way» (HUTCHESON 1929, 278).

Therefore, Hutcheson seems to believe that in hard cases the legal mind does not rely on the application of legal rules and the use of rational reasoning schemes: it is the domain of hunch. He further observes that this kind of dissociation of decision-making and rationality is confirmed by the fact «all of us have known judges who can make the soundest judgments and write the dullest opinions on them» (HUTCHESON 1929, 287). *A contrario*, it would seem that easy and straightforward cases do not require hunch and can be dealt with through the utilisation of "logical algorithms".

From the perspective of the view embraced in this essay, the way Hutcheson treats easy and straightforward cases is based on an illusion. Similarly to hard cases, they are decided by intuition. However, since they quickly "fit" into the institutional framework of the law, one is unable to *observe* the work of intuition here. No play of imagination is needed; poor judges do as well as extraordinary do. Let us observe that the change of perspective I argue for makes things much simpler. One does not need to assume that there are two different mental faculties at work in dealing with legal cases: hunch when hard cases are considered, and 'logical engine' when an easy case is solved. There is a continuity between easy and hard, and both kinds of cases are solved in the same instinctive way.

Another illustration I would like provide is the understanding of one of the most popular and intriguing conceptions of practical rationality discussed today, which is usually called the "Reasons First Approach" (cf. WEDGWOOD 2015). It is advocated, in different versions, by Joseph Raz, John Skorupski, Thomas Scanlon, or Mark Schroeder, and posits that the irreducible concept of reason is central to any successful explanation of rationality and normativity. But what are reasons? According to Raz, they are certain facts which «constitute a case for (or against) the performance of an action» (RAZ 2011, 36). The fact that I am hungry is a reason for eating something; the fact that my uncle is seriously ill is a reason to visit him; the fact that I have made a promise to a colleague to help him paint his house is a reason to do it.

It is not our power of rational thinking which makes facts into reasons; rather, «they are reasons because rational creatures can recognise and respond to them» (RAZ 2011, 85). In other words, reasons are "out there"; they exist independently of whether they are identified as such or not. Moreover, as Raz repeatedly observes, finding an appropriate response to a reason does not necessarily involve our rational faculties or abilities—it is possible to do it «without the mediation of rational power» (RAZ 2011, 85). As Raz puts it, «with experience we learn to identify and respond to reasons instinctively, though in ways which depend on and presuppose first, reliance on past reflection, and second, the monitoring presence of rational powers which control and stand ready to correct misidentifications or misdirected responses» (RAZ 2011, 86).

The "Reason First Approach" is troublesome when seen from a perspective of some traditional conceptions of rationality. First, it is difficult to accept that simple, instinctive behaviour—such as eating when one is hungry—should be seen as a "response" to reason. Second, it is equally difficult to agree that one can respond to reason and act rationally in an unconscious way. Third, the ontological status of reasons is questionable: why should we assume that there exist designated "state of affairs" or "facts" called reasons, which are independent of our minds? Where do those mysterious entities come from? Such declarations make one think of sharpening the Occam's Razor.

These troublesome aspects of the "Reason First Approach" become less controversial, or even disappear altogether, when the view embraced in this essay is taken to replace the traditional conception of rationality. If every decision we make is instinctive, then it is not surprising that deciding to eat when one is hungry is not much different from solving a complex mathematical puzzle or a legal hard case. If one utilises intuition even when one consciously entertains a problem, then it is not surprising that we make rational decisions (i.e., we respond to reasons) also in an unconscious way. If all that is rational is "out there", outside of the human mind, in the social structures which the evolutionary forces have shaped, speaking of reasons as independently existing facts seems much more innocent. It may not be the best conceptual scheme there is, but at least reasons are no more mysterious entities emerging out of nowhere. It seems therefore, that there is much to defend the "Reasons First Approach" with; the fault of its proponents is not that they are too revolutionary, but that they are not revolutionary enough, as when Raz insists that responding to reasons requires «reliance on past reflection, and [...] the monitoring presence of rational powers which control and stand ready to correct misidentifications or misdirected responses» (Raz 2011, 86).

My final illustration comes from the domain of institutional design and revolves around the concept of nudge. It was popularised in 2008 in the book by Richard Thaler and Cass Sunstein (cf. Thaler & Sunstein 2019). Adopting Kahneman's conception that the human mind operates with two systems: the conscious and deliberative System 2, and the unconscious intuitive System 1, Thaler and Sunstein argue that people make a vast part of their decisions in an unconscious way. This fact, they claim, should be reflected in the way we think of designing social institutions. Rather than shaping them for a rational and deliberative individuals, the policymakers should put in place such frameworks, which recognise the importance of unconscious decision-making and "nudge" people to behave in a way which is beneficial for them and the society.

The key thesis behind the idea of nudge is that it is unreasonable—or, in fact, impossible—for human beings to behave *only* in a rational and deliberative way. Because of that, no social institutions designed *solely* for rational individuals will lead to beneficial outcomes; they must necessarily include "nudges" which serve as "obstacles" or "incentives" in the environment and which are recognised by the unconscious mind and lead it to beneficial behaviour. The key in this setting is, of course, to design the system of nudges in such a way that the behaviour they generate is indeed beneficial.

The difference between the original conception of nudge and the conception advanced in this essay is that the existence of two different systems: 1 and 2, is rejected here. What we have is *only* the intuitive (although not necessarily unconscious) System 1. As a consequence, *all* institutional design should be nudge-esque. This claim immediately leads to the final question I need to shortly address. If all our decisions are intuitive, how can we *rationally* design anything? The designers of social institutions are human beings, which means that they do *not* think in a conscious deliberative way, but rather make their own intuitive decisions. How is it possible that such decisions have led to the emergence of a social structure which embodies some form of rationality?

The only answer to this question lies in the mechanisms of variation and selection associated with cultural evolution. Once put into such an evolved social framework, people learn by

experience how to intuitively react to various situations: they learn the tricks to earn money, solve legal cases, or write books on institutional design. They also learn how to construct arguments and present the process of deliberation as if it was a purely conscious and rule-governed effort. In this framework, new tricks may emerge and people may learn them; but they appear as any evolutionary novelty - through chance or error.

The law is a part of this complex social structure. It is not an ideal system of norms, although we have learnt the trick to present it as such. Legal decisions are not deliberative rational acts: they are, as all our decisions, intuitive; but we have also learnt the trick how to disguise them into something they are not. We put the disguise on instinctively, and tearing it away is difficult. It goes against all the tricks we have learnt. But are we old dogs to be afraid of new tricks?

References

FINE C. 2006. *Is the Emotional Dog Wagging Its Rational Tail, or Chasing It? Unleashing Reason in Haidt's Social Intuitionist Model of Moral Judgment*, in «Philosophical Explorations», 9, 83 ff.

GIGERENZER G. 2001. *The Adaptive Toolbox*, in GIGERENZER G., SELTEN R. (eds), *Bounded Rationality: The Adaptive Toolbox*, MIT Press, 37 ff.

HAIDT J. 2001. *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgement*, in «Psychological Review», 108, 814 ff.

HURLEY M., DENNETT D., ADAMS R. 2011. *Inside Jokes: Using Humor to Reverse-Engineer the Mind*, MIT Press.

HUTCHESON J. 1929. *The Judgment Intuitive: The Role of the "Hunch" in Judicial Decision*, in «Cornell Law Review», 14, 274 ff.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

KANT I. 2002. *Groundwork of the Metaphysics of Morals*, Yale University Press.

LEVY N. 2006. *The Wisdom of the Pack*, «Philosophical Explanations», 9, 99 ff.

PIZARRO D., BLOOM P. 2003. *The Intelligence of the Moral Intuitions: A Reply to Haidt*, in «Psychological Review», 110, 193 ff.

RAZ J. 2011. *From Normativity to Responsibility*, Oxford University Press.

SCHAUER F. 2008. *A Critical Guide to Vehicles in the Park*, «New York University Law Review», 83, 1109 ff.

SETMAN S. 2022. *Teaching an Old Dog New Tricks: Intuition, Reason, and Responsibility*, in «Revista de Humanidades de Valparaíso», 19, 85 ff.

SMITH V. 2008. *Rationality in Economics: Constructivist and Ecological Forms*, Cambridge University Press.

THALER R., SUNSTEIN C. 2019. *Nudge*, Penguin.

TVERSKY A., KAHNEMAN D. 1974. *Judgment under Uncertainty: Heuristics and Biases*, in «Science», 185, 1124 ff.

WEDGWOOD R. 2015. *The Pitfalls of "Reasons"*, in «Philosophical Issues», 25, 123 ff.

WIKIPEDIA CONTRIBUTORS 2023. *Lies, Damned Lies, and Statistics*, in *Wikipedia, The Free Encyclopedia*, 2023, March 14. Available on https://en.wikipedia.org/w/index.php?title=Lies,_damned_lies,_and_statistics&oldid=1144481014.

Mindreading in Law

ŁUKASZ KUREK

1. *Introduction*

"Mindreading" refers to the human cognitive capacity to attribute mental states - such as thoughts and emotions - to other people and oneself. People use this capacity to make sense of the behaviour of the target of mental state attribution. Mindreading is used in various domains which include, but are not limited to, everyday life and law. The capacity in question is studied by cognitive scientists who push forward our understanding of it by describing the psychological mechanism which underpins it. Although there are still some disagreements about how this mechanism works, what we already know about should be of interest to those who want to better understand legal reasoning about human behaviour. The goal of this chapter is to survey the cognitive-scientific research on mindreading and discuss its implications for law.

2. *What is mindreading for?*

As it was mentioned, mindreading is used to interpret behaviour. To illustrate the relationship between mindreading and behaviour interpretation consider the following example. Imagine there is a person, Jones, who, upon leaving a party, took someone else's umbrella instead of his own. Both umbrellas were very similar. In fact, both umbrellas were two exemplars of the same model and the only noticeable difference between them was that the opening mechanism in Jones' umbrella was malfunctioning. However, before leaving the party, Jones did not open the umbrella that he took.

At least two competing interpretations of Jones' behaviour are available. According to the first interpretation, which is more favourable to Jones, he made an honest mistake. He was simply unaware that the umbrella he took belonged to someone else. This interpretation is corroborated by the fact that both umbrellas were very similar and the only noticeable difference between them could be observed upon their opening—and Jones did not open the umbrella he took upon leaving the party. According to the second interpretation, which puts Jones' in a far less favourable light, he stole the umbrella. This would have been the case if had known that he was taking someone else's umbrella and he just had it with his own, broken one.

Deciding between these two interpretations requires deciding which mental state to attribute to Jones. The interpreter needs to determine whether Jones believed that he was taking someone else's umbrella or whether he believed that the umbrella is his own. This is because stealing requires that the person who takes an object belonging to someone else is aware of this. A person cannot steal an object she believes to be her own.

Given the above, we may conclude that mental state attribution is essential if our goal is to determine whether someone made a mistake. Although this is true, the role of mindreading is

* I would like to thank the anonymous reviewer for offering insightful comments and raising hard questions which helped me strengthen this chapter.

much more prominent than this. We mindread in order to interpret behaviour for various practical purposes which fall into a broad category of behaviour management (MALLE 2004, 63-82; HUTTO 2008, 23-40; SPAULDING 2018, 42-61). Managing behaviour consists in, for example, attributing moral or legal responsibility, criticizing or praising as well as administering more substantial rewards or punishments.

3. *Cognitive science of mindreading: Some preliminary remarks*

Mindreading can be approached from three perspectives. First, mindreading is a cognitive *capacity*. Second, it was mentioned that cognitive scientists are interested in the features of the *information-processing* associated with this cognitive capacity. Finally, it was claimed that the capacity to attribute mental states is exercised in various domains which means that there is a diverse range of concrete *cases* in which we people use mindreading. Some readers may want to know a little bit more about the relationship between cognitive competences, information-processing and concrete cases in which these cognitive capacities are exercised as well as about the relationship between mindreading, behaviour interpretation and behaviour management.

Cognitive capacities tell us what a mind can do. Typically, they are individuated by their goal—by spelling out what minds endowed with these capacities can achieve. Thus, the most immediate goal of mindreading is to attribute mental states. Cognitive capacities are often nested, which means that having them is linked with having other, lower-level capacities. For example, the capacity to interpret behaviour—the most immediate goal of which is to make sense of behaviour—is linked with having mindreading. Behaviour interpretation is itself a lower-level cognitive capacity in comparison to the capacity to manage behaviour. To illustrate this, recall Jones, who had taken someone else's umbrella instead of his own. Our capacity to interpret Jones' behaviour—to decide whether he stole the umbrella or just made a mistake—is linked with our capacity to mindread, in the sense that the former involves being able to attribute to Jones a belief with a particular content. Going further, our capacity to decide whether Jones' should be punished involves our capacity to interpret his behaviour, in the sense that administering this punishment involves being able of making sense of what he did.

One way of describing how the mind processes information when a particular cognitive capacity is exercised consists in providing an algorithm—a set of instructions—for achieving the goal of this capacity. Below, we will spend some time discussing the information processing associated with mindreading. This is where the cognitive-scientific research on mindreading will be of particular relevance.

Finally, it is worthwhile to consider concrete cases in which people use mindreading. First, concrete cases in which people attribute mental states supply most of the observable phenomena in every discussion about mindreading. Not so long ago these cases supplied the only observational phenomena in this context but this state of affairs changed when cognitive scientists began to investigate how mindreading is implemented in the human brain. Notice that in concrete cases in which people attribute mental states we do not actually observe how mindreading works. What we observe in concrete cases is human behaviour—linguistic or non-linguistic—which suggests that people who engage in this behaviour mindread.

For example, we may hear Smith saying to Brown something like 'Jones did not steal your umbrella. He took your umbrella because he thought it was his and this may be taken to suggest that in order for Smith to utter this sentence, it is not enough that Smith knows the English language—that he knows what to say if he wants to communicate to Brown that Jones took Brown's umbrella by mistake. Instead, we may suggest that in order for Smith to utter this sentence Smith needs to be capable of solving a cognitive, not a linguistic problem—the former

consisting in attributing to Jones a belief with a particular content. This is precisely what cognitive scientists are suggesting in their study of mindreading.

Although what people say when they mindread is an important source of data in these studies, when cognitive scientists investigate mindreading they do not target a linguistic competence but a cognitive one. The drawback of this approach is that cognitive competences and the information-processing which is associated with them are non-observable phenomena in the sense that their existence and their features need to be inferred from what can be directly observed.

Second, it is worthwhile to take a closer look at concrete cases in which people use mindreading because there are good reasons to think that there will be differences in the features of this cognitive capacity depending on the domain in which it is exercised. In particular, we will discuss differences between how mindreading is exercised for legal purposes and in everyday situations.

4. *Is mindreading modular?*

Cognitive science is often described as an interdisciplinary attempt to explain how (human) cognition works (THAGARD 2005, IX; BERMÚDEZ 2014, 88-90). This description is not inaccurate, but it is also not very informative. Specifically, it does not put into view an important assumption made by cognitive scientists according to which cognition is not a unified phenomenon: it is not only useful, but also correct to break down cognition into different cognitive capacities such as the capacities to perceive colours, shapes and movements, the capacity to mindread or the capacity to interpret behaviour. These distinctions are useful because, instead of tackling one general and quite complex issue, cognitive scientists study more specific and manageable ones.

Most of all, however, this partitioning of cognition seems to be a correct assumption about how the mind works. Numerous empirical and theoretical studies of the mind suggest that the mind is divided into parts which are sometimes called "mental modules" (CARRUTHERS 2006). We will focus on three features of mental modules: specialization, inaccessibility and automatism. Specialization means that each module deals with problems of a particular type. Inaccessibility refers to the fact that what happens within a module cannot be monitored from the outside. The upshot of inaccessibility is automaticity: once modules are provided with input, they process it until completion without any outside influence. These features are not an all or nothing matter. A particular module may be more or less specialized in the sense that the problems that it deals with may be more or less diverse. More inaccessible modules will be less available for external monitoring in comparison to more accessible modules. Finally, it will be more difficult to influence what happens inside highly automatic modules in comparison to modules that operate less automatically.

The more modular information processing associated with a given cognitive competence is, the less flexible this competence will be, in the sense that it will adapt more poorly to circumstances in which too much is demanded of it. Consider vision, for example. We know very well that reality consists in much more than medium-sized objects which lie around us. Still, our knowledge about reality cannot change the fact that we perceive the world in this particular way. The problem with perception is that it is not very flexible and it adapts poorly to investigations aiming to determine what kinds of things there are in the world different from medium-sized objects which lie around us. You need to go beyond what you perceive in order to construct an adequate image of reality. In the case of mindreading, the less flexible this competence is, the more poorly it will adapt to circumstances in which too much is demanded of it. So it is worthwhile to investigate whether there is any evidence that the information processing associated with mindreading is modular. If such evidence exists, it will provide us with preliminary reasons and a conceptual framework to

investigate a more specific issue: how mindreading works if it is used for the purposes of legal interpretation and management of behaviour.

However, mindreading does not look to be associated with information-processing of a modular nature. On the contrary, this cognitive capacity looks to be closely related to general reasoning. General reasoning is one of our most flexible capacities and the information processing that is associated with it does not look to be specialised, inaccessible or automatic. This is because the information processing in question deals with problems of a diverse nature, it can access information stored in various parts of the mind and we may control it. Information processing associated with general reasoning looks to be a good example of a non-modular process.

Discussing the relationship between mindreading and general reasoning in more detail, it may appear that mindreading is just a special case of general reasoning. It may appear that when we attribute mental states we use our general reasoning capacity—that is, the reasoning capacity we use to solve problems in various domains—for the purpose of mental state attribution. The reasoning required to decide whether Jones—who took someone else's umbrella—was aware that he was taking someone else's umbrella does not look very different from the reasoning required to decide, for example, which animal left the footprints that we observed when walking through a forest—and in the latter case there is no need for mental state attribution. Thus, if mindreading is only a subspecies of general reasoning, then the information processing associated with the former is not specialized.

What is more, mindreading looks to be associated with information processing which is accessible from the outside. When we attribute mental states we can consciously access what we know about the relationship between mind and behaviour. Mental state attribution does not look to be automatic either. If an actor in a play convincingly pretends despair, initially you may truly believe that this person actually suffers despair. But if you remember that what you see is a play you may override your initial reaction and attribute to the actor the correct mental state which is pretend despair.

In sum, it appears that information processing associated with mindreading does not seem to be modular, which means that this cognitive capacity is a rather flexible one. Thus, we should not worry very much that mindreading will be particularly prone to fail if too much is demanded from it in comparison to cases in which we demand too much from one of our most flexible cognitive capacities which is general reasoning. However, if we go beyond these appearances concerning mindreading, we will observe that things look very differently. To illustrate that mindreading and general reasoning capacity are associated with different kinds of information processing—which means that they are separate cognitive capacities—and that the information processing associated with the former is much more modular than it looks at face value, we will briefly discuss the cognitive-scientific research on two issues related to mindreading: its neural basis and the autism spectrum disorders.

Research on the neural basis of mindreading indicates that the brain regions processing information associated with mindreading are different from the brain regions processing information associated with general reasoning capacity. The brain regions associated with the former include, in particular, dorsomedial prefrontal cortex (DMPFC) and temporoparietal junction (TPJ). The brain regions associated with the latter include, in particular, lateral prefrontal cortex (LPFC) and lateral posterior parietal cortex (LPPC) (Saxe & Powell 2006; Lieberman 2013, 116 f.).

Broadly speaking, to determine which brain regions process information associated with a given cognitive competence, researchers measure which brain regions are more active when people solve problems requiring this competence. These measurements are made with the aid of neuroimaging devices, notably functional imagining devices which measure brain activity in a particular period of time. A typical problem that involves mindreading consists in deciding whether Jones—who took someone else's umbrella—did it by mistake or on purpose. As for the

problems which require general reasoning, cognitive scientists often identify these problems as those which require that a person consciously holds and updates numerous pieces of information. The cognitive capacity to consciously hold and update numerous pieces of information is called "working memory". Although not all problems which involve working memory involve general reasoning—the latter cognitive capacity is more specific as it is associated with consciously holding and updating numerous pieces of information in an ordered manner: forming conclusions from premises—, all problems which involve general reasoning involve working memory.

One of the early neuroimaging studies illustrating the dissociation between information processing associated with mindreading and information processing associated with working memory involved participants reading three types of sentences (FLETCHER et al., 1995; other studies showing differences in how the brain processes linguistic information about mental states and linguistic information about non-mental phenomena include ROTTSCHY et al. 2011, TURKELTAUB et al. 2003, CASTELLI et al. 2002). The first type, "mindreading stories", described what happened to various fictional persons and their understanding required mindreading. One such group of sentences told a story about a burglar who dropped a glove while running past a police officer. The police officer yelled at the burglar to stop so that the burglar could take his glove back. However, the burglar assumed that the police officer wanted to arrest him and gave himself up. To understand the burglar's behaviour, participants needed to attribute to him the false belief that the police officer knew about his crime and wanted to arrest him. Understanding the other two types of sentences—"physical stories" and "unlinked sentences"—did not involve mindreading. Among physical stories was a story about a burglar breaking into a jeweller's store who steps on something soft which turns out to be an animal. The animal runs away and sets off the alarm. Examples of unlinked sentences included «Jill repeated the experiment, several times» or «She took a suite in a grand hotel». To control whether participants understood sentences belonging to each of the three types they were asked questions about what they read.

When participants read the sentences, they were scanned by a neuroimaging device which showed that understanding the three types of sentences was associated with three different patterns of brain activity. The most significant difference in brain activity was observed in the case of mindreading stories and unrelated sentences. Brain activity was more similar in the case of mindreading stories and physical stories, but in the case of mindreading stories increased activity was observed in a highly circumscribed region of the brain: DMPFC and TPJ. These regions were quiet in the case of physical stories and unrelated sentences. What is more, brain regions associated with working memory, LPFC and LPPC, were relatively quiet in the case of mindreading stories in comparison to physical stories and unrelated sentences.

Recall the three features of modular information processing: specialization, inaccessibility and automatism. The above-mentioned empirical research suggests to what extent the information processing associated with mindreading shares these features. First, it looks as though this information processing is specialized. That is, it looks as though there is a cognitive module that deals only with mental state attribution. This is corroborated by the fact that a highly circumscribed part of the brain shows greater activity when people solve problems of a particular type that involve mindreading and this brain region remains quiet when people solve problems of the same particular type but which do not involve this cognitive capacity. For example, this brain region is active when people read written sentences the understanding of which involves mindreading and it remains quiet when people read written sentences the understanding of which does not involve this cognitive capacity. This suggests that the cognitive module associated with mindreading comes online only in very specific circumstances: when the task at hand involves mental state attribution.

Second, the above-mentioned empirical study suggests that the cognitive module associated with mindreading is, at least to a certain extent, both inaccessible and not subject to the control of general reasoning. Recall that when participants in the above-mentioned experiment solved

problems involving mental state attribution their working memory was not activated. If this was the case, then they did not consciously monitor the process which resulted in, let's say, their attribution to the burglar the mistaken belief that the policeman knows about the robber's crime. This accounts for the claim that the mindreading module may be partially inaccessible to general reasoning—perhaps participants did not monitor this process because it was inaccessible to their general reasoning. Going further, if participants' working memory was offline when they exercised their mindreading capacity, then they did not influence what happened inside their mindreading module via their general reasoning. This accounts for the claim that the mindreading module may not be subject to the control of general reasoning and in this sense automatic. For it may be the case that participants did not influence what happened inside their mindreading modules via their general reasoning capacity because it was not possible for them to influence this module in this way.

It may appear that accessibility to general reasoning and controllability by means of general reasoning are associated with each other in the sense that the former is a necessary condition of the latter: in order for general reasoning to influence what happens inside the mindreading module, a person needs to be able to consciously monitor what happens there. This needs not to be the case, however. A module may not be accessible for conscious monitoring and still remain under the influence of general reasoning. In such cases, the influence of general reasoning will not be direct—in the sense that the premises in general reasoning will not concern what happens inside the mindreading module because this would require accessibility to this module—but it will be indirect. Indirect influence of general reasoning on the mindreading module consists merely in the fact that results of general reasoning are available to this module as an input. That requires that the mindreading module has the capacity to access the information processing associated with general reasoning and monitor what happens inside.

The drawback of the experimental study discussed above is that the mental state attribution problem that the participants needed to solve in order to understand the stories that they read was relatively simple. It may well be the case that it is because of its simplicity that the participants did not engage into general reasoning to determine, for example, what was the robber thinking when he gave up. Other, more complicated mindreading problems may not be possible to deal with unless one explicitly reasons about which mental state needs to be attributed.

These more complicated mindreading problems are often encountered when mindreading is used for the purpose of legal interpretation and management of behaviour. Think of a lawyer who wants to determine whether someone, let's say Wilson, committed fraud. This type of crime requires that Wilson intentionally deceives his victim in order to unlawfully obtain some gain. Thus, interpreting Wilson's behaviour as an instance of fraud requires attributing to him several mental states such as: a belief that what he told his victim was false, a belief that by deceiving his victim he will obtain a particular gain and a desire to obtain this particular gain. In effect, substantiating the claim that Wilson committed fraud requires an explanation of why it is appropriate to attribute to Wilson these particular mental states. Explaining this point requires that one engages in careful consideration about why this attribution is supported by the available evidence. This consideration will heavily draw on the working memory and general reasoning capacity of the interpreter of Wilson's behaviour.

If in more complex cases mindreading is linked with general reasoning, then this could be taken to show that information processing associated with mindreading is not modular. For in these more complex cases, it looks as though there are no two separate information-processing mechanisms underpinning mental state attribution—one associated with mindreading and the other associated with general reasoning—, but only one such mechanism which looks to be accessible to general reasoning and subject to its control. What is more, this mechanism does not look to be domain-specific, because it encompasses information processing associated with general reasoning which is not, by definition, limited in its purposes. One could also argue that there is no qualitative

difference between simple and difficult mindreading problems in the sense that in simple cases information processing associated with mindreading is accessible to general reasoning and subject to its control, but because these cases are simple general reasoning is not required. In the case of simple mindreading problems people could, however, access the information processing mechanism associated with their mental state attribution if they were explicitly asked to do so.

The idea that information processing associated with mindreading is not modular in the sense that it can be accessed and controlled via general reasoning is tempting if only for its simplicity. The upshot of this idea is that there is no truly separate mindreading capacity. In this view, complex mindreading problems, such as the problems encountered in the domain of legal interpretation and management of behaviour, require using general reasoning capacity in a specific domain which is mental state attribution. However, there are empirical findings that undermine this idea. These findings concern autism spectrum disorder (ASD) which is associated with impaired mental state attribution.

In one such experiment, there were four groups of participants: high-functioning children with autism, children with general intellectual impairment, normally developing 8-years old children, and adults (Abell et al. 2000). The three groups of children were matched for their verbal age, which means that children in the first and second group were 3-4 years older than children in the third group. Participants were shown animations of geometrical figures some of which interacted with each other as if one figure responded to the mental state of another figure. For example, in one such animation one triangle was trying to persuade another triangle to leave an enclosure. This means that the former triangle was aware that the latter triangle was reluctant to leave and the former made an attempt to change the latter's mind. After watching the animations participants were asked: "What happened in the cartoon?" and their descriptions were evaluated for accuracy. In the case of animations that required mindreading, participants' descriptions were evaluated for the accuracy of mental state attribution.

Results have shown a marked difference between high-functioning children with autism and the other three groups. More specifically, the accuracy of mental state attribution in the case of children with autism was significantly lower in comparison to all the other groups. For example, in the case of the animation showing one triangle trying to persuade another triangle to leave the enclosure, one child with autism said that «they are trying to push each other and want to kiss one another». What is more, results of this experiment corroborated previous findings showing that a characteristic feature of high-functioning individuals with autism is not that they use mentalizing descriptions less frequently in comparison to non-autistic individuals, but that the former are less accurate in their mental state attributions than the latter.

These results are of significance to our considerations because they go a long way towards showing that mindreading and general reasoning are separate capacities and that the information processing of the former is modular in the sense that it is domain-specific as well as inaccessible to general reasoning and not subject to its control. Observe that the mindreading problems that the participants in this experiment encountered were complex in the sense that these problems required general reasoning. Participants needed to come up with an explanation that best accounted for the available evidence on their own. However, even though the general reasoning capacity of the children with autism who took part in the experiment was significantly higher in comparison to the general reasoning capacity of the children with general intellectual impairment—as it was attested by the IQ tests the children were administered—the former group of children still shown a marked inaccuracy in their mental state attributions. This suggests that in the case of children with autism a more domain-specific information processing is impaired and that this domain-specific information processing is associated with mindreading. What is more, these children could not compensate for their impaired mindreading capacity and inaccurate mental state attributions with their superior general reasoning capacity which suggests that information processing associated with their mindreading capacity was inaccessible to their general reasoning and not subject to its control.

5. *Reflexive and reflective mindreading systems*

Taking stock of the above considerations, we may hypothesize that the information processing associated with mindreading is composed of two systems: a reflexive mindreading system and a reflective one. The reflexive system is highly modular in the sense that it is domain-specific—it is designed to deal with mental state attribution only—and it is highly inaccessible to general reasoning and largely not subject to its control. The reflexive system processes information whenever you encounter problems that require mental state attribution. If these problems are simple, the mental state attributions made by this system will be sufficient to solve them. It is the reflexive system that is responsible for the fact that you immediately make sense of information that you hear or read that requires mindreading to understand. The reflective system comes online when more complex mindreading problems are encountered and it uses the information processing associated with general reasoning. The reflective system is less modular than the reflexive one because the former uses resources associated with a domain-general capacity. This means that in the case of the reflective system the process which leads to mental state attribution is accessible to general reasoning and subject to its control. Importantly, however, the reflective system is not capable of dealing with mindreading problems on its own. It always requires input from the reflexive system as the latter is responsible for providing the necessary semi-finished products to the former. These semi-finished products are intuitive attributions—that is, mental states which are attributed by default when mindreading problems are encountered. We will call these outputs of the reflexive system "mindreading intuitions" or, in short, M-intuitions. How important M-intuitions are in the case of mental state attribution is illustrated by the fact that if the reflexive system is somehow impaired and its outputs are flawed it may be difficult for the reflective system, provided with these flawed M-intuitions, to correct them and produce accurate mental state attributions—as it is the case with individuals with autism spectrum disorder.

In the light of above considerations, our immediate concern should be the reflexive mindreading system. In particular, we should be concerned with its algorithm—the set of instructions that this system follows in order to generate M-intuitions. The reason for this concern is that even formally immaculate reasonings will lead to false conclusions if their premises are false. Thus, even the most advanced reflective mindreading system—correctly utilizing complex algorithms for deductive or probability reasoning—will generate inaccurate mental state attributions if the M-intuitions that it is provided with are flawed. Another reason why we should be concerned with the reflexive system is that, due to its modularity, we may expect that it will be difficult to monitor and influence this system so that it no longer produces flawed M-intuitions.

There is a broad consensus among cognitive scientists that the reflexive system has two components—a theory of mind component and a mental simulation component—which work together in order to generate M-intuitions (Stich & Nichols 2003; Goldman 2006; Stueber 2006). A theory of mind (ToM) is a body of knowledge about mind and behaviour that is stored in memory and used by default to solve mindreading problems. This body of knowledge is not innate and it grows as the individual develops and learns new information relevant to attribute mental states.

One of the most interesting findings pertaining to the development of ToM is that in the course of early childhood it grows in a strikingly similar fashion across different individuals. One such finding is that 4-years old children begin to understand that other people have beliefs and that these beliefs may be wrong. Younger children, on the other hand, tend to think that others always conceive the world as it actually is. This is illustrated by the fact that these younger children tend to think that if they themselves have seen an event happening, then other people will also be aware that this event happened—even individuals who cannot be aware of this because the event happened in their absence. Later on, however, increasing individual differences may be observed in the content of the ToMs of different subjects and in the ways in

which they use it. Some of these differences are culturally driven. For example, individuals grown up in Far East cultures—such as Chinese, Japanese and Korean—are less likely to interpret others' behaviour by appealing to their mental states in comparison to individuals raised in Western cultures, such as Americans and Europeans. Individuals raised in Far East culture will be more likely to make sense of others' behaviour by appealing to the features of the environment which influenced this behaviour.

Individual differences in the content of ToM and in the way that it is used do not need to result from different cultural backgrounds. We may hypothesize that some of these differences will be observed across individuals belonging to much smaller groups—perhaps even to such a small group as people working in a particular profession. This is because the differences in question may result from the fact that individuals who work in certain professions will be required to add to their ToMs new information which will be of little relevance outside this profession. In effect, it will be unlikely that this information is included in ToMs of individuals who work in other professions. What is more, the new information which some individuals are required to add to their ToMs may be even at odds with ToMs of individuals working in other professions. These considerations are relevant to our purposes because legal professions require adding to one's ToM new information which may not only be of little relevance outside the legal domain but sometimes it may even be at odds with ToMs of non-lawyers.

6. *Mindreading and the legal interpretation and management of behaviour*

A well-known piece of information included in the ToMs of many individuals who work in a legal profession that goes beyond and is at odds with the ToMs of non-lawyers is that in certain cases people may be attributed oblique intentions. Oblique intention is a mental state which, if attributed to an individual, means that an individual intended to bring about a particular event because he recognized that this event may be one of the consequences of his action. The upshot of introducing oblique intention into law is that, from the legal point of view, an individual may be attributed an intention to bring about an event even though he did not want to bring about his event.

To illustrate in more detail what is essential to oblique intention, consider *Hyam v. DPP* (1975), a famous case in English law (Ormerod & Laird 2020, 98). Mrs. Hyam, who was jealous of her former partner's new fiancé, Mrs. Booth, poured gasoline into Mrs. Booth's letter box, ignited it and left home without warning anyone about the fire. Unfortunately, the fire spread, burned down Mrs. Booth's house and killed her two daughters. Mrs. Hyam was tried for murder. In the trial, Mrs. Hyam claimed that her intention was only to frighten Mrs. Booth and cause her to leave the town. Thus, according to Mrs. Hyam's defense, she had no intention of hurting anyone. Mrs. Hyam was convicted, however, and appealed. Eventually, the case was decided by the House of Lords where the appeal was refused. Lord Hailsham claimed that in order to attribute an intention it is sufficient that the defendant

> «knew there was a serious risk that death or serious bodily harm will ensure from his acts and he commits those acts deliberately and without lawful excuse with the intention to expose a potential victim to that risk as the result of those acts. It does not matter in such circumstances whether the defendant desires those consequences or not».

Although oblique intention is a common feature of contemporary criminal law systems, non-lawyers will be uncomfortable with the above claim by Lord Hailsham. This is because according to non-lawyers' ToMs—which they use in ordinary, everyday attributions of intention—intention requires that an individual desires a particular event to occur and believes that this event will be the consequence of his action (Malle & Nelson 2003, 573). On this

account, undesired side-effects cannot be intended. According to a non-lawyer's ToM, it makes even less sense to attribute intention for an unexpected side-effect in the case of which the individual is consciously unaware that his action will bring it about. Yet, despite the oblique intention's peculiarity from the non-legal perspective, lawyers will often intuitively attribute this mental state to an individual if the relevant facts of the case at hand are similar to the facts in *Hyam v. DPP*.

The second component of the reflexive mindreading system is mental simulation. Mental simulation consists in using one own's decision-making mechanism in the production of M-intuitions. In the case of mental simulation, this decision-making mechanism is used in an offline manner which means that the M-intuitions that it produces do not influence how the simulator behaves. Instead, these M-intuitions are decoupled and attributed to someone else or provided to the reflexive mindreading system for further processing.

In the early days of cognitive-scientific research on mindreading many authors thought that the reflexive mindreading system is driven exclusively by a ToM. However, various kinds of considerations—grounded in empirical findings and conceptual developments—suggested that this simple view is difficult to hold. To mention one of such considerations, observe that if the reflexive mindreading system were limited to using only a ToM, it would be very difficult for this system to deal with even simple mindreading problems. To be more specific, if this were the case, then many mindreading problems would be computationally quite demanding in the sense that solving them in real-time would require a lot of resources such as time, memory or processing power.

To illustrate this point, consider again *Hyam v. DPP*. It was mentioned that lawyers will often intuitively attribute oblique intention to an individual if the relevant facts of the case at hand are similar to the facts in *Hyam v. DPP*. But in *Hyam v. DPP* oblique intention attribution to Mrs. Hyam was by no means intuitive. The judges in this case carried out explicit and detailed considerations whether it is appropriate to attribute oblique intention to Mrs. Hyam. And the jury was carefully instructed to consider this issue as well. This indicates that it was the reflexive mindreading system that was responsible for the attribution of oblique intention to Mrs. Hyam. However, a closer look at the court's considerations suggests that there was a different mental state which was intuitively attributed to Mrs. Hyam: the knowledge that her actions make it highly probable that anyone who lived in Mrs. Booth's house may get seriously injured. Recall that Mrs. Hyam denied that she intended to hurt anyone. However, she also said that when she burned the letterbox she realized that what she did was dangerous to anyone living in the house and added that she thought that the house was empty. Mrs. Hyam's statement that when she committed the action she realized how dangerous its consequences may be was the ground for the intuitive attribution to her of the above mentioned knowledge that her actions make it highly probable that anyone staying in Mrs. Booth's house may get seriously injured.

What suggests that this attribution was intuitive is that it is made without any explicit consideration—the court does not offer any explanation why it is appropriate to attribute to Mrs. Hyam this knowledge apart from mentioning that she admitted being aware that what she did was dangerous to anyone living in the house. This intuitive attribution was provided as an input to the reflexive mindreading systems of the judges and the jury for further processing where it played a key role in the carefully considered attribution to Mrs. Hyam the oblique intention to kill Mrs. Booth's children.

Consider how the reflexive mindreading system would need to operate to generate the M-intuition that Mrs. Hyam knows that it is highly probable that her actions may cause serious injuries to Mrs. Booth's children if this system was driven exclusively by a ToM. If this was the case, then this system would need to make a series of inferences that ultimately resulted in this attribution. The premises in these inferences would be, on the one hand, the relevant psychological generalizations stored within the ToM—that is, generalizations that link mind

and behaviour relevant for attributing the above-mentioned type of knowledge to Mrs. Hyam—and, on the other hand, the facts of the case which relate to these generalizations. According to one such generalization, for example, if someone says that when she acted she realized that her action may harm someone else, then this person acted knowing that it is highly probable that her action may harm someone else.

A generalization of this kind was probably used in *Hyam v. DPP* in order to attribute to Mrs. Hyam that she knew about the consequences of her action. This generalization is not explicitly mentioned in the court's considerations about what Mrs. Hyam knew. Instead, it is an implicit premise in the inference from what Mrs. Hyam said to what she knew. However, taking into account the facts of the case, it is by no means obvious that this generalization is applicable. Notice that Mrs. Hyam mentioned that she thought that the house is empty. This undermines the claim that she knew that it is highly probable that her actions may cause serious injury. If she believed that the house was empty, then she could not simultaneously believe that she may harm someone inside the house. To deal with this apparent inconsistency between Mrs. Hyam's beliefs, an appeal to further generalizations is required which would exclude Mrs. Hyam's belief that the house is empty from the set of premises in the inference aiming to determine what she knew. What is more, Mrs. Hyam explicitly denied that she intended to kill or hurt anyone and she also denied that she was aware that she could kill anyone. What she admitted is only that she realized that her actions may be dangerous to those inside the house. This is further evidence that it is by no means clear that Mrs. Hyam knew that it is highly probable that her actions may cause serious injury or even kill those who were staying in fact inside the house. Even more generalizations are required to show why it is appropriate to attribute to Mrs. Hyam this knowledge.

In short, if the reflexive mindreading system was driven exclusively by ToM, then this system would need to apply numerous generalizations stored within ToM to finally attribute to Mrs. Hyam the mental state under discussion. This, however, would place a heavy burden on this system in the sense that it would need to process a lot of information before inferring Mrs. Hyam's mental state. To sum up, assuming that the reflexive mindreading system is driven exclusively by ToM it is surprising that the judges and the jury in *Hyam v. DPP* were capable of intuitively—that is, without an explicit consideration—attributing to Mrs. Hyam the knowledge that that it is highly probable that her action may seriously injure those inside the house.

According to a more plausible explanation of how this intuitive mental state attribution was made, the reflexive mindreading system did this in a much less complicated way. Namely, this system accessed and used the same information-processing resources which are used when the individuals who do the mindreading experience the kind of mental state which they attribute to Mrs. Hyam. In this way, the cognitive burden on the reflexive mindreading system is significantly diminished. This system no longer needs to compute every psychological generalization, stored into ToM, relevant to the case at hand: it just generates, in those who engage in mindreading, the pretended or imaginary mental states which are supposed to correspond to those of the target of the simulation. In other words, mental simulation allows the mindreader to realize what his own mental states would be if he were to find himself in the situation of his target. The idea behind mental simulation is that instead of using an elaborate ToM, the simulator takes a shortcut and uses her own mind as a model of the mind of the simulation's target. Thus, to attribute to Mrs. Hyam the knowledge that it is highly probable that her actions may injure or kill someone, the jury and the judges took a shortcut and solved a simpler mindreading problem: whether they themselves would be aware that it is highly probable that their actions may injure or kill anyone, if they were in Mrs. Hyam situation. The result of this mental simulation was then attributed to Mrs. Hyam.

Some researchers found it tempting to suggest that mental simulation is all there is if you want to explain how mindreading works. This view, however, is also problematic. The main difficulty here is that if the results of mental simulation are to be plausible, the simulator needs

to adjust for the differences between himself and the target of the simulation. For example, in order to simulate another individual, the simulator needs to adjust for the possible difference between how he views a particular situation and how this situation is viewed by his target. What is more, the simulator needs to adjust for the difference between what he would want to achieve in this situation and what his target wanted to achieve. In short, the simulator needs to adjust for any relevant differences in beliefs, desires, emotions and other mental states between himself and his target. These adjusted mental states serve as the input to the simulation process which generates output in the form of a pretended set of mental states. However, the adjustment in question cannot proceed without the simulator knowing which differences in mental states between himself and his target are relevant to the case at hand as well as knowing what do these differences amount to. For example, adjusting for a difference in beliefs between the simulator and his target requires that the simulator knows what his target believes. Thus, if a mental simulation is to generate plausible outcomes, it needs to be augmented with a ToM in order for the simulator to adjust for the differences between himself and his target.

Taking into account the proposal that both mental simulation and ToM are involved in intuitive mental state attributions, we may propose the following algorithm for these attributions:

(1) adjust for the differences between yourself and the target of your mental state attribution,
(2) imagine what mental states you would have if you found yourself in the situation of your target,
(3) attribute the imagined mental states to your target (STUEBER 2006, 120).

Thus, the above-mentioned claim that, in order to attribute to Mrs. Hyam the knowledge that it is highly probable that her actions may injure or kill someone, the jury and the judges imagined whether they themselves would be aware that this would be the case if they were in her situation, does not completely account for this attribution. What is missing in this description is that their reflexive mindreading systems needed to adjust for the differences between themselves and Mrs. Hyam. One such difference could be that—to give a perhaps not so hypothetical example—the discrepancy between their belief that if they were in her situation, it would be obvious to them that their actions could lead to the deaths of those living in the house and Mrs. Hyam's lack of this belief or even her belief that her actions will surely not cause anyone to die. Noticing differences of this kind requires that the reflective mindreading system applies information stored in its ToM and attributes to the target the discrepant mental states.

7. *Why and how mindreading fails*

The previous discussion concerning how intuitive mental attributions are made can be used to predict when these attributions will be incorrect. This happens when the interpreter fails to adjust for some important difference between his own mental states and the mental states of his target. In the case of such a failure, the unadjusted mental states will be provided as input to the simulation process and this process will generate simulated mental states which do not correspond to the mental states of the target person. Still, despite their lack of correspondence, these mental states will be intuitively attributed by the interpreter to his target.

Taking into account that adjusting for the difference between the interpreter and his target involves applying the information stored in a ToM, this adjustment will fail if there is information about mind and behaviour that is relevant in a particular case of mental state attribution and which is not included in this interpreter's ToM. One important type of information about mind and behaviour that may not be included in the ToM of a person who

uses mindreading for the purposes of legal interpretation and management of behaviour is information provided by scientific research.

It is perhaps not too far-fetched to assume that the two dominant sources of information about mind and behaviour stored in this person's ToM are ordinary or folk psychology and its legal counterpart. The former consists of numerous psychological generalizations recognized by most people, such as the principle according to which a person has beliefs that may be mistaken, desires which motivate her to act or the principle that if a person says that when she carried out a particular action she realized that the consequences of this action may be dangerous to others, then she acted with a belief that the consequences of her action may be dangerous to others. These folk psychological generalizations are used in everyday situations as well as for the purposes of legal interpretation and management of behaviour. On the other hand, the legal counterpart of folk psychology consists in generalizations designed specifically for legal purposes. Many of these legal-psychological generalizations are in essence elaborations of posits of folk psychology such as the legal-psychological generalization that strong emotional disturbance inhibits control over action. Other legal-psychological generalizations appear to be refinements or developments of folk psychology such as the legal distinction between premediated and unpremeditated actions. In some cases, however, generalizations for mental state attribution of a legal nature may be at odds with what folk suggests as it is the case with oblique intention.

A characteristic feature of both folk psychology and its legal counterpart is that for the most part their principles are fixed and developed through linguistic considerations about how words associated with mind and behaviour ought to be used. That is, because in ordinary cases it makes sense to say that if someone admitted that he realized what will result from his actions, then he knew what will result from his actions—and it makes little sense to deny him this knowledge unless there are some circumstances which undermine his admission—, we think that we are allowed to attribute to him this knowledge. And because for the purposes of legal interpretation and management of behaviour it makes sense to say that if someone knew that it is highly probable that his action will result in someone else's injury, then he intended this injury, we think that we are allowed to attribute to him this intention.

What scientific research shows, however, is that there is a lot of information about how mind and behaviour are related which will go unnoticed even for those who are most careful in their linguistic considerations of these matters. As it was mentioned, a lot of this research focuses directly on how cognition works instead of studying people's competence with ordinary words associated with psychology. In the case of mindreading, a lot of this research is focused on the features of the reflexive mindreading system which is not really available for investigation for someone equipped only with linguistic competence. To support this claim with an example, consider the psychological research on memory—in particular, the research on misinformation effect.

Misinformation effect occurs when a person's memory of an event is distorted due to information that she acquired after this event occurred. Much of the research on this phenomenon was concerned with how false or misleading information acquired after experiencing an event influenced eyewitness testimony. In one of such experiments, participants were shown a film of a car accident (Loftus 1979). Afterwards, the participants were asked to describe what they have seen in the film. One group of participants was asked the question «About how fast were the cars going when they smashed into each other?». Another group was asked the question «About how fast were the cars going when they hit each other?». Subjects in the first group estimated that the car was moving faster than subjects in the second group. What is more, one week after seeing the film both groups were asked the following question about the film: "Did you see any broken glass?"—even though there was no broken glass in the film. Still, there were participants in both groups which responded affirmatively to this question and participants in the first group were

much more likely to provide this answer. As we can see, participants' memories were influenced by how the questions were formulated and what they were about.

Empirical findings pertaining to the misinformation effect were surprising because memory was not considered to be malleable to information acquired after the event—or at least it was not considered to be malleable in this way and to this extent. However, we can explain the surprising nature of these findings by appealing to the way in which our mindreading works—in particular, to the operation of the reflective mindreading system. Observe that deciding whether someone's testimony is reliable is an exercise in mental state attribution. This decision consists in attributing to the witness a set of beliefs about what he experienced in the past—a set of beliefs that correspond to his testimony—and determining whether these beliefs are true or false. Judging this last issue involves appealing to various generalizations stored in one's ToM about how memory works. One such generalization, which is probably shared by most, is that recent events tend to be better remembered than events that occurred in the distant past. Thus, if one is to judge whether a particular testimony is reliable, an issue under consideration will be how recent was the event that the testimony is about. If the event occurred in the distant past, we will be more willing to question the witness' reliability in comparison to cases in which the event was recent. Perhaps this willingness may even be observed at the level of the reflexive mindreading system. This means that intuitively—that is, without explicitly considering this issue—we will be more willing to judge a witness' reliability in accordance with the generalization that recent events tend to be better remembered than events that occurred in the distant past. However, even if this is will not always be the case—for example, because this generalization is not stored in the ToM which is accessible to the reflexive mindreading system of a particular person—it may be possible to apply this generalization via the reflective mindreading system and assess whether the intuitive attribution is correct.

Misinformation effect proved to be surprising which illustrates that the psychological generalization according to which a person's memory of an event may be distorted due to information that she acquired after this event occurred was not stored in the ToM's of most mindreaders. This concerns ToMs which were applied both intuitively and reflexively. The observation that before the empirically oriented research on memory revealed its shortcomings the legal systems relied too much on eyewitness memory shows that this generalization was not recognized by the law as well. At the present moment, however, this state of affairs appears to change and knowledge about how post-event interventions influence memory formation is more commonly applied in legal settings. We may expect that future empirical research on mind and behaviour will result in discoveries of more generalizations which will need to be included in the ToMs of those who mindread for the purposes of legal interpretation and management of behaviour.

Further readings on mindreading in law:

BROWN T. R. 2022. *Demystifying Mindreading for the Law*, in «Wisconsin Law Review», 1, 1 ff.

GREELY H.T. 2013. *Mind Reading, Neuroscience, and the Law*, in MORSE S.J., ROSKIES A.L. (eds.), *A Primer on Criminal Law and Neuroscience. A Contribution to the Law and Neuroscience Project, Supported by the MacArthur Foundation*, Oxford University Press.

GREGORY D. 2019. *Judging the Mental States of Others: 'Mindreading' in Legal Decision-Making*, in «Jurisprudence», 11, 1, 48 ff.

References

Abel F., Happe F., Frith U. 2000. *Do Triangles Play Tricks? Attribution of Mental States to Animated Shapes in Normal and Abnormal Development*, in «Cognitive Development», 15, 1 ff.

Bermúdez J.L. 2018, *Cognitive Science. An Introduction to the Science of the Mind*, Cambridge University Press.

Castelli F., Frith C., Happé F., Frith U. 2002. *Autism, Asperger Syndrome and Brain Mechanisms for the Attribution of Mental States to Animated Shapes*, in «Brain», 125, 8, 1839 ff.

Carruthers P. 2006. *The Architecture of the Mind. Massive Modularity and the Flexibility of Thought*, Oxford University Press.

Fletcher P. C., Happe F., Frith U., Baker S. C., Dolan R. J., Frackowiak R. S., Frith C. D. 1995. *Other Minds in the Brain: A Functional Imaging Study of "Theory of Mind" in Story Comprehension*, in «Cognition», 57, 109 ff.

Goldman A. 2006. *Simulating Minds. The Philosophy, Psychology, and Neuroscience of Mindreading*, Oxford University Press.

Hutto D. 2008. *Folk Psychological Narratives. The Sociocultural Basis of Understanding Reasons*, MIT Press.

Lieberman M. 2013. *Social. Why Our Brains Are Wired to Connect*, Oxford University Press.

Loftus E. 1979. *Eyewitness Testimony*, Cambridge University Press.

Malle B. 2004. *How the Mind Explains Behaviour. Folk Explanations, Meaning and Social Interaction*, MIT Press.

Malle B., Nelson S. 2003. *Judging* Mens Rea: *The Tension Between Folk Concepts and Legal Concepts of Intentionality*, in «Behavioral Sciences and the Law», 21, 563 ff.

Ormerod D., Laird K. 2020. *Smith, Hogan, & Ormerod's Text, Cases, & Materials on Criminal Law*, Oxford University Press.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Eickhoff, S. B. 2011. *Modelling Neural Correlates of Working Memory: A Coordinate-Based Meta-Analysis*, in «NeuroImage», 60, 830 ff.

Saxe R., Powell L., 2006. *It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind*, in «Psychological Science», 17, 692 ff.

Spaulding S., 2018. *How We Understand Others. Philosophy and Social Cognition*, Routledge.

Stich S., Nichols S. 2003. *Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford University Press.

Stueber K. 2006. *Rediscovering Empathy. Agency, Folk Psychology and the Human Sciences*, MIT Press.

Turkeltaub P. E., Gareau L., Flowers D. L., Zeffiro T. A., Eden G. F. 2003. *Development of Neural Mechanisms for Reading*, in «Nature Neuroscience», 6, 7, 767 ff.

Thagard P. 2005. *Mind. Introduction to Cognitive Science*, MIT Press.

PART III.

The Nature of Law and Normative Phenomena

Origins of Human Normativity

PHILIPPE ROCHAT, NIKITA AGARWAL

1. *Origins of human normativity*

What do we understand by normativity and what are the origins of norms in human development? As developmental psychologists, those are the questions we want to address in this chapter. Our goal is to show that standards of behavior and expectations are deeply rooted in development. Standards and expectations find their roots in infancy, at an implicit level. They develop to become explicit, re-described with the emergence of language, self-consciousness, the sense of reputation and theory of mind. We want to show also that the explicit re-description of implicit standards and expectations in development operates in parallel with the increase in moral autonomy (PIAGET 1932; KOHLBERG 1981) and the emergence of an ethical stance manifested from around 3-5 years of age (ROBBINS & ROCHAT 2011). By the time children start socializing in schools, they begin to internalize, enforce, follow, but also protest rules and norms. From this point on, children become explicitly "moral" or normative in relation to standards that are shared with others. What is particularly interesting is the fact that with this explicit re-description, new concepts emerge with complex shared values attached to them. We try to show that these concepts are just veneer extensions or explicit re-descriptions of primordial motives guiding behavior from infancy. They are expressed by newborns at an implicit level, including the notions of "trust", "promise", as well as "obligation". We try to make the case that these concepts find their roots in the implicit motives that guide much of newborns' behavior and cognition.

The plan of the chapter is as follows. We first try to define what we understand by normativity. We then review progress of the past fifty years in infant studies, focusing on the way early competencies were discovered by researchers using new and clever experimental paradigms that revolutionized the field of developmental psychology. We briefly discuss this scientific leap showing that it was made possible by tapping into the natural propensity of infants, who from birth engage in renewed exploration when something deviates from what they expect, what they putatively perceive and memorize as "standards". From the outset, there is indeed a natural inclination of the mind to detect sameness and deviation from it in relation to standards. We propose that this inclination may be the natural roots of human normativity, its "cognitive cradle". For the rest of the chapter, we discuss how early standards of behavior develop to become eventually explicit with the emergence of language and how rules and norms may start to be spontaneously enforced as well as a source of protests in childhood, for better (good trouble) or for worse (bad trouble).

2. *What is normativity?*

In the most generic sense, a norm is a standard that is used to judge and evaluate. It refers to a benchmark representation that can be explicitly codified as in a book of laws (thou shall not kill),

but also implicitly expressed in gut feeling reactions like the disgust and rejection of spoiled, poisonous food. Being normative, thus, is expressing such standards in our evaluation and judgement of things, be they implicit or explicit. Accordingly, *normativity* is the phenomenon that surrounds any reference to standards in the evaluation of actions and outcomes as good, permissible, laudable; or inversely as bad, intolerable, condemnable, etc and human morality is centered around adhering to the former and avoiding the latter. Aside from being either implicit or explicit, standards of judgement and evaluation are of different kinds. They may be cultural (e.g. eating fish on Friday); logical (e.g., A is larger than C if it is also larger than B and B is larger than C); societal (e.g. all have to pay taxes); or legal (first degree vs. second degree homicide). Thus norms may vary and may be arbitrary. They may be considered as more or less rigid and changeable, more or less viewed as set in stone, hence more or less debatable. Thus, from a psychological viewpoint, normativity also refers to how one lives and abide to norms, question and feel the obligation to follow norms. Inversely, it also refers to how one may be judged and evaluated by others based on what they see as obligations, promises, and expectations. Being normative is having expectations and knowing that others may expect us to behave and think in a certain way. Likewise, being morally good is living up to these expectations.

As mentioned above, normativity may be implicit or explicit, expressed in uncontrollable approach/avoidance gut feeling reactions, or cogently articulated and reasoned as in a political argument or in the defense of a lawyer in front of a jury. Normativity covers all of these levels of "standard" awareness, against which value judgments are expressed either verbally or implicitly enacted in approach/avoidance behaviors. Looking at normativity in development illustrates and helps us to sort out the multi-layer aspect of what it means to be normative and abide to standards.

3. *From embodied to abstract normativity*

It can be said that normativity is pervasive at all levels of nature. Norms and standards are expressed everywhere, across all that is living, as any organism survives by maintaining homeostasis or stable balance within the internal and in relation to the external environment. At all levels of the living, there are set points or "standards" against which balance is maintained. It also seems kind of costly and redundant to invent new rules and standards for the millions of transactions and encounters we make everyday at an individual level. In that sense, the set benchmark and rule serves as quick heuristic to get by. Take the regulation of hunger or thirst for example. Deviation from a given homeostatic standard triggers complex motivated behavior triggering goal-oriented behavior in the organism in search of food or liquid. Drinking and eating brings back the system to a standard point of equilibrium in the organism, analog to a thermostat regulating ambient temperature.

However, there is obviously a fundamental difference between physiological, embodied normativity and cognitive normativity, i.e., the normativity that shapes our decisions (implicitly or explicitly) and that we may either enforce or protest. The latter is incommensurably more open and variable depending on the age of individuals and their cultural circumstances, in general depending on experience. True that the need to drink or eat also varies with age (metabolism) and culture (food habits), but not to the extent that an individual or groups of individuals may abide or not abide to rules and normative laws of their culture. Basic functional analogy aside, there is clearly a fundamental difference between the two. The developmental question is then: how does cognitive normativity emerge in human ontogeny, on top or in parallel to the physiological, embodied normativity (homeostasis) that allows any living organisms to survive. That is the question of we want to address, i.e., "the origins of human normativity".

4. *Sameness detection*

It is now well established that newborn infants, even fetuses are not just reflex machines functioning as close-loop systems that would be comparable to thermostats. Even before birth, at least by the third trimester of gestation, fetuses learn and memorize (DeCasper & Fifer 1980). They habituate to repeated stimulations, showing for example significantly less startling responses to a repeated loud sound. More importantly, they tend to recover attention as indexed by renewed startling responses when the sound is suddenly novel, with a different pitch or frequency for example (Lecanuet et al. 1988).

This simple, highly reliable observation found in the context of audition, but also smell or repeated tactile (vibro-acoustic) stimulation, indicates that fetuses are de facto already learning to discriminate the old (same) from the new (different). They somehow notice the difference of the novel sound as an event that deviates from what they learned before, via habituation. It is novel by virtue of the fact that it implicitly violates a learned standard (loud sound of a particular pitch). It also demonstrates that already prenatally, we are capable of building implicit expectations about what is of the same. Inversely, from infancy onward, we are programmed to notice what *deviates* from the standard we learned. Such detection of a deviation from learned standard is a fundamental law of learning and memory, from the most elementary (habituation, conditioning), to the most complex (mental inference, deductive and logical reasoning). At the most basic level, they all entail some comparison with a standard that has somehow been stored in memory.

As Willam James writes, "Sameness detection is the backbone of the mind". Inversely, such detection necessarily also entails the ability to detect what deviates from what would be normally expected (i.e., sameness). The rudimentary learning process of habituation and dishabituation that is expressed in most, if not all living creatures, including human fetuses, demonstrates a natural inclination to compare against an implicitly memorized standard. Without such comparison, no *learning* could take place. However, and this is crucial, what changes in the perspective of both phylogeny and ontogeny, is the nature of the standard, namely the level of its representation by the organism. This representation may be just sensory in the beginning. However, with development, it becomes perceptual and entails much higher cognitive processes, such as those contained in the practice of law, protests, explicit rule enforcement. We are interested here to capture the different levels of standard representation (i.e. levels of normativity) expressed in the development of children from birth to approximately 8 years of age.

Sameness detection is an active process from birth, not just taking place in passive contemplation of the surrounding world. Infants from birth show a propensity to actively imitate others (Meltzoff & Moore 1977), presumably not just by simple contagion as in the case of yawning for example, but with what is documented as a deliberate attempt at reproducing sameness of behavior, others used as the standard they try to mirror. Neonatal imitation is an embodied primary precursor of the explicit manifestation of conformity that is well documented as emerging by 3-5 years of age as part of what we will describe here as self-consciousness and *tertiary normativity* (see below).

5. *Primary (embodied) normativity*

Infants are born surrounded by implicit and explicit values. These values are held by their caretakers in terms of good parenting and child care, values and beliefs that may greatly vary across cultures.

Aside from values associated with the particular culture of their surroundings, newborns are innately driven by values that are part of their evolved, genetically determined biological make-

up. From birth, they manifest highly predictable approach or avoidance toward particular stimulation as well as complex features of the environment. All newborns react negatively, with aversive emotional expressions (disgust) when dispensed with a drop of salty water on their tongue. Likewise, they manifest avoidant head turns and negative facial expressions when a cotton swab impregnated with citrus scent is brought close to their nostrils (ROSENSTEIN, OSTER 1988). Inversely, they engage in sucking with a relax expression when a drop of water with 5% sucrose is poured on their tongue or when they smell the sweet odor of their mother's milk. Animal models demonstrate that sweet taste and sweet smell are associated with the triggering of the brain's endorphin/opioid system that has analgesic power associated with a powerful experience of pleasure. As a case in point, circumcised newborns are documented to cry significantly less during circumcision if given drops of sweet water prior to surgery. Specific sensory and experiential values drive infants from birth within a simple approach-avoidance polarity of action. Infants are thus born biased and oriented toward certain qualities of experience in their encounter with the environment outside of the womb. They are born attracted to some aspects of the environment and avoidant of others. This approach avoidant polarity is not expressed only in relation to proximal sensory stimulations like touch or taste, but also in relation to complex perceptual features like face or eyes. Newborns are shown to pay particular attention and track more canonical face-like displays, and significantly less to faces with scrambled eyes, nose, and mouth. Immediately after birth, they prefer to look at a face looking straight in the eyes than avoiding their gaze, sensitive to pupil to pupil contacts (FARRONI et al. 2002). All these sensory and perceptual inclinations on display immediately after birth demonstrate that infants are not born as a "tabula rasa" in need of experience to learn values. We are born with values that are built-in the organism, expressed at birth and even prior, while still in the confine of pregnancy. These values are part of newborns' inherited preparedness evolved by the species. They motivate, orient, and jumpstart early development. They also constitute a primordial *or primary normativity*, an ensemble of motives and standards that all animals must possess in order to survive.

6. *Secondary (experience-based) normativity*

Within the first 6-9 months, infants quickly enrich their embodied primary normativity they inherit from birth with the elaboration of new, experience-based standards, here referred to as secondary normativity. Beyond birth, infants learn to discriminate and group things encountered in the environment based not only on their direct surface resemblance or direct experiences they might trigger, but also based on indirect, more abstract inferences such as whether it is more or less familiar or unfamiliar, pro-social or anti-social, intentional or accidental. They start to infer increasingly abstract non-obvious characteristics with the elaboration of new implicit standards helping them to sort out and chunk into distinct categories the zillions of new objects and things they keep encountering in the environment. In the language domain, for example, it is now well established that up to 7 months, infants are capable of discriminating and learning to discriminate almost any phonemes of any spoken languages (MAURER & WERKER 2014). However, by the end of the first year, infants are shown to lose such capacity as they learn to discriminate only among the limited class of phonemes that are relevant to the language of their culture. By the end of the first year, for example, Japanese infants, become deaf to the difference between phonemes like /RA/ and /LA/, as for their native Japanese speaker parents. They do detect the difference early in the first year. An analog of such perceptual narrowing during the first year is also documented in relation to faces and ethnicity. Young infants are first capable of discriminate any faces, even the faces of other species (monkeys). By the end of the first year, however, they narrowed their ability to faces of their own ethnicity. Implicit perceptual standards have changed and infants'

grouping of speech sounds as well as faces is modified based on exposure. Primary normativity is enriched by experience-based categorization, a secondary kind of normativity that blossoms in the first months after birth.

By being categorical, shifting their grouping criteria (standards of same/different), infants express implicit normativity, yet a normativity that goes above and beyond the primordial, embodied normativity that infants are born with and that jumpstarts post-natal development with a set of pre-determined and motivated action systems (feeding system, protective system, temperature regulation system). Early categorization is an experience-dependent learning of new standards and norms, what we label *secondary normativity*.

7. *Tertiary (symbolic and self-conscious) normativity*

Symbolic functioning and language acquisition are the cardinal developmental threshold separating human infancy from childhood. By becoming symbolic children also become self-conscious in the sense that they start representing how they are perceived and evaluated by others. With self-consciousness children start to contemplate and assess their own value in the mind of others. They begin to care about their own *reputation*, literally calculating (from the Latin verb *putare*) how they present themselves to others in public and to themselves privately. As they acquire language and become symbolic with the work of their own imagination, they also develop what we can coin a *reputable sense of self* or *identity*.

In this process, the standards against which children start to measure themselves become essentially subjective, leaving room for much delusion regarding how they are being perceived, and how they perceive others as evaluators of them. Standards and norms become symbolic. They stand for tertiary constructs like obligation, promise, or trust (see below). The normativity of the child becomes also moral. In the same way that experience-based categorical *secondary* normativity enriched the implicit *primary* normativity of newborns by developing new standards of grouping and same/different implicit judgments, self-conscious *tertiary* normativity enriches the latter by developing more abstract and conceptual standards of judgments that with the emergence of language become explicit. Standards and norms are now explicitly stated and enforced by authorities inside and outside the family environment (e.g., school). Entering school in particular, or any extrafamilial group activities, children find themselves surrounded by explicit new rules and norms they memorize and treat as standards, avoiding sanctions from immanent adult authority (PIAGET 1932; KOHLBERG 1981). Eventually, children will realize that all rules are not set in stone and that there is always a relative "arbitrary-ness" attached to rules that may be revised or rejected, even protested in open negotiation with others. As the child's mind grows to become symbolic, tertiary normativity emerges, opening up standards and norms to politics, political judgments, jurisprudence, and other highly codified and abstract moral assessments. It opens up the potential for questioning standards, eventually protest and decisions to engage in "good" or "bad" troubles with the establishment (see § 5).

8. *Obligation, promise and trust*

As already mentioned, with tertiary normativity, standards and norms become highly abstract and subjective. They are more than just based on direct perceptual inference as in the case of early speech sound or face categorization (see above experienced based categorization and secondary normativity). Tertiary standards and norms, although potentially expressed at a "gut" level as in well documented racial or gender stereotypes and other implicit biases, they are mainly explicit, symbolically represented in language and in our minds. That is why the main

characteristic of tertiary standards is that they are negotiable in exchange and in collaboration with others. They are negotiable via debate and politics as in any parliament or any court of law legislating obligation, promise, and trust of the individual toward society—and inversely—of society toward the individual.

With the emergence of self-consciousness and symbolic *tertiary* normativity, as they cross the symbolic threshold in their development, children start in synchrony to express new feeling experiences such as the emotions of guilt, shame, pride, hubris or embarrassment. Those correspond to what is described as "self-conscious" emotions, all typically starting to be expressed from the middle of the second year as children begin to recognize themselves in mirrors and start using personal pronouns and adjectives like, I, me, and especially "mine!" to assert possession and control over things (this is mine, hence not yours...). With this, they manifest a new conceptual level of self-awareness which is less embodied and more abstract, an explicit self-concept that extends to possessions (ROCHAT 2014; ROCHAT 2018).

With self-concept and self-conscious emotions, the relation of the child with others is re-described to allow for new levels of collaboration and exchanges. Children will start to cooperate with others in accomplishing shared goals that imply shared intentionality in order to accomplish such goals (TOMASELLO & CARPENTER 2007). They begin to have an explicit understanding of rules that are shared with others and they may begin to protest when the rule is broken. In fact, young children (from 2 or 3 years of age) not only follow norms and rules in their actions but also enforce these same norms on others by spontaneously and normatively sanctioning mistakes through third-party protest, critique, and teaching in response to norm transgressions (RAKOCZY et al. 2008; RAKOCZY & SCHMIDT 2012; SCHMIDT et al. 2012; RAKOCZY et al. 2009; WYMAN et al. 2009). They correct others if they violate an agreement all *should* abide to. We may say that from this point on (2-3 years), children become moral proper, starting their moral career in relation to standards and norms that are now explicit, and enforceable, the ground for judgments and moral reasoning. From this point on, in the context of new collaborations children develop an explicit sense of obligation, promise, and trust: the triumvirate abstract constructs of human morality.

These constructs depend on both language and self-consciousness emerging by the middle of the second year. They are symbolic enrichment of primary and secondary normativity developing in infancy. Obligation corresponds to what one implicitly and explicitly "ought to do". Promise corresponds to what one is "expected to do". Trust is what one can count on others to do, accordingly. This triumvirate of moral constructs correspond to the foundation of further highly abstract moral rules and explicit social norms that children develop primarily by collaborating with others as they play, learn to share, cooperate, and engage in group learning with adults and peers (school).

9. *Equity, fairness, and collaboration*

Natural observations of family life demonstrate that the great majority of conflicts among siblings surrounds issues of possession and sharing ("Why did you get a larger piece of cake?" "No, this is my toy!" etc.). It is typical for parents and adults to intercede and impose their rule and rationale for justice distribution, not unlike judges in a court of law. From 2-3 years of age, children are prompt to detect and explicitly complain about what they perceive as an unfair distribution of resources or being unfairly treated. Although they may not be able yet to articulate it, the spontaneous, primary detection of what is unjust indicates that already by 2 years, the child has some implicit notions of fairness and how resources should be distributed. The question, however, is what may constitute such abstract notion and where does it originate from? Is it adult pressure and interventions, or is it something that may be more instinctive and

innate? It makes sense to think that the sense of possession and unfairness is a pre-requisite for adults to intervene and apply pressure on the child in order to resolve conflicts. It is after the fact that parents may become explicit regarding fairness principles. Thus conflicts necessarily precede the fairness rationale expressed by adults in their intervention. So, what drive young children to engage in possession conflicts and to be so prompt at detecting what they experience as "unfair"? What kind of implicit sense of fairness hides behind such prompt detection? Asking these questions gets to the heart of what might be the original source of the moral normativity surrounding the fundamental notion of equity.

From around 3 years of age, as children become more fluent speakers, they start also to engage in more tit-for-tat bartering and sharing of objects for which they claim ownership. At this early age, while they are prompt to detect and complain about unfair distribution, when they themselves are asked to share, they tend systematically to give more to themselves. Across all cultures, 3 year-olds show a strong propensity toward self-maximizing (ROCHAT et al. 2009). They create the state of unfairness yet, they are so prompt to detect when they are on the short end of a distribution ("why did she get a larger piece of the cake?").

By 5 years, children start to take a more equitable and ethical stance as both recipients and actors of resource distribution. Compared to 3 year-olds, they are more coherent and more vocal in the manifestation of an *inequity aversion*. They take an explicit ethical stance, for example, by refusing to distribute unequal quantities of food or stickers to third party protagonists, be they peers or puppet dolls. They are even willing to sacrifice some of their own resources to assert principles of equity and fairness. They begin to engage in costly punishment. Furthermore, and more telling of a genuinely moral normativity behind their expression of inequity aversion, from 7-8 years of age children start to express an equal aversion toward inequity even when it is advantageous to them. They choose to distribute equally between themselves and another child although they are presented with the option of distributing more to themselves and less to the other (advantageous inequity) and reject distributions that are advantageous to them because it is unfair to the other (BLAKE & MCAULIFFE 2011).

In short, it appears that the sense of inequity is deeply rooted in child development, expressed as children develop a conceptual sense of themselves (self-consciousness) and an explicit sense of what they own (possessions). The development of inequity aversion between 3 and 8 years of age is transcultural. It might also be a universal foundation of human moral normativity from which the triumvirate feelings of obligation, promise, and trust may derive. From an instinctive (innate) aversion to unequal distribution, de facto a deviance from the ingrained detection of "sameness" that infants express from birth (see above), children may develop an explicit sense of what one is naturally owed (obligation), what one should naturally expect (promise), and what one can naturally count on (trust). However, such development takes place within the larger context of growing joint actions of the child with others.

There is indeed a necessary precursor to the act of sharing and the expression of inequity aversion. This precursor is the drive to perform with others; to collaborate and engage together in joint actions, the drive to affiliate. It is also the drive to share resources, share rules in a game, to join force in order to achieve a goal that could not be reached alone, to engage in reciprocal barter exchanges with others, etc. It is in the context of developing joint actions that inequity aversion may find its roots and become solidified in its expression, that inequity will be felt if one for example gets same rewards for less efforts. Free loaders may be detected and a basic sense of injustice may be first naturally felt and reasoned by the child in the context of developing collaboration and joint actions with others (TOMASELLO 2016).

By engaging in joint actions, trying to solve problems with others or playing according to rules, the child is increasingly co-conscious of shared goals and intentions. It is in this collaborative context that from 2-3 years children would naturally derive a sense of mutual obligation, hence also a sense of mutual promise and trust.

The development of collaboration and co-consciousness may thus be a primary terrain for the growth of both implicit and explicit moral normativity, the main soil for the development of the mental constructs of the moral triumvirate that are obligation, promise, and trust. It may be the main soil for the growth of what one ought to do, is expected to do, and can count on others of doing. It is also within this collaborative context that the growth of inequity aversion, both advantageous and disadvantageous, may find their roots.

Beyond 5 years, children begin to manifest increasing autonomy in their moral judgments, not simply abiding to the rules that are dictated by the authority of an adult or a majority. They develop to stand on their own moral principles and defend their own moral values, what they consider should be standards of obligation, promise, and trust. They begin to understand the relativity and arbitrary dimension of rules and norms, an understanding that may invite them to engage in good or bad troubles with the establishment.

We now turn to this development, starting from around 5 years of age but continuing all through the lifespan as we judge, defend values as well as the interests of our own community, making daily ethical decisions and navigating the politics of our social worlds.

10. *Good and bad troubles in children*

Explicit normativity help children adapt to the local contexts they are placed in, including their families, schools and culture. They come to internalize and acquire norms, and enforce them on others while adjusting the domain of applicability—some rules and norms that apply in school don't apply at home and vice-versa, one cannot wear swimwear to a funeral or harming someone is bad everywhere (Turiel 1983; Nucci 2001; Smetana et al. 2012; 2018; Jambon & Smetana 2019). Even though normativity finds its roots in collaboration and development of co-consciousness, children do not perceive rules and norms to be a process of co-creation until 5-6 years of age. They see them set in stone and transgressions of any kind to be met with sanctions and punishment. So strict are their principles and expectations of normativity from themselves and others that they rigidly apply norms on others and follow them at the cost of their own preferences and desires (Bernard et al. 2015; Li et al. 2021). However, at 5, like we discussed above, there emerges an autonomous morality and a nuanced understanding of rules and norms—they are seen as co-created by people and hence, both arbitrary and flexible. When a couple of 5 year olds are put together to formulate rules for a game, they negotiate, deliberate upon and cooperate in the task and when they are asked to teach these rules to novices, children use normative language (should, ought to) to express the rules (Göckeritz et al. 2014; Hardecker et al. 2016). This tells us two crucial things—that children understand rules are made upon consent and hence are changeable yet, once formulated they are normatively binding.

This shift in the kind of normativity—from strict rule following and avoiding sanctions to an autonomous ethical stance that children come to take at around 5-6 years of age, is cardinal to the development of normativity and normative reasoning, the latter bringing with it the emergence of *good trouble* in children. Good Trouble is simply challenging the status quo i.e. existing rules and norms if they are unfair. In short, it is making trouble for good, something better. Children at 3 years of age do question unfairness, however, it is only when they receive a smaller share. They find themselves surrounded by rules and norms, dictated by adults and hence, do not deem it necessary to make trouble to change the status quo, or find themselves incapable of doing so. This begins to change at about 5-6 years of age, when children have shown to make a sacrifice (give up on their own valued objects) to punish another who has violated fairness norms. In addition to that, children also express a preference to restore stolen items to their rightful owners when asked between punishing the thief or restoring the objects to the victim (Yang et al. 2021).

We are now starting to probe what may trigger the perception and evaluation of *good trouble* in children (AGARWAL & ROCHAT, in preparation), Because children may begin to conceive rules as not set in stone, they may also start showing appreciation for someone who challenges an unfair rule. In other words, they may start to value those protesting rules that are arbitrary and unfair, beginning to value good as opposed to bad trouble.

In an on-going study, we present children with stories based on a fictional town occupied by two groups of people, both depicted as strict rule abiders. In a series of different vignettes, children are told that the town's rules vary. For example, one of the rules is that one group gets less food than the other. Alternatively, the rule may be that both should get the same amount of food (egalitarian control condition). There are thus instances (experimental condition) where the ruling brings one of the 2 groups at a blatant disadvantage regarding privileges and resource distribution. Following each of the various vignettes, the child is then told that a member of the disadvantaged group is protesting, expressing strong *disagreement* with the established rule. Likewise, is also told that a protagonist from the advantaged group is expressing strong *agreement* with the rule, thus countering the protester of the other group by insisting that the rule must be followed because it is the rule. The child is then asked to evaluate each of the two (protesting vs. rule abiding) protagonists. Their evaluation is used as an index of their relative value of good vs. bad trouble. Preliminary findings confirm that there is a significant age-related shift from around 6 years of age, a majority of children starting to value *good* as opposed to *bad* trouble. They tend to evaluate more positively those who are ready to question authority and oppose arbitrary rules that they judge unfair.

11. *Summary and conclusion: From implicit to explicit normativity in children*

Human social life rests upon an edifice of rules and norms. Based on common language, a norm refers to a standard or a benchmark for comparison. From this generic definition, we reasoned that at the root of normativity lies the idea of expectations—what we think we and others *ought* to do in a given situation, in other words an obligatory force to follow norms and rules. With that in mind, norms and normativity beg the question of their developmental origins. In this chapter, we tried to address the question of when and how do children come to acquire norms and shared behavioral standards. What is the authority behind a norm and what may be the source of the general consent that makes it become a benchmark of *standard* for social and other comparisons?

As a first step, we considered how normativity, as defined, may manifest itself *implicitly* early in life, at birth or even before. Based on established empirical evidence, we proposed that implicit norms and reasoning around norms (implicit normative reasoning) is an early fact of life, for humans but also for any creatures that are capable of memory and learning.

From birth and even prior to birth, we showed that infants and fetuses are capable of memorizing standards and to discriminate at an implicit level between familiar (standard) and unfamiliar (non-standard) perceptual events. A substantial amount of evidence from experiments show that long before they speak, children show expectations for familiar events and react to those that deviate from the familiar standard. Studies with infants demonstrate that from birth, they react with regained attention and surprise to things that are unfamiliar relative to what they store as representation of past memorized experiences. Thus, from the outset, we tried to show that they are rudiments of norms and normative reasoning, at least at an implicit (i.e., non linguistic) level of abstraction.

In general, implicit norms early in life are primarily dictated by evolved built-in mechanisms and action systems babies are endowed with from birth with attached to them an implicit grammar of *approach and avoidant values*. We insisted that babies are indeed born in a world of implicit values. These values are attached to their evolved preparedness to act and respond to

the world in order to survive. We discuss the implicit norms and standards expressed by newborns and how they rapidly develop in the course of the first 18 months of life (from *primary* embodied to *secondary* experience-based normativity), prior to linguistic fluency and self-consciousness (*tertiary* normativity).

The second step in constructing a developmental model of norms and normativity is to describe major changes occurring from birth and in particular from the middle of the second year (18 months) as children, in parallel to their acquisition of symbolic and syntactic language, manifestation of an explicit care for their own reputation and self-concept (the emergence of social emotions like pride, hubris, or guilt). From this point on, we tried to show that children express more than the internalization of implicit standards. They do become explicit about what is right or wrong, just or unjust, correct or incorrect in reference to standards they start to articulate in both implicit and explicit communication with others. We reviewed research showing that from 3 years children rigidly apply norms in social interactions and become explicit in protesting when someone does not abide to an agreed upon rules or deviate from them.

We further discussed that between 3 and 5 years, normative thinking and reasoning undergoes changes—from avoiding sanctions to gaining increasingly autonomous moral standards, starting to take an ethical stance even if it is at a personal cost. We discussed this transition toward children's progressive moral autonomy in their judgments and abiding to norms that lead them eventually to value "good" as opposed to "bad" troubles in the face of either unfair or fair rules and norms. This last step, linked to the growth of moral autonomy in children's reasoning about values. It opens up a whole new realm of exchanges and engagement with others, consensus building and negotiation around flexible rules that in many ways are analogous to the fundamentals of politics, including jusisprudence and adult legal reasoning.

References

AGARWAL & ROCHAT, in preparation. *Emerging Sense of 'Good Trouble' in Children.*

BERNARD S., CLÉMENT F., KAUFMANN L. 2015. *Rules Trump Desires in Preschoolers' Predictions of Group Behavior*, in «Social Development», 25, 2, 453 ff. Available on: https://doi.org/10.1111/sode.12150.

BLAKE P.R., MCAULIFFE K. 2011. *"I Had So Much It Didn't Seem Fair": Eight-Year-Olds Reject Two Forms of Inequity*, in «Cognition», 120, 2, 2011, 215 ff. Available on: https://doi.org/10.1016/j.cognition.2011.04.006.

DECASPER A.J., FIFER W.P. 1980. *Of Human Bonding: Newborns Prefer Their Mother's Voices*, in «Science», 208, 1174 ff.

FARRONI T., CSIBRA G., SIMION F., JOHNSON M. H. 2002. *Eye Contact Detection in Humans from Birth*, in «Proceedings of the National Academy of Sciences», 99, 14, 9602-9605.

GÖCKERITZ S., SCHMIDT M.F., TOMASELLO M. 2014. *Young Children's Creation and Transmission of Social Norms*, in «Cognitive Development», 30, 81 ff. Available on: https://doi.org/10.1016/j.cogdev.2014.01.003.

JAMBON M., SMETANA G. 2019. *Socialization of Moral Judgments and Reasoning*, in LAIBLE D.J, CARLO G., PADILLA-WALKER L.M. (eds.), *The Oxford Handbook of Parenting and Moral Development*, Oxford University Press, 374 ff.

HARDECKER S., SCHMIDT M.F.H., TOMASELLO M. 2016. *Children's Developing Understanding of the Conventionality of Rules*, in «Journal of Cognition and Development», 18, 2, 163 ff. Available on: https://doi.org/10.1080/15248372.2016.1255624.

KOHLBERG L. 1981. *The Philosophy of Moral Development: Moral Stages and the Idea of Justice (Essays on Moral Development, Volume 1)* (1st Ed.), Harper & Row.

LECANUET J.P., GRANIER-DEFERRE C., BUSNEL M.C. 1988. *Fetal Cardiac and Motor Responses to Octave-band Noises as a Function of Central Frequency, Intensity and Heartrate Variability*, in «Early Human Development», 18, 81 ff.

LI L., BRITVAN B., TOMASELLO M. 2021. *Young Children Conform More to Norms Than to Preferences*, in «PLOS ONE», 16, 5, e0251228. Available on: https://doi.org/10.1371/journal.pone.0251228.

MAURER D., WERKER J.F. 2014. *Perceptual Narrowing During Infancy: A Comparison of Language and Faces*, in «Developmental psychobiology», 56, 2, 154 ff.

MELTZOFF A.N., MOORE M.K. 1977. *Imitation of Facial and Manual Gestures by Human Neonates*, in «Science», 198, 4312, 75 ff.

NUCCI L.P. 2001. *Education in the Moral Domain*, Cambridge University Press.

PIAGET J. 1932. *The Moral Judgment of the Child*, Harcourt Brace.

RAKOCZY H., BROSCHE N., WARNEKEN F., TOMASELLO M. 2009. *Young Children's Understanding of the Context-Relativity of Normative Rules in Conventional Games*, in «British Journal of Developmental Psychology», 27, 2, 445 ff. Available on: https://doi.org/10.1348/026151008x337752.

RAKOCZY H., SCHMIDT M.F.H. 2012. *The Early Ontogeny of Social Norms*, in «Child Development Perspectives», 7, 1, 17 ff. Available on: https://doi.org/10.1111/cdep.12010.

RAKOCZY H., WARNEKEN F., TOMASELLO M. 2008. *The Sources of Normativity: Young Children's Awareness of the Normative Structure of Games*, in «Developmental Psychology», 44, 3, 875 ff. Available on: https://doi.org/10.1037/0012-1649.44.3.875.

ROBBINS E., ROCHAT P. 2011. *Emerging Signs of Strong Reciprocity in Human Ontogeny*, in «Frontiers in Psychology», 2. Available on: https://doi.org/10.3389/fpsyg.2011.00353.

ROCHAT P. 2014. *Origins of Possession: Owning and Sharing in Development*, Cambridge University Press.

ROCHAT P. 2018. *The Ontogeny of Human Self-Consciousness*, in «Current Directions in Psychological Science», 27, 5, 345 ff.

ROCHAT P., DIAS M.D.G., GUO L., BROESCH T., PASSOS-FERREIRA C., WINNING A., BERG B. 2009. *Fairness in Distributive Justice by 3- and 5-Year-Olds across 7 Cultures*, in «Journal of Cross-Cultural Psychology», 40, 3, 416 ff.

ROSENSTEIN D., OSTER H. 1988. *Differential Facial Responses to Four Basic Tastes in Newborns*, in «Child development», 59, 6, 1555 ff.

SCHMIDT M.F., RAKOCZY H., TOMASELLO M. 2012. *Young Children Enforce Social Norms Selectively Depending on the Violator's Group Affiliation*, in «Cognition», 124, 3, 325 ff. Available on: https://doi.org/10.1016/j.cognition.2012.06.004.

SMETANA J.G., ROTE W.M., JAMBON M., TASOPOULOS-CHAN M., VILLALOBOS M., COMER J. 2012. *Developmental Changes and Individual Differences in Young Children's Moral Judgments*, in «Child Development», 83, 2, 683 ff.

SMETANA J.C., BALL C.L., JAMBON M., YOO H.N. 2018. *Are Young Children's Preferences and Evaluations of Moral and Conventional Transgressors Associated with Domain Distinctions in Judgments?*, in «Journal of Experimental Child Psychology», 173, 284 ff.

TOMASELLO M. 2016. *A Natural History of Human Morality*, Harvard University Press.

TOMASELLO M., CARPENTER M. 2007. *Shared Intentionality*, in «Developmental science», 10, 1, 121 ff.

TURIEL E. 1983. *The Development of Social Knowledge*, Cambridge University Press

WYMAN E., RAKOCZY H., TOMASELLO M. 2009. *Normativity and Context in Young Children's Pretend Play*, in «Cognitive Development», 24, 2, 146 ff. Available on: https://doi.org/10.1016/j.cogdev.2009.01.003.

YANG X., WU Z., DUNHAM Y. 2021. *Children's Restorative Justice in an Intergroup Context*, in «Social Development», 30, 3, 663 ff. Available on: https://doi.org/10.1111/sode.12508.

On the Cognitive Foundations of Legal Reality

CORRADO ROVERSI

1. *Introduction*

There is a traditional distinction, discussed among others by Norberto BOBBIO (2011, chap. 3), among three main problems for philosophy of law. The first problem is ontological, and it has to do with the analysis and understanding of our basic legal concepts: What is law? What are legal facts and legal institutions? This is the traditional problem of the concept and nature of law. The second problem is methodological, and it relates to the criteria of correctness for legal reasoning and the epistemological status of legal science: What counts as a justification in the legal domain? The third problem is labeled by Bobbio as deontological, and it is the problem of justice and its nature. As Bobbio notes, this last problem connects legal philosophy with political philosophy. One could note, however, that also the other two problems connect legal philosophy with other philosophical disciplines: metaphysics, social philosophy, logic, epistemology—and, by the way, a fourth problem identified by Bobbio, that of the relation between law and society, clearly connects legal philosophy with sociology. Perhaps this is one of the reasons why legal philosophy is such a broad-ranging, encompassing, and inspiring discipline.

Philosophers, and hence legal philosophers, have their conceptual tools and methods to address these problems. A typical way to proceed for them is called "conceptual analysis": when considering the question "what is law", for example, one could start from the ordinary concept of law (a concept that, as speakers of a given language, we all know), derive from that concept some paradigmatic instances of it (judges issuing a ruling, parliaments enacting provisions, etc.), and consider whether the concept and its instances suggest some necessary conditions that something must fulfill to be law. Conceptual analysis is standardly associated with the strand of philosophical thought that is labeled "analytic philosophy" and that emerged in the second half of the 20th century. Still, in a sense, this kind of method—considering and problematizing our notions about a given thing, asking for its essential elements, taking into account mental experiments to investigate into the boundaries of concepts—has been *the* philosophical method since its origins.

In parallel with standard philosophical conceptual analysis, however, some philosophers consider empirical sciences important and relevant to address philosophical problems. Depending on the kind of relation you have in mind between philosophy and science, you can qualify yourself in different ways. For example, one could consider empirical findings to complement standard conceptual analysis or to completely substitute it. If you go in the second direction, you can qualify yourself as a "reductionist" of some sort and describe your research project as a sort of "naturalization" (of legal science, of normativity, of truth, etc.). If you

instead opt for the first alternative, you can use empirical knowledge of the paradigms of a given concept to complement and possibly correct our ordinary understanding of that concept. This chapter will adopt this last attitude.

The kind of empirical knowledge considered here falls under the domain of cognitive science. "Cognitive science" is a broad expression, encompassing various disciplines from cognitive psychology to neurophysiology. Basically, it includes the empirical sciences that aim to explain the machinery of human cognition: how we perceive, believe, and know things. The methods that cognitive sciences adopt are experimental but can vary considerably, going from the statistical study of the behaviors of human subjects under certain controlled conditions to the observation of neurological mechanisms activated in certain controlled situations. Here, I will be quite ecumenical in using the findings of cognitive science: if relevant to my topic, I will use studies in cognitive psychology, observations made by developmental psychologists, or research about how our brain works.

What is my topic? I will deal with how cognitive science can enrich our understanding of the ontological problem of legal philosophy by identifying and describing the cognitive foundations of law, legal institutions, and legal facts. "Cognitive foundation" or "root" could be seen as a misleading metaphor (see CHIASSONI 2021, 498), so let me clarify what I am aiming at with that label. I want to understand which cognitive features of human beings are necessary for them to act in light of legal facts they believe exist and hence to support the existence of these facts and their connected legal institutions. To my knowledge, the amount of work devoted to how cognitive sciences can enrich the ontological inquiry in legal theory is much smaller than that dedicated to the methodological and the deontological problem[1]. When it comes to the methodological problem, and hence to legal reasoning, the question of how that can be affected by theories of human decision-making has produced much discussion in recent years, particularly related to a systematic exploration of biases and heuristics in judicial decision-making[2]. Problems of justice typical of the deontological problem are instead raised, for example, by the question of whether cognitive-based kinds of manipulation should be admitted and regulated by law when used by private actors (like in the case of so-called "neuro-marketing") or public institutions (in the form of "nudge-like" regulations: THALER, SUNSTEIN 2008). And, of course, deontological is also the question whether our findings on how human intentional activity works should impact criminal law, possibly changing our overall theory of punishment (see for example HAGE & WALTERMANN 2021). Instead, I will try to see if we can put cognitive sciences to use in perfecting our answer to the ontological problem of philosophy of law.

My main thesis will be that our knowledge of how the human mind develops from childhood to adulthood, as well as our hypotheses on the evolution of human cognitive capacities in relation with those of our evolutionary ancestors (primates in particular), can significantly contribute to a better understanding of how legal facts and legal institutions can exist and have a role in our social life, because law's existence depends on human minds. This thesis is not new: Scandinavian legal Realists traceable to the Uppsala School (Axel Hägerström, Karl Olivecrona, Alf Ross) and Polish-Russian legal realists (Leon Petrażycki in particular) have insisted on the inherent connection between law and psychology since the first half the 20th century[3]. However, these scholars had in mind a kind of psychology very different from everyday cognitive science, which has strengthened the rigor of its experimental methodology and, even more importantly, has undergone the revolution of the neurosciences.

[1] With some notable exceptions, particularly focused on legal normativity: BROŻEK 2013; BRIGAGLIA & CELANO 2018; BRIGAGLIA 2022.

[2] See for example GUTHRIE et al. 2001; GUTHRIE et al. 2007; WISTRICH & RACHLINSKI 2017; HOFFMAN 2021.

[3] See for example HÄGERSTRÖM 1953; OLIVECRONA 1971; PETRAŻYCKI 1955; see also in this regard PATTARO 2016 and FITTIPALDI 2016.

At least three methodological objections could be raised against the research project whose current results I present in this chapter. The first is that such a project is inevitably speculative, both because many of its conclusions about legal institutions can have very different explanations, all of them in some way coherent with the available empirical data, and because the empirical theories assumed as starting points are in their turn the object of strong debate in the respective disciplines. It seems to me that a similar objection can be raised against any kind of scientific endeavor: different explanations are available, and assumptions can be debated. The crucial point is the extent to which the proposed theory and its assumptions resist and accommodate potential counterexamples. Hence, I will propose my view, but I will also try to clarify the possible alternatives and the problems at stake. In this way, if the reader finds the final reconstruction not sufficiently reliable, she will at least have a clear conceptual map of some necessary steps that must be taken to build her alternative.

The second objection addresses not the empirical assumptions of my reconstruction of the cognitive underpinnings of legal facts and institutions but rather the viability of proposing a unified theory of the nature of legality. One could say, for example, that it is not possible to identify the necessary features of law on conceptual grounds because the historical reality of law is so complicated and diversified that different things have inevitably fallen under the scope of this concept in different cultures and/or at different times: every theory of the nature of law will inevitably be a projection of our parochial, limited view (see SCHAUER 2012; TAMANAHA 2017a). This objection can be understood either as a kind of skepticism against armchair conceptual philosophy ("You cannot understand the essential features of law a priori, you must study history and sociology to do so!") or as a general thesis about law's metaphysics ("there are no essential features of law, any theory of this kind will be false!"). In specific relation to my work, Pierluigi Chiassoni argues that the phrase «'Metaphysics of (the) law' fatally evokes the premodern idea of a set of 'principles' or 'essential features' calling for a philosophical enquiry capable of grasping the very 'nature' (the very 'ontology') of legal phenomena by delving into their depths in ways that are not available to different forms of investigation» and that we should «abandon the phrase» in favour of «some less obscure, less misleading, and plainer expressions, like, for example, 'law' or 'legal reality'» (CHIASSONI 2021, 497). As Chiassoni rightly points out, the terms "metaphysics" and "ontology" of law, that I use, should not evoke in the reader the idea that this research is a kind of essentialist theory closed to empirical revision: it seems to me that metaphysics does not imply dogmatism, and I agree with the idea that conceptual philosophy must be complemented by empirical disciplines (indeed, this is exactly the methodology underlying this chapter). However, I think that dismissing the possibility of essential, or at least typical, features of law is the outcome of a dogmatic attitude, not so different from that which is assumed by scholars who think that conceptual armchair analysis is perfectly self-sufficient. How can one rule out a priori the possibility of an encompassing description and explanation of law? It is certainly possible to falsify by way of counterexamples the explanatory conjectures to refute and improve them: this entails, however, participating in the endeavor of (science-based) legal ontology rather than dismissing it.

Finally, a third possible objection to this research could be that the legal domain is so complex and includes so many different kinds of entities that a single and unified ontological theory to explain it in its entirety could not be possible (legal reality could be "disunified": PLUNKETT & WODAK 2022). This objection puts forward an important caution rather than a thesis: we should take seriously the possibility that not all legal entities can be explained by a single theory. Here, too, one should consider the proposed alternatives to show this is the case. My theory, however, will have to do with legal institutions and rule-constituted legal facts: so, even if it does not capture the whole of legally-relevant entities, it is aimed at explaining a significant part of it.

Let me present how I intend to proceed in this chapter. Section 2 will form its bulk, explaining why legal facts depend on human cognitive capacities and what these capacities are. In the first part of that section, I will start with social institutions, of which legal institutions are an instance (Sections 2.1-2.3). In the second part, I will instead introduce some peculiar features of legal institutions, putting forward three theses about them (Section 2.4). In Section 3, I will explain how the proposed analysis, which is theoretical, can have significant practical consequences: in particular, I will argue that if legal institutions and legal facts require some cognitive capacities and dispositions, weakening the latter could entail weakening the former. This can lead to a cognitive-based theory of legal pathologies, the possibility of which jurists should be aware of. Finally, in Section 4, I will conclude by summarizing the theory and putting forward a normative question that, in light of it, deserves serious reflection: If there are institutional frameworks that we cherish as political ideals, shouldn't we consider the nurturing of its cognitive underpinnings as a political priority?

2. *The cognitive underpinnings of legal facts, step by step*

2.1. *Legal facts are a subset of institutional facts*

Facts in the world are typically not dependent on humans: the fact that it rained today or that Jupiter revolves around the sun depends on how physical reality is framed. Other kinds of facts, however, could not exist without humans, particularly human society: workers protest for their rights during a strike, colleagues have a conversation in front of a coffee machine, a country falls into an economic recession, Fascists ruled Italy starting from 1924. These facts involve entities that become possible only when human beings organize and conceptualize their social life: groups of humans (workers, Fascists), speech acts (conversation), social events (strikes, economic recessions), social entities (the Western World, Italy) depend on a way of "carving the world" that is impossible without the emergence of society. This is why they are called social facts, and the philosophical discipline that deals with the nature of these facts (with the metaphysics of society) is called social ontology (see on this EPSTEIN 2018). I assume that legal facts are included among social facts, and hence, when legal philosophy deals with the ontological problem, it is a subset of social ontology.

Among social facts, some are peculiar. Consider the fact that workers protest for their rights during a strike, for example. The fact that there is a group of workers or that they protest is slightly different from the fact that they perform a strike and even more different from the fact that they have a right to do so. Legitimate strikes, and hence the right to strike, are regulated by the law, and the law attaches to them both proper conditions of performance (for example, that they are allowed only under certain conditions) and normative consequences (for example, that workers do not get their salary, but their behavior does not count as a breach of contract). Strikes, chiefs within a tribe, walls that are the boundaries of a community's territory, or even wooden pieces that have a meaning within a shared game, all these entities and facts we may call institutional. Why "institutional"? Because all these facts involve statuses, which require a sort of fixed normative framework, a set of rules, and in most cases, these rules are general. A social group can decide that a specific individual, say, Mr. Rex, has the status of the King. Still, typically members of the group will develop a rule about it: *in general*, if someone is chosen/is the son of.../is ordained by priests, then he will have the status of the King, which means that it will have some kind of power. This is an institution in that group, and the facts made possible by the existence of that institution are, therefore, institutional facts. To be sure, the notion of institution used here is very broad because it can include normative frameworks going from religion to politics and games as well. But, independently of how broad should the notion of

institution be, it is evident that legal facts are a subset of institutional facts. When we make a contract, elect the President, enact a statute, find someone guilty, counter-examine a witness, or give a grade to students, we do acts or instantiate facts that are institutional because they have a status connected with normative consequences defined by the rules of an institution. I borrow the elements of this description of institutional facts from the social philosophy of John Searle, which is assumed here as the starting point of my analysis[4].

2.2. *Social institutions have a common structure*

The question, then, is whether institutional facts (and legal facts among them) have a recognizable common structure apart from the generic reference to rules. When we attribute a status connected with normative consequences to an entity (e.g., someone is the President), a fact (two persons reaching an agreement), or an event (our reaching the age of 18), we do so by way of rules that are slightly different in function from other, more ordinary kinds of rules. Typically, when we regulate an activity, for example, by stating that smoking is not allowed in public places, the regulated activity can exist even if the rule does not exist. Indeed its previous, independent existence is why we want to regulate it. So smoking is something that some people do whether or not there is a rule about it, and we create a rule exactly because we do not want people to harm the well-being of other people. In the case of statuses and institutional facts, however, these cannot exist without rules. Rules create them and make them possible: without constitutional provisions, you could not have a President of the Republic, just as you could not have a contract or the age of majority without a civil code. You could not have a game, nor the activity of game-playing, without the game's rules. This is the main peculiarity of institutions and institutional facts, namely, that the rules making up the institution are constitutive of the facts that are regulated by them. In contrast, in the case of ordinary, regulative rules, the relevant facts are already there and are perfectly possible even in the absence of the rule. Institutions thus require *constitutive* rules, namely, rules that "constitute and regulate"[5].

Constitutive rules are often described in the literature as having a common structure in the form "X counts as Y in context C", or better, "X counts as Y, which implies Z in context C" (this is known as the "XYZ formula": HINDRIKS 2005, 123 ff.). This is a useful formula to appreciate the capability of these rules to constitute statuses connected with normative consequences: these rules indeed state the conditions, consequences, and the relevant context of institutional statuses. Thus, the basic structure of the constitutive rule of the President of the Republic can be described as "a person with some properties elected under certain circumstances (X) counts as the President of the Republic (Y) in the Italian constitutional system (C), which entails a specific institutional function and a set of duties and powers (Z)". Similar rules can be formulated for contracts, the age of majority, game pieces and more in general for all institutional statuses. The formal, "count as" structure, however, should be considered more as a useful mnemonic than as a strict requirement. Constitutive rules can have a seemingly regulative form, as in the case of "professors must teach at least one course at the University" or "bishops must move in diagonal in chess" (see CONTE 1995), and in most cases only the overall system of rules about a given institutional status can be framed in the "count as" form. Conversely, not all rules having a "count as" structure are constitutive: if we say, for example, that «Deputies engaged in activities outside the Chamber premises [...] shall be counted as present for the purpose of establishing the presence of a quorum» (Art 46, par. 2 of the *Rules of*

[4] For further details, see SEARLE 1995, 2010; a further distinction concerns constructivist facts, which will not be treated here: see HAGE 2022.

[5] On constitutive rules see, among many others, SEARLE 1995; SEARLE 2010; HAGE 2018; LORINI & ŻEŁANIEC 2018; RAMÍREZ LUDEÑA & VILAJOSANA 2022; ROVERSI 2022.

Procedures of the Italian Chamber of Deputies), this rule states simply an equivalence and does not constitute any kind of institutional fact.

Institutions, conceived as the framework that makes institutional facts possible, are sets of rules, some constitutive. If a given community frames only regulative rules, you can have a social regulation of behavior but not institutional facts: for example, the fact that people do not smoke in public is not institutional, whereas the fact that they marry is, and indeed it requires a constitutive rule about what counts as marriage. Some argue that institutions can be possible without rules. Human behavior can also be regulated by way of standards, for example, and more in particular by referring to paradigmatic examples of correct behavior and institutional statuses: one can define what the king is by pointing to important kings of the past, and some communities can have constitutive paradigms rather than constitutive rules (see on this RUST 2021). This is an important qualification, needed to stress that constitutive rules cannot be traced to a single, common structure and that the "counts as" form is simply a rough generalization covering many kinds of regulation. This said, even a regulation by constitutive standards is, in the end, a regulation by rules: to create an institutional status, people will have to infer a rule from standards. And typically, this activity will be open to many interpretations because standards will provide very vague answers to the question of how a given institutional status can be instantiated and its precise normative powers. Hence, a regulation by constitutive standards eventually turns out to be a regulation by competing, and most often alternative, constitutive rules derived from those standards.

Institutions also have a function—one could say a purpose or a point (see MACCORMICK 2007, 36 f.)—in the form of a given social role for which they have been built: this purpose shapes the overall institutional structure, and both the conditions and the normative consequences of the institutional elements are framed accordingly. So, the function of the President of the Republic in the Italian Constitutional system is to represent the Nation's unity and to safeguard the Constitution. For this reason, the election of the President must be performed within specific boundaries that enhance the role's representativeness and the President's powers are framed to enable its role as a constitutional guarantee. It is important to note that the institution's point is not constituted but is rather presupposed by the constitutive rules: it forms an important part of those rules' *ratio*, namely, the purpose they serve. Hence, the overall point of the institution is *meta*-institutional rather than institutional, properly speaking: you cannot have an institutional structure if not inscribed within a meta-institutional, pre-existing social framework, just as you could not have the game of chess, with all its rules, if the idea and concept of game-playing did not exist in society (see on this, among others, LORINI 2000, ROVERSI 2018).

Some scholars find this idea of constitutive rules rather mysterious. It is argued, in particular, that it seems possible to describe the structure of institutions using simpler, more intuitive notions like those of regulative rules plus definitions of terms. Hence, an institutional concept like "property" can be seen simply as the aggregate of certain conditions and certain consequences, like, for example, «[i]f a person has lawfully acquired a thing by purchase, judgment for recovery shall be given in favor of the purchaser against other persons retaining the thing in their possession» (ROSS 1957, 819; see also HINDRIKS & GUALA 2015) Property, under this analysis, is simply defined as the overall set of these conditions and consequences, and in the end is nothing else than the outcome of regulative rules in the form "do this... if you want to achieve...": more generally, institutions and institutional concepts are considered here as being nothing else than "shortcuts" to systematize sets of regulative rules. Analyses of this kind result from skepticism towards the idea of facts and entities "made possible" by rules: How can rules make possible reality? Where should these institutional facts exist if they are not concrete, empirical facts? They seem to be the outcome of an unwarranted hypostatization, namely, the fallacious attitude of treating as real things that are not.

I cannot enter here into this reductionistic attitude towards constitutive rules, but in general, I think that, though intuitively attractive, it is ultimately misguided (see ROVERSI 2021). Hypostatization is not a fallacy but a feature of human minds, of how humans talk and categorize the world. We hypostatize when we talk about games, fiction, ideas, theories, projects, values, history, and of course, when we talk about rules and their content. In general, as humans, we can consider the product of our linguistic practices as an object, an entity out there, something that exists given our social framework. Reductionists insist that these things cannot be said to exist, and of course, they do not exist as part of the natural world; they would not be part of the universe if humans did not exist. But the same could be said of many other things, such as buildings, screwdrivers, databases, cars, weapons, bottles, etc. These are artifacts built by humans to serve their purposes. Among these, there are symbolic artifacts, namely, entities that serve their purpose by a meaning we assign to it: some of them are at least partly concrete, such as jewelry or neckties or crowns, some of them are completely immaterial, like works of art, games, fairytales, and social institutions. Just like material artifacts, immaterial artifacts are meant for interaction with other human beings: Dante's *Inferno*, for example, is a set of words in vulgar Italian meant to express something to other people who are supposed to understand those words. In the case of institutional artifacts, you need rules that must be practiced, not simply sentences that somebody must understand: to have the institution of property, you need to state by way of rules the conditions under which somebody can own something, the normative consequences of owning, and, more in general, the meaning of the status "owner". But, just like the sentences of Dante's *Inferno* are constitutive of the poem, the rules are constitutive of property. They create something that remains "out there", in the domain of our symbolic artifacts, that populate the human world just like screwdrivers and cars. Hence, I do not see anything mysterious in social institutions, institutional facts, and constitutive rules: they are simply an instance of the general capacity of human beings to build artifacts by way of language (on the "artifact theory of law" see BURAZIN et al. 2018).

2.3. *The grammar of cooperation: collective intentionality*

Institutions would be only on paper if people did not collectively intend to accept and practice the relevant rules. But what do we mean by "collective intention and acceptance", and how can humans have them, given their cognitive makeup?

There are several ways in which we can do something together with others. We could, for example, walk in a park where other people are walking: but, in this case, we would be doing something together in a very weak sense because we would merely be co-present in performing an activity. Or we could agree to read something together: but, here too, apart from the explicit agreement and perhaps our meeting at a certain time, the activity would be performed by each of us in parallel, so to say, without much coordination or interdependence. The most central case of "doing something together" is rather when two or more individuals intend to cooperate to bring out an activity, as when we decide to collectively paint a house or cook a recipe together.

Despite its simplicity at first sight, this idea of collective intention is not so easy to analyze. Intending to cook together is not simply a sum of individual intentions: I can intend to cook the recipe, and my wife can intend to do the same, but if we do not conceive this intention as collective, we simply end up performing activities in parallel, as in the reading example above. The point is not that *I* intend to cook the recipe *and* that my wife intends to, but rather that *we* intend to do it together, cooperatively. Several analyses in contemporary social ontology try to explain how a collective intention is different from an individual one, going from a more collectivistic to a more individualistic approach. According to a purely individualistic approach, it is sufficient that individuals share a goal and have individual intentions (see, for example, MILLER 2001) plus some mutual beliefs about others' intentions; these individual intentions

must, however, interlock so that the general plan of the shared activity is specified along sub-plans the individuals intend to realize, and these sub-plans effectively mesh to realize cooperation[6]. Some scholars, however, criticize the idea that individual, interlocking intentions and mutual beliefs are sufficient. They rather argue that sharing intentions with others changes the nature of our intention: we intend as part of a group, in a we-mode (TUOMELA 2016), and this can be a modality of intending encoded in the human brain, just as the capacity to intend on an individual level (see SEARLE 1995, 23-6). All these approaches, however, share at least two basic elements about explaining collective intentions, namely, that human beings can share *attention* over a given activity and coordinate their intentions with that of others by doing their part and supporting their partners in doing theirs. This is the foundation of human cooperative endeavors and legal institutions.

Cooperation typically requires commitment. When my wife and I cook together, we give for granted that we will not give up without explanations and do our best to perform our part and support our partners. Our cooperative attitude presupposes these commitments, giving rise to specific expectations: My wife expects me to whisk the eggs with cream at the right time, for example. If I do not, she will react by protesting or urging me to do what I am expected to do. Intending to do something together has, therefore, an inherently normative structure[7].

However, this basic, normative layer of cooperation must be enriched and elaborated for social life to be effectively organized, and this can happen because humans can collectively intend and commit to supporting a set of rules, regulative and constitutive. Apart from cooking, my wife and I commit to the rule that teeth should be brushed before going to bed (regulative rule), and we would very much like our children to do the same. Or our kids and we commit to the rule that Friday evening is our "family cinema evening" and that everyone in the family is supposed to stay home, watch a movie, and comment on it (constitutive rule). Members of an association could intend to support and commit to the rule that they should adhere to a formal dress code when attending monthly meetings (regulative rule), or that all presidents of the association who have served in that capacity for at least five years are to be recognized as honorary founders of the association (constitutive rule).

In moving from the mere normativity inherent in cooperation to the acceptance of social norms, a passage in perspective must be clarified. If I commit to cooperating with my wife, my duty depends on her and my relationship with her. This normativity is based on a second-person perspective: I must do X because you expect me to do this. But in accepting and sharing a social norm, we all must do what is required by the norm because there is a norm, not in virtue of the expectations of a specific community member. We enter here into a third-person perspective: norms are objectified and become the source of duties. In this perspective, my normative pressure on others becomes possibly disinterested, it is not connected with my relationship with a cooperative partner but rather with the group's existence. We are not cooperating only to fulfill a specific task but to support the group by enforcing its norms. Even in this case, however, the basic normative structure of social cooperation in terms of collective intentions and commitments forms the bedrock of normativity: hence, of all kinds of social institutions, the legal ones being (as we saw) a kind of these last (see TOMASELLO 2016, 67-70, 73-5).

One could say that this passage is too quick and rough: one thing is two persons cooking together a cake or a small association having an internal code of behavior; another is 60 million people sharing a legal system. Do I expect every citizen of Italy to intend to cooperate with other citizens in supporting the law? Political systems, and the legal institutions that derive from them,

[6] See BRATMAN 1992, 2014; for an application of this idea to institutions, see BRATMAN 2022.

[7] See GILBERT 1989, 2000, 2006, 2014, 2018; Bratman does not agree on the idea that joint intention can by itself generate normativity: BRATMAN 1999; recent discussion on normativity and joint action can be found in GOMEZ-LAVIN, RACHAR 2019 and 2021, LÖHR 2022.

are the outcome of all kinds of historical contingencies, which very often show a high degree of violence, dominance, and fear of sanction—in a word: the actual *power* of chiefs, rather than the peaceful collaboration of all community members—as the real backbone of social regulation (see CANALE 2014 for a powerful formulation of this objection). Moreover, the more a community becomes large, the less likely it is that a significant part of its members have not a strong sense of social cohesion and cooperation, and this is even more true in large, liberal, multicultural countries where different worldviews co-exist, and the life of individuals is much more impacted by the mechanisms of the capitalistic global market than by political identity. In this context, people can go along passively with a very rough idea of the content of legal provisions they are subject to: In what sense do they cooperate with others in supporting the system?

This is an important objection, and perhaps legal theory can be helpful here. In describing the structure of a legal system in his classic masterpiece *The Concept of Law*, of 1961, H.L.A. Hart drew a sharp distinction between two classes of people that are relevant for the existence of that system: ordinary citizens, or the mass of the population, on the one hand, and legal officials or experts of the legal system, on the other hand. Hart argued, in particular, that, though an ideal legal system would be grounded on the acceptance of all members of the community, in practice this is not required: rather, it is sufficient that ordinary people, in general, conform to the legal norms of conduct, whereas officials and experts must have an internal point of view, and hence a supportive attitude, about both the norms of conduct and what Hart calls the "secondary rules" of the system, which are power-conferring and constitutive (see HART 2012 [1961], 60 f., 114-6). This view makes it possible to reply effectively to the objection just raised. Even though one should not assume a cooperative attitude in all citizens of complex, multicultural, hierarchical political systems, even in these systems, there is at least a group of people where this supportive, cooperative attitude must be assumed for legal institutions to exist: the group of officials and experts of law who apply, enforce, and practice the norms of the legal system. Hence it is true that cooperation cannot be everything that matters for the law's existence, but it is its ultimate foundation.

Apart from officials, how can we describe the attitude of other community members, then? This is a form of collective recognition rather than intention. Recognition is a weaker attitude than intention: you can recognize or accept, in a weak sense, that something happens even if you do not actively cooperate with someone to make that happen. In this sense, even simply going along with an existing social framework without criticizing or endorsing it is a kind of recognition (SEARLE 2010, 56-8). And, of course, you can have degrees of recognition, going from this kind of indifference to a disposition to support and cooperate, but only if it is necessary and if explicitly requested to do so. Hence, even if it is true that in complex societies most community members cannot be said to have a collective intention to support the institutional framework—this is an attitude that only legal officials are supposed to have—they nevertheless can be described as collectively recognizing (or accepting in a weak sense) it[8].

What about mere dominance? Could not legal systems be the outcome of hierarchical arrangements where the subordinates simply obey the commands of people in power? Undoubtedly, there is an element of fear of sanction in the attitude of acceptance that most people hold towards a legal system in force, and it is certainly true that there are, and have been in history, systems of law where this element of fear was pushed to extremes to achieve obedience. This may not be true for democratic constitutional systems, but it is when we consider dictatorships or totalitarian regimes. However, this does not show that there is no cooperation at the core of law: it shows that the contrary is the case. If we keep in mind Hart's

[8] Elaborate discussions about how to apply the concept of collective intentionality to Hart's description of law can be found in SÁNCHEZ BRIGIDO 2010 and TOH 2022.

distinction between officials and ordinary citizens, we can realize that legal systems based on hierarchical dominance require even stronger cooperation on the part of officials than in the case of democracies, this precisely to ensure the actual application of sanctions and ubiquitous surveillance of all aspects of social life. In these systems, officials must be trained and indoctrinated to become cooperative supporters of the status quo to the point of fanaticism. Hence, cooperation is necessary here, too: only it is restricted to officials and possibly to other supporters of the status quo in the social community, a group that the regime typically supports, nurtures, and aims to expand.

Some scholars believe that rules do not play such a fundamental role in the construction of social institutions: They are rather the outcome of something more basic, namely, of the typical ways in which individual preferences can converge. These patterns of convergence of individual preferences are called "game-theoretic equilibria" in contemporary economic theory, and an example of them can be the following. Imagine that we must collectively decide how to drive our cars: our main individual preference will be that of being able to drive without constantly risking a car accident, so we will need to agree on a rule. In this case, the kind of rule that we agree on—whether to drive on the left or the right—is not particularly important, but we need a rule that we all accept. This shows that, from this perspective, rules are simply devices to coordinate individual preferences: the reason why institutions understood as sets of constitutive rules emerge is not ultimately explained by our acceptance of rules but by the most efficient combination of our underlying interests[9]. I doubt that this "strategic" account of institutions can really explain the contingent and chaotic character of real-life society: even if there may be an element of efficiency in their emergence, social institutions are the outcome of unpredictable factors that are highly contextual and historically dependent. However, it is important to note that, even if one accepts the view that social institutions are accepted for strategic, individual reasons, collective intentionality does not fade away from the picture. Rather, it is still the main machinery needed to keep in place the social institution that represents the equilibrium. Cooperation, commitment, constitutive rules, and norms do not disappear from this account: they are not the "most basic" thing but are still crucial[10].

Let me summarize what I have been maintaining so far. Social institutions are structured along rules, some of which are regulative and others constitutive of statuses. These rules must be collectively recognized by the population and supported—in the sense of an active, collective intention to support cooperatively—by officials. A collective commitment to these rules is part of creating a social community; it is a central element of the glue that gives the group a sense of membership and cohesion. There are, therefore, three progressively more complex layers in the construction of social, institutional facts: (1) intending and committing to cooperate on an activity, (2) intending and committing to cooperate on the support of shared rules, (3) intending and committing to cooperate on the support of shared constitutive rules about statuses. My question is: how are these three layers possible in human beings? Part of the answer to this question is psychological: humans can do these things because their minds and brains are framed accordingly. Let us see, then, what contemporary cognitive science can tell us about humans' mental capacities and dispositions that make these shared activities possible.

[9] Game-theoretic analyses of the emergence of conventions and social norms trace back to LEWIS 1969 and ULLMANN-MARGALIT 1977, and significant developments of these models have more recently been provided by BINMORE 2005 and BICCHIERI 2006, among others. A deep and thought-provoking application of this approach to social ontology, and to the ontology of social institutions in particular, can be found in GUALA 2016.

[10] See on this GUALA 2016, 71 ff., which explains institutions as solutions to problems of coordination in the form of rules, even though he rejects constitutive rules: but see ROVERSI 2021 for a counterargument to this last point.

2.4. *Social institutions require certain cognitive capacities*

2.4.1. *The ontogenesis of social institutions*

An important preliminary distinction for this analysis is between "ontogenesis" and "phylogenesis", two concepts used in the biological sciences. Both concepts are connected with the idea of development, but they apply this idea to two different entities: given a biological species, ontogenesis is the development of one instance of that species from birth to maturity, whereas phylogenesis is the development, or better, the evolution of the species itself. Hence, for example, ontogenesis is the process through which a single human goes from birth through childhood to adulthood. In contrast, phylogenesis is the process through which the human species evolved through time. Biological features have, therefore, this double aspect: you can analyze them from an ontogenetic perspective, asking how they emerge in the growth of an individual, or from a phylogenetic perspective, considering the problem of how they emerged as the outcome of natural selection, in the overall evolution of the species that individual belongs to. A description of cognitive features is twofold in the same way. My question is, what are the cognitive capacities of human beings that make social institutions and institutional facts possible? I will first address this question from an ontogenetic perspective: how do these capacities emerge in humans?

Human children can share attention and a simple goal with a caregiver since they are six months old, so this is a very basic ability: already at the age of twelve to fifteen months, they show active interaction and capacity to coordinate actions with a caregiver (see TOLLEFSEN 2004, TOMASELLO et al. 2005), and they understand the basic normativity of joint commitment, typically attempting to reengage their partner if the shared activity is interrupted abruptly (see WARNEKEN & TOMASELLO 2009). At this stage, some argue that human children already have a primitive capacity to attribute beliefs about the shared activity to their partner because they are surprised if their behavior is inconsistent with the belief they have hypothesized (ONISHI & BAILLARGEON 2005, BAILLARGEON et al. 2013). Emotions, and particularly social emotions, play an important role in the development of collective intentionality: the capacity to cooperate with others for human children around two years of age is connected with the capacity to detect the emotions of others and be touched by them; hence with the development of empathy, because if we had to rely only on strategic rationality, our default mode towards shared activities should be cautious and non-cooperative, and it would bear high cognitive costs (see MICHAEL 2011, MICHAEL & PACHERIE 2015). On the contrary, what we observe in human children already at that stage is that sensitivity to the expectations of others, as well as an attitude of expecting others to cooperate reinforced by emotions like anger and shame in case of non-compliance, is the default (see MICHAEL et al. 2016, 8 f.). At three years, this default disposition to cooperate shows a full development: explicit commitments are taken, and hence expectations become stronger, along with the capacity to resist temptations to give up. If a child wants to leave the joint activity, at this stage, he or she feels the responsibility to communicate it to the partner and make some amends (GRÄFENHAIN et al. 2009). In general, all the participants in the joint activity are supportive even if they have already earned their reward, so for the sake of cooperation rather than merely for individualistic gain (HAMANN et al. 2012, GRÄFENHAIN et al. 2013).

This full-fledged, not necessarily self-centered, supportive attitude towards cooperation that emerges at three years of age is at the core of the second level of the construction of social institutions, namely, commitment and support towards shared norms. We saw above that this passage requires the shift from a second-person perspective, where normativity depends on the expectations of others, to a third-person perspective, where norms are objectified and normative pressure can become disinterested, that is, can be in place even when there are no expectations of a specific person involved, rather being the outcome of interest towards the existence of the

group in itself. Children start to internalize normative behavior by way of imitation: already at the age of three, they are particularly disposed to imitation, to the point that they imitate behavior even when it is not causally efficient (see HORNER & WHITEN 2005) when this can require them to sacrifice a strategy they previously considered efficient (see HAUN & TOMASELLO 2011), and when warned not to imitate actions that seem "silly" (see LYONS et al. 2007)—an attitude called "over-imitation". Children consider the behavior they imitate to be normative also from a third-person perspective (see KENWARD 2012), and for this reason, they intervene to correct deviant behavior and to protect the rights of others (for example, property rights), even when they are not directly affected (see ROSSANO et al. 2011; on the significance of imitation for the emergence of normative behavior, more in general, see BROŻEK 2013). The emergence of these normative attitudes is connected with the idea of a group, of the importance of belonging to it, and hence with reputational concerns, which are explicit in human children at the age of eight and perhaps implicit already at the age of five (see SHAW et al. 2013) and which are, even in adulthood, crucial in deciding whether to cooperate (see ROCKENBACH & MILINSKI 2006). In the passage from the age of three to the age of five, human children also reinforce their third-person perspective by understanding that some norms are less conventional than others, particularly those related to physical assault (see in this regard TURIEL 1983, but also KELLY et al. 2007 for skeptical remarks on this point). They show a transcultural conception of fairness, modulated by cultural parameters only at a later stage (HOUSE et al. 2013, 14590). The emergence of language in humans, starting from age two, plays an important role in the objectivization of norms. The third-person perspective towards norms is considered absent or, at best, embryonic in primates, while present in human children, even by some of those theorists who underscore the strong cooperative and normative attitude of non-human primates (DE WAAL 2006, 54; DE WAAL 2014, 197-200), and this could be indirect support to the idea that those processes are reinforced, if not constituted, by linguistic practices: However, caution is needed about this connection, because the lack of evidence, for example, on third-person punishment in chimpanzees could be due to flaws in experimental settings (see ANDREWS 2020, 49). Shared language plays an important role in the way young children communicate social norms to their peers and in the way they treat them as objective (see GÖCKERITZ et al. 2014), and also, more in general, in defining the boundaries of the relevant group: children prefer to trust and imitate more people whose language they understand (see KINZLER et al. 2011, BUTTELMANN 2013).

Moreover, language becomes crucial in the emergence of the third level of the construction of social institutions, which focuses on commitment and support of shared norms constituting institutional statuses. To understand the idea that someone must be regarded as the King or the President, it is necessary to have the cognitive capacities to endow physical entities with a value they do not have in virtue of their physical features: a symbolic, not merely instrumental value (see on this BRIGAGLIA & CELANO 2021)—and, of course, language is a system of symbolic values, in which sounds and marks are given meaning. Human children engage in symbolic games of pretend-play with objects since the age of two[11], when language also emerges, and they understand the normative structure of the game; that is, they protest if its constitutive rules are not respected by a partner (RAKOCZY et al. 2008). To have all the elements of institutions in place, however, it is also necessary (as we saw above) to have in place a network of mutual beliefs: we all accept and believe that someone elected by the Parliament under certain circumstances is the President of the Republic, I accept this and believe that you accept it, and I believe that you believe that I accept it, and so on. This capacity to attribute beliefs to others and to understand that others' beliefs can be false and yet guide their behaviors—this theory of

[11] See RAKOCZY et al. 2005a, 2005b; see also, on "pretensive shared reality", KAPITANY et al. 2022.

mind, and false beliefs—requires a high-level perspective taking that starts to develop in human children only at the age of four to five[12], an age at which they start to understand semantics, hence the idea that language is based on rules, to the point that the two capacities may be conceived as connected (see DOHERTY & PERNER 1998). From that point, the normative, constitutive framework becomes the preferred way to understand social roles and to predict social behavior according to them (see KALISH & LAWSON 2008), and institutional objects like banknotes are conceptualized as standard artifacts, namely, perfectly objective. Only at a later stage, starting from the age of eight, do children understand that these objects—as well as much of their constitutive normative framework—depend on shared beliefs and intentions within the community (NOYES et al. 2018).

My picture of the ontogenesis of legal institutions can thus be summarized as follows. To understand institutional facts, human children must first understand social cooperation and its normative, second-person framework, which they start to practice at the age of one and intertwine with structured social emotions starting from two. At the age of three, a third-person perspective over social norms and a disposition to support them emerges in connection with a strong interest in imitation, group belonging, and reputation within the group, factors facilitated by the progressively stronger capacity to use a shared language. From the age of four to five, children develop a theory of mind and false beliefs and an understanding of semantics, two factors that make it possible for them to attribute symbolic meaning to objects and hence to understand the constitution of institutional objects. At first, they conceptualize institutional entities as ordinary, objective artifacts, while they understand the conventional and symbolic nature of these things only later, from age eight. From that point on, all the cognitive elements necessary for social institutions to exist in the perspective of a single human individual are in place. Now a further question is: apart from the individual perspectives, how did these capacities evolve in the human species as a whole? This is the question regarding phylogenesis.

2.4.2. *The phylogenesis of social institutions*

The human brain is made up of layers, so to say, and these layers are the outcome of evolution: the lower and deeper the layer is, the more ancient it is in evolutionary terms. To put it in a very simplified way: the brain stem is the primitive part, sometimes called also the "reptilian brain", which mediates sensations, bodily perceptions, and the default reactions concerning survival, namely, fight, flee, or freeze; the limbic systems, which coordinate adaptive responses to the environment (among which the amygdala, crucial in the elaboration of emotions), developed later with the emergence of the first mammals; the cortex, which is the more external and recent part of the brain, developed in various ways in mammals (see FRANCHINI 2021), but in its most developed form is necessary for abstract thinking, self-reflection, and self-consciousness, mental capacities that are typically considered to be only human. There is a parallelism between how the brain develops in childhood and how it developed by adding progressively more complex layers during the evolution of the human species: ontogenesis, at least when it comes to understanding "what came first", reproduces phylogenesis. This is an important point because, as we saw, human children become able to understand all the elements of institutional facts only between 5 and 8 years of age, which means that at least the most cognitively complex among these elements are, from an evolutionary point of view, a more recent outcome than those regarding social emotions and basic normativity. And, of course, this should not come as a surprise: humans of the species *Homo sapiens* have institutions, whereas

[12] See WELLMAN et al. 2001, WELLMAN 2018; but see also ONISHI & BAILLARGEON 2005, BAILLARGEON et al. 2013, and SLAUGHTER 2014, who attribute basic, implicit mind-reading capacities and understanding of false beliefs also to infants and toddlers.

other animals, including primates, do not, at least not at a comparable level of complexity. So, a possible story that can be told here is one in which, given the cognitive abilities we found when analyzing the ontogenesis of social institutions, at least some of them evolved particularly in human beings and not in our evolutionary ancestors, namely, primates. But where should we find this turning point that makes it possible for us, but not for chimpanzees, to have Presidents? This is a very complex question, and its answers are inevitably speculative, based on different interpretations of the available empirical data.

Here is a possible account, obtained mostly by combining insights drawn in large part from Dubreuil 2010, Tomasello 2016, and Wrangham 2019. To have evolutionary success and survive, humans had to cooperate, and they found an extremely elaborate way to cooperate, namely, by way of norms and, eventually, of social institutions. This process of self-regulation and the progressive abandonment of dominance as the typical way of regulating relationships started with *Homo habilis* (approximately 2 million years ago) in a context where climatic changes made competition for food harsher than normal. Cooperation was, of course, crucial in group hunting big mammals, but it also became important to share the responsibilities of children breeding among females, thus giving them some time to gather food and other resources. Individuals started to commit to cooperation, sharing their prey and protesting if others did not comply. A reputation connected with cooperation gradually became crucial for survival, eventually becoming a selective trait: cooperative agents could get better support in a group and thus have a higher success rate in reproduction. The process, by prompting anatomical and neurological changes like the enlargement and re-organization of the cortical areas of the brain and reduced dimorphism between males and females, gradually became self-reinforcing because it required even more cooperation: the brain took longer to develop in children, and this required more support to mothers in pregnancy and after birth; moreover, the process of domestication of human males made them more oriented towards the expectations of others.

The first migrations out of Africa, attested with *Homo erectus* (approximately 1.8 million years ago), made it possible for humans to perform a huge leap forward in the competition for food with other animals: they became able to explore the environment by overcoming distances of a completely new scale, something which required even more structured and strong cooperation and ability to communicate. The first normative notions emerged: individuals had expectations towards their partners which gradually produced the idea of cooperative duties, these duties became connected with the roles and tasks they fulfilled in the cooperative activity, and progressively these normative ideas, at first connected only with cooperative partners, became constitutive of the very idea of a group organized around collective activities, tasks, and roles. The second-person perspective became a third-person perspective: the norms started to be perceived as objective and independent from other individuals, and from personal interests, full-fledged normativity reinforced the cooperative bonds of the group. Inhibition of selfish reactions, risk assessment in cooperation, and social and emotional integration are capacities required in this process that involve elements of the brain cortex, so a further increase in brain size must be hypothesized at this stage, and it is attested in *Homo heidelbergensis*.

The crucial elements to arrive at the stage of *Homo Sapiens* are the remaining cognitive capacities that we have seen necessary for social institution properly speaking, namely, symbolization and the capacity to read others' beliefs and thus take their perspective. These are conjectured to be at the core of a cognitive evolution between 300.000 and 100.000 years ago that required an expansion of the temporal and parietal cortices, which resulted in the reshaping and globularization of the human cranium. Objects acquired symbolic value, as in the case of shell beads painted and considered to be ornaments or other objects painted to have a ritual value in burial sites, thus opening the possibility of endowing concrete things with a meaning that goes beyond their mere physical nature. This required shared representations based on the capacity to read others' beliefs (like in "we all believe that this thing X means/has value Z"),

representations made possible by an increase of phonological working memory and the emergence of semantics ("we all believe that this sound or mark X means/stands for Z"), which in turn formed the background to build symbolic, institutional hierarchies ("we all believe that X counts as Y, which has power Z")[13]. At this point, a new potential for social cooperation is opened: *Homo Sapiens* eventually became the only dominant human species.

2.4.3. *Some challenges to this theory*

I have provided above an account of the development and evolution of the cognitive features underpinning social institutions in human beings, but, as said, this is only one possible story, and one which underlines the discontinuity between humans, on the one hand, and animals—primates in particular—on the other. This discontinuity could need significant revision if we include in the picture recent theories about the possibility of (at least naïve) normativity in animals (see ANDREWS 2020; see also LORINI 2018). If mammals, and primates in particular, have developed normative capacities even in the absence of meta-representations of rules, it seems unlikely that normative capacities evolved as a cooperative break in the species *Homo*, but rather more plausible to assume a gentler continuum between the cooperative capacities of *Homo erectus* and those of our closest ancestors. Some authors argue that several elements of this account, like a sophisticated form of cooperation, recognition of the role and status of other members of the group, a capacity to attribute beliefs and intentions to others, and even a strong sense of fairness and reciprocity in interpersonal dealings is already present in chimpanzees (see DE WAAL 1998, 2006, 2014, 2016; KRUPENYE et al. 2016), while others deny it, arguing that great apes can at most do the same thing together, but not act under a shared plan in which everyone has its role (see TOMASELLO 2016, 20 ff.) and they cannot inhibit their cognition in light of the beliefs they attribute to others (see TOMASELLO & MOLL 2013, see also TOMASELLO 2020). This debate is important for our purposes because it is clear that primates do not have legal institutions in all their complexities; hence, the cognitive elements that theorists consider to be necessary and sufficient for institutions to exist cannot already be present in our evolutionary ancestors, at least not to the same degree: but what exactly is lacking? If legal institutions are argued to be based on symbolic behavior, and this last is hypothesized to require a structural reshaping of the brain that can be found only in *Homo Sapiens*, like the globularization of the human cranium, then legal institutions, as well as symbolic behavior, should be absent in animal species that do not possess those structural features. A similar problem emerges concerning the other species of the genus *Homo*: In this sense, any evidence of proto-legal practices and symbolic behavior in *Homo Neanderthalensis* (and, of course, *a fortiori* in non-human animals) would require us to significantly qualify this theoretical model (see for example LEDER et al. 2021 for recent evidence).

The ones just mentioned are empirical questions and hence possible sources of falsification for the theory, which will then have to be adapted meaningfully or discarded. However, more general issues can be raised about the *prima facie* plausibility of this reconstruction. One is connected with the problem of cognitive overload: If social institutions always require such a complex network of mutual beliefs and intentions, should we assume this cognitive machinery is present in the minds of all community members? Wouldn't this be an enormous cognitive cost for normal individuals, who have a lot of other things to do and think about in their own lives apart from supporting the community's institutional framework (for a similar objection to some theories of collective intentionality, see PACHERIE 2011)? This is an important question,

[13] See on this also ROMEO 2011, who instead phylogenetically connects symbolic behavior with normativity more in general.

which highlights the fact that collective intentionality at the basis of social institutions will inevitably end up having two features: it will never be completely collective, and it can very well be tacit and/or dispositional. In every community, a significant part of its members will not enthusiastically endorse the relevant institutions, and in most cases, they will not have a complete understanding of it: in this sense, Hart's already-mentioned distinction between officials and ordinary citizens (see above, 2.3.1) can be generalized as a distinction between a portion of the community which supports the institution by understanding and practicing it and a portion which merely goes along without resisting (as noted, the notion of "collective acceptance" can also be understood in a very weak sense to include also this passive attitude). Moreover, it is not necessary to assume that the mental states required by collective intentionality be explicit and always present in a subject's consciousness: they can rather be thought to be tacit or dispositional, not actual properties of individuals, namely, subjects could be ready to activate the relevant mental states only if the circumstances required them to do so.

A related general concern about this approach is that, while it stresses too much the role of conscious acceptance and beliefs, it seems to completely rule out a very significant portion of the inner lives of individuals, namely, emotions. Under this model, cooperation seems to emerge as a prudent evaluation of other people's cognitive states and institutions as the outcome of an endeavor that we simply decide to participate in together. But, as the primatologist Frans De Waal shows, cooperation is often the outcome of emotional empathy, namely, the outcome not so much of an understanding of others' beliefs but of feeling their frustration, rage, and possible sources of joy. Under this reading, emotions create the background for and give place to the more developed cognitive empathy and perspective-taking required for full-fledged third-person morality and legal institutions (see on this DE WAAL 2006)[14]. The main consequence of such an approach is to significantly extend the degree of cooperative capacities that humans and primates are argued to have in common. If cooperation is based on emotional contagion and not on a complex web of cognitive mental states, it is much more plausible to assume that humans are less a discontinuity in evolution than it may seem at first sight. I do not have an answer to this significant challenge, and I explicitly stated (above, 2.3.2) that the role of social emotion can very well be crucial. Still, I think that this is not a lethal blow to the overall theory. If emotional contagion is at the core of the kind of cooperation that is necessary for legal institutions to exist, it is nevertheless evident that primates do not have the kind of legal institutions and symbolic behavior that humans have. Indeed, this is a point that De Waal himself seems to concede (see above, 2.3.2).

Further, it could be argued that the one provided here is an account of the emergence of norms and social institutions that stresses too much the role of human cooperation and not sufficiently the role of domination, competition, and conflict. As Richard Wrangham conjectures, for example, normative attitudes in humans could have arisen because non-cooperative individuals were executed by coalitions of cooperative ones (the so-called «execution hypothesis»: see WRANGHAM 2019, ch. 7): in this view, cooperation, goodness, tolerance would be nothing else than the outcome of the distinctively human, amazing capacity to organize violence in more elaborated, planned forms—what Wrangham calls the «goodness paradox». As previously mentioned (see above, 2.3.1), however, the presence of hierarchies in human societies is not a counter-example but rather a support to the idea that human institutions are based on social cooperation: cooperation must also be ensured by those who created and maintained the hierarchies, among those who inflicted violence based on complex plans and rituals. The idea that human social institutions require cooperation does not in any way entail that they are inherently peaceful or egalitarian or that they cannot include regulation of violence[15].

[14] An elaborate analysis that argues for a strong connection between emotions and the emergence of norms can also be found in FITTIPALDI 2022.

[15] Wrangham himself acknowledges the crucial role of joint intentionality in his account: see WRANGHAM 2019, ch. 13.

Finally, of course, it is possible to conceive alternative explanations of the emergence of social institutions, particularly if we connect this problem with the rise of normative behavior, a much more explored topic in the literature. Apart from the already-mentioned proposals of De Waal and Wrangham, which insist on social emotions and aggressiveness, respectively, I should mention at least the recent proposals of Jonathan Birch (BIRCH 2021) and Evan Westra and Kristin Andrews (WESTRA & ANDREWS 2022). According to Birch, normative cognition is the evolutionary result of technical cognition, namely, the human capacity to build and use tools. In his view, the competitive advantage for humans in developing the ability to build complex and efficient tools required the evolution of a cognitive capacity to control performance, an affective disposition to react negatively (by way of shame or anger) in case of failure, as well as standardization of procedures. These elements, in turn, set the background for normative behavior. Hence, Birch argues, a sort of technical cooperation, namely, cooperation in developing and teaching technical skills, provided the cognitive toolbox to make normative behavior and social institutions possible. Westra and Andrews, on the other hand, advocate a pluralistic theory of the psychology of normativity: in their view, normativity is an ambiguous concept. They propose to replace the concept of norm, conceived psychologically, with that of a normative regularity, which can be based on different kinds of phenomena, cognitive underpinnings, and evolutionary history. Though Westra and Andrews argue for pluralism by isolating wonderfully different aspects of normative behavior, the most cognitively complex among these aspects seem traceable to human cooperative capabilities.

2.5. *Legal institutions involve certain additional cognitive capacities*

2.5.1. *A concept of law: legal institutions involve authority, sanctions, and validity*

I have so far provided an account of the ontogenesis and phylogenesis of social institutions. As mentioned, within these social institutions, all kinds of institutional facts can happen: facts, entities, or events endowed with a status connected with normative consequences. This characterization is very broad, however, and can include facts about games, religious rituals, social customs, and perhaps also particularly formalized versions of morality. What, then, is the peculiarity of legal facts among social-institutional facts? Is it possible to draw the boundaries of law within this vast domain? A concept of law is needed to do this, and for sure, identifying it is not easy, given that legal philosophers have been working on this problem for centuries. What I will try to do, then, is to provide a minimal concept able to avoid some of the most common pitfalls and counterexamples identified in philosophy of law, with the understanding, however, that assuming this concept as a starting point cannot but be done tentatively, and to a certain extent in a stipulative manner. The reader, then, will eventually find the overall analysis reliable to the extent that he or she will consider this starting point to be safe, and it is necessary to point out that the possibility of defining the essential features of law is a matter of debate in legal theory (for critical approaches see LEITER 2011, SCHAUER 2012; TAMANAHA 2017a; see also GIUDICE 2015).

I propose to define law as a social practice consisting of following a set of formally-valid rules that regulate social behavior through serious social pressure (typically in the form of sanctions) and that constitute the authority to create rules and apply them. This definition—Hartian in spirit—is less demanding than it may seem at first sight. First, one could object that there can be instances of law without any kind of centralized, state-like authority, as in the case of multiple coexisting sources of legal authority in medieval Europe up to the 18th century (see TAMANAHA 2017b, 105 ff.), but this is not inconsistent with the proposed definition. Law is said to organize authority, possibly also a plurality of competing legal authorities, and similarly, even though reference is made here to criteria of validity, it does not entail that these criteria must be unique and supreme. Second, we can imagine a society where law is applied to regulate

disputes but where there is no violation and hence sanctions are not necessary (a "society of angels": see RAZ 1999 [1975], 159 f.; but see HIMMA 2020, ch. 10 for a counter-argument), and this is why primary reference is made in my definition to the application of law and only "typically" to sanctions as a way to apply it[16]. Third, and finally, it could be argued against this definition that law has primarily a moral function, such as that of coordinating free choices in a coherent plan or creating a morally desirable community, and hence that institutions are neither necessary nor sufficient for law to exist: but this view—which we could label as a generic form of non-positivism—would, in any case, require rules to specify the relevant moral ideal and authorities endowed with the task of realizing it. The definition seems to be sufficiently broad to avoid several traditional problems and alternatives that have emerged in the legal-theoretical discussion on the nature of law.

One could wonder at this point whether that definition is too broad, to the point of being overinclusive: you can have rules supported by serious social pressure and authority to create and to apply them in social contexts which are different from law, such as religion, or games, or even private associations (see again, on this, TAMANAHA 2017b, 44 ff.). This, in my view, is not a counterargument to the proposed definition but rather a way of stressing an important fact, namely, that the domain of rituals, religious justification, games, and in general of social organizations are connected, from a historical and anthropological point of view, with that of law. Legal authority originally emerged from a religious background. However, religion necessarily has a transcendent and supernatural objective, and it regulates social life given this objective, which is primary in the sense that you can have completely transcendent religious practices without any kind of regulation of social life. And in the case of games, typically, game-playing practices are performed for fun, and violation of their rules may lead to protest and some kind of normative reaction, which, however, cannot be seen as organized social pressure. Private associations can be seen to have "their own" internal law, and deciding whether this is law only derivatively, namely, only when connected with State's law, is a matter of debate between legal pluralism and statualism. Modern State law typically claims superiority over other kinds of normative arrangements, which, however, does not entail, in my definition, that this is the only kind of law possible.

2.5.2. *Revenge is cognitively ancient, sanctions are not*

Given the definition of law I have assumed, we can now specify the cognitive underpinnings of those aspects that make legal institutions peculiar. One central aspect, as we saw, is serious social pressure, typically through organized sanctions. At a very basic level, punishment finds its root in personal reaction to damage or goal frustration, a reaction that can be found in other animals and that is ancient and deep from a neurological point of view, being based on the emotions of rage and disgust generated by the limbic system (see PINKER 2011, ch. 8). Basically, we have a natural attitude, that we share with primates and other animals, to react aggressively when threatened, and we also share a disposition to be disgusted by certain kinds of behaviour whose specific criteria are certainly cultural, but whose cognitive grounds can be universal (see HAIDT 2012, ch. 7, sec. 5). However, as Richard Wrangham shows, humans are comparatively less aggressive than chimpanzees, for example, at least when reactive aggressiveness is concerned, and the reason could be that reactive aggressiveness is inherently disruptive for social cooperation: so a capacity to plan reaction, rather than react instinctively, evolved as an important trait to maintain social bonds (see WRANGHAM 2019). This seems to be coherent with findings according to which the degree of reaction in humans is strongly and peculiarly

[16] On "characteristic", rather than "essential" features of law see recently POSTEMA 2022, 40.

connected with expectations; namely, it escalates only when a threshold of unexpected unfairness is surpassed, thus showing a tendency to tolerate violation—and to temper punishment—if one could expect it (see DUBREUIL 2010, 23-27).

Humans are still quite animal in formulating judgments about punishment: they may be calculators when it comes to crimes in general and in judging punishments in the abstract, but they are still emotional deontological retributivists when that crime affects them directly (see GREENE 2008), and they also tend to overestimate the damage received and to underestimate the damage done (see SHERGILL et al. 2003). However, perhaps differently from other mammals and also primates, humans show a capacity for indignation when they perceive that a norm is violated even when this violation does not directly affect them: in these cases, the motivation to punish is less strong (see FEHR & FISCHBACHER 2004), but humans show that they can adopt a third-person perspective, namely, to play the role of a third party who is not directly involved in the relationship between an offender and an offended, but who can nevertheless feel to be involved "in the name of the group" (see also above, 2.3.2 on third-person normativity). Of course, the capacity to adopt this kind of third-party perspective is crucial to experience the passage from mere reaction, or even revenge, and what we can intuitively label as a legal coercion or punishment, which also involves a kind of delegation of reaction.

As mentioned (above, 2.3.4), Richard Wrangham put forward a complex and articulated description of how this human tendency to organize sanctions and engage in cooperative, third-party support in punishing even when we are not directly affected may have emerged in the evolutionary history of humanity. According to this reconstruction, in ancient human groups, males who were more able to control their impulses and thus cooperate reacted strongly against males prone to reactive aggressiveness and violence: by way of planning and cooperation—a kind of aggressiveness that Wrangham calls "proactive", which requires strong inhibition of the limbic system on the part of the pre-frontal cortex (see WRANGHAM 2019, ch. 2)—these latter became able to form coalitions to kill the former, and this explains why a cooperative attitude towards punishment, a disposition to delegate sanctions to the group, general attention to group reputation and norm conformity was selected as an adaptive trait in humans (humans basically "auto-domesticated" themselves in this process: see WRANGHAM 2019, ch. 3, 6). This general, evolutionary explanation provides a description of organized sanctions as something disposition for which is not culturally dependent but rather genetically determined, at least in its basic coordinates, and this seems confirmed by studies according to which the propensity to punish wrongdoers, and also the identification of some core crimes such as physical aggression, takings without consent, and deception in exchange is highly cross-cultural (see ROBINSON & KURZBAN 2007, also compare HERMANN et al. 2008 for cross-cultural variations).

Thus, even though reactive aggressiveness and second-person reaction are certainly a core element of the cognitive machinery behind sanctions that we share with other animals, sanctions in the legal sense, as something centralized, delegated, organized, and performed from a third-person perspective involve impulse control, capacity to plan, trust for cooperation, a narrative about group belonging and about the legitimacy of the power which coercion is delegated to, all elements that involve complex cognitive activities that depend on cortical areas of the brain. In this sense, the human disposition to organize sanctions has a twofold cognitive foundation, resulting from dialectics between emotional and impulsive reactions located in the deeper layers of the brain, on the one hand, and complex representations based on its higher layers, on the other hand. It is important for lawyers to be aware of this double cognitive nature when discussing the nature and function of punishment. While the normative question of the role that punishment must serve in our legal system cannot be reduced to the descriptive question of what we think and perceive when processing punishment in our cognition, it seems reasonable to assume that the function of punishment in our social life cannot be detached completely from the cognitive conditions for our thinking about it.

2.5.3. *Authority is not cognitively connected with fear of sanctions but with group belonging*

Authority is a status: hence, it requires status attribution, which involves the complex cognitive machinery we saw above, including a collective, normative symbolization and high-level mind-reading. This means that authority as a core element of law is a completely human phenomenon from the cognitive perspective, requiring a great deal of high-level cortical activities. Apart from symbolism, which is indeed a cognitive precondition of authority, there is another conceptual element of this notion that was described with great analytical clarity by Joseph Raz (see RAZ 1979) and that requires an explanation in cognitive terms: delegation of power, namely, the mechanism by which authoritative pronouncements can become "exclusionary" in the reasoning of human agents, excluding other reasons for the very fact that they come from authority (provided that the authority is perceived to be legitimate). How is this mechanism possible?

Ontogenetically, there are certainly some passive elements involved in the construction of authority: preschool children tend to conflate the notion of obligation, and hence (in legal-theoretical terms) that of a "strong" reason for action, with the idea of an authority's desires (see KALISH & CORNELIUS 2007), paternal and maternal authority being of course, in this case, the paradigmatic examples, but significant cross-cultural variations have been found in the way preschoolers defer to adults' assessments and choices (see HARRIS & CORRIVEAU 2013). In adults, however, respect for authority is connected not so much with mere conformity nor with fear of sanctions but with a narrative for the authority's justification and legitimacy. Hence, the passive aspect of conformity must be complemented with an active aspect regarding the identification with a group and a set of purposes and values (see TYLER 1997, TYLER 2006). Experiments made within the paradigm provided by Stanley Milgram (see MILGRAM 1974, BURGER 2009), in which the experimenter requests subjects to deliver potentially lethal electric shocks on other humans "for the sake of science", show the effects of perception of legitimacy on behavior, and they can be interpreted in terms of identification within a group (see REICHER et al. 2012). Moreover, analyses of the famous "Stanford prison experiment" (see HANEY et al. 1973a, 1973b), in which participants were selected to play the role of prison guards against other participants acting as prisoners, argue that considerations of social identity played a strong role in justifying the impressive escalation of cruelty that guards showed (HASLAM et al. 2019). Hence, in human adults, deference to authority seems to be an essential part of a given social and normative identity—we delegate judgment to authority because we believe it to be a constitutive part of our social community and consequently of what we are. In this sense, Raz's insistence on the conceptual connection between authority, deference, and legitimation seems to be on the right track from a cognitive point of view.

It is important to bear in mind, however, that conformity to authority can also be, at least to a certain extent, a matter of unreflective and automatic habitual behavior, so even the psychology of habits can be relevant to understand this "alienated" aspect of conformity to the law[17]. The reader may recall in this connection the well-known dialectics in legal theory, traceable to Hart's *The Concept of Law*, between Austin's explanation of sovereignty as based on "habits of obedience" and Hart's alternative perspective on authority as based on the internal point of view towards rules (see HART 2012 [1961], ch. 4). This opposition can serve here, rather than simply as a legal-philosophical debate between two views that claim superiority, as a

[17] That of "habit" is a very complex notion, both philosophically and psychologically: see BARANDIAN & DI PAOLO 2014 for a conceptual map, and RAMÍREZ-VIZCAYA & FROESE 2019. secs. 1-3 for useful references. For an interesting perspective about how habits can be connected with social norms from a neurological point of view, see LORINI & MARROSU 2018.

guideline to highlight two aspects that may be complementary in understanding the psychology of legal authority.

Philogenetically, authority and hierarchies were the outcome of an evolution that has been described as distinctively U-shaped, consisting in a peak with high level of dominance based on bullyism on the part of stronger, alpha males in primitive humans closer to apes, low level of dominance in proto-egalitarian bands of hunter-gatherers, and finally, the highest level of dominance in symbolic, normative authority based on a linguistic narrative for its legitimacy in *Homo Sapiens* (see BOEHM, ch. 6). Some authors have conjectured that authority ranking is an elementary form of human relation for all social groups (see FISKE 1991) and that it emerged as an answer to the problem of groups' enlargement: given that, in bigger groups, sanctions could not be consistently applied by the collective, sub-groups were created with "chiefs" or "leaders" as their representatives, who guaranteed for the trustworthiness of the other members of the group but also had the power, delegated by the community, to punish them in case of violation (see DUBREUIL 2010, 164 ff.). Other, less functionalist and more conflictualist readings, such as the already-mentioned one provided by Richard Wrangham (above, 2.3.4, 2.4.2), trace authority simply to a kind of monopoly of violence held by a set of cooperative males over all the other members of the group—what Wrangham, tracing back to Ernest Gellner, calls «tyranny of the cousins» (see WRANGHAM 2019, ch. 8, 10). One could also conjecture, drawing from Birch's skill-centered approach (above, 2.3.4), that normative authority emerged from epistemic authority over technical skills. Alternative explanations are possible, and perhaps all of them may capture an element of truth. For sure, authority in the full-fledged human sense emerged by virtue of its organizing role. It was grounded on symbolic meta-representations, and it was backed both by a capacity to sanction deviation and an appeal to the superior epistemic and practical skills of those entrusted with power. This is shown clearly by the fact that legal authority was originally based on magical and cosmological grounds, which provided a story about the actual, superior concrete powers of those who held normative power: Marc Bloch's well-known description of the "thaumaturge kings" (*Les Rois thaumaturges*) in the Middle Ages is a vivid and wonderfully described example of this ancient, conceptual connection between normative powers and factual powers (a kind of metaphoric transformation: see ROVERSI 2016, 251-3 on this).

2.5.4. *Validity is a matter of categorization*

Law is in large part a matter of formal properties: directives must come from a certain source to be "legal", not any norm can count as law. A fundamental element of any legal reasoning consists of normative qualification, namely, in the subsumption of an individual act or fact under a general and abstract notion. From the cognitive point of view, this process, by which, for example, we come to say that this agreement counts as a valid contract for the purposes of civil law or that a given behavior counts as sexual harassment under criminal law, is not different from ordinary classification of things or events under a given category: this thing is a cat, that thing is a bird, this object is a biscuit and not a cake. Hence, cognitive theories about categorization play a crucial role in the domain of law: of course, these theories are relevant in general for any kind of epistemic activity, but in the legal domain, serious and immediate practical consequences can follow from different categorizations of the same thing.

Categorization is a huge field of research for contemporary cognitive psychology, and several different conceptions of it are available[18]. The so-called "Classical Theory of Concepts", which

[18] See MARGOLIS & LAURENCE 1999, 2019 for an introduction to the topic; MARGOLIS & LAURENCE 2015 for recent developments; see KALISH 2015 on normative concepts.

has been dominant in philosophy at least until the first half of the 20th century, is very close to the standard legal picture: concepts, in this view, are constituted by a set of necessary and sufficient conditions for their application, specified by rules in the form of definitions. This is a kind of top-down approach, in which concepts are defined in general terms as sort of platonic entities and then applied to concrete things: against this view, several other theories have emerged which showed that human categorization very likely works the other way around, namely, as a bottom-up rather than top-down process. This is the case, for example, with prototype theories (see ROSCH & MERVIS 1975, ROSCH 1978, LAKOFF 1987), according to which conceptual categorization is made by assessing degrees of similarity (or even metaphorical connections: see LAKOFF & JOHNSON 1980) between the categorized entity and some prototypical entities which are taken to be standard instances of the class; or with exemplar-based views (see MEDIN & SHAFFER 1978, WILLS et al. 2015) in which concepts are represented simply through specific instances. Another view that is radically different from the standard, classic conception is that put forward by embodied cognition theories (see BARSALOU 1999, 2008), according to which categorization consists of a re-activation of sensory and motor neural patterns that are activated when interacting with an instance of the category: in this perspective, categorization is not a kind of subsumption of the concrete under the general, nor an assessment of similarity between representations of concrete things, but rather a re-enactment of an experience of sensory and motor interaction, and this raises the problem—particularly relevant for the legal domain, where many crucial concepts are abstract—of how this experience can be sensory and interactive, rather than purely linguistic when dealing with abstract concepts (see BORGHI & BINKOFSKI 2014).

Several legal theorists have studied the effect of prototype theory on the categorization of legal acts and transactions (see PASSERINI GLAZEL 2005) and on legal interpretation in general (see ZEIFERT 2022, 2023); others have studied conceptual metaphor theory in connection with legal reasoning (see WINTER 2001, JOHNSON 2007, SARRA 2010, WOJTCZAK 2017), with the development of legal institutions (see ROVERSI 2016), or also with specific reference to single concepts (such as the concept of "standing" in U.S. constitutional law: see WINTER 1988) or specific legal domains (such as copyright law: LARSSON 2017). Recently, an experimental study has been performed within the paradigm of embodied cognition to highlight some differences between legal conceptualizations in experts and non-experts (ROVERSI et al. 2022), and embodied cognition has also been connected with the problem of the ontology of legal concepts (see JAKUBIEC 2021). Moreover, experimental jurisprudence studies show that legal officials can categorize ordinary concepts differently depending on the tools they use for assessing semantic content, whether dictionaries or linguistic corpora (see TOBIA 2020).

Should we assume that a single theory of categorization can account for all cognitive processes involved in legal reasoning? Probably not: at first sight, any subsumption of a fact under a legal concept defined by a provision seems to involve a top-down, criterion-based mechanism, whereas application of a precedent or filling a gap in the legal system seems to require a kind of analogical, prototypical reasoning from the bottom-up, but in reality, both kinds of reasoning involve elements taken from both a top-down and bottom-up approach. Perhaps the best way to describe legal categorization in cognitive terms is by way of a kind of dialectics between rule-dependent criteria and prototypical exemplars, a mechanism by which we both construe a prototype of the possible application of a provision, by reasoning about the definitions set forth in that provision, and then we assess the similarity of the case at hand with that rule-dependent prototype. In this sense, recent hybrid categorization models influenced by both rules and exemplars (see THIBAUT et al. 2018) seem best suited to account for legal, normative qualifications.

Moreover, in law categorization is almost never a "pure" cognitive mechanism but is also mediated by consequentialist reasoning about the practical outcomes of a given solution: as Hart

famously showed in his discussion of mechanical jurisprudence, that of an unreflective and automatic categorization is not an ideal that we should cherish if we want to avoid unreasonable consequences, because any judgment about how to subsume a possible case under a legal concept should be balanced against the practical purpose of the provision in which that concept is included (see HART 2012 [1961], ch. 7). This kind of balancing between conventional categories and purpose seems to be something that humans learn to perform quite early: young children tend to assume that categories in general are natural kinds, but it also seems that already since such an early stage they can also understand that some categories are conventional and can be construed in different ways depending on the goal one is aiming to achieve (see KALISH 1998). No doubt this is a crucial cognitive capacity to "think like a lawyer", using Frederick Schauer's well-known phrase (see SCHAUER 2009).

3. *Arguments for a theory of the cognitive pathologies of legal institutions*

My overview of the cognitive foundations of legal institutions and legal facts is thus complete. To summarize, legal reality is possible only when agents are capable of:

1) joint intention and joint commitment;
2) adoption of a third-personal normative perspective based on a sense of group belonging;
3) mind-reading and perspective-taking;
4) symbolic behavior and status attribution;
5) control of reactive impulses and delegation of proactive aggressiveness;
6) conceptual categorization.

Although some of these cognitive capacities—joint intention, some degree of mind-reading, inhibition of reactive impulses, and perhaps of third-person normativity—can be attributed to other animals, particularly primates, the combination of these and their most complex elements are typically human. This is why law is a distinctively human phenomenon: only we humans have the combination of cognitive features that make it possible. Other animals may show a strong degree of social behavior, even be better than us at behaving as a collective entity, but they cannot have the law.

One could say at this point that this conclusion is not striking at all—we know intuitively that other animals do not have courts, tribunals, and legislators—and that the analysis provided in this paper does nothing else than provide evidence for an obvious conclusion. One could also say that this research may have some relevance from a philosophical point of view but not from a legal point of view: after all, what difference can the cognitive grounds of legal institutions make when it comes to discussing cases in courts or to advising people about their enforceable rights or duties under an actual, positive legal system? Lawyers focus on the content of conventional norms, the "rules of the game", not the capacities of the players' minds.

As said at the beginning (above, 1), I think this kind of inquiry is very relevant for legal philosophy, particularly the traditional, ontological inquiry into the nature of law and legal practices. How can we understand what law is if not by understanding how we can create it, given that law is essentially created by the human mind? In this sense, any lawyer interested in this traditional and millenary question should consider this inquiry seriously: this does mean that he or she should accept my conclusion, of course, but at least consider "the cognitive foundations of law" as a relevant legal-philosophical topic. To be honest, if this paper sufficed to convince the reader about the necessity of an interdisciplinary approach to the ontological problem in legal theory, I would already have reached my main objective. Perhaps, however, I can give some elements to suggest that even a practical lawyer should be open to this kind of research.

To put it straight: if legal institutions and legal facts are based on the cognitive capacities of humans, then if these capacities are impaired and weakened, the legal domain gets weakened to the point of losing relevance for humans. And, given that legal facts are the kind of facts practical lawyers are experts of—the kind of facts that they must be able to describe accurately to earn their pay—if these facts gradually become less vivid and relevant for humans because their cognitive capacities to understand and support them are weakened and impaired, then this can become a problem, even for a practically-oriented lawyer. This is not surprising because ontological problems always lay the grounds for any other discussion: if these discussions' underlying ontological grounds are stable, they can go on as if ontology was irrelevant, but when those grounds change, everything else that is built on it changes. "But how can ontology change?" One may ask. "Isn't metaphysics the kind of thing that is supposed to remain stable and necessary?" Let me reply using metaphysical terminology in a somewhat mouthful way: Not the laws of grounding law may change, but the grounding base of law. What I mean is that even if one identifies the kind of human attitudes and capacities necessary for legal rules, institutions, and systems to exist, nothing rules out the possibility that those attitudes and capacities get weakened and even go out of existence.

What I have in mind here is similar to what Hart called a "pathology" of a legal system. In Chapter 6 of the *Concept of Law* (see HART 2012 [1961], 117 ff.), Hart describes his theory of the foundations of law and presents in detail his concept of legal validity based on the well-known "rule of recognition", which consists of a social rule and hence of a kind of behavioral attitude of legal officials. Also, in the same chapter, Hart explains at length how the existence of a complex legal system requires that some social facts obtain, namely, at least that officials, for the most part, adopt an internal point of view toward the rule of recognition and that ordinary people, for the most part, obey primary legal rules of obligation. In case these facts do not obtain, or in case some ambiguities arise on the fact that they obtain, Hart says, a situation emerges that can be labeled as "pathological" for law: «a breakdown in the complex congruent practice which is referred to when we make the external statement of fact that a legal system exists. There is here a partial failure of what is presupposed whenever, from within the particular system, we make internal statements of law. Such a breakdown may be the product of different disturbing factors» (HART 2012 [1961], 117 f.). I submit that, just as in Hart a change in the sociological factors that underlie the legal system can represent a pathology for it, so can a change, weakening, or impairment in the cognitive capacities that make legal institutions possible. In these cases, we could talk about "cognitive pathologies" of law, legal systems, and legal institutions.

What could these "cognitive pathologies" be? Of course, it depends on whether the analysis provided above is reliable. Still, if it is, it provides an interesting and somewhat illuminating guide about what we should nurture and protect as lawyers, apart from normative rights, rules, and policies. A cognitive pathology of law could be a widespread incapacity on the part of members of the legal community to understand that in supporting the law— their legal community, or even the international or global legal community, depending on context—they are involved in a collective endeavor, and that they are so involved for a reason that relates to what they are, to their normative identity as persons. Another cognitive pathology of law could be a reduced ability to understand the symbolic nature of rituals, objects, and roles: the fact that some objects, persons, and behaviors can *mean* much more than what they concretely are and that they do so mean because we collectively support that meaning. Finally, an incapacity to control our reactive impulses, delegate reactions at a collective level, and take the perspective of others is certainly pathological for law. As we have seen, our understanding of normative frameworks crucially depends on our ability to see what others think, to see what they believe and intend to do, and to understand that they can act on beliefs that we consider false but they take to be true. This requires a capacity to control the impulse to make our perspective

completely dominant in our mind, focus on the social dimension as an integral part of our individual reasoning, and of course, postpone the immediate pleasure of reaction and revenge.

4. *Conclusion… with a normative question*

In this work, I have provided an analysis of the cognitive underpinnings for the existence of legal facts and legal institutions. I have argued that, for legal facts to obtain, there must necessarily be humans capable of joint intention, joint commitment, capacity to understand the group narrative of the legal community, symbolic thinking and thus status-attribution, mind-reading and perspective-taking, self-control and inhibition of reactive aggression. I have described how these capacities emerge in human children and how they may have emerged in the evolution of the genus *Homo*, from *Homo erectus* to *Homo sapiens*: thus, of the ontogenesis and phylogenesis of legal institutions. Finally, I have concluded that if these cognitive capacities are necessary for legal institutions to exist, a weakening or impairment of these cognitive capacities can undermine their existence and that a theory of the cognitive pathologies of law is, therefore, possible and useful for lawyers.

These are mostly descriptive questions, but they introduce a normative one: Should we avoid, as a collective, nurturing the cognitive pathologies of law? Of course, this depends on how much we value our legal framework and, more in general, law as a kind of social organization. If we assume that these things have a value—and lawyers, for the most part, make this assumption—then there are some statements framed as a "should" that seem to follow from the analysis I have provided, statements that hold for lawyers first, and more in general for all people who find legal institutions in their community to be valuable.

First, given that an incapacity for joint intentions, commitment, and perception of the group narrative is a cognitive pathology of legal institutions, lawyers should work to protect and cultivate the sense of a "we" behind a legal community: A purely formalistic attitude, that reduces the law to a set of normative structures, procedures, and rules that are meant to be treated scientifically, so to say, or technically, could thus in the long run weaken, or even impair a perception that is crucial for the law to exist. Such a formalistic attitude should at least be complemented with a narrative about why law, and its formal procedures, should be accepted as an integral part of our collective and individual normative identity.

Second, given that symbolic behavior and status attribution is crucial for legal institutions, lawyers should avoid, and even fight against, any kind of reductionistic attitude about symbols, because the legal domain is, in part, the outcome of a hypostatization of symbols. Symbols are there for a reason, and this reason relates to a group narrative, which again can be the constitutional narrative of our legal community or even, depending on context, of a broader, international, and global community. In any case, a purely instrumental attitude, according to which all legal meanings are eventually reduced to a practical outcome in terms of loss and gains, could turn out to be pathological for law in a cognitive sense, namely, because it could elicit a mode of thinking that in the long run threatens the very existence of legal institutions. Moreover, lawyers should remember that law is made of meanings, hence of language, and thus should support all the collective endeavors that protect linguistic culture, high linguistic capacities, as well as capacities for abstract and symbolic thinking in their legal community: conversely, they should resist and problematize all social changes that may impair these capacities, for example by gradually replacing writing and speaking with pictorial, image-oriented modes of thinking. Mere visual perception is insufficient to convey symbolic legal meanings: from a merely visual perspective, banknotes and contracts are only pieces of paper, and a Parliament is nothing but a building or a bunch of people. Thinking in terms of words, and not simply in terms of images, is a cognitive ability that is

essential for the existence of law, an ability that we should nurture rather than dismiss if we want to keep legal institutions strong.

Finally, given that mind-reading, perspective-taking, and inhibition of aggressive, reactive impulses is essential for legal institutions, lawyers should fight against any social and economic tendency to nurture reactive impulses and disseminate an image of self-control as a kind of weakness or incapacity to enjoy life. In this sense, «limbic capitalism», conceived as «a technologically advanced but socially regressive business system in which global industries [...] encourage excessive consumption and addiction» (COURTWRIGHT 2019, 6), can also be interpreted—and perhaps not so intuitively—as a threat for the very existence of law and legal institutions.

These normative conclusions are certainly tentative, speculative, and they highly depend on the reliability of the theory I have provided. The general argument, however, seems to me sound: if indeed it is possible to frame a theory of the cognitive foundations of law and legal institutions, and if we consider these last to be valuable, then we should protect the cognitive abilities that are necessary for them to exist. In a recent and wonderful book, Gerald Postema argued that an essential part of the ideal of the Rule of Law is people's fidelity to it, namely,

> «a general willingness to submit to law's governance and to give deference to its limits and requirements [...]. It is not enough that people believe in the rule of law and see it as "a necessary and proper aspect of their society." Most crucially, the rule of law needs, in addition, the active engagement of officials and citizens in holding each other to their responsibilities under the law» (see POSTEMA 2022, 66).

Understanding the Rule of Law is a complex cognitive phenomenon because it requires us to make sense of the highly symbolic, abstract, and intangible idea that Laws—and not dangerous persons able to threaten people by way of punishment—are Sovereign. If Postema is right, and if the analysis of the cognitive foundation of legal institutions that I have provided is accurate, at least in its essentials, and, finally, if we consider the Rule of Law to be a value, then it seems that the battle for fidelity to the law must be fought not simply on a political and social level. It must be fought, at a deeper level, in people's minds.

References

ANDREWS K. 2020. *Naïve Normativity: The Social Foundation of Moral Cognition*, in «Journal of the American Philosophical Association», 6, 1, 36 ff.

BAILLARGEON R., HE Z., SETOH P. 2013. *False-Belief Understanding and Why It Matters*, in BANAJI M., GELMAN S. (eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, Oxford University Press, 88 ff.

BARANDIAN, X.E., DI PAOLO E.A. 2014. *A Genealogical Map of the Concept of* Habit, in «Frontiers in Human Neuroscience», 8, Article 522.

BARSALOU L. W. 1999. *Perceptual Symbol Systems*, in «Behavioural Brain Sciences», 22, 577 ff.

BARSALOU L. W. 2008. *Grounded Cognition,* «Annual Review of Psychology», 59, 617 ff.

BICCHIERI, C. 2006. *The Grammar of Society*, Cambridge University Press.

BINMORE, K. 2005. *Natural Justice*, Oxford University Press.

BIRCH, J. 2021. *Toolmaking and the Evolution of Normative Cognition*, in «Biology and Philosophy», 36, 4 ff.

BOBBIO N. 2011. *Giusnaturalismo e positivismo giuridico*, Laterza. (Originally published in 1965)

BOEHM C. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behaviour*, Harvard University Press.

BORGHI A., BINKOFSKI F. 2014. *Words as Social Tools: An Embodied View on Abstract Concepts*, Springer.

BRATMAN M. 1992. *Shared Cooperative Activities*, in «The Philosophical Review» 101, 327 ff.

BRATMAN M. 1999. *Shared Intention and Mutual Obligation*, in ID., *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge University Press, 130 ff.

BRATMAN M. 2014. *Shared Agency: The Planning Theory of Acting Together*, Oxford University Press.

BRATMAN M. 2022. *Shared and Institutional Agency: Toward a Planning Theory of Human Practical Organization*, Oxford University Press.

BRIGAGLIA M. 2022. *Regole: un saggio di psico-deontica*, «L'Ircocervo», 21, 210 ff.

BRIGAGLIA M., CELANO B. 2018. *Reasons, Rules, Exceptions: Towards a Psychological Account*, iin «Analisi e diritto 2017», 131 ff.

BRIGAGLIA M, CELANO B. 2021. *Constitutive Rules: The Symbolization Account*, in «Ratio Juris», 34, 241 ff.

BROŻEK B. 2013. *Rule-Following: From Imitation to the Normative Mind*, Copernicus Center Press.

BURAZIN L., HIMMA K. E., ROVERSI C. (eds.) 2018. *Law as an Artifact*, Oxford University Press.

BURGER J. 2009. *Replicating Milgram: Would Still People Obey Today?*, in «American Psychologist», 64, 1 ff.

BUTTELMANN D. 2013. *Selective Imitation of In-Group Over Out-Group Members in 14-Month-Old Infants*, in «Child Development», 84, 422 ff.

CANALE D. 2014. *Is Law Grounded in Joint Action?*, in «Rechtstheorie», 45, 289 ff.

CHIASSONI P. 2021. *The Law and Cognitive Sciences Enterprise: A Few Analytic Notes*, in BROŻEK B., HAGE J., VINCENT N.A. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 490 ff.

CONTE A. G. 1995. *Paradigmi di analisi della regola in Wittgenstein*, in ID., *Filosofia del linguaggio normativo. Vol. 2*, Giappichelli, 265 ff.

COURTWRIGHT D.T. 2019. *The Age of Addiction: How Bad Habits Became Big Business*, Harvard University Press.

DE WAAL F. 1998. *Chimpanzee Politics: Power and Sex among Apes* (2nd revised Ed.), Johns Hopkins University Press.

DE WAAL F. 2006. *Primates and Philosophers: How Morality Evolved*, ed. by S. Macedo, J. Ober, Princeton University Press.

DE WAAL F. 2014. *Natural Normativity: The 'Is' and 'Ought' of Animal Behavior*, in «Behaviour», 151, 185 ff.

DE WAAL F. 2016. *Apes Know What Others Believe*, in «Science», 354, 39 ff.

DOHERTY M., PERNER J. 1998. *Metalinguistic Awareness and Theory of Mind: Just Two Words for the Same Thing?* «Cognitive Development», 13, 279 ff.

DUBREUIL B. 2010. *Human Evolution and the Origins of Hierarchies: The State of Nature*, Cambridge University Press.

EPSTEIN B. 2018. *Social Ontology*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Available on: https://plato.stanford.edu/entries/social-ontology/.

FEHR, E., FISCHBACHER U. 2004. *Third Party Punishment and Social Norms*, «Evolution and Human Behavior», 25, 63 ff.

FISKE A. P. 1991. *Structures of Social Life: The Four Elementary Forms of Human Relations*, Free Press.

FITTIPALDI E. 2016. *Leon Petrażycki's Theory of Law*, in PATTARO E., ROVERSI C. (eds.), *Legal Philosophy in the Twentieth Century: The Civil Law World. Tome 2: Main Orientations and Topics*, Springer, 443 ff.

FITTIPALDI E. 2022. *Norma. Una proposta di concettualizzazione per la sociologia del diritto e le altre scienze sociali*, LED Edizioni universitarie.

FRANCHINI L.F. 2021. *Genetic Mechanisms Underlying Cortical Evolution in Mammals*, in «Frontiers in Cell and Developmental Biology», 9, 591017.

GILBERT M. 1989. *On Social Facts*, Princeton University Press.

GILBERT M. 2000. *Social Rules: Some Problems with Hart's Account, and an Alternative Proposal*, In ID., *Sociality and Responsibility: New Essays in Plural Subject Theory*, Rowman and Littlefield Publishers.

GILBERT M. 2006. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*, Clarendon Press.

GILBERT M. 2014. *Joint Commitment: How We Make the Social World*, Oxford University Press.

GILBERT M. 2018. *Rights and Demands: A Foundational Inquiry*, Oxford University Press.

GIUDICE M. 2015. *Understanding the Nature of Law: A Case for Constructive Conceptual Explanation*, Edward Elgar Publishing.

GÖCKERITZ S., SCHMIDT M.F.H., TOMASELLO M. 2014. *Young Children's Creation and Transmission of Social Norms*, in «Cognitive Development», 30, 81 ff.

GOMEZ-LAVIN J., RACHAR M. 2019. *Normativity in Joint Action*, in «Mind & Language», 34, 97 ff.

GOMEZ-LAVIN J., RACHAR M. 2021. *Why We Need a New Normativism about Collective Action*, in «Philosophical Quarterly», 72, 2, 478ff.

GRÄFENHAIN M., BEHNE T., CARPENTER M., TOMASELLO M. 2009. *Young Children's Understanding of Joint Commitments*, in «Developmental Psychology», 45, 1430 ff.

GRÄFENHAIN M., CARPENTER M., TOMASELLO M. 2013. *Three-Year-Olds' Understanding of the Consequences of Joint Commitments*, in «PLoS ONE», 8, e73039.

GREENE J.D. 2008. *The Secret Joke of Kant's Soul*, in SINNOTT-ARMSTRONG W. (ed.), *Moral Psychology. Vol. 3. The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, The MIT Press, 35 ff.

GUALA F. 2016. *Understanding Institutions: The Science and Philosophy of Living Together*, Princeton University Press.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A.J. 2001. *Inside the Judicial Mind*, in «Cornell Law Review», 86, 777 ff.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A.J. 2007. *Blinking on the Bench: How Judges Decide Cases*, in «Cornell Law Review», 93, 1 ff.

HAGE J. 2018. *Two Concepts of Constitutive Rules*, in «Argumenta», 4, 21 ff.

HAGE J. 2022. *Rules and the Social World*, in «L'Ircocervo», 21, 2, 16 ff. Available on: https://lircocervo.it/wp-content/uploads/2023/01/02-Hage_Rules-and-the-Social.pdf.

HAGE J., WALTERMANN A. 2021. *Responsibility, Liability, Retribution*, in BROŻEK B., HAGE J., VINCENT N.A. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 255 ff.

HÄGERSTRÖM A. 1953. *Inquiries into the Nature of Law and Morals* (ed. by K. Olivecrona, trans. By C.D. Broad), Almqvist & Wiksells.

HAIDT J. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, Penguin.

HAMANN K., WARNEKEN F., TOMASELLO M. 2012. *Children's Developing Commitments to Joint Goals*, in «Child Development», 83, 137 ff.

HANEY C., BANKS C., ZIMBARDO P.G. 1973a. *Interpersonal Dynamics in a Simulated Prison*, in «International Journal of Criminology and Penology», 1, 69 ff.

HANEY C., BANKS C., ZIMBARDO P.G. 1973b. *Naval Research Reviews: A Study of Prisoners and Guards in a Simulated Prison*, Office of Naval Research.

HARRIS P.L., CORRIVEAU K.H. 2013. *Respectful Deference. Conformity Revisited*, in BANAJI M., GELMAN S. (eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, Oxford University Press, 122 ff.

HART H.L.A. 2012. *The Concept of Law* (3rd Ed.), Clarendon Press. (Originally published in 1961)

HASLAM S.A., VAN BAVEL J.J., REICHER S.D. 2019. *Rethinking the Nature of Cruelty: The Role of Identity Leadership in the Stanford Prison Experiment*, in «American Psychologist», 74, 809 ff.

HAUN D.B.M., TOMASELLO M. 2011. *Conformity to Peer Pressure in Preschool Children*, «Child Development», 82, 1759 ff.

HERMANN B., THONI C., GÄCHTER S. 2008. *Antisocial Punishment Across Societies*, in «Science», 319, 1362 ff.

HINDRIKS F. 2005. *Rules and Institutions: Essays on Meaning, Speech Acts and Social Ontology*, Haveka BV.

HINDRIKS F., GUALA F. 2015. *Institutions, Rules, and Equilibria: A Unified Theory*, in «Journal of Institutional Economics», 11, 459 ff.

HOFMANN M.B. 2021. *The Psychology of the Trial Judge*, in BROŻEK B., HAGE J., VINCENT N.A. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 165 ff.

HORNER V., WHITEN A.K. 2005. *Causal Knowledge and Imitation/Emulation Switching in Chimpanzees (Pan Troglodytes) and Children*, in «Animal Cognition», 8, 164 ff.

HOUSE B. R., SILK J. B., HENRICH J., BARRETT H. C., SCELZA B. A., BOYETTE A. H., HEWLETT B. S.,

McElreath R., Laurence S. 2013. *Ontogeny of Prosocial Behaviour across Diverse Societies*, in «Proceedings of the National Academy of Sciences of the United States of America», 110, 14586 ff.

Jakubiec M. 2021. *Legal Concepts as Mental Representations*, in «International Journal for the Semiotics of Law», 35, 1837.

Johnson M., 2007. *Mind, Metaphor, Law*, in «Mercer Law Review», 58, 845 ff.

Kalish C. W. 1998. *Natural and Artifactual Kinds: Are Children Realists or Relativists about Categories?*, in «Developmental Psychology», 34, 376 ff.

Kalish C.W. 2015. *Normative Concepts*, in Margolis E., Laurence S. (eds.), *The Conceptual Mind: New Directions in the Study of Concepts*, The MIT Press, 519 ff.

Kalish C.W., Cornelius R. 2007. *What is to be Done? Children's Ascriptions of Conventional Obligations*, in «Child Development», 78, 859 ff.

Kalish C.W., Lawson C.A. 2008. *Development of Social Category Representations: Early Appreciation of Roles and Deontic Relations*, in «Child Development», 79, 577 ff.

Kapitany R., Hampejs T., Goldstein T.R. 2022. *Pretensive Shared Reality: From Childhood Pretense to Adult Imaginative Play*, in «Frontiers in Psychology», 13, 774085.

Kelly D., Stich S., Haley K., Eng S., Fessler D. 2007. *Harm, Affect, and the Moral/conventional Distinction*, in «Mind and Language», 22, 117 ff.

Kenward B. 2012. *Over-imitating Preschoolers Believe Unnecessary Actions Are Normative and Enforce Their Performance by a Third Party*, in «Journal of Experimental Child Psychology», 8, 195 ff.

Kinzler K.D., Corriveau K.H., Harris P.L. 2011. *Children's Selective Trust in Native-accented Speakers*, in «Developmental Science», 14, 1, 106 ff.

Krupenye C., Kano F., Hirata S., Call J., Tomasello M. 2016. *Great Apes Anticipate that Other Individuals Will Act According to False Beliefs*, in «Science», 354, 110 ff.

Lakoff G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press.

Lakoff G., Johnson M. 1980. *Metaphors We Live By*, University of Chicago Press.

Larsson S. 2017. *Conceptions in the Code: How Metaphors Explain Legal Challenges in Digital Times*, Oxford University Press.

Leder D, Hermann R, Hüls M., et al. 2021. *A 51,000-Year-Old Engraved Bone Reveals Neanderthals' Capacity for Symbolic Behaviour*, in «Nature Ecology & Evolution», 5, 9, 1273 ff.

Leiter B. 2011. *The Demarcation Problem in Jurisprudence: A New Case for Scepticism*, in «Oxford Journal of Legal Studies», 31, 663 ff.

Lewis D. K. 1969. *Convention: A Philosophical Study*, Harvard University Press.

Löhr G. 2022. *Recent Experimental Philosophy on Joint Action: Do We Need a New Normativism about Collective Action?*, in «The Philosophical Quarterly» 72, 3, 754 ff.

Lorini, G. 2000. *Dimensioni giuridiche dell'istituzionale*, CEDAM.

Lorini G. 2018. *Animal Norms: An Investigation of Normativity in the Non-Human Social World*, in «Law, Culture, and the Humanities», 18, 3, 652 ff.

Lorini G., Marrosu F. 2018. *How Individual Habits Fit/Unfit Social Norms: From the Historical Perspective to a Neurobiological Repositioning of an Unresolved Problem*, in «Frontiers in Sociology» 3, Article 14.

Lorini G., Żełaniec W. (eds.) 2018. *The Background of Constitutive Rules*, «Argumenta: Journal of Analytic Philosophy», Special issue. Available on: https://www.argumenta.org/article/background-constitutive-rules-introduction/.

LYONS D.E., YOUNG A.G., KEIL F.C. 2007. *The Hidden Structure of Overimitation*, in «Proceedings of the National Academy of Sciences», 104, 19751 ff.

MACCORMICK N. 2007. *Institutions of Law: An Essay in Legal Theory*, Oxford University Press.

MARGOLIS E., LAURENCE S. 1999. *Concepts and Cognitive Science*, in MARGOLIS E., LAURENCE S. (eds.), *Concepts: Core Readings*, The MIT Press.

MARGOLIS E., LAURENCE S. (eds.) 2015. *The Conceptual Mind: New Directions in the Study of Concepts*, The MIT Press.

MARGOLIS E., LAURENCE S. 2019. *Concepts*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). Available on: https://plato.stanford.edu/entries/concepts/.

MEDIN D.L., SHAFFER M.M. 1978. *Context Theory of Classification Learning*, in «Psychological Review», 85, 207 ff.

MICHAEL J. 2011. *Shared Emotions and Joint Action*, in «Review of Philosophy and Psychology», 2, 355 ff.

MICHAEL J., PACHERIE E. 2015. *On Commitments and Other Uncertainty Reduction Tools in Joint Action*, in «Journal of Social Ontology», 1, 89 ff.

MICHAEL J., SEBANZ N., KNOBLICH G. 2016. *The Sense of Commitment: A Minimal Approach*, in «Frontiers in Psychology», 6, 1968.

MILGRAM S. 1974. *Obedience to Authority: An Experimental View*, Harper and Row.

MILLER S. 2001. *Social Action: A Theleological Account*, Cambridge University Press.

NOYES A., KEIL F.C., DUNHAM Y. 2018. *The Emerging Causal Understanding of Institutional Objects*, in «Cognition», 170, 83 ff.

OLIVECRONA K. 1971. *Law as Fact* (2nd Ed.), Stevens.

ONISHI K. H., BAILLARGEON R. 2005. *Do 15-Month-Old Infants Understand False Beliefs?*, in «Science», 308, 255 ff.

PACHERIE E. 2011. *Framing Joint Action*, in «Review of Philosophy and Psychology», 2, 173 ff.

PASSERINI GLAZEL L. 2005. *La forza normativa del tipo: teoria della categorizzazione e pragmatica dell'atto giuridico*, Quodlibet.

PATTARO E. 2016. *Axel Hägerström at the Origins of the Uppsala School*, in PATTARO E., ROVERSI C. (eds.), *Legal Philosophy in the Twentieth Century: The Civil Law World. Tome 2: Main Orientations and Topics*, Springer, 319 ff.

PETRAŻYCKI A. 1955. *Law and Morality*, ed. by N.S. Timasheff, Harvard University Press.

PLUNKETT D., WODAK D. 2022. *The Disunity of Legal Reality*, in «Legal Theory», 28, 235 ff.

POSTEMA G. 2022. *Law's Rule: The Nature, Value, and Viability of the Rule of Law*, Oxford University Press.

RAKOCZY H., TOMASELLO M., STRIANO T. 2005a. *On Tools and Toys: How Children Learn to Act on and Pretend with 'Virgin' Objects*, in «Developmental Science», 8, 57 ff.

RAKOCZY H., TOMASELLO M., STRIANO T. 2005b. *How Children Turn Objects into Symbols: A Cultural Learning Account*, in NAMY L. (ed.), *Symbol Use and Symbol Representation*, Erlbaum.

RAKOCZY H., WARNEKEN F., TOMASELLO M. 2008. *The Sources of Normativity: Young Children's Awareness of the Normative Structure of Games*, in «Developmental Psychology», 44, 875 ff.

RAMÍREZ LUDEÑA L., VILAJOSANA J.M. (eds.) 2022. *Reglas constitutivas y derecho*, Marcial Pons.

RAMÍREZ-VIZCAYA S., FROESE T. 2019. *The Enactive Approach to Habits: New Concepts for the Cognitive Science of Bad Habits and Addiction*, in «Frontiers in Psychology», 10, 301.

REICHER S.D., HASLAM A., SMITH J.R. 2012. *Working Toward the Experimenter: Reconceptualizing Obedience within the Milgram Paradigm as Identification-Based Followership*, in «Perspectives on Psychological Science», 7, 315 ff.

ROCKENBACH B., MILINSKI M. 2006. *The Efficient Interaction of Indirect Reciprocity and Costly Punishment*, in «Nature», 444, 718 ff.

ROBINSON P.H., KURZBAN R. 2007. *Concordance and Conflict in Intuitions of Justice*, in «Minnesota Law Review», 91, 2007, 1829 ff.

ROMEO F. 2011. *Some Aspects of the Evolution of Legal Norms in the Lower Pleistocene*, in «JusIT», 2011, 1 ff.

ROSCH E. 1978. *Principles of Categorization*, in ROSCH E., LLOYD B.B. (eds.), *Cognition and Categorization*, Hillsdale, 27 ff.

ROSCH E., MERVIS C.B. 1975. *Family Resemblances: Studies in the Internal Structure of Categories*, in «Cognitive Psychology», 7, 573 ff.

ROSSANO F., RAKOCZY H., TOMASELLO M. 2011. *Young Children's Understanding of Violations of Property Rights*, in «Cognition», 121, 219 ff.

ROSS A. 1957. *Tû-tû*, in «Harvard Law Review», 70, 812 ff.

ROVERSI C. 2016. *Legal Metaphoric Artefacts*, in STELMACH J., BROZEK B., KUREK Ł. (eds.), *The Emergence of Normative Order*s, Copernicus Center Press, 215 ff.

ROVERSI C. 2018. *Constitutive Rules and the Internal Point of View*, in «Argumenta: Journal of Analytic Philosophy», 4, 139 ff.

ROVERSI C. 2021. *In Defense of Constitutive Rules*, in «Synthese», 119, 14349 ff.

ROVERSI C., UBERTONE M., VILLANI C., D'ASCENZO S., LUGLI L. 2022. *Alice in Wonderland: Experimental Jurisprudence on the Internal Point of View*, in «Jurisprudence: An International Journal of Legal and Political Thought», DOI: 10.1080/20403313.2022.2109884.

SÁNCHEZ BRIGIDO R. 2010. *Groups, Rules and Legal Practice*, Springer.

SARRA C. 2010. *Lo scudo di Dioniso. Contributo allo studio della metafora giuridica*, Franco Angeli.

SCHAUER F. 2009. *Thinking like a Lawyer: A New Introduction to Legal Reasoning*, Harvard University Press.

SCHAUER F. 2012. *On the Nature of the Nature of Law*, in «Archiv für Rechts- und Sozialphilosophie», 98, 457 ff.

SEARLE J.R. 1995. *The Construction of Social Reality*, The Free Press.

SEARLE J.R. 2010. *Making the Social World*, Oxford University Press.

SHAW A.W., LI W., OLSON K.R. 2013. *Reputation is Everything*, in BANAJI M., GELMAN S. (eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, Oxford University Press, 220 ff.

SHERGILL S.S., BAYS P.M., FRITH C.D., WOLPERT D.M. 2003. *Two Eyes for an Eye: The Neuroscience of Force Escalation*, in «Science», 301, 187 ff.

SLAUGHTER V. 2015. *Theory of Mind in Infants and Young Children: A Review*, in «Australian Psychologist», 50, 169 ff.

TAMANAHA B. 2017a. *Necessary and Universal Truths about Law?*, in «Ratio Juris», 30, 3 ff.

TAMANAHA B. 2017b. *A Realistic Theory of Law*, Cambridge University Press.

THALER R.H., SUNSTEIN C.S. 2008. *Nudge. Improving Decisions about Health, Wealth, and Happiness*, Yale University Press.

THIBAUT J.-P., GELAES S., MURPHY G. L. 2018. *Does Practice in Category Learning Increase Rule Use or Exemplar Use—or Both?*, in «Memory and Cognition», 46, 530 ff.

TOBIA K. 2020. *Testing Ordinary Meaning*, in «Harvard Law Review», 134, 726 ff.

TOH K. 2022. *Collectivity and Law*, in MARQUES T., VALENTINI C. (eds.), *Collective Action, Philosophy and Law*, Routledge, 25 ff.

TOLLEFSEN D. 2004. *Let's Pretend! Children and Joint Action*, in «Philosophy of the Social Sciences», 35, 75 ff.

TOMASELLO M. 2016. *A Natural History of Human Morality*, Harvard University Press.

TOMASELLO M. 2020. *The Many Faces of Obligation*, in «Behavioral and Brain Sciences», 43, E89.

TOMASELLO M., CARPENTER M., CALL J., BEHNE T., MOLL H. 2005. *Understanding and Sharing Intentions: The Origins of Cultural Cognition*, in «Behavioural and Brain Sciences», 28, 675 ff.

TOMASELLO M., MOLL H. 2013. *Why Don't Apes Understand False Beliefs?*, in BANAJI M., GELMAN S. (eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, Oxford University Press, 81 ff.

TUOMELA R. 2016. *Social Ontology: Collective Intentionality and Group Agents*, Oxford University Press.

TURIEL E. 1983. *The Development of Social Knowledge: Morality and Convention*, Cambridge University Press.

TYLER T.R. 1997. *The Psychology of Legitimacy: A Relational Perspective on Voluntary Deference to Authorities*, in «Personality and Social Psychology Review», 4, 323 ff.

TYLER T.R. 2006. *Why People Obey the Law*, Princeton University Press.

ULLMANN-MARGALIT E. 1977. *The Emergence of Norms*, Clarendon Press.

WARNEKEN F., TOMASELLO M. 2009. *Varieties of Altruism in Children and Chimpanzees*, in «Trends in Cognitive Science» 13, 2009, 397 ff.

WELLMAN H.M. 2018. *Theory of Mind: The State of the Art*, in «European Journal of Developmental Psychology», 15, 6, 728 ff.

WELLMAN H.M., CROSS D., WATSON J. 2001. *Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief*, in «Child Development», 72, 655 ff.

WESTRA E., ANDREWS K. 2022. *A Pluralistic Framework for the Psychology of Norms*, in «Biology and Philosophy», 37, 40 ff.

WILLS A.J., INKSTER A.B., MILTON F. 2015. *Combination or Differentiation? Two Theories of Processing Order in Classification*, in «Cognitive Psychology», 80, 1 ff.

WINTER S. 1988. *The Metaphor of Standing and the Problem of Self-Governance*, in «Stanford Law Review», 40, 1387 ff.

WINTER C.L. 2001 *A Clearing in the Forest*, The University of Chicago Press.

WOJTCZAK S. 2017. *The Metaphorical Engine of Legal Reasoning and Legal Interpretation*, Beck.

WRANGHAM, R. 2019. *The Goodness Paradox: The Strange Relationship between Virtue and Violence in Human Evolution*, Pantheon Books.

ZEIFERT M. 2022. *Rethinking Hart: From Open Texture to Prototype Theory—Analytic Philosophy Meets Cognitive Linguistics*, in «International Journal for the Semiotics of Law», 35, 409 ff.

ZEIFERT M. 2023. *Basic Level Categorisation and the Law*, in «International Journal for the Semiotics of Law», 36, 227 ff.

The Nature of Law and Constructivist Facts

JAAP HAGE

1. *Introduction*

The cognitive sciences are not only relevant for understanding the mental processes in individual persons; they also provide insight in how minds play a role in the construction of social reality and, as a special case, of law. Law is based on events such as legislation and judicial decisions. Even where law has grown 'spontaneously' in social interactions, such as the technical standards for the Internet (CALLIESS & ZUMBANSEN 2010), it only has become law in the full sense because it was adopted in legislation, court decisions, or both. At the same time, values and reason have always played some role in legal reasoning. In the 13th century, the theologian and philosopher Thomas Aquinas had defined natural law as a prescript of reason. Human legislation would only be binding if it is in accordance with this natural law. An important 20th century legal philosopher argued that 'law' that does not even aspire to be just, is not law at all (ALEXY 1992; ALEXY 2002).

Historically, there has been a close connection between law and reason, often in the shape of morality and justice. At the same time, law has a firm foundation in what people do, in social practices. A heavily debated question is what the law is if our social practices appear to be unreasonable or unjust. As we will see, some authors—for instance HART 2012—have argued that our social practices are decisive for what the law is. If these practices are unjust, we have unjust or bad law. Just like a bad car is still a car, unjust law would still be law. These authors are sometimes called *legal positivists* (GREEN & ADAMS 2019). Others have claimed that (very) unjust law is not law at all (RADBRUCH 1945; ALEXY 1992; ALEXY 2002), or at best a defective kind of law (FINNIS 1980, 364). The latter authors call themselves *non-positivists*, or—if they also believe that there is a natural law to which the positive law must conform—*natural law theorists*.

The debate between positivists and non-positivists may seem to be merely verbal. Why not stipulate that the word *law* will only be used for rules that are reasonable, or that the word will be used for social phenomena of a particular kind, whether they are reasonable or not? However, if the debate were merely verbal, it is difficult to understand why it has continued for such a long time. There must be more to it than a mere disagreement about the use of a word. From a jurisprudential perspective, the central question of this contribution is *why there can be an ongoing debate about the role of reason in law*. An answer to this question will also increase our insight into the nature of law.

The first part of this contribution, sections 2 and 3, focuses on a question that has been central in legal philosophy since the 20th century. It is the question of how law can have an

* The author thanks the European Union for sponsoring the Recognise project which facilitated the research for this contribution and the anonymous reviewer whose comments made it possible to improve the draft version.

independent existence, next to morality, while at the same time legal reasoning seems to have moral aspects.

The second part, sections 4 and 5, argues for an answer to this question. This answer is that law is a part of social reality that consists of what will be called *constructivist facts* and that this characteristic explains the ambiguous nature of law. To justify this answer, it will be necessary to explain how social reality and constructivist facts exist. Here, it will become clear how minds play a crucial role in the construction of social reality and law.

This contribution contains advanced teaching materials. Therefore, it provides many links to debates in the philosophy of law of the 20th century and to the growing theory of how social reality exists (social ontology). At the same time, this contribution strives for a clear overview of what is at stake. This means that its argument must sometimes remain on the surface, sacrificing scientific precision to clarity of exposition.

2. *The background of the debate*

2.1. *Hume's guillotine*

For a proper understanding of many debates in legal philosophy, including the debate about the nature of law, it is necessary to understand *Hume's guillotine*. Simply stated, this guillotine boils down to the thesis that it is impossible to derive judgments about what ought to be done from only statements about the facts (HUME 1978, 469). For example, it is not possible to derive that John ought to be punished from the statement that John is a thief. The metaphorical guillotine separates what ought to be done from the facts. To connect the two, a premise needs to be added. In the example, this premise might be that thieves ought to be punished. However—and this is important—the additional premise is an ought judgment and not an ordinary statement of fact.

Hume's guillotine is a central element of an elaborate theory about the world. According to this theory, the world consists of a large collection of facts, but these facts only concern what is the case, not what ought to be done. In the world itself, there is no ought. What ought to be done is added to the world by human beings—or, for some, by God. Moreover, this addition does not depend on what the facts in the world are. Suppose that the world contains the fact that John is a thief. Humans can add both that John ought to be punished, or that John ought not to be punished, or nothing at all. What ought to be done is logically independent of what is the case. Any ought can be combined with any set of facts. And therefore, a description of the world, however elaborate, cannot determine what ought to be done.

The impossibility to derive Ought from Is—because that is what we are talking about—is the necessary consequence of a particular view of the world. According to this view, the world may contain many facts, but it does not determine what ought to be done. What ought to be done is a human addition. In an argument, this human addition takes the form of an ought judgment, such as the judgment that thieves ought to be punished (RAILTON 2006).

2.2. *The open question argument*

A similar story can be told about value judgments. The world may contain many facts, but these facts do not include any evaluation. So, it may be a fact in the world that this is a football match with many goals, but it cannot be a fact that this is a good match. The world contains facts, but not the valuations thereof. What is good or what is bad, is not given with the world. Just like what ought, or ought not to be done, valuations are human additions, and these additions are not determined by the facts.

The addition of value to the world sometimes takes the shape of accepting standards. In the case of football matches, a simplistic standard may be that matches with many goals are good. If this standard is adopted, matches with many goals are also good matches. However, any standard is only provisional, or—to state it with different words—contestable (GALLIE 1956), or open to debate. The standard for good football matches does not define what 'a good football match' means. In contrast, the meaning of a word is a matter of convention. If people agree on what a word means, the word has this meaning. If somebody contests that this kind of animal is a horse, he[1] thereby shows that he does not know what horses are, or—which boils down to the same thing— what the word 'horse' stands for. If a football lover does not wish to count matches with many goals as good ones, he does not exhibit a lack of knowledge of what 'good football match' means; he merely disagrees with the standard in use.

This insight—that the standards for valuation do not give the meanings of evaluative words—underlies the so-called *open question argument* (MOORE 1903, sections 6-17; RIDGE 2019). An open question is a question that may sensibly be asked without showing a misunderstanding of the topic at stake. It makes sense to ask whether a football match with many goals is 'really' good. However, it makes no sense to ask whether this kind of animal, a typical horse, really is a horse.

An open question, in the technical sense at issue here, shows possible disagreement with a standard for evaluation. Similar open questions are also possible regarding ought judgments: any concrete ought judgment based on a general rule can be questioned by someone who disagrees with the rule. Later, we will see that some also consider questions about the content of law as open questions in this sense.

2.3. *Objective facts and objective knowledge*

For a proper understanding of the debates in legal philosophy, the stories about Hume's guillotine and the open question argument need to be supplemented by a story about objective knowledge. This third story starts from a philosophical doctrine about the world and our knowledge of it. According to this doctrine, the world exists independent of what we believe about it. The world is, to state it with a technical term, *mind-independent*. For instance, the world contains the fact that Mount Everest is more than 8,000 metres high, and this is a fact whether people believe it or not. The mind-independence of the world and the facts in it explains why the world does not contain an ought or valuation, as what ought to be done or what is good or bad depends on human minds.

The philosophical doctrine at issue is called *ontological realism*.[2] This is the doctrine that the facts do not depend on what we believe them to be. It is the natural way to think about physical objects and their characteristics. That is the reason why most people are ontological realists about the physical facts. We consider physical facts to be objective, where 'objective' stands for the same mind-independence that ontological realism advocates.

Assuming that there are objective facts in the world, we would like to have objective knowledge about them. Objective knowledge is knowledge that depicts the facts as they really are, without distortion (REISS & SPRENGER 2020). Distortion can easily occur, for example if we 'look' at the universe by means of a radio telescope. Such a telescope provides us with data which need to be processed, to become meaningful to human beings. Such processing is nowadays performed by a computer. If it takes place in a wrong way, the result is distorted, and our knowledge is not objective.

[1] To avoid cumbersome formulations such as 'he/she' or 'her/him', I will use pronouns that reflect the gender of the author, in my case therefore the male form. I can only encourage female authors to use female pronouns.

[2] An accessible, although also contestable, description of ontological realism can be found in SEARLE 1995, 149-157. See also MILLER 2021.

Another example of knowledge that is not objective is knowledge about the social world. If we follow the famous sociologist Max Weber, true knowledge of the social world presupposes a recognition—*Verstehen*—of the meanings that social events have (WEBER 1921-22). For example, you can only understand why people dress in black when they visit a place on the border of a town if you know that they attend a funeral and that the colour black symbolizes mourning. However, this meaning of dressing in black is not mind-independent and our meaning-based understanding of a funeral cannot be a mind-independent representation of the facts. If Weber was right, objective knowledge is not possible in the social sciences.

This example about a funeral also illustrates that there are some facts for which ontological realism seems less attractive. Facts with meaning in the social world are a case in point, but perhaps also mathematical and moral 'facts', or facts about the law. The very idea of objective knowledge, knowledge that represents the facts as they really are, only makes sense if it is presupposed that this knowledge concerns objective, mind-independent facts (HAGE 2022b). If such facts are not available, objective knowledge is impossible. On the assumption that ought judgments do not express facts, it is not possible to have objective knowledge of what ought to be done. Not because our knowledge of what ought to be done is distorted, but because there is not even a fact to obtain knowledge about.

Facts that are not objective are often called 'subjective'. In section 4, we will see that the opposition between what is objective and what is subjective is not as simple as suggested by the picture sketched in the present section.

3. *The debate in the 20th century*

With the doctrines about the separation of Is and Ought (Hume's guillotine), the open question argument, and the objectivity of the world and our knowledge of it in place, we can have a look at the work of four important 20th century legal philosophers.

3.1. *Kelsen and the Grundnorm*

At the beginning of the 20th century, Hans Kelsen developed a view of law, the *Pure Theory of Law*, in which he tried to set out the proper domain of legal science in contrast to ethics and to the social sciences (KELSEN 1934). The contrast with ethics lies in the fact that legal norms express a special kind of 'ought', namely a legal ought. Ethics, which studies morality, would deal with norms that express a moral ought.

The contrast with the social sciences is, according to Kelsen, that law consists of norms. Norms are not the proper domain of the social sciences, which—as sciences—can only deal with facts. Legal facts, such as the fact that John's act is a case of theft, are not objective. These facts have a legal meaning which does not exist in the objective world. The meaning stems from the classificatory norm that taking away a good that belongs to somebody else has the legal meaning of being theft.

Some actions, such as legislation, or—in the common law—issuing a court decision, have the meaning of creating a legal norm. They derive this meaning from a 'higher' norm, which gives the law creating actions their specific meaning. For instance, the norm that the municipality council can issue norms for the municipality gives some actions of the council the legal meaning of creating norms. If the council has created a norm, the norm exists. However, this legal meaning can only exist because of another norm which gave the decision of the community council the meaning of creating a norm. Legal meaning does not exist in the objective world.

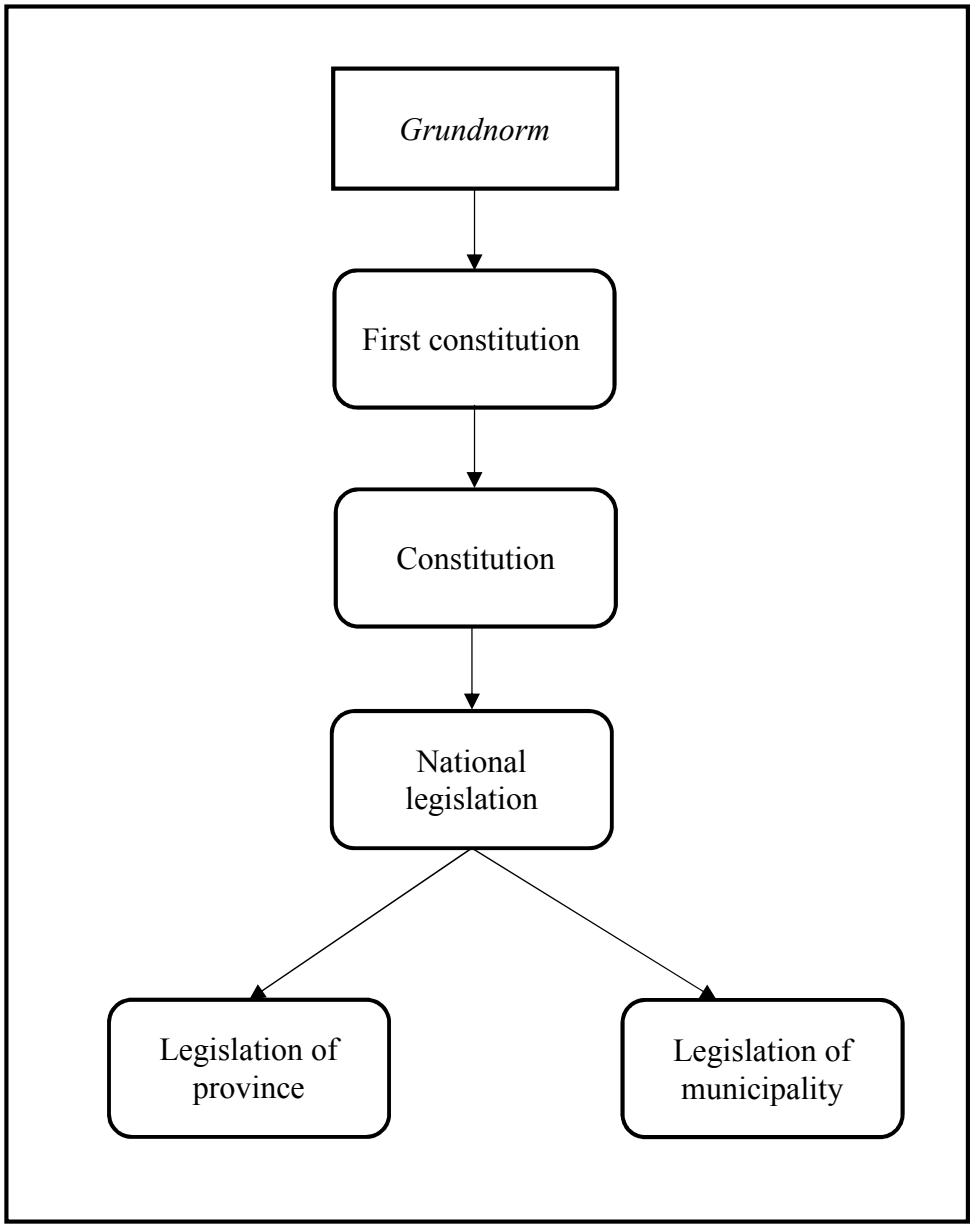


Figure 1: A legal system, based on a Grundnorm

According to Kelsen, created norms can only exist because of a 'higher' norm which attributes actions the meaning of creating a new norm. All legal norms have been created in this way. In other words, the existence of every legal norm presupposes the existence of a 'higher' norm.

All norms that stem from one and the same 'higher' norm belong to the same legal system. For instance, all norms that were created by the community council belong in this way to the same legal system. The norm that attributes the community council the power to create new norms may itself be based on a norm from the national legislator which attributes legislative powers to 'lower' legislative bodies. The latter norm was created by legislation of the national legislator. Moreover, the national legislator derived its power to create this norm from the constitution. The maker of the constitution derived his power from an earlier constitution, the maker of which in turn derived her power from an earlier constitution, and so on... All the norms that stem from the same constitution belong to the same legal system. See figure 1.

In the Kelsenian picture of a legal system, all norms were made on the basis of some other norm. It is easy to see that this picture has a problem: who created the 'highest' norm of the system and which norm gave this action the legal meaning of norm creation? The first of these two questions can still be answered: the 'highest' norm was formulated in the first constitution and was therefore created by the makers of this first constitution. However, which norm gave these makers the legal power to make the constitution?

Kelsen's answer to this question was original, but also controversial. According to Kelsen, the norm that empowered the makers of the first constitution was not created at all. It is a presupposition of the whole legal system. A legal system can only exist if there is a 'highest' empowering norm which makes the creation of the other norms of the system possible. Kelsen called this 'highest' norm the *Grundnorm* of the system. No legal system can exist that does not have such a *Grundnorm*. If the people of a country take it that they have a legal system, they are committed to the existence of a *Grundnorm* which validated the creation of all other norms of the system. The legal system then consists of precisely those norms that stem from this *Grundnorm*.

According to Kelsen, legal systems consist of norms. These norms are not facts, and all the facts in the world by themselves cannot determine what the norms are. To have norms, it is necessary to add normativity to the world of facts. The function of the *Grundnorm* in Kelsen's theory is to add this normativity to the world of facts. However, Kelsen did not really answer the question how this addition took place. He only stated that if people assume that there are norms, they are committed to some normative input. In a hierarchical legal system, this input can only be provided by a 'highest' norm that provides all the norm creating actions with their status of norm creation. Such a *Grundnorm* cannot really exists as a matter of objective fact. Still, according to Kelsen, there cannot be legal norms without a *Grundnorm*. Therefore, the belief that there are legal norms commits to the belief that there is such a *Grundnorm*.

We see that Kelsen's distinction between facts and norms becomes a separation. All the facts in the world cannot determine what the norms are. This reflects Hume's guillotine and its underlying picture of the world. Because the facts nevertheless seem to determine the norms, the question what the norms are remains open. Kelsen does not answer the question what normativity is, and neither does he explain where normativity stems from.

3.2. *Alf Ross*

Alf Ross, a student of Kelsen, found the Kelsenian construction of a *Grundnorm* that adds normativity to the world of facts but which itself does not exist as a matter of fact, problematic. In several books (Ross 1946, 1959), he criticised Kelsen's view. Ross adopted Kelsen's view that legal norms provide facts with a specific legal meaning. However, where Kelsen considered this meaning as something that lies outside the world of facts, Ross gave it a firm position inside the world of facts. In a central passage of his book *On Law and Justice*, Ross wrote: «A national law system, considered as a valid system of norms, can accordingly be defined as the norms which actually are *operative in the mind of the judge*, because they are *felt* by him to be socially binding and therefore obeyed» (Ross 1959, 34; italics added).

This emphasis on what goes on in the mind of a judge, and on what a judge feels about norms, is a major step away from Kelsen. Kelsen would emphasize that mental events and feelings are facts in the world—with which Ross would agree—and can therefore not provide the meaning which characterizes legal facts. This latter step is exactly what Ross rejects. In Ross' view, the meaning of legal facts precisely *is* what goes on in the minds of judges.

The facts do have meanings in the real world, be it that the meanings lie in the psychological states of human beings, which are then projected onto the external world. If this sounds as a strange doctrine, a parallel with colours may elucidate Ross' view. The objective—in the sense of mind-independent—world does not contain colours, but 'only' electro-magnetic waves of frequencies

within a specific range. These waves, if they impinge on our eyes, cause our experiences of colour. Moreover, we ascribe these colours to the objects that reflected the waves to our eyes. Colours are the result of the interaction between our minds and the objective world and are in that sense mind-dependent. Nevertheless, we ascribe the colours to the objects in the world. Ross can be interpreted as claiming the same for legal meaning: it is mind-dependent, but nevertheless ascribed to—projected onto—the world. He refused to subscribe to the world view according to which the world consists of meaningless facts. This is a crucial step towards recognising the role of the cognitive sciences in understanding what law is. *I will return to it in the second part of this contribution, when recognition is discussed as a condition for the existence of basic social facts.*

3.3. *Hart's rule of recognition*

Another step was taken by Hart. In his book *The Concept of Law*, he reacted to both Kelsen's theory about the *Grundnorm* and Ross' psychological alternative. First, he criticised the psychological theory for ignoring the normative aspect of legal rules. To this purpose, Hart introduced the example of a gunman threatening a passer-by: give me your money, or I will shoot you! Because of this threat, the passer-by is obliged to hand over the money. However, as Hart emphasizes, she is not under an obligation to hand over the money. The threat creates a psychological necessity (obliged), not a duty or obligation. Similarly, if a judge feels bound by a set of rules, or knows that legal subjects feel bound, this does not suffice for the legal validity of these rules. Law is not the gunman situation writ large (HART 2012, 83 f.).

However, the Kelsenian account, with a presupposed *Grundnorm*, would not be realistic either. Law has a foundation in social reality, even though this foundation is not that judges or legal subjects feel obliged to comply with legal rules. Hart's alternative is a two-step test (HART 2012, 100-109):

1. Legal rules are identified by means of a rule of recognition, which points out what the sources of law are. Rules that stem from a thus identified source are legal rules; other rules are not.
2. The rule of recognition is itself not a valid legal rule and does not stem from a source of law. Instead, it is a *social rule*.

The first step is familiar to all lawyers. Legal rules stem from a source of law, such as legislation, conventions (treaties), or judicial decisions. The second step addresses the question of why precisely those are the sources of law. Hart claims that it depends on the legal system at issue what the sources are. Every legal system is defined by its own rule or recognition. This rule exists because the officials of the system—read: the courts—use it to identify the sources of law. This use involves both application of the rule, as well as a critical reflective attitude towards the behaviour of all officials including oneself. The user of a social rule is motivated to use the rule, encourages relevant others to use the rule, and criticizes himself and these others in the case of unjustified non-use. This normative social practice distinguishes Hart's rule of recognition from the individual psychology of judges proposed by Ross. *It will return to it in the second part of this contribution, in the discussion of rule-based social facts.*

3.4. *Dworkin's criticism*

In the 1970s, Dworkin criticized Hart's views by arguing that there is more law than what is identified by the rule of recognition (DWORKIN 1978, 14-45). To this purpose he used a case that was decided by a New York court (*Riggs v. Palmer*, 115 N.Y. 506, 22 N.E. 188 (1889)). The case was about a young man called Elmer, who murdered his grandfather to prevent the grandfather

from changing his last will and disinheriting Elmer. The question in law was whether Elmer could still inherit from the grandfather he murdered.

At the end of the 19th century, New York law had no rules that specifically dealt with such situations. The rules that could be based on the local rule of recognition only indicated that Elmer could inherit because of his grandfather's last will, which obviously had not been modified. Nevertheless, the New York court refused Elmer his inheritance. This decision was justified by invoking the principle that nobody should benefit from his own wrong. As Dworkin pointed out, legal principles such as the principle used by the New York court, are not identified as law by a rule of recognition. So, it seems that Elmer's case empirically refuted Hart's theory.

However, that conclusion can only be drawn if the principle invoked by the court was law indeed. An alternative interpretation of the event is that the New York court refused to apply the law, as the law would be wrong. The court itself adopted the former interpretation of what it did: it was applying New York law and the principle was part of the law that was then valid in New York. However, the self-interpretation of the court is not necessarily decisive for what happened.

Both the 'empirical' refutation of Hart's theory that all law stems from a rule of recognition which exists as a social rule, and the alternative interpretation that the court only refused to apply valid law because it was unjust will return in the second part of this contribution, in the discussion of constructivist social facts.

4. *Law as a part of social reality*

4.1. *Intermezzo*

The 20th century debate that was described in section 3 deals with a central question: to what extent can law be a social phenomenon and to what extent is an infusion of 'normativity' (morality) necessary. Kelsen claimed that the normativity of law could not stem from the world of facts, but that the infusion of normativity could not be an input from morality either, because the legal ought differs from the moral ought. The input of normativity could only stem from a presupposed *Grundnorm*. Ross, on the contrary claimed, that the normativity of law is nothing other than mental phenomena projected onto the external world. The world of facts includes the normativity of law, be it that this normativity is in the end nothing other than mental facts. Hart amended Ross' view by replacing the role of mental facts by a social practice of rule following. However, if law is a social practice of rule following, the question remains how this social practice can be contested. How can the social practice of identifying law by means of a rule or recognition be criticized as being against the law? Apparently, the view that law is a rule-based social practice is contradicted by this practice itself. Dworkin rejected Hart's view for ignoring this insight.

In the second part of this contribution, the debate of the 20th century will be considered from the perspective of what is presently called *social ontology*. Social ontology deals with the question of how social reality exists[3]. If law is a part of social reality, social ontology should provide us with answers to at least some of the questions that legal theorists raised.

The following account of social ontology starts from basic social facts, building up to basic rule-based facts and facts based on rule-based rules. The account ends with the introduction of constructivist facts, which are fundamentally open to questioning. The jurisprudential debate about the role of morality in law can to a large extent be interpreted as a debate on whether legal facts are constructivist facts.

[3] Important general work on social ontology comprises the studies of the great classical sociologists (a.o. Marx, Weber, Durkheim, Mead, Parsons and Luhmann), BERGER & LUCKMANN 1966, GILBERT 1989, SEARLE 1995 and SEARLE 2010, TUOMELA 2002. Especially relevant from a legal perspective are a.o. RUITER 1993 and 2001, LAGERSPETZ 1995, RAZ 1979 and RAZ 1986, MARMOR 2009.

4.2. *Objective, subjective, and social facts*

As a first indication of what social facts are, it is useful to contrast them to objective facts and subjective 'facts'. Objective facts are facts that exist independent of whether people believe in their existence. Examples of objective facts are the facts that the Pacific Ocean contains water, that 3+5 equals 8, or that most human beings have lungs. Because objective facts do not depend on what people believe, they are the same for everybody, even if people have different beliefs about them.

Subjective 'facts' are completely personal. Many people would not even want to call them 'facts', because of their personal nature. Examples are that Carla is a pretty woman, that chocolate tastes good, or that Johann Sebastian Bach was a better composer than Paul Hindemith. Subjective facts depend completely on what individual people believe or find, and in this sense, everybody has her own subjective 'reality'.

Social facts take an intermediate position between objective and subjective facts. They depend on what people recognise, but not on individual recognition. For every individual person, social facts are like objective facts: they exist whether they recognise them or not. If a single Belgian would not recognise Philippe as the king, he would be wrong if almost all other Belgians recognise Philippe. However, for a whole group, social facts are like subjective facts: it depends on what the members of the group believe these facts are. So, if (almost) all Belgians would change their minds about their king, Philippe would not be the king of the Belgians anymore.

This means that different groups of people may disagree on what the social facts are, for instance on whether international law allows the Russians to claim the Ukraine as belonging to Russia. Different countries may have different systems of positive law. Somebody who is guilty of the crime of 'unnatural sexual behaviour' in one country, may be innocent of a crime in another country.

A brief digression is in place here. It may be argued that something is objectively a crime according to the law of country A, but not a crime according to the law of B. Facts that are relative to a social group become objective by mentioning that group in the description of the fact ('a crime according to French law').[4] On this line of thinking, social and even subjective facts may be made objective by including the standard on which they are based in the description of the facts. However, if this approach is taken, a useful distinction for understanding the world collapses. Here, this approach will only be mentioned, but not explored any further.

4.3. *Basic social facts*

Many social facts are the products of rules—they are rule-based, or institutional facts (MacCormick & Weinberger 1986)—but for a proper understanding of social reality it is better to start with *basic social facts*. As a first approximation, basic social facts exist in a group if sufficiently many members of that group believe that they exist. For instance, Henrike is the leader of the Maastricht Cycling Club (MCC) if most members of that club believe that Henrike is their leader.

For some kinds of basic social facts, it does not suffice if sufficiently many members of a group *believe* them. Suppose, for instance, that all members of MCC believe that Henrike is their leader, but do not attach any consequences to this belief. If Henrike proposes to take a break during a cycling trip, this proposal is treated like the proposal of any other club member. To have a leader for a club means not only that sufficiently many club members believe that she is the leader, but also that they attach the relevant consequences to this believed leadership. What these consequences are, depends on how the notion of leadership is given content in the

[4] Thanks to Manuel Atienza, who emphasized the relevance of this point for me.

group, but there cannot be leadership without any consequences. If a person is to be the leader of a social group, the members of the group must accept her leadership. We find that for some kinds of basic social facts mere belief satisfies, while other kinds of basic social facts can only exist if their consequences are accepted. It is convenient to introduce a technical term that stands for belief if that is all that is required for the existence of a basic social fact, but that includes acceptance if that is an additional requirement. I propose to use the words 'recognise' and 'recognition' to this purpose.

Suppose that all members of MCC recognise that Henrike is the leader of the club but are not aware that the other members do the same. Individually, they are all disposed to follow the directives of Henrike about destinations, breaks, or departure times, but at the same time they are surprised that the others also follow Henrike's suggestions. In such a case, we cannot say that Henrike is (already) the leader of the group. More is needed, and this more includes that the group members should be aware that Henrike fulfils the same function for most other members that Henrike fulfils for them personally. It should not only be the case that the members of MCC recognise Henrike as their leader, they should also believe that the other members also recognise Henrike as leader of the group, and that these other members have the same beliefs about their fellow cyclists. In other words, a group member P should not only have beliefs about Henrike, but also about what her fellow group members recognise, including what her fellow group members believe about the beliefs of P herself. The beliefs of P should include indirectly—namely via the beliefs of her group fellows—meta-beliefs about her own beliefs. I will call these beliefs *reflexive meta-beliefs*; reflexive because these meta-beliefs are also about the believer's own beliefs (LEWIS 1969, 52-57; RUBEN 1985, 105-119; LAGERSPETZ 1995; TUOMELA 2010, 66).

These reflexive meta-beliefs mark the difference between the views of Ross and Hart about the existence of rules. Ross focused on the beliefs and feelings of individual persons, while Hart emphasized the importance of a social practice. Part of what defines such a social practice is the existence of reflexive meta-beliefs, which represent the step from individual mental states and individual psychology to social practices, institutional reality, and sociology.

Reflexive meta-beliefs are crucial for the existence of basic social facts, but there are other important meta-beliefs. Imagine we are in the early Middle-Ages, a time at which most people believed Earth to be flat. Moreover, these people not only believed Earth to be flat, but also that others believed both that Earth is flat and that all others believed the same. In other words, sufficiently many members of the group consisting of all humanity believed Earth to be flat and also entertained the reflexive meta-beliefs necessary for Earth's being flat to exist as a basic social fact. Nevertheless, we cannot say that in the early Middle-Ages, Earth was flat as a matter of basic social fact. The reason is that another important meta-belief was lacking. Most likely, also in the early Middle-Ages, people believed that whether Earth is flat is a matter of objective fact. In those days, people would not only believe that Earth was flat, but also that if in a different society people would believe that Earth was round, those people would be wrong, objectively wrong. People who thought that the flatness of Earth was something to be established by shared recognition in a group would not understand what kind of fact the shape of Earth was.

Something similar would hold if one member of MCC claimed that, even though all other members recognise Henrike as their leader, Henrike was objectively speaking not their leader. This single person would show by his claim that he did not understand that being the leader of an informal club is a matter of basic social fact and that it does not make sense to talk about 'objective leadership'.

The point that these examples are meant to illustrate is that something can only exist as a particular kind of fact (objective, social, or subjective) *if it is a social fact that it belongs to this kind of fact*. So, Henrike can only be the leader of MCC if sufficiently many members of MCC recognise her as the leader *and* if they consider leadership to be a matter of social fact.

4.4. *Rule-based rules and their consequences*

Rules attach 'new' facts to existing facts or to events[5]. For instance, they attach the fact that Biden is commander-in-chief of the USA army to the fact that he is the president of that country. They attach the fact that Meryl owns a book to the event that this book was validly transferred to her. Or they attach the obligation that Kristin pays €150 to Ove to the event that Kristin broke Ove's vase. I will call these new facts 'rule-based facts'.

There are two kinds of rule-based facts, namely facts based on social rules and facts based on rule-based rules. The most basic form of existence for rules is existence as a social rule.[6] A social rule exists in a group if sufficiently many members of the group are disposed to recognise the rule consequence if they believe the facts of the rule conditions. For instance, if most people in Belgium are disposed to recognise the person to whom a property was validly transferred as the (new) owner of the property, then the social rule exists in Belgium that the person to whom a property was validly transferred has become the (new) owner of this property. Another example deals with a duty-imposing (mandatory) rule. If most members of the Maastricht Cycling Club normally recognise the duty to do what the leader of the club told them to, the social rule exists in MCC that if the leader says that you must do something, you have a duty to do it.

Notice that this definition of when a social rule exist applies to both mandatory rules, which prescribe behaviour, and other rules, such as classificatory rules, or competence conferring rules. Notice also that the definition does not mention anything about the way the rule came into existence. It may have been created in a legislative procedure and then the rule is both a social rule and a rule-based rule; these two do not exclude each other. The rule may also have grown in a social practice with the rise of dispositions to associate facts and the rise of mutual expectations.

A direct consequence of this characterisation of social rules and their existence is that members of a group who do not recognise the consequences of social rules make a mistake. Rationally speaking, they ought to recognise these rule-based facts. For example, somebody makes a mistake if he answers 8 to the question what the sum of 4 and 2 is. Or somebody who greets another person by taking off her shoes makes a mistake, at least in Europe. Such a perceived mistake may lead to social pressure to recognise the rule-consequences, and to (self)criticism (HART 2012, 88 f.). On one hand, social rules depend for their existence on the recognition of their consequences, but on the other hand they determine for individual group members what they ought to recognise.

This is a second aspect of the transition from mental states and individual psychology to social practices and sociology that marks the difference between the views of Ross and Hart. Ross emphasized individual dispositions to act in a particular way, while Hart focused on the social interactions that characterize social rules.

One special kind of rule-based fact are the facts which involve that a particular thing exists. The things that exist because their existence is a rule-based fact, will be called 'rule-based things'. Let us assume that in football the rule exists that the team that has scored the most goals is the winner of the football match. On the assumption that Team A scored more goals than Team B, this rule brings about that Team A is the winner of the football match. The winner exists because of the mentioned rule, and the winner is, therefore, a rule-based 'thing'.

There are many rule-based things, such as the Dean of the Law School, a bank note, a property right, a legal obligation, and the United Nations. However, for our present purpose,

[5] A more elaborate account of rule-based facts and rules can be found in HAGE 2022c.

[6] The present characterisation of social rules deviates substantially from the classic circumscription in HART 2012, 55-61.

the most relevant rule-based things are rules that exist because they were created. Rule-based (or institutional) rules can exist anywhere in a society where there are rules that specify how new rules can come about. For instance, MCC may have the rule that the finance committee of the club can make rules on the yearly contribution. If the finance committee uses this power, it makes rule-based rules. However, most rule-based rules exist in law.

In contrast to social rules, rule-based rules can generate consequences which are not broadly recognised. In the Netherlands such a rule arguably exists regarding traffic lights for pedestrians. If a traffic light for pedestrians is red, pedestrians are not allowed to cross the street. However, many pedestrians ignore the rule. They do not even feel bound by it and make their crossing behaviour dependent on the intensity of the traffic, or the presence of a police officer. Despite this massive lack of recognition, the rule about traffic lights does create legal duties and a pedestrian who crosses the street while the traffic light is red violates the law and becomes liable to be punished. This liability can exist without being recognised.

More in general, facts based on rule-based rules can exist without being recognised, although this is a relatively exceptional phenomenon. The phenomenon is theoretically important, however, because it illustrates how facts in social reality can exist without being recognised.

4.5. *Ought-facts*

Objective facts are, by definition, mind-independent. In contrast, social facts depend on the minds of the members of a social group. Of course, it is possible to define facts in a narrow way, to allow only objective facts as 'real' facts, but that would also exclude the existence of the United Nations as a fact, as well as the facts that John the thief is *punishable* (a disposition), that it is *certain* that the train will be late, or that Harry *cannot* play chess (both modalities). Many more kinds of fact than we may initially think are mind-dependent, including the existence of all dispositions and all modalities[7]. These facts are not objective, and if we disallow them to be facts, the notion of a fact will become much narrower than our present use of it.

As soon as it is recognised that facts do not need to be objective, but that they can also be mind-dependent, it should become easier to accept the idea that some facts are about what ought to be done. Ought-facts are mind-dependent, but they are nevertheless facts. Perhaps some ought-facts are purely subjective, for instance if I impose some duties on myself. However, most ought-facts exist independent of whether you personally recognise them. Those facts are social facts; they exist in social reality (HAGE 2022a).

The refusal to recognise ought-facts as facts in the world is characteristic for Kelsen's view on law and distinguishes his view from the views of Ross and Hart. Ross recognised projected mental states as parts of reality, while Hart did the same for social practices. Because of his refusal, Kelsen was forced to adopt the view that the normativity of law is only presupposed. Both Ross and Hart could account for the existence of law as a normative practice, by recognising mind-dependent elements in the world.

However, if some facts depend on what people recognise as existing, how is it possible to seriously question these facts? If you know that practically all members of MCC recognise Henrike as their leader, you cannot seriously question whether Henrike is the leader. However, in law this seems to be different. Even if there is a social practice according to which Meryl owns this copy of *Pride and Prejudice*, it seems still possible to question whether Meryl really owns the book. Perhaps the legal practice is wrong, but how is that possible if law exists as a social practice? Here constructivist facts enter the picture.

[7] On modalities and their mode of existence, see WHITE 1975.

5. *Constructivist facts*

5.1. *Introducing constructivist facts*

Suppose that the members of MCC take a vote on what was the best cycling trip they made this year and that they decide unanimously that the trip to the castle gardens in Arcen was the best trip. Does this mean that the Arcen trip really was the best trip? No, even if all club members agree on what was the best trip, this does not mean that it *really* was the best trip. Perhaps all club members mistakenly felt that the trip with the best weather was the best trip, while some reflection would have made them prefer the trip with the largest number of flat tires, because that trip created the strongest personal ties between the members.

There seems to be a difference between what most or even all members of the group accept as the best trip and what really was the best trip. It is interesting to compare this with the leadership of the club. Suppose that MCC does not have a leader that was designated by a rule, but that all members of the club recognise Henrike as their informal leader. In this case, Henrike *is* their informal leader, independent of the quality of the reasons why she is recognised as the leader. In this case of informal leadership, bad reasons bring about that Henrike is the leader for bad reasons. In the case of the best cycling trip, bad reasons for preferring one trip over another mean that the preferred trip was not really the best one. Such facts, which depend on the best reasons rather than on broad recognition, are called *constructivist facts*[8].

5.2. *The possibility of questioning*

Constructivist facts are social facts that exist because they are recognised as existing or are based on a rule-based rule, but which are nevertheless open to questioning[9]. How is this combination possible? The answer is easy: the social practice of a group does not only recognise the existence of these facts, but also the possibility to question them. The possibility of questioning is a social fact, just as much as the questionable fact itself. If the possibility of questioning a particular kind of fact does not exist in social reality, facts of that kind are perhaps social facts, but not constructivist facts.

This is quite abstract, and it may be useful to consider some examples. Let us start with the best cycling trip and assume that, if interviewed about it, all members of MCC would mention the trip to the castle gardens of Arcen as last year's best trip. Moreover, they do not believe this to be an objective matter, because it depends on their appreciation of the trip, but neither do they believe the issue is purely subjective. And, finally, they do not only consider this trip the best one of the year but they also believe that the other club members feel the same and also know that their feelings are shared by the others. In short, it is a basic social fact in MCC that the trip to the castle gardens of Arcen was the best trip of the year. That is: in first instance, because they also agree and know that the others also agree that, theoretically speaking, everybody might be mistaken. If somebody came up with convincing reasons why another trip was even better, this other trip would turn out to be even better and their original judgment would turn out wrong. The value judgment about the best cycling trip expresses a constructivist fact.

A second example deals with rule-based duties. Assume that the legal system of France has the rule that car drivers must halt at red traffic lights. This rule was created by means of a statute and exists as a matter of rule-based fact. Zala drives her car and approaches a red traffic

[8] The idea of constructivist facts has mainly been developed in the philosophy of mathematics (IEMHOF 2020), moral theory (BAGNOLI 2021), and legal philosophy (DWORKIN 1986; SOETEMAN 2009).

[9] This definition captures most constructivist facts, but not all. In theory, a constructivist fact also exists if it logically ought to be recognised. Implicitly, this will be explained at the end of this subsection.

light. The rule imposes on her a duty to halt; at least that would be the normal situation. However, Zala brings a severely injured person to the hospital, and it is important that they arrive there as soon as possible. This is a reason to make an exception to the general traffic rule and to conclude that Zala should not stop for the traffic light (assuming, of course, that she does not cause a collision by driving on)[10]. This example differs in two respects from the former one about the best cycling trip. First, it deals with a normative judgment—what should Zala do?—rather than with a value judgment. And second, the normal social fact, namely that Zala should stop, would be a rule-based fact, rather than a basic social fact. However, these two differences do not make a difference; it is still possible to argue that the normal situation does not occur, by adducing convincing reasons to this effect.

Let us abstract from these examples. Constructivist facts are characterized by the possibility to have a *serious* debate about them. 'Serious' means in this connection that the participants in the debate believe that it is possible to disagree about these facts without thereby showing a misunderstanding of what the debate is about. For instance, if Joanna and Frédéric disagree whether red wine is better than white wine, while they also believe it is just a matter of taste, they consider the issue at stake to be a merely subjective one and their difference of opinion not to be serious. If two members of MCC disagree about whether Henrike is their leader, while both know that practically all members of the club accept Henrike as their leader, their disagreement is not serious either. The reason is that not believing that Henrike is the leader while also believing that 'everybody' recognises Henrike as the leader, shows misunderstanding of the conditions for leadership. The seriousness of the debate becomes manifest in the assumption of all participants that there is a right answer to some question, even though it is not a matter of objective fact, and that this does not change if people disagree about the right answer.

The reasons that can be adduced in a debate about a constructivist fact determine what the fact is. A slightly too simple way of stating the point is that the facts are what the best possible argument claims they are. This constructive role for arguments in determining what the constructivist facts are justifies their name as being *constructivist* facts. By adducing arguments for why the trip that everybody initially thought to be the best one was not really the best one, it turned out to be possible to 'change' the facts about the best cycling trip.

This 'change' can be interpreted in at least two different ways. First, it may be claimed that because of the arguments that were adduced, (almost) everybody changed her mind, and now the other trip has become the best one because 'everybody' recognises it as the best one. This interpretation has the advantage that it does not need the introduction of constructivist facts as a separate category. It suffices to see that adducing reasons may change what people recognise and thereby also what the social facts are. However, this interpretation has a serious drawback. If people adduce reasons why a claim that was broadly recognised is not correct, they do not merely want to change the social facts; they claim that the social facts already were different from what 'everybody' believed. Interpreting the event as a mere change of opinions does not do justice to what occurred, namely that people came to see that they were wrong from the start.

That is the reason why the second interpretation should be preferred. Given what the group members already recognise and given what the objective facts are, it is shown why another trip is 'really' the best one. The possibility to do this suffices for making the other trip the best one from the beginning. If this second interpretation is adopted, constructivist facts depend, not on what is actually recognised in a group, but on what *ought to be recognised* given the objective facts and what already is (or ought to be) recognised in the group.

[10] This is an example of so-called *defeasible reasoning*. More about defeasible reasoning in Prakken & Sartor 1997 and Ferrer Beltrán & Ratti 2012, and more about exceptions to rules in Bartels & Paddeu 2020.

Dworkin's criticism of Hart can be understood from the view that Dworkin was a constructivist with regard to legal facts. He claimed, in one phase of his career, that legal questions have one right answer. However, at the same time he criticised the view that the law is a purely conventional matter (as Hart claimed). The latter criticism is a special case of the view that constructivist facts are always open to questioning. The former view can be explained from the fact that a debate about constructivist facts is serious: the debate is conducted as if there were a correct answer to the issue at stake which is the same for everybody.

5.3. *Meta-constructivism and legal theory*

Finally, an important complication deserves mentioning. I wrote that whether a kind of fact is constructivist depends on the social practice in a group. This needs to be amended: it is not only a social fact whether a particular kind of fact is constructivist; it is even a constructivist fact. For instance, it is possible to seriously question whether ought-judgments are constructivist. Alternative views would be that they are purely conventional (non-constructivist social facts), or that they are objective, or purely subjective. The debates in ethical theory about the nature of moral judgments illustrate this very phenomenon. Some believe that moral judgments are objective, and that there are objective moral facts (SAYRE-MCCORD 2021). Others believe that they are expressions of individual preferences and that they are merely subjective or that the moral facts are relative to social groups and that what is a moral fact in one group is no such a fact in another group (GOWANS 2021). And, finally, there are those who believe that the moral facts are the conclusions of the best possible moral arguments (BAGNOLI 2021). The debates between the ethical theorists expressing these views are serious, and apparently the fact of the matter in this debate is constructivist. So, if legal judgments are considered to describe constructivist facts, this is itself a constructivist fact.

Constructivism bites its own tail: whether facts of some kind are constructivist, is itself a matter of constructivist fact. This explains the existence of *meta-constructivism*: constructivism about the issue of whether a particular kind of facts is constructivist. In legal theory there is a debate between hard (exclusive) positivists, soft (inclusive) positivists and non-positivists. Hard positivists claim that the law is a matter of social convention (basic social fact) and that the convention cannot refer to morality or anything other which is not objective or conventional (RAZ 1996; GARDNER 2001; MARMOR 2002). Soft positivists defend the view that the law is a matter of social convention but that the convention can refer to morality or to some other standard that is not conventional (HIMMA 2002). Non-positivists, finally, argue that law is not conventional in the first place, although there may be rational grounds for using conventions for the purpose of legal certainty (HAGE 2019). This debate has dominated much of 20th century legal philosophy, and it is not expected to end soon. The reason why the debate is so persistent is that it is a meta-constructivist debate about whether law is a matter of constructivist fact.

6. *Cognitive sciences and the nature of law*

Cognitive sciences deal with cognition in a broad sense. Cognition includes having true descriptions of the world. It also includes the possession of tools for making good inferences, and for giving useful explanations. The same holds for mechanisms for evaluating the world in a manner that is adequate to the ends of the evaluator. On the side of the entity that performs cognitive tasks, the scope of the cognitive sciences ranges from human beings, over other animals, to robots and immaterial cognitive agents such as organisations and artificially intelligent software.

This contribution has focused on a role of cognition that is seldom emphasized[11], namely its role in producing the social world. In contrast to the objective world, which is assumed to be mind-independent, the social world depends in a complex manner on minds. Moreover, the mental processes that create the social world are cognitive processes in the broad sense outlined above. This makes the mental creation of the social world a research topic for the cognitive sciences.

One important part of the social world is law. There are many legal facts which are, in a sense, brought about by legal rules. There are also many legal 'things' such as judges, courts, tax inspectors, property rights, criminal suspects, competences, and seizures, which are also rule-based. For a good understanding of how these legal facts and things have entered existence, the cognitive sciences play an important role.

However, this contribution has focused on a different topic concerning law and the cognitive sciences. Not only legal facts and things are products of cognitive activities, the same holds also for theories about the nature of law. In this contribution it was shown how jurisprudence has turned different aspects of social reality into different theories about the nature of law. Of course, the legal theorists of the 20th century—and they are who we are talking about—did not formulate their views in the terminology of social ontology or cognitive science. That terminology was not available yet. However, it is possible to translate the jurisprudential debate into the terminology of the cognitive sciences. And that opens new insights, such as the insight that the ongoing debate about the role of morality in law illustrates that the nature of law is a matter of constructivist fact.

[11] An exception is BERGER & LUCKMANN 1966. See also HEIDEMANN 2021, 2-22.

References

ALEXY R. 1992. *Begriff und Geltung des Rechts*, Karl Alber GmbH.

ALEXY R. 2002. *The Argument from Injustice. A Reply to Legal Positivism*, Oxford University Press.

BAGNOLI C. 2021. *Constructivism in Metaethics*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). Available on: https://plato.stanford.edu/archives/win2017/entries/constructivism-metaethics/.

BARTELS L., PADDEU F. (eds.) 2020. *Exceptions in International Law*, Oxford University Press.

BELTRÁN J.F., RATTI G.B. (eds.) 2012. *The Logic of Legal Requirements*, Oxford University Press.

BERGER P.L., LUCKMANN T. 1966. *The Social Construction of Reality*, Penguin.

CALLIES G.-P., ZUMBANSEN P. 2010. *Rough Consensus and Running Code. A Theory of Transnational Private Law*, Hart Publishing.

DWORKIN R. 1978. *Taking Rights Seriously* (2nd Ed.), Duckworth.

DWORKIN R. 1986. *Law's Empire*, Fontana Press.

FINNIS J. 1980. *Natural Law and Natural Rights*, Clarendon Press.

GALLIE W.B. 1956. *Essentially Contested Concepts*, in «Proceedings of the Aristotelian Society», 56, 167 ff.

GARDNER J. 2001. *Legal Positivism: 5 1/2 Myths*, in «American Journal of Jurisprudence», 46, 199 ff.

GILBERT M. 1989. *On Social Facts*, Princeton University Press.

GOWANS C. 2021. *Moral Relativism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). Available on: https://plato.stanford.edu/archives/spr2021/entries/moral-relativism/.

GREEN L., THOMAS A. 2019. *Legal Positivism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). Available on: https://plato.stanford.edu/archives/win2019/entries/legal-positivism/.

HAGE J. 2019. *The Limited Function of Hermeneutics in Law*, in DUARTE D., MONIZ LOPES P., JORGE SILVA SAMPAIO J. (eds.), *Legal Interpretation and Scientific Knowledge*, Springer, 1 ff.

HAGE J. 2022a. *Constructivist Facts as the Bridge Between Is and Ought*, in «International Journal for the Semiotics of Law». Doi: https://cris.maastrichtuniversity.nl/en/publications/constructivist-facts-as-the-bridge-between-is-and-ought

HAGE J. 2022b. *Objectivity of Law and Objectivity about Law*, in Villa-Rosas G. and Fabra-Zamora J.L. (eds.), *Objectivity in Jurisprudence, Legal Interpretation and Practical Reasoning*, Edward Elgar Publishing, 31 ff.

HAGE J. 2022c. *Rules and the Social World*, in «L'Ircocervo», 21, 2, 16 ff. Available on: https:///lircocervo.it/?p=5346.

HART H.L.A. 2012. *The Concept of Law* (3rd Ed.), Oxford University Press.

HEIDEMANN C. 2022. *Hans Kelsen's Normativism*, Cambridge University Press.

HIMMA K. E. 2002. *Inclusive Legal Positivism*, in COLEMAN J., SHAPIRO S. (eds.), *The Oxford Handbook of Jurisprudence and Philosophy of Law*, Oxford University Press, 125 ff.

HUME D. 1978. *A Treatise of Human Nature*, Oxford University Press. (Original text 1739/40.)

IEMHOFF R. 2020. *Intuitionism in the Philosophy of Mathematics*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). Available on: https://plato.stanford.edu/archives/fall2020/entries/intuitionism/.

KELSEN H. 1934. *Reine Rechtslehre: Einleitung in die rechtswissenschaftliche Problematik*, Franz Deuticke.

LAGERSPETZ E. 1995. *The Opposite Mirrors*, Kluwer Academic Publishers.

LEWIS D. 1969. *Convention: A Philosophical Study*, Harvard University Press.

MACCORMICK N., WEINBERGER O. 1986. *An Institutional Theory of Law*, Reidel.

MARMOR A. 2002. *Exclusive Legal Positivism*, in COLEMAN J., SHAPIRO S. (eds.), *The Oxford Handbook of Jurisprudence and Philosophy of Law*, Oxford University Press, 104 ff.

MARMOR A. 2009. *Social Conventions. From Language to Law*, in Princeton University Press.

MILLER A. 2021. *Realism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Available on: https://plato.stanford.edu/archives/win2021/entries/realism/.

MOORE G.E. 1903. *Principia Ethica*, Cambridge University Press.

PRAKKEN H., SARTOR G. (eds.) 1997. *Logical Models of Legal Argumentation*, Kluwer.

RADBRUCH G. 1945. *Fünf Minuten Rechtsphilosophie*, in «Rhein-Neckar-Zeitung», 12 September 1945.

RAILTON P. 2006. *Humean Theory of Practical Rationality*, in COPP D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, 265 ff.

RAZ J. 1979. *The Authority of Law*, Clarendon Press.

RAZ J. 1986. *The Morality of Freedom*, Clarendon Press.

REISS J., SPRENGER J. 2020. *Scientific Objectivity*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition). Available on: https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/.

RIDGE M. 2019. *Moral Non-Naturalism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Available on: https://plato.stanford.edu/archives/fall2019/entries/moral-non-naturalism/.

ROSS A. 1946. *Towards a Realistic Jurisprudence*, Einar Munksgaard.

ROSS A. 1959. *On Law and Justice*, University of California Press.

RUBEN D.H. 1985. *The Metaphysics of the Social World*, Routledge and Kegan Paul.

RUITER DWP 1993. *Institutional Legal Facts,* Kluwer Academic Publishers.

RUITER DWP 2001. *Legal Institutions,* Kluwer Academic Publishers.

SAYRE-MCCORD G. 2021. *Moral Realism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition). Available on: https://plato.stanford.edu/archives/sum2021/entries/moral-realism/.

SEARLE J.R. 1995. *The Construction of Social Reality*, The Free Press.

SEARLE J.R. 2010. *Making the Social World*, Oxford University Press.

SOETEMAN A. 2009. *Rechtsgeleerde waarheid*, Valedictory address Vrije Universiteit Amsterdam.

TUOMELA R. 2010. *The Philosophy of Sociality*, Oxford University Press.

WEBER M. 1921-22. *Wirtschaft und Gesellschaft*, Mohr.

WHITE A. R. 1975. *Modal Thinking*, Basil Blackwell.

PART IV.

Legal reasoning and Cognitive Biases

Moral Character Judgments and Motivated Cognition in Legal Reasoning

NIEK STROHMAIER, SOFIA DE JONG

1. *Introduction*

The idea that legal judgments do not always follow the ideal mechanistic pattern in which a judge or jury applies the relevant laws to the facts in a particular case to render a verdict is not new. At the beginning of the 20th century, legal realism already posited that judges «decide not primarily because of law, but based (roughly speaking) on their sense of what would be "fair" on the facts of the case» (LEITER 2003, 50). Moreover, realists have argued that «legal rules and reasons figure simply as *post-hoc* rationalizations for decisions reached on the basis of non-legal considerations» (LEITER 2003, 50; see also GREEN 2005). Psychological research has also devoted ample attention to factors that can affect legal judgments and decision making in ways that are typically deemed undesirable. Consider for example the role of emotions[1], political orientation (e.g., BRAMAN 2009; EPSTEIN & KNIGHT 2013; FURGESON et al. 2008), and irrelevant defendant characteristics like race and gender (e.g., LINDHOLM & CEDERWALL 2010; YOURSTONE et al. 2008). The role of bias in legal decision making is also addressed in other chapters in this book.

In this chapter, we zoom in on two particular phenomena that are directly related to biases in legal judgments. The first is *motivated reasoning* or *motivated cognition*. Through this process, certain irrelevant information or unconscious motives drive the sense-making process of legal decision makers in such a way that this allows them to reach their preferred outcome. The second phenomenon discussed in this chapter is the role of moral character inferences in legal judgments. How, when, and why do people in general and legal decision makers in particular judge the moral character of defendants and other parties involved in legal disputes? We discuss how the judgments about a defendant's moral character drives legal decision makers to a preferred outcome, which then, through motivated cognition, drives their legal analyses in a biased fashion.

[1] For a review of the literature on the role of emotion in legal decision making, see FEIGENSON 2016.

The structure of this chapter is as follows: First, we introduce the relevant theories surrounding motivated cognition and how this process can affect legal decision making (sect. 2). Second, we discuss recent theorizing in moral psychology about the central role of character inferences in moral judgments (sect. 3) and the role of moral character inferences in motivated legal reasoning (sect. 4). Finally, we briefly review research on potential debiasing techniques and the possible implications for legal practice (sect. 5).

2. *Motivated cognition in (legal) reasoning*

2.1. *What is motivated cognition?*

Motivated cognition or motivated reasoning can be described as the process in which decision makers engage in «inadvertently biased processes for accessing, constructing, and evaluating beliefs» (SOOD 2013, 309; KUNDA 1990, 480) when they unconsciously have a preference regarding the outcome of an evaluative task. For example, when a judge is faced with a defendant charged with manslaughter and that judge has a strong and automatic intuition—not based on the evidence—that the defendant is guilty, this judge could be more likely to perceive and weigh the evidence in such a manner that it allows for a conviction. In contrast, if the judge has a gut feeling that the defendant is innocent, the theory of motivated cognition would suggest that the judge would perhaps be more critical of incriminating evidence, more receptive to exculpating circumstances, or spend more time critically analyzing the forensic evidence proffered by the prosecution. In other words, the desired outcome (conviction or acquittal) will guide the sensemaking process and the weighing of the evidence by the judge or jury.

This is supported by research showing that decision makers have strong preferences for one conclusion over another. For example, preferences stemming from the evaluation of people as immoral or moral, and thus having «an affective stake in reaching a certain conclusion», can strongly alter reasoning processes in order to align the moral judgment with that desired conclusion (DITTO et al. 2009). Furthermore, these preferences can, for example, strongly influence the evidence a decision maker uses for an evaluation and the criteria for the evaluation itself (DITTO et al. 2009).

It is important to highlight the unconscious nature of motivated cognition. Decision makers are deemed to be under "the illusion of objectivity" when analyzing information and reasoning their way towards a desired conclusion (SOOD 2013, 309). Hence, legal decision makers may not be consciously aware of their own intuition regarding a specific defendant, nor that they may unconsciously be working towards a certain outcome. In a landmark paper describing the phenomenon of motivated cognition, Kunda describes the unconscious nature of motivated cognition as follows (KUNDA 1990):

> «People do not realize that the process is biased by their goals, that they are accessing only a subset of their relevant knowledge, that they would probably access different beliefs and rules in the presence of different directional goals, and that they might even be capable of justifying opposite conclusions on different occasions».

It is therefore clear that the process of motivated reasoning does not entail the conscious or strategic updating of attitudes, beliefs, or convictions in order to justify holding a position that serves the holders' interest. Conversely, legal decision makers can typically be deemed to act in good faith and thus to strive towards analyzing the relevant facts as unbiasedly and objectively as possible. Yet, despite their best efforts, research suggests that motivated reasoning processes are frequently at play in legal decision making. The next section discusses evidence for

motivated cognition in legal reasoning. It will become clear that motivated cognition plays a role in several facets of legal reasoning, in fact it is a widespread phenomenon.

2.2. *Motivated cognition in legal reasoning*

2.2.1. *Culpable Causation*

Several studies have investigated motivated cognition in legally relevant contexts. In Alicke's classic study, although the term motivated cognition was not used *per se*, he did show how moral judgments based on legally irrelevant information can subsequently drive legally relevant judgments concerning causality and responsibility (ALICKE 1992). He presented jury-eligible participants with a case involving a traffic accident in which a speeding car T-boned another car when crossing an intersection, resulting in significant injuries for the driver in the other car. Several other contributing factors were at play, such as a stop sign that was largely covered by overhanging tree branches, an oil spill making the road slippery, and the fact the T-boned car ignored a red traffic light. Participants were asked to what extent the speeding driver was the cause of the accident and whether this driver was responsible for the injuries suffered by the other driver.

The crux of the experiment was the reason *why* the speeding driver was exceeding the speed limit while approaching the intersection. Half of the participants read that John (the driver of the speeding car) had to get home quickly to hide a vial of cocaine before his parents could see it. The other half of the participants were told that John had to get home quickly to hide an anniversary present for his parents. In other words, half of the participants believed John had a socially desirable motive for speeding whereas the other half believed he had a socially undesirable motive.

The results showed that participants in the socially undesirable motive condition of the experiment believed more strongly that John was the primary cause of the accident and also that John was judged as more responsible, compared to those in the socially desirable motive condition. This study provides strong evidence for the notion that moral considerations can drive legally relevant judgments. More specifically, if a person is presented as "morally bad" their role will be evaluated as being causally more important.

2.2.2. *Cognitively cleansing tainted evidence*

A series of studies by Sood provide more direct evidence of motivated reasoning in legal judgments. She tested whether jury-eligible participants would judge a particular piece of illegally obtained evidence as admissible if that evidence would help convict a defendant who engaged in an egregious crime, despite it formally being inadmissible based on the exclusionary rule (SOOD 2015). The exclusionary rule under US law holds that evidence obtained in an unlawful manner is inadmissible in court. An exception to this rule is when illegally obtained evidence would have inevitably come to light at a later stage through lawful means.

In the case presented to participants, police officers conducted an illegal search of a car, in which they found large quantities of drugs. Importantly, however, half of the participants were told that the police had discovered « bags of marijuana that the defendant had been selling to terminally ill cancer patients to ease their suffering», whereas the other half were told that the police had discovered «bags of heroin and needles that the defendant had been selling to high school students». Participants in the heroin version of the case gave significantly higher punishment recommendations than those in the marijuana version of the case. Moreover, the heroin group believed more strongly that the illegally obtained evidence should be admissible in court. They based their answer on their conviction that the evidence would ultimately have been discovered lawfully, and were thus using the inevitable discovery exception to justify their belief that the incriminating evidence should be admissible. This study again clearly shows that

when people have a strong moral reaction towards a person having engaged in an egregious crime, they are motivated to construe the available information in such a way that it permits punishment of the criminal act.

2.2.3. *The harm principle*

Motivated cognition has been also observed in the application of the harm principle. The harm principle states that «the State should use its powers to regulate individual conduct only if doing so is necessary to prevent harm to others»[2]. Sood tested whether people motivated to criminalize a particular act would report that the act does causes harm to others, even though the act clearly does not cause any harm.

Participants were presented with a case involving a naked man going to the supermarket. In pretests it had been established that this was perceived as not causing any harm to others, but that it deviated from social norms to the degree that people would want to criminalize the act regardless of the absence of harm (SOOD & DARLEY 2012). Half of the participants were told that in order to criminalize a certain behavior, it would need to be established that the behavior causes harm, while the other half of the participants received no instructions.

It was expected that all participants would want to criminalize this socially unacceptable behavior, but that only those participants who were informed about the harm principle would be motived to construe the act as causing harm. The result confirmed the hypotheses, showing that all participants criminalized the act of going to the supermarket naked to the same degree, but the informed group reported higher degrees of harm than the group unaware of the harm principle. These results provide further evidence for the notion that motivated cognition can cause people to perceive elements of a case in such a way that it allows them to justify their punitive inclination, even if that justification requires construal of harm.

2.2.4. *Motivated enforcement of phantom rules*

Another example of motivated cognition has been found in the selective enforcement of legal rules. Legal rules should be universally applicable to all subjects within a particular jurisdiction. However, given the same transgression of a certain rule, it is not uncommon that some are punished whereas others are let off the hook. A particular subclass of legal rules where unequal treatment is frequently observed are "phantom rules"; rules that are «frequently broken and rarely enforced» (WYLIE & GANTMAN 2023, 2). Examples of these are rules on things like jaywalking and downloading music. Punishment in cases where these rules are violated is largely at the discretion of legal decision makers who can take the relevant circumstances into consideration in their judgment. The resulting ambiguity opens the door for motivated reasoning in such a way that these phantom rules are more likely to be enforced when decision makers are motivated by reasons that should not enter any legal analysis.

In a series of experiments, Wylie and Gantman tested this hypothesis. They presented participants with a case involving a man smoking marijuana on a park bench. Smoking marijuana is considered to violate a phantom rule, given its use is illegal, yet this law is infrequently and inconsistently enforced. Half of the participants learned that this man tried to provoke a stranger, thereby breaking a social norm but not a legal norm. The other half learned that the man politely asked a stranger what time it was, thus neither breaking a social nor legal norm. All participants were then asked whether it would be justified for the police to start an interaction with the man and for him to be punished.

[2] SOOD & DARLEY 2012, 1357; they refer to John Stuart Mill's *On Liberty* (1859) when discussing the harm principle.

The hypothesis was that participants informed of the provocation would be more likely to enforce the phantom rule and thus indicate that it would be justified for the police to engage with the man and punish him. The results confirmed that those participants believed it was more justified to enforce the phantom rule for the social norm violating man than for the man who did not break any social norm. Collectively, the experiments show that motivated cognition can cause people to use an opening in the law to punish a person for unrelated behavior such as breaking social norms.

2.2.5. *Motivated cognition and the advocacy bias in lawyers*

One of the key players in the legal domain is the lawyer. Lawyers are obliged to serve their clients' needs and to act in their interest. Therefore, an inherent part of lawyers' work is the motivation to advocate for the cause that meets their clients' needs. However, lawyers are taught that they can and have to form and maintain unbiased beliefs while advocating their client's cause (MELNIKOFF & STROHMINGER 2020, 1261) so a "blind" adoption of their clients' perspective would be detrimental to an accurate assessment of their chances in court. The question therefore is whether professional lawyers can remain impartial and unbiased or whether lawyers are also susceptible to motivated cognition processes in such a way that advocating a certain position can alter their true beliefs (also called *advocacy bias*). In other words, are lawyers able to control when to stay objective and when to advocate for a certain position?

Melnikoff's and Strohminger's research shows that professional lawyers are not as "in control" as they are taught and thought to be (MELNIKOFF & STROHMINGER 2020). Across a series of experiments, participants were asked to either act as a defense or prosecuting attorney and were given a range of different cases. Ultimately, the experiments aimed to test the automaticity and controllability of the advocacy bias, by (i) using weak stimuli to test whether a minimally immersive context can already cause advocacy bias, (ii) focusing on strongly held beliefs to test whether advocacy bias can alter such strong beliefs, (iii) by instilling a motivation to be accurate and objective to test whether advocacy bias can be eliminated or reduced when participants were required to be unbiased, and (iv) by using participants trained to be impartial (i.e., professional lawyers). Collectively, the results provide robust evidence for the notion that advocating for a certain position can automatically and uncontrollably alter even strongly held beliefs, even among professional lawyers.

2.3. *Motivated cognition in laypeople versus legal experts*

Thus far, we have seen that motivated cognition can affect judgments concerning for example causality, responsibility, and punishment. Furthermore, it can motivate the recruitment of harm if punishment requires it, it can result in the cognitive cleansing of tainted evidence, and it can affect the enforcement of phantom rules. Above all, we have seen that motivated cognition happens automatically and is hard to control.

However, the extent to which these findings can be generalized to legal experts has to be questioned, given that the majority of the studies used jury-eligible participants (i.e., laypeople). This is particularly relevant for the many jurisdictions that do not have jury trials and thus fully rely on judges. To what extent can legal expertise be a safeguard against the unwarranted influence of legally irrelevant information on legal judgments?

A number of studies on biases in legal judgments suggest that judges may be less biased than laypeople. For example, Rachlinski et al. found that specialized judges (bankruptcy judges) were unaffected by omission bias, debtors' race, or apologies, all of which have been shown to affect laypeople (RACHLINSKI et al. 2007; see also RACHLINSKI et al. 2006). In a similar line of research, it was shown that even though judges do rely heavily on their intuitive faculties, thus

making them susceptible to biases, they did manage to resist hindsight bias (i.e., the overestimation of the foreseeability of past events)[3].

However, despite the limited number of studies showing that judges' training and experience help them avoid certain biases, other studies show that legal experts (e.g., lawyers and judges) are affected by cognitive biases to a similar extent as non-experts. For example, the study by Melnikoff and Strohminger on the automaticity and uncontrollability of the advocacy bias found strong evidence for this bias among professional lawyers (MELNIKOFF & STROHMINGER 2020). Moreover, Guthrie and colleagues investigated whether judges are swayed by classic biases such as anchoring effects, framing effects, and hindsight bias, and found this to be the case (GUTHRIE et al. 2000). Similar results were found for professional arbitrators (HELM et al. 2016) and for insolvency law experts (STROHMAIER et al. 2021). Furthermore, research has shown that judges struggle to a similar extent as jurors with ignoring inadmissible information, even when they are reminded about the inadmissibility of certain evidence and even when they themselves ruled the evidence as inadmissible[4].

In a study directly comparing laypeople with professional judges in terms of their susceptibility to bias, both groups were presented with a case in which the mayor of a beach town commissioned the construction of a new highway (KNEER & BOURGEOIS-GIRONDE 2017). However, half of each group were told that as a side effect, the surrounding environment would be harmed, whereas the other half were told the environment would actually benefit. Both groups were informed the mayor did not care that the environment would be harmed/helped and that he just wanted the new road to be built.

A common finding is that people attribute intentionality for harmful side effects, but not for beneficial side effects, Thus, the question put to the participants was whether they believed the mayor intentionally harmed/helped the environment. The authors found the same pattern for both laypeople and judges, providing evidence for the notion that judges are also susceptible to this so-called "side-effect" effect. Results of a follow-up study varying the severity of the harm caused to the environment as a result of building the new road showed that judges were affected by outcome bias. That is, judges believed the mayor intentionally harmed the environment more strongly when the environment was harmed to a more extreme extent versus when the environment was harmed relatively mildly. As a final example of motivated cognition in legal experts, a recent study investigated whether experts' legal judgments concerning foreseeability, blame, and punishment in the context of directors' liability were affected by irrelevant information about the character of company directors. Results showed that both laypeople and legal experts were equally affected by this irrelevant information, even though laypeople did appear to be more punitive (STROHMAIER et al. forthcoming).

To conclude, it is safe to assume that professional legal decision makers such as judges are not exempt from the pervasive influence of motivated reasoning processes in legal judgments. Be that as it may, future research could help delineate under what circumstances the risk of motivated cognition is highest and when legal expertise may actually provide for some kind of buffer. These insights will also help design effective interventions. In the next section, we review existing knowledge on the boundary conditions of motivated cognition.

2.4. *Reality constraints*

From the first sections, the reader may feel that judges and jurors reach the conclusion they desire, unconsciously and automatically construing and/or altering whatever they need to

[3] GUTHRIE et al. 2007; for an introductory paper on hindsight bias, see ROESE & VOHS 2012; for a paper on hindsight bias in a legal context, see, e.g., STROHMAIER et al. 2021.

[4] WISTRICH et al. 2005; for another study on judges and inadmissible evidence, see LANDSMAN & RAKOS 1994.

justify their preferred outcome. However, research shows that motivated cognition is limited to what can still be considered reasonable and realistic. In other words, decision makers only alter their beliefs and attitudes to reach a desired outcome that can be reasonably justified to themselves and others (HSEE 1996). Sood also argues that legal decision makers do not ignore the law in order to reach their desired conclusions and that the process of motivated cognition occurs mostly if the law leaves room to do so, for example if the law is unclear or if open norms need to be interpreted and applied (SOOD 2013, 311).

Motivated cognition also appears to be "need-based", meaning that beliefs, attitudes, and perceptions are only 'stretched' to the extent needed to reach the preferred conclusion (BOINEY et al. 1997). People generally strive to act as rationally and objectively as possible, at least when acting in good faith, and it is therefore unlikely that legal decision makers bend reality to the extent that it conflicts with limits set by law or their inner desire to act rationally. Or, as Sood notes, legal decision makers «engage in motivated cognition—unknowingly processing information in an outcome driven manner—to achieve their desired result, seemingly within the terms of the given legal doctrine» (SOOD 2015, 1562).

However, in a recent study, the researchers examined whether having a strong goal or motive to adopt a certain belief could perhaps break through these reality constraints (STROHMINGER & MELNIKOFF manuscript). The authors assigned participants to either the role of defense attorney or prosecuting attorney and presented them with a case involving a defendant who was unambiguously guilty. Since the experiment left no "wiggle-room" regarding the defendant's guilt, the idea of reality constraints on motivated reasoning would predict that all participants would agree that the defendant was guilty. However, if the goal of advocacy was so strong that it could override reality constraints, perhaps participants assigned to advocate the position of the defense attorney would still question the defendant's guilt.

The results showed that advocacy, as a strong motivator, can affect perceptions regarding seemingly unambiguous facts, even when questioning the defendant's guilt required giving at least some weight to absurd theories, such as that the defendant was abducted and cloned by aliens and it was the clone who committed the crime, or that a shape-shifting creature took on the defendant's form and then committed the crime. Hence, even though motivated cognition generally remains within the bounds of what can be considered reasonable, when there is a strong motivation to reach a certain conclusion, it seems that reality constrains the motivated construction of certain beliefs less than once assumed.

2.5. *Summary and concluding remarks*

Thus far, we have introduced the theory of motivated cognition and given examples of studies in the legal domain showing that motivated reasoning can automatically and unconsciously alter legal decision maker's beliefs and attitudes, which then ultimately can affect legally relevant judgments concerning causality, foreseeability, responsibility, and blame. We also discussed whether there is a meaningful difference between laypeople and experts when it comes to their susceptibility to motivated cognition, and concluded that it is safe to assume that legal expertise and training provide insufficient protection. Finally, we described the constraints typically put on motivated cognition, as decision makers tend to stay within the limits of the law and also within the boundaries of what can be considered reasonable. At the same time, though, we noted recent research that shows that these reality constraints can be "overruled" under certain circumstances, further highlighting the pervasive impact of motivated cognition.

Further investigations into what exactly causes that initial desire to reach a certain conclusion in the first place are needed, answering questions like, what exactly is the instigator of motivated cognition? As we have seen thus far, the starting point for motivated reasoning processes is often a strong moral reaction in response to misconduct or other perceived moral transgressions. It appears

that it is primarily our need to blame and punish wrongdoers and thus our initial moral judgment regarding someone's character that drives these processes. That is, does someone have a good or bad moral character and is this person therefore a worthy recipient of praise or blame? To better understand how these moral character inferences can motivate legal reasoning, in the next section we discuss how moral judgments take shape, firstly exploring moral judgments in a general sense and then by focusing on the role of moral character inferences.

3. *Moral judgments and the role of moral character inferences*

3.1. *Introduction*

When we judge the blameworthiness of people's actions, we typically take several factors into account that we consider relevant. Take the example of Pete who harmed Henry, a poodle puppy, while playing fetch. One relevant factor in determining Pete's blameworthiness would be whether Pete intentionally harmed Henry, or whether it was a mere accident. Another relevant factor would be whether it was actually the stick that Pete threw to Henry that hit the puppy in the head, or whether it was the stick thrown by another dog owner playing with their dog nearby. Yet another relevant factor would be whether it was reasonably foreseeable that playing fetch with a stick could result in Henry being hit in the head. As a final example, a relevant factor would be whether it was within Pete's power to avoid the harm from occurring, for example by choosing a different object to play fetch with or using a smaller stick. In other words, when determining a person's blameworthiness, ideally we carefully analyze relevant factors such as causality, intentionality, foreseeability, and controllability.

Early models of moral judgment adhered to a prescriptive approach as they described how moral judgments are ideally made (SHAVER 1985). Since then, however, theories of moral judgment have sought to better understand how moral judgments are actually made in real life and key factors such as causality, mental state (e.g., was the harm intentional?), and preventability were identified (see, e.g., CUSHMAN 2008).

In the next sections, we introduce models of moral judgment that incorporate both automatic, intuitive processes, as well as more deliberate and cognitively intense processes (so-called *dual-process* models). We then review models that focus primarily on the quick, affective, and intuitive reactions towards moral transgressions, which generally attribute less significance to more conscious deliberations. Finally, we discuss recent developments in research on the role of moral character inferences in moral judgment and highlight the centrality and importance of such inferences We close with a discussion on the empirical support for the prominent role of moral character inferences in moral judgment.

3.2. *Dual-process models of moral judgment*

Prominent models of moral judgment attribute qualities to both reason and emotion. Specifically, dual-process models stipulate that moral judgments can be the result of two distinct modes of operation. On the one hand there is the quick, automatic, affective mode, and on the other there is the more deliberate, slow, cognitive mode. A well-known dual-process model is one put forward by Greene, which is based on an extensive line of research involving moral dilemmas (GREENE et al. 2001; see also GREENE & HAIDT 2002).

Greene presented participants with classic moral dilemmas such as being able to save five lives by having to push another person onto a train track, thereby stopping the train from running over the five people laying on the track. Both behavioral and neurological data show that more affective modes of processing typically result in deontological judgments (i.e.,

following clear moral rules such as "killing is wrong") regarding the permissibility of killing one person to save five. In contrast, more cognitive, deliberate modes of processing tend to result in consequentialist judgments, meaning that the permissibility of an action is judged based on the consequences, rendering the killing of one person to save five morally permissible.

However, it has been argued that Greene's model is better equipped to predict judgments of moral permissibility and less so for attributing (legal) blame (MONROE & MALLE 2017, 123). Moreover, recent insights in neuroscience suggest that there are more than two processes in moral judgment, meaning that dual-process models focusing on the distinct tracks of intuition and emotion and on deliberate reasoning are an oversimplification of how moral judgments are actually formed in real life (VAN BAVEL et al. 2015). Rather than being separate streams, it seems more plausible that rational deliberation and emotion interact in a variety of different ways (HELION & PIZARRO 2015).

The Path Model of Blame, coined by Malle, Guglielmo, and Monroe, permits a more complex interplay between emotion and reason, with moral judgments at times being fast and automatic, and at other times more effortful and deliberate[5]. The model's main aim is to describe *what* exactly is being processed, rather than through which *mode* (i.e., automatic vs. deliberate). The model posits that upon the detection of a norm violating act, the first factor to be analyzed is whether the agent is the cause of a certain harm. If not, no blame is assigned. If causality is established, the next step is to determine the degree of intentionality. If an agent intentionally caused a certain harm, the final degree of blame assigned to the agent is dependent on the reasons for this intention. If the agent did not intend to cause the harm, the degree of blame is dependent on the extent to which the agent had the duty to prevent the harm and whether the agent was capable of doing so.

A limitation of this model is that it does not account for the severity of a transgression (GOODWIN 2014). Ample research on outcome bias has shown that with all else being equal, moral transgressions are judged harsher when the outcome is more severe (see, e.g., KNEER & MACHERY 2019; KNEER & SKOCZEŃ 2023; KNEER 2022). Moreover, in the model, the degree of blame assigned to a person is the final "product" of processing the factors included in the model. Alternative theories, that will be discussed next, claim that the attribution of blame actually happens relatively fast, automatically, and intuitively, and subsequently affects perceptions of causality, intent, and other factors deemed relevant by the Path Model of Blame (GOODWIN 2014, 217 f.). A key criticism of the Path Model of Blame, therefore, is that it assigns insufficient weight to the role of motivated reasoning processes initiated by quick and intuitive blame judgments (NADLER 2014). We now focus on several theories of moral judgment that give more primacy to quick and automatic moral intuitions and subsequent motivated reasoning processes.

3.3. *Biased information models of moral judgment*

Several theories of moral judgment assign more weight to the role of emotion and intuition than the dual-process models discussed thus far. These "emotion heavy" theories do not go as far to say that conscious reasoning plays no role in moral judgment, but, in the more "extreme" versions of these theories, conscious deliberation mostly serves to *post-hoc* rationalize the judgment that was instantly formed through affective and intuitive processes. A prominent theory in this domain is that of Jonathan Haidt, called "social intuitionism" (HAIDT 2001).

In its essence, the social intuitionist model holds that moral judgment is mostly based on intuitions instead of on moral reasoning (HAIDT 2001, 1024). The idea is that when learning of a moral transgression, people experience a quick and automatic moral intuition, which is then

[5] MONROE & MALLE 2017, 124; for their first paper positing the model see, MALLE et al. 2014.

potentially, but not necessarily, followed by motivated reasoning in order to rationalize the initial intuition (HAIDT 2001, 817). Hence, according to Haidt's theory, there is a direct link between a person's moral intuition and their moral judgment, without an intermediary phase of conscious reasoning. Haidt states that, in contrast to the rationalist models, «judgment comes first, based on educated intuition; justification is undertaken next» (HAIDT 2012, 868). Consistent with this theory, studies on "moral dumbfounding" show that people can have a strong moral reaction towards certain acts and also condemn these acts, without being able to give a good explanation for their moral judgment (see e.g., HAIDT & HERSH 2001). Evidence for Haidt's theory comes largely from the studies on motivated cognition discussed in Section 2.

Alicke's culpable control model, which we have touched on earlier, is largely compatible with Haidt's theory. Alicke suggests that social groups have a need to hold transgressors accountable in order to maintain social order (ALICKE 2000). This need is reflected in humans' natural inclination to blame. Alicke found that this inclination to blame affects the evaluation of culpable control. When people are motivated to blame an individual, they are more likely to conclude that that individual had causal control over a certain outcome. This is why the speeding driver who wanted to hide a vial of cocaine being judged to be more causally responsible for the traffic accident than the speeding driver trying to get home to hide his parent's anniversary present. In Alicke's theory, it is the affective reaction and spontaneous evaluations of an agent and their behavior that directly affects blame judgments, while at the same time allowing for an indirect link between affect and blame through the evaluation of mental state and causal control.

Thus far, we show that moral judgments can, for an important part, be the result of automatic, affective reactions triggered by witnessing a harmful event. The theories differ in terms of the role of conscious deliberations being either limited to mere *post-hoc* justification of the initial blame judgment, or to rational deliberation that is influenced by affective processes through a complex interplay between the two. Hence, moral judgments are at the very least influenced by quick and automatic affective reactions, and are potentially their sole determinant. But what is the source and focus of those initial affective reactions and moral intuitions? What precisely triggers people to have a certain moral reaction? Recent theorizing in moral psychology suggests that moral character inferences take center stage, to which we turn next.

3.4. *Person-centered approach to moral judgment*

The theories discussed thus far focus mainly on moral judgments in response to a certain act. Hence, the blameworthiness of an agent is derived from elements of the act, such as whether someone acted intentionally, whether the harm caused by the act was foreseeable, whether the act was in fact the cause of a certain harm, etc. Such act-based theories of moral judgment, however, cannot account for a range of studies, some of which have already been discussed, that show that the same act is judged differently based on factors independent from the actual norm-violating act. For instance, Alicke's study, in which he shows that the same act (speeding whilst approaching an intersection, resulting in a collision) is judged differently depending on the social acceptability of the driver's motive for speeding.

As an alternative to the act-based theories of moral judgment, there are the more recent person-centered theories, where people primarily ask themselves whether a person is good or bad, rather than whether a certain act is right or wrong (UHLMANN et al. 2015). From an evolutionary perspective, it makes sense that we are quick to judge whether someone is a friend or a foe, as it is adaptive to quickly know whether someone will either be an ally in achieving our goals or a threat to our survival. Indeed, evaluating a person's moral character can be crucial and it is therefore important to judge correctly. Think about the horrific opening scene of the movie *Silence of the lambs* in which a girl kindly accepts to help a man who turns out to be a serial killer. Albeit extreme, this example shows the practical importance of getting our moral

character evaluations right. Given the evolutionary advantage of assessing people's character swiftly and accurately, we have evolved to do so rather aptly.

This idea of moral character inferences taking center stage in our moral judgments is consistent with findings from research on impression formation. Studies on impression formation show that evaluating a person's moral character is the first thing we do when forming impressions of others, and that this process occurs automatically and can already be observed at an early age and across cultures[6]. Recent research has even shown that inferences of a person's moral character are a more important driver in impression formation than attributes such as a person's warmth and competence, which until recently were believed to be the two key dimensions in person evaluations (BRAMBILLA et al. 2021). Based on the primacy of moral character inferences, person-centered theories argue that descriptive models of moral judgment should include moral character evaluations as a central feature (see GOODWIN et al. 2014; GOODWIN 2015).

In short, based on recent theorizing, it appears that people form quick and automatic judgments of a person's moral character, and these judgments are expected to then guide their sensemaking process. In this way, a person with a bad moral character is expected to have acted more intentionally when causing a certain harm than one with a good moral character. Hence, rather than focusing on the blameworthiness of the act, the act is used to infer a person's moral character, which then informs our moral judgments.

3.5. *Empirical support for the person-centered approach to moral judgment*

Empirical support for the centrality of moral character inferences in moral judgments is plentiful. We have repeatedly referred to Alicke, who demonstrates that judgments about blame and causality are strongly influenced by moral attributes of the speeding driver, even though all the factors considered relevant in prescriptive models of moral judgment were kept constant across the different experimental conditions. It appears that participants judged the moral character of the speeding driver with the socially undesirable motive to be worse than that of the other with the socially desirable motive. This evaluation of the speeding driver's moral character is expected to then have influenced participants' perceptions of the causal role of the driver in the accident.

3.5.1. *Empirical support in the context of criminal law*

In the context of criminal law, research has been conducted on whether jurors follow the ideal model of legal decision making in which relevant factors such as causality, intentionality, and foreseeability are carefully weighed, or whether jurors are affected by irrelevant information about a defendant's moral character, as predicted by the person-centered theories of moral judgment (NADLER & MCDONNELL 2011). In a series of experiments, participants were presented with a case in which all legally relevant factors were constant, such as the (severity of the) harm that occurred and the mental state of the defendant, but in which the moral character of the defendant varied.

For example, in one of the experiments, participants read a case about a woman living with her two dogs that sometimes escaped from her fenced yard and behaved aggressively towards children. At a certain point, the dogs escape again, reap havoc in the neighborhood, and ultimately attack two young boys, one of whom dies from his injuries. Those in the good moral character condition read that the owner of the dogs is a very social person who maintains a healthy lifestyle and has many close friends and adores her two young nieces. The other half read that the owner does not socialize much and spends much of her time watching trash tv

[6] For further relevant citations on the topic, see UHLMANN et al. 2015, 74.

while smoking and eating junk food, and does not like to spend time with her two nieces.

After reading the case, participants were asked to what extent the dogs' owner was responsible for the death of the young boy, to what extent she was the cause of the death, how much blame she deserved, and how foreseeable the death of the boy was from her perspective. It is clear that the participants who read the good moral character version of the case will have a more favorable impression of the woman's character[7], and that the provided character information should have no bearing on any legally relevant judgment. Still, the results showed that the overall responsibility for the boy's death was rated higher by those who read the bad moral character version of the case and the causal role of the dogs' owner was believed to be higher.

In a follow-up study, participants were presented not with just one moral character version of the case (i.e., either the good character version or the bad character version), but with both (NADLER 2012). This was done to test whether jurors believe that moral character inferences *should* in fact influence their legal judgments, or whether they believe that normatively speaking, character information should be disregarded. If for example a participant first read the aggressive dog case in which the owner has a good moral character and, directly after, read the same case but then with the owner having a bad moral character, and the participant still believes the morally bad dog owner deserves more blame, it would then appear that jurors believe moral character information is somehow relevant for their legal judgment. If, however, the effect of moral character on legally relevant judgment disappears when jurors read both versions, this would suggest that jurors agree that irrelevant character information is irrelevant for legal judgments. The results of this study consistently showed that the biasing effect of moral character inferences disappears when participants are presented with both the good and the bad moral character version of the case, suggesting that jurors can be affected by irrelevant character information, but rightly believe they ought not to (see also Alicke & Zell 2009).

3.5.2. *Empirical support in the context of civil law*

A recent study was set up to extend the findings discussed thus far to the context of civil law instead of criminal law, and to legal experts instead of laypeople (STROHMAIER et al. forthcoming). This study also found evidence for irrelevant moral character inferences biasing legal judgments (STROHMAIER et al. forthcoming). Specifically, legal professionals were presented with a case concerning a company in financial distress where the director was faced with a range of difficult decisions whilst trying to save his company from bankruptcy. In the end, the company fails, and questions arise about the director's liability for damages incurred by creditors.

In this specific legal context, the foreseeability of the company's bankruptcy is important for determining liability, as well as the director's actual awareness of the likelihood the company was going to fail. Importantly, as in the studies discussed thus far, participants either read that the director had a bad or good moral character, for example he was a bad husband and absent father, or in contrast a very loving husband very much involved in his children's upbringing. The information provided about the director's character was clearly irrelevant from a legal point of view.

The results showed that, relative to the legal professionals who read the good moral character version of the case, those who read the bad moral character version believed that the company's bankruptcy was more foreseeable, that the director was aware of the likelihood of bankruptcy, and also believed the director deserved more blame for the damages suffered by creditors. It thus seems that the influence of moral character inferences in legal judgments stretches beyond the domain of criminal law and also affects legal experts.

[7] The data confirmed that participants in the good moral character condition indeed viewed the woman's character more positively than participants in the bad moral character condition, see NADLER & MCDONNELL 2011, 286.

3.5.3. *The side-effect effect in legal judgments*

Another line of research in which judgments of an agent's moral character have been shown to affect legally relevant judgments, such as those of intentionality and blame, concerns the so-called "Knobe effect", also termed the "side-effect effect". In the original study, participants read a short case about a chairman of a company's board being approached by an executive about a new program designed to significantly increase profits, whilst also affecting the environment (for the original study, see KNOBE 2003). The chairman then expresses a total disregard for the environment and states to be solely interested in making as much profit as possible. The new program is launched and the environment is affected as predicted. Crucially, however, half of the participants were told that as a side effect of running the new program, the environment would be harmed. The other half were told that the environment would actually benefit from the program. Either way, the chairman did not care. Depending on the version of the case given to them, participants were then asked to indicate whether they believed the chairman intentionally harmed or helped the environment. Respondents generally agreed that the chairman intentionally harmed the environment, but did not believe the chairman intentionally helped the environment. This asymmetry is puzzling, given that it is clearly stated that the chairman is only interested in increasing profits and does not care about how the new program affects the environment. Explaining the results through the lens of moral character inferences, it seems reasonable to assume that people generally formed a negative impression of the chairman's moral character, given we typically believe people ought to at least have some regard for the environment and put environmental concerns in the scale when making business decisions. The fact that the participants were probably motivated to blame the chairman for harming the environment (albeit as a side effect of his true intentions) and reluctant to give praise for helping the environment, it makes sense they indicated the chairman did intentionally harm the environment, but were reluctant to say the chairman intentionally helped the environment. Hence, rather than intentionality judgments serving as key input for blame judgments, it appears that evaluations of an agent's moral character trigger the need to blame the agent, which then drives perceptions of intentionality[8].

3.6. *Summary and concluding remarks*

In this section, we reviewed the workings of moral judgments in order to better understand how legal decision making can be affected by motivated reasoning processes. We introduced several models of moral judgments, ranging from those in which conscious and deliberate processes are given significant weight, to those in which emotions and automatic moral intuitions take center stage and in which conscious deliberations are primarily for *post-hoc* rationalizations. Furthermore, we discussed the dominant role of moral character inferences in our moral intuitions. We argue that it is typically these quick and automatic moral character inferences that trigger the need to blame morally bad people and that it is this need to blame that subsequently drives important constituents of blame (e.g., intentionality, causality, control, etc.). Thus, instead of moral and legal judgments being the final product of a careful weighing of relevant factors, it is more likely that automatic inferences about people's moral character bias our sensemaking processes in such a way that it allows us to blame and hold liable morally bad people, and to go easy on people we perceive to be morally good.

Having explained how motivated cognition can unconsciously drive legal reasoning, and having discussed how moral character inferences are often the key drivers behind motivated

[8] For empirical support for the claim that moral character inferences are an important driver of the side-effect effect, see, e.g., SRIPADA & KONRATH 2011. See also SRIPADA 2012.

legal reasoning, an important question arises: how can we improve legal decision making in order to limit the unwanted influence of moral biases and motivated cognition? The final part of this chapter addresses this question.

4. *Debiasing the influence of moral character inferences and motivated legal reasoning*

The question of what can be done to limit or even prevent the unwanted influence that moral character information can have on legal reasoning is difficult to answer. Current research on potential debiasing techniques has not yet provided direction, however, the problems outlined in this chapter beg for at least some consideration of the current state of knowledge on these techniques.

4.1. *Requirements to reduce moral bias in judgments*

In order for people to reduce the influence of moral biases in moral and legal judgments, at least four requirements need to be met (NADELHOFFER 2006, 212). First, awareness must be created about a particular psychological process a person needs to protect themself against. Second, a person needs to be motivated to counter the unwanted effects. Third, in order to effectively correct for the unwanted effect, people need to be aware of «the direction and magnitude of the bias» (WILSON & BREKKE 1994, 118). And finally, people must have sufficient cognitive control to effectively prevent a particular bias from exerting an effect.

For legal practitioners, the first requirement is often already absent, as they are generally unfamiliar with the body of literature on unconscious bias in legal reasoning and decision making. Furthermore, legal professionals commonly believe that their expertise and experience protects them from being affected by biases. This first problem could in theory be relatively easily overcome through training and education; when it comes to decision making by judges or jurors, the first three steps could be achieved through proper training in combination with clear (jury) instructions, even though the effects of the latter have been mixed[9]. The major hurdle to overcoming bias is the fourth requirement, as people generally lack the cognitive control to consciously counter effects of bias (NADELHOFFER 2006, 212).

4.2. *Debiasing through procedural constraints*

Therefore, rather than relying on the motivation or willpower of people to overcome their own biases, it may be worthwhile to search for solutions in the procedural domain.

4.2.1. *The prohibition of character evidence*

An example of constraints in legal procedures that are (in part) meant to prevent irrelevant information from biasing legal decision making is the prohibition of character evidence in legal proceedings. In many common law jurisdictions (e.g., US, UK, Australia), there are restrictive rules for the admissibility of character evidence, based on the assumption that such evidence may be given disproportionate weights by juries[10]. Unfortunately, however, these regulations have many loopholes, meaning that in practice character evidence is often deemed admissible (CULBERG 2009; MELILI 1998). And even when character evidence is ruled inadmissible, research

[9] For studies in the effectiveness of jury instructions, see, e.g., HALVERSON et al. 1997; CONKLIN 2021; INGRISELLI 2014; DAFTARY-KAPUR et al. 2010; SAUERLAND et al. 2020; SOOD 2015; GREENE & DODGE 1995.

[10] For a comprehensive historical account, see SEVIER 2019.

has shown that it is difficult to ignore this inadmissible evidence once it has been registered (DAFTARY-KAPUR et al. 2010; STEBLAY et al. 2006). Moreover, and as discussed, people cannot help but form quick and intuitive judgments about people's moral character, meaning that even in the absence of character evidence as such, it is likely that legal decision making will form an opinion of a defendant's character as either good or bad.

4.2.2. *Linear sequential unmasking*

A more promising avenue of procedural constraints can be found in the field of forensic science where a procedure has been introduced called "linear sequential unmasking" (DROR et al. 2015; STOEL et al. 2014). In short, a first investigator (the case manager) has access to all information available in a particular case, part of which is irrelevant and has biasing potential. This first investigator then filters that information and only provides the relevant information to a second investigator needed to conduct a particular part of an investigation. For example, when analyzing a fingerprint found on a crime scene, information that can unconsciously motivate an investigator to find matching elements with the fingerprint of a particular suspect ought to be filtered out and thus not passed on the investigator conducting that analysis. Applied to the context of legal decision making, a form of trial bifurcation in which an initial judge who is given access to all case files and other sources of (irrelevant) information filters the irrelevant information for a second judge may prove promising. After all, information with biasing properties that is never registered cannot exert any effect. A clear downside is that any form of trial bifurcation puts significant demands on courts that typically have limited resources and a significant backlog of cases.

4.2.3. *Limiting court access of litigating parties*

An extreme example of procedural constraints would be to limit judges' and jurors' access to information about the litigating parties by, for example, not allowing parties to appear in court. In some (parts of) legal proceedings, parties do not appear in court and judges/jurors only base their legal decisions on written documents. This mostly applies to civil cases[11]. However, typically, both parties do appear in court, meaning the judge and jury can see what they look like, how they dress, how they talk, how they conduct themselves, etc. All of these factors evidently affect decision makers' perceptions and judgments. Seen from the goal of accuracy and objectivity, it would be better if judges and jurors would have no access to this type of information that might bias them in a certain direction.

From a procedural fairness point of view, however, it is typically considered to be important and desirable for parties to be present in court and to have their say; this is considered a fundamental principle of procedural law. Moreover, it can be desirable and more efficient for judges/jurors to be able to ask questions directly instead of having to request a reaction in writing. It is therefore unlikely that the right (or obligation) to appear in court will ever be revoked. However, in line with the notion of linear sequential unmasking, perhaps the veil of ignorance could be lifted later in the process to allow the majority of a legal proceeding to take place with as little biasing information as possible. Still, as explained above, it would also be necessary to filter some information from the written documents, as written information also has biasing potential.

[11] For example, the Dutch Supreme Court typically does not see parties in person, but only decides on their written statements.

4.3. *Individual differences in susceptibility to bias*

In addition to looking at procedural interventions as a way to debias, to what extent can we derive valuable insights from studies looking at individual differences in susceptibility to biases? Why are some people more affected by certain biases than others? What can we learn from these individual differences?

4.3.1. *Intellectual humility*

Recently, increased attention has been devoted to the notion of "intellectual humility". This «involves recognizing that there are gaps in one's knowledge and that one's current beliefs might be incorrect» (PORTER et al. 2022). What makes intellectual humility (IH) conceptually distinct from, for example, perspective-taking and general modesty is that IH focuses on recognizing a person's own fallibility and ignorance. In simple terms, people high in IH typically acknowledge that there are many things they do not know and that they may even be wrong about the things they believe they do know.

The past decade has seen a steep increase in research on this topic and, for the present purposes, an important consequence of IH is thought to be a reduced susceptibility to certain biases (PORTER et al. 2022; BOWES et al. 2022). This makes sense, given that IH allows a person to be intellectually flexible and not to hold strictly to certain beliefs or attitudes. It follows that people high in IH will experience a reduced motivation to reach a certain outcome, relative to those low in IH. Therefore, it might be a solution to select people high in IH or train people in IH. However, a possible drawback is that it is likely to be hard for someone low in IH to become high in IH, if only because it would require a significant time investment[12].

4.3.2. *Actively open-minded thinking*

A construct similar to IH, and one that can serve as a prescriptive model for rational thinking that can be consciously adopted, is that of "actively open-minded thinking" (AOT). AOT, as a style of thinking, «includes the tendency to weigh new evidence against a favored belief, to spend sufficient time on a problem before giving up, and to carefully consider the opinions of others in forming one's own» (HARAN et al. 2013, 189). Or as others have defined it:

> «a thinking disposition encompassing the cultivation of reflectiveness rather than impulsivity; the desire to act for good reasons; tolerance for ambiguity combined with a willingness to postpone closure; and the seeking and processing of information that disconfirms one's beliefs» (STANOVICH & TOPLAK 2019, 156).

Importantly, AOT has been associated with a reduced susceptibility to (among others) belief bias (see e.g., MACPHERSON & STANOVICH 2007), which is roughly similar to confirmation bias and myside bias, both of which entail a tendency to evaluate evidence in such a way that it favors a person's existing opinions and beliefs.

For an important part AOT is considered to be a predisposition, meaning that is a relatively stable trait and thus that some are more likely to adopt this style of thinking and reasoning than others. However, it has been argued that AOT can also be consciously adopted by adhering to a set of principles. These include, for example, principles such as "correct conclusions are more likely when more than one possible conclusion is evaluated", "evaluations of possible

[12] For a discussion of ways to improve one's intellectual humility, see PORTER et al. 2022, 531 f.

conclusions will be more accurate when both positive and negative evidence is sought in a balanced way", and "evidence, once found and evaluated, should be used in a balanced way, that is, independent of whether it favors or opposes the front runner". As such, descriptive elements of AOT can be used in a prescriptive way and thus serve as a good standard for evaluating rational thinking in oneself and others.

4.3.3. *Free will beliefs*

In a recent paper on another, related example of individual differences in susceptibility to bias, it was shown that legal professionals who believed more strongly in the notion of free will also showed an increased hindsight bias relative to those who were more skeptical of the idea that humans have free will (STROHMAIER et al. 2021). The proposed mechanism underlying this finding was thought to be that believing in free will correlates with punitive inclinations and reduced IH. Thus, those scoring high on the belief in free will scale were believed to have a stronger need to condemn a potential wrongdoer and would be less comfortable with the uncertainty and ambiguity resulting from withholding judgment.

The idea of withholding judgment and engaging in further exploration has recently also find support in popular media, as evidenced by Julia Galef's book *The Scout Mindset* (GALEF 2021). The author argues that we can prevent ourselves from falling into the trap of bias and in particular from motivated reasoning through adopting a scout mindset. A scout mindset is loosely defined as deriving pleasure from seeking new information, from challenging even our deeply held beliefs, and as being comfortable with being wrong as our self-worth is then not tied to our being right or wrong. These ideas are largely based on the literature discussed in this section.

Hence, it appears that in order to reduce susceptibility to (moral) biases and motivated cognition, it helps to withhold moral judgment, gather additional information, stress test our convictions, and only be confident in our judgment once we have actively and thoroughly engaged in a fact-finding mission. In other words, the scout mindset, intellectual humility, and actively open-minded thinking (as overlapping concepts) can help protect against a range of moral biases and prevent engaging in motivated (legal) reasoning.

4.4. *Summary and concluding remarks*

To summarize, due to our limited cognitive control, it is difficult for us to overcome the automatic and unconscious processes of motivated moral cognition. Thus, even when we are made aware of the risk of unconscious bias and are motivated to correct for them, actually doing so may prove difficult. Solutions are therefore more likely to be found in changing legal procedures, particularly by adopting those that prevent biasing information from being observed in the first place. Additionally, striving towards becoming more intellectually humble and adopting an actively open-minded thinking style can help reduce the tendency of holding on to previously held beliefs. This may, in turn, stimulate us to actively engage in a critical review of a range of different conclusions, ultimately reducing the chance of failing pray to motivated reasoning.

5. *Conclusion*

In this final part of this chapter, we conclude by highlighting the important process of motivated reasoning in legal judgments and the crucial role of moral character inferences in motivated legal reasoning. The key takeaways are the following.

First, legal decision making is unlikely to follow the mechanistic ideal as much as is often believed or hoped. Instead, it seems more likely that motivated reasoning plays an important part in how judges and jurors reach their verdicts.

Second, an important starting point for these motivated reasoning processes is the automatic and instant evaluations of defendants' moral characters. When being presented with information about a defendant's character (e.g., through seeing them in court, pre-trial publicity, character evidence, etc.), legal decision makers cannot help but form an opinion of a defendant's moral character automatically and swiftly. This character evaluation then biases sensemaking processes in such a way that they construe a case in a way that allows them to allot blame to a morally bad defendant and to go easy on a morally good defendant, for example through their attributions of intentionality, causality, control, and foreseeability (i.e., through motivated reasoning).

A third takeaway is that legal expertise is unlikely to prove effective in providing any protection against the automatic and unconscious influence that moral character inferences and blame intuitions have on legal judgments (through motivated reasoning). We reviewed the literature focusing on finding differences between laypeople and legal experts in terms of their susceptibility to a range of different biases, and we can safely conclude that legal experts are equally affected by unwanted psychological influences in their reasoning and judgments. This is an important point as legal professionals still brush off the risk of bias in their decision making and hide behind the argument that they are professionals and therefore not a likely to fall victim to processes affecting the general public. This position is thus not supported by insights derived from scientific research.

Finally, it should have become clear that debiasing the automatic and pervasive influence of motivated reasoning stemming from moral character inferences is a great challenge. It is unlikely we can rely on people's willpower and cognitive control to simply will their way out of biases affecting judgments. Instead, it is probably better to search for solutions in the domain of procedural adjustments that limit the amount of biasing information available to legal decision makers. However, procedural adjustments are costly and will put extra pressure on the heavy workload of courts.

As a more explorative avenue for debiasing potential, we discussed the potential of exploring individual differences in thinking styles and intellectual humility. Limiting motivated cognition is a major challenge as it requires a change within the decision makers themselves. Increasing their intellectual humility and adopting an actively open-minded thinking style may significantly limit the degree to which they engage in motivated reasoning. However, given that change is difficult for anyone as well as being time consuming, it remains to be seen how feasible this strategy really is. Awareness of motivated reasoning processes surely is the first step, and we hope that this chapter will inspire a future generation of legal scholars and practitioners to take up the challenge of improving legal judgments by incorporating insights from behavioral sciences.

References

Alicke M.D. 1992. *Culpable Causation*, in «Journal of Personality and Social Psychology», 63, 3, 368 ff.

Alicke M.D. 2000. *Culpable Control and the Psychology of Blame*, in «Psychological Bulletin», 126, 4, 556 ff.

Alicke M.D., Zell E. 2009. *Social Attractiveness and Blame*, in «Journal of Applied Social Psychology», 39, 9, 2089 ff.

Boiney L.G., Kennedy J., Nye P. 1997. *Instrumental Bias in Motivated Reasoning: More When More Is Needed*, in «Organizational Behavior and Human Decision Processes», 72, 1, 1 ff.

Bowes S. M., Costello T.H., Lee C., McElroy-Heltzel S., Davis D. E., Lilienfeld S.O. 2022. *Stepping Outside the Echo Chamber: Is Intellectual Humility Associated with Less Political Myside Bias?*, in «Personality and Social Psychology Bulletin», 48, 1, 150 ff.

Braman E. 2009. *Law, Politics, and Perception: How Policy Preferences Influence Legal Reasoning*, University of Virginia Press.

Brambilla M., Sacchi S., Rusconi P., Goodwin G.P. 2021. *The Primacy of Morality in Impression Development: Theory, Research, and Future Directions*, in «Advances in experimental social psychology», 67, 187 ff.

Conklin M. 2021. *I Knew It All Along: The Promising Effectiveness of a Pre-Jury Instruction at Mitigating Hindsight Bias*, in «Baylor Law Review», 74, 307.

Culberg D. 2009. *The Accused's Bad Character: Theory and Practice*, «Notre Dame Law Review», 84, 3, 1343 ff.

Cushman F. 2008. *Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment*, in «Cognition», 108, 2, 353 ff.

Daftary-Kapur T., Dumas R., Penrod S.D. 2010. *Jury Decision-Making Biases and Methods to Counter Them*, in «Legal and Criminological Psychology», 15, 1, 133 ff.

Ditto P.H., Pizarro D.A., Tannenbaum D. 2009. *Motivated Moral Reasoning*, in «Psychology of Learning and Motivation», 50, 307 ff.

Dror I.E., Thompson W.C., Meissner C.A., Kornfield I., Krane D., Saks M., Risinger M. 2015. *Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making*, in «Journal of Forensic Sciences», 60, 4, 1111 ff.

Epstein L., Knight J. 2013. *Reconsidering Judicial Preferences*, in «Annual Review of Political Science», 16, 11 ff.

Feigenson N. 2016. *Jurors' Emotions and Judgments of Legal Responsibility and Blame: What Does the Experimental Research Tell Us?*, in «Emotion Review», 8, 1, 26 ff.

Furgeson J.R., Babcock L., Shane P.M. 2008. *Do a Law's Policy Implications Affect Beliefs about Its Constitutionality? An Experimental Test*, in «Law and Human Behavior», 32, 219 ff.

Galef J. 2021. *The Scout Mindset: Why Some People See Things Clearly and Others Don't*, Penguin.

Goodwin G.P. 2014. *How Complete Is the Path Model of Blame?*, in «Psychological Inquiry», 25, 2, 215 ff.

Goodwin G. P. 2015. *Moral Character in Person Perception*, in «Current Directions in Psychological Science», 24, 1, 38 ff.

Goodwin G.P., Piazza J., Rozin P. 2014. *Moral Character Predominates in Person Perception and Evaluation*, in «Journal of Personality and Social Psychology», 106, 1, 148 ff.

Green M.S. 2005. *Legal Realism as Theory of Law*, in «William & Mary Law Review», 46, 6, 1915 ff.

Greene E., Dodge M. 1995. *The Influence of Prior Record Evidence on Juror Decision Making*, in «Law and Human Behavior», 19, 1, 67 ff.

Greene J.D., Haidt J. 2002. *How (and Where) Does Moral Judgment Work?*, in «Trends in Cognitive Sciences», 6, 12, 517 ff.

Greene J.D., Sommerville R.B., Nystrom L.E., Darley J.M., Cohen J.D. 2001. *An fMRI Investigation of Emotional Engagement in Moral Judgment*, in «Science», 293, 5537, 2105 ff.

Guthrie C., Rachlinski J.J., Wistrich A.J. 2000. *Inside the Judicial Mind*, in «Cornell Law Review», 86, 777 ff.

Guthrie C., Rachlinski J.J., Wistrich A.J. 2007. *Blinking on the Bench: How Judges Decide Cases*, «Cornell Law Review», 93, 1 ff.

Haidt J. 2001. *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, in «Psychological review», 108, 4, 814 ff.

Haidt J. 2012. *Moral Psychology and the Law: How Intuitions Drive Reasoning, Judgment, and the Search for Evidence*, in «Alabama Law Review», 64, 867 ff.

Haidt J., Hersh M.A. 2001. *Sexual Morality: The Cultures and Emotions of Conservatives and Liberals*, in «Journal of Applied Social Psychology», 31, 1, 191 ff.

Halverson A.M., Hallahan M., Hart A.J., Rosenthal R. 1997. *Reducing the Biasing Effects of Judges' Nonverbal Behavior with Simplified Jury Instruction*, in «Journal of Applied Psychology», 82, 4, 590 ff.

Haran U., Ritov I., Mellers B.A. 2013. *The Role of Actively Open-Minded Thinking in Information Acquisition, Accuracy, and Calibration*, in «Judgment and Decision Making», 8, 3, 188 ff.

Helion C., Pizarro D.A. 2015. *Beyond Dual-Processes: The Interplay of Reason and Emotion in Moral Judgment*, in Clausen J., Levy N. (eds.), *Handbook of Neuroethics*, Springer, 109 ff.

Helm R.K., Wistrich A.J., Rachlinski J.J. 2016. *Are Arbitrators Human?*, in «Journal of Empirical Legal Studies», 13, 4, 666 ff.

Hsee C.K. 1996. *Elastic Justification: How Unjustifiable Factors Influence Judgments*, in «Organizational Behavior and Human Decision Processes», 66, 1, 122 ff.

Ingriselli E. 2014. *Mitigating Jurors' Racial Biases: The Effects of Content and Timing of Jury Instructions*, in «Yale Law Journal», 124, 1690 ff.

Kneer M. 2022. *Reasonableness on the Clapham Omnibus: Exploring the Outcome-Sensitive Folk Concept of Reasonable*, in Bystranowski P., Janik B., Prochnicki M. (eds.), *Judicial Decision-Making: Integrating Empirical and Theoretical Perspectives*, Springer Nature, 25 ff.

Kneer M., Bourgeois-Gironde S. 2017. *Mens Rea Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, in «Cognition», 169, 139 ff.

Kneer M., Machery E. 2019. *No Luck for Moral Luck*, in «Cognition», 182, 331 ff.

Kneer M., Skoczeń I. 2023. *Outcome Effects, Moral Luck and the Hindsight Bias*, in «Cognition», 232, 105258.

Knobe J. 2003. *Intentional Action and Side Effects in Ordinary Language*, in «Analysis», 63, 3, 190 ff.

Kunda Z. 1990. *The Case for Motivated Reasoning*, in «Psychological Bulletin», 108, 3, 480 ff.

Landsman S., Rakos R.F. 1994. *A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation*, in «Behavioral Sciences & the Law», 12, 2, 113 ff.

Leiter B. 2003. *American Legal Realism*, In Edmundson W., Golding M. (eds.), *The Blackwell Guide to Philosophy of Law and Legal Theory* (2nd Ed.), Blackwell.

LINDHOLM T., CEDERWALL J.Y. 2010. *Ethnicity and Gender Biases in the Courtroom*, in GRANHAG P.A. (ed.), *Forensic Psychology in Context: Nordic and International Approaches*, Routledge, 228 ff.

MACPHERSON R., STANOVICH K.E. 2007. *Cognitive Ability, Thinking Dispositions, and Instructional Set as Predictors of Critical Thinking*, in «Learning and Individual Differences», 17, 2, 115 ff.

MALLE B.F., GUGLIELMO S., MONROE A.E. 2014. *A Theory of Blame*, in «Psychological Inquiry», 25, 2, 147 ff.

MELILLI K.J. 1998. *The Character Evidence Rule Revisited*, «Brigham Young University Law Review», 4, 1547 ff.

MELNIKOFF D.E., STROHMINGER N. 2020. *The Automatic Influence of Advocacy on Lawyers and Novices*, in «Nature Human Behaviour», 4, 12, 1258 ff.

MONROE A.E., MALLE B.F. 2017. *Two Paths to Blame: Intentionality Directs Moral Information Processing along Two Distinct Tracks*, in «Journal of Experimental Psychology: General», 146, 1, 123 ff.

NADELHOFFER T. 2006. *Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality*, in «Philosophical explorations», 9, 2, 203 ff.

NADLER J. 2012. *Blaming as a Social Process: The Influence of Character and Moral Emotion on Blame*, in «Law and Contemporary Problems», 75, 2, 1 ff.

NADLER J. 2014. *The Path of Motivated Blame and the Complexities of Intent*, in «Psychological Inquiry», 25, 2, 222 ff.

NADLER J., MCDONNELL M.H. 2011. *Moral Character, Motive, and the Psychology of Blame*, in «Cornell Law Review», 97, 255 ff.

PORTER T., ELNAKOURI A., MEYERS E.A., SHIBAYAMA T., JAYAWICKREME E., GROSSMANN I. 2022. *Predictors and Consequences of Intellectual Humility*, in «Nature Reviews Psychology», 1, 9, 524 ff.

RACHLINSKI J. J., GUTHRIE C., WISTRICH A. J. 2006. *Inside the Bankruptcy Judge's Mind*, in «Boston University Law Review», 86, 1227 ff.

RACHLINSKI J.J., GUTHRIE C., WISTRICH A. J. 2007. *Heuristics and Biases in Bankruptcy Judges*, in «Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft», 163, 167 ff.

ROESE N.J., VOHS K.D. 2012. *Hindsight Bias*, «Perspectives on Psychological Science», 7, 5, 411 ff.

SAUERLAND M., OTGAAR H., MAEGHERMAN E., SAGANA A. 2020. *Allegiance Bias in Statement Reliability Evaluations Is Not Eliminated by Falsification Instructions*, in «Zeitschrift für Psychologie», 228, 3, 210 ff.

SEVIER J. 2019. *Legitimizing Character Evidence*, in «Emory Law Journal», 68, 3, 441 ff.

SHAVER K.G. 1985. *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*, Springer.

SOOD A.M., DARLEY J.M. 2012. *The Plasticity of Harm in the Service of Criminalization Goals*, in «California Law Review», 100, 5, 1313 ff.

SOOD A.M. 2013. *Motivated Cognition in Legal Judgments—An Analytic Review*, in «Annual Review of Law and Social Science», 9, 307 ff.

SOOD A.M. 2015. *Cognitive Cleansing: Experimental Psychology and the Exclusionary Rule*, in «Georgia Law Journal», 103, 1543 ff.

SRIPADA C.S. 2012. *Mental State Attributions and the Side-Effect Effect*, in «Journal of Experimental Social Psychology», 48, 1, 232 ff.

Sripada C.S., Konrath S. 2011. *Telling More Than We Can Know about Intentional Action*, in «Mind & Language», 26, 3, 353 ff.

Stanovich K.E., Toplak M.E. 2019. *The Need for Intellectual Diversity in Psychological Science: Our Own Studies of Actively Open-Minded Thinking as a Case Study*, in «Cognition», 187, 156 ff.

Steblay N., Hosch H.M., Culhane S.E., McWethy A. 2006. *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, in «Law and Human Behavior», 30, 4, 469 ff.

Stoel R., Berger C.E.H., Kerkhoff W., Dror I. 2014. *Minimizing Contextual Bias in Forensic Casework*, n Hickman M.J., Strom K.J. (eds.), *Forensic Science and the Administration of Justice. Critical Issues and Directions*, SAGE Publications, 67 ff.

Strohmaier N., Pluut H., Van den Bos K., Adriaanse J., Vriesendorp R. 2021. *Hindsight Bias and Outcome Bias in Judging Directors' Liability and the Role of Free Will Beliefs*, in «Journal of Applied Social Psychology», 51, 3, 141 ff.

Strohmaier N., Zehnder M.A., Kneer M. forthcoming. *Moral Character Inferences Bias Legal Judgments of Jurors and Experts Across Jurisdictions*.

Strohminger N., Melnikoff D. manuscript. *Breaking Reality's Constraints on Motivated Cognition*. Online pre-print. Available on: https://psyarxiv.com/qnda3/download?format=pdf.

Uhlmann E.L., Pizarro D.A., Diermeier D. 2015. *A Person-Centered Approach to Moral Judgment*, in «Perspectives on Psychological Science», 10, 1, 72 ff.

Van Bavel J.J., Feldman Hall O., Mende-Siedlecki P. 2015. *The Neuroscience of Moral Cognition: From Dual Processes to Dynamic Systems*, in «Current Opinion in Psychology», 6, 167 ff.

Wilson T.D., Brekke, N. 1994. *Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations*, in «Psychological Bulletin», 116, 1, 117 ff.

Wistrich A.J., Guthrie C., Rachlinski J.J. 2005. *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, in «University of Pennsylvania Law Review», 153, 1251 ff.

Wylie J., Gantman A. 2023. *Doesn't Everybody Jaywalk? On Codified Rules That Are Seldom Followed and Selectively Punished*, in «Cognition», 231, 105323.

Yourstone J., Lindholm T., Grann M., Svenson O. 2008. *Evidence of Gender Bias in Legal Insanity Evaluations: A Case Vignette Study of Clinicians, Judges, and Students*, in «Nordic Journal of Psychiatry», 62, 4, 273 ff.

When "a Citizen" Becomes Little Mary J. The Abstract-concrete Effects in Legal Reasoning and the Rule of Law

PRZEMYSŁAW PAŁKA, PIOTR BYSTRANOWSKI, BARTOSZ JANIK, MACIEJ PRÓCHNICKI

1. *Introduction*

March 14 was a day of celebration. Little Mary J. would receive her extremely expensive cancer treatment. After an emotional court battle, the judge hearing her case ordered the national healthcare provider (NHP) to fund the overseas trip and the next-generation procedure and drugs. In the abstract, her case seemed doomed from the start. As the NHP statute makes clear, decisions on what can be refunded are made by the provider based on a clear-cut algorithm, taking into account the available public resources and the effectiveness of each procedure. In the case of Mary, who suffered from a very rare form of cancer, the decision had to be negative. Yet, as Mary's parents managed to hire a celebrity lawyer, who delivered a stellar speech about justice, little teddy bears, Mary's passion for horse riding, and how poor she was, the judge decided to rule based on high-level principles of equity and protection of human dignity enshrined in the constitution. Who could have denied treatment to this beautiful, unfortunate, yet brave little girl?

During the same week, the same treatment was denied to Bobby, Chris, Danielle, and other similar children in need whose names we do not even know. They were not fortunate enough to be represented by experienced lawyers. Their cases never ended up being about them—the real people with real voices and faces—and remained administrative decisions about the numbers, names of drugs, and specifics of cancer treatment. And, had they not been denied the treatment, the NHP would go bankrupt, rendering treatment for thousands of other patients impossible. If you manage healthcare for the whole nation, you make hard choices, choices in the abstract. In an individual case, each little boy and girl obviously is an exceptional patient.

This hypothetical case[1] illustrates a familiar tension: Many of us, facing a specific individual in need, would agree that society should do everything possible to help such a person. However, when asked a more general question, we might argue that scarce public resources should be allocated efficiently, which means that some people in need must be denied help. Such a seemingly inconsistent set of preferences, depending on the level of abstraction, is not limited to

* The research leading to these results has received funding from the Norwegian Financial Mechanism 2014–2021, project no. 2020/37/K/HS5/02769, titled "Private Law of Data: Concepts, Practices, Principles & Politics" (Przemysław Pałka). The work of Piotr Bystranowski and Maciej Próchnicki was supported by the European Research Council (ERC) under the H2020 European Research Council research and innovation program (agreement 805498). The work of Bartosz Janik was supported by the National Science Centre, Poland, grant no. 2020/36/C/HS5/00111. We would like to thank Tomasz Gizbert-Studnicki and Noel Struchiner for providing useful feedback on the earlier drafts.

[1] Even though this is a hypothetical case, it is representative of a class of actual court rulings across jurisdictions. See, for example, the case law of the Brazilian Superior Court of Justice, as discussed by STRUCHINER & HANNIKAINEN 2016.

any single context. Indeed, over the last couple of decades, social scientists—psychologists, economists, or experimental philosophers—have documented multiple areas in which people change their minds about a given problem when more (or less) irrelevant detail is provided (JENNI & LOEWENSTEIN 1997; SCHELLING 1968; SINNOTT-ARMSTRONG 2008). These abstract-concrete effects (ACEs), as we call them, are present in people's reactions to deeply philosophical questions, such as the issue of determinism and free will (NICHOLS & KNOBE 2007) but also in their everyday actions, such as charitable donations (SMALL et al. 2007).

Recently, many studies have indicated the presence of abstract-concrete effects in the legal domain. This should not come as a surprise. Switching between abstract and concrete perspectives lies at the core of the legal process. Judicial decision-making, the paradigm of the application of law, typically requires the judge to apply an *abstract* rule to resolve a *concrete* dispute. The tension between the two aspects of every case may not seem new, as it has long been recognized that the strict application of a given rule without considering the overall picture may lead to gross injustice. Legal reasoning is defeasible; the presence of given factors may limit the application of a general rule (BROŻEK 2004; BILLI et al. 2021). However, the presence of ACEs would indicate that this process is even more complex than traditionally assumed[2]. It is not limited to situations in which, as traditionally discussed, one has a legally, or at least morally relevant reason to not apply the binding rule (which ideally would have been thought of when drafting the rule). In contrast, here we focus on situations in which providing legal decision-makers with legally and morally irrelevant factors, which could make a given problem more concrete, may determine how a given case is decided. An analogous argument applies to cases in which there is a collision of two different legal principles, and one ought to perform a balancing procedure (ALEXY 1985). While we cannot solve a case of collision between two constitutional values fully in the abstract, it is puzzling that the addition of irrelevant details, such as the name of the plaintiff, may influence how the collision is resolved.

The widespread presence of ACEs in legal reasoning poses challenges to the ideal of the Rule of Law[3]. Indeed, the Rule of Law requires the presence of general and prospective norms, and officials base their decisions on the law, not their individual opinions (WALDRON 2008). In addition, it requires that every individual should have a right to argue their case in a fair procedure (TASHIMA 2008). In a legal system committed to the Rule of Law, both these aspects—formal and procedural—are important. Scholars have pointed out that *sometimes* these two aspects are in tension (WALDRON 2020). However, given the empirical evidence generated by psychology and cognitive sciences, we argue that this tension is likely much more commonplace than usually assumed.

In this chapter, we summarize existing research on a class of psychological phenomena that we jointly label ACEs (Section 2), give an overview of the Rule of Law requirements (Section 3), and analyze why the tension between the abstract and the concrete is problematic from the point of view of the Rule of Law (Section 4). We conclude with some ideas for potential policy intervenitons (Section 5) and further research (Section 6).

2. *Abstract-concrete effects: What are they?*

Abstract-concrete effects is an umbrella term we introduce to refer to a family of psychological effects. Although those effects have been studied within different sub-disciplines of social science, oftentimes without much cross-referencing, all of them, on the most basic level, are

[2] For the discussion on the tension between general rules in legislation and the justice in a given case, see for example SCHNEIDER 1998.

[3] We capitalize "the Rule of Law" after WALDRON 2008, in order to distinguish it from a rule of law (i.e., a particular rule in a legal system, such as the prohibition to drink and drive).

instances of the impact that the level of abstraction or concreteness has on human judgment. As these effects have been studied within different frameworks, and slightly different explanations and normative assessments have been proposed for each of them, we will first introduce them separately in this section.

Perhaps the most famous and striking instance of ACEs is the *identifiability effect*, which was first discussed by SCHELLING 1968. He pointed out the contrast between the willingness to help financially in an emotionally engaging case of a girl dying of cancer and the reluctance to be taxed to support the budget of a local hospital, even though the latter action could likely save a larger number of lives. Therefore, the effect is mostly known as the "identified victim effect," as it was usually analyzed in the context of helping victims in need, but it also extends to other contexts, such as blaming and punishing identifiable individuals (KOGUT 2011a; LOEWENSTEIN et al. 2005; SMALL & LOEWENSTEIN 2005). The more general term, identifiability effect, has the advantage of covering the effect of a setting in which one has not yet been identified, but it will be possible to identify in the future (LEWINSOHN-ZAMIR et al. 2016; SMALL & LOEWENSTEIN 2003).

The essence of this phenomenon relates to people's shifting preferences, depending on whether they are provided with the details that allow them to pinpoint a particular individual affected by their choices. Identified groups may not trigger similar reactions (see KOGUT & RITOV 2005b; LEE & FEELEY 2016), which constitutes another phenomenon: the singularity effect (KOGUT & RITOV 2005b; MOCHE et al. 2022)., also known as individuation (STRUCHINER et al. 2020). However, the effect may persist if a group is perceived as a unit (SMITH et al. 2013). A variety of information can be treated as allowing the identification of an individual, including a name, photograph, or even age or ID number (FRIEDRICH & MCGUIRE 2010; KOGUT & RITOV 2005a, 2005b; SMALL et al. 2007; SMALL & LOEWENSTEIN 2003). The identifiability effect seems to affect choices even if people are asked to attribute responsibility to a single individual identified only by a number rather than to an unspecified individual (SMALL & LOEWENSTEIN 2005).

A number of theories have been proposed to explain the identifiability effect. One explanation emphasizes the affective aspect of the process. Emotional reactions, such as compassion or sympathy or feeling the distress of the other, are directed *via* empathy toward the suffering fellow human (see also SABATO & KOGUT 2021). Interestingly, the effect may be moderated by the perception of the individual as a member of one's own group or an outsider (KOGUT 2011a, KOGUT 2011b; KOGUT & RITOV 2007). Other approaches stress the potential cognitive aspect: people may perceive their actions as more effective or having a greater impact if they are addressed at an identified individual (DUNCAN 2004) or have a greater responsibility toward such an individual (BASIL et al. 2006). Some models (CHAIKEN 2003) argue that both affective and cognitive factors may be at play due to different kinds of information processing about specific and abstract targets[4].

When it comes to real-life studies, the strength of the effect is disputed (LESNER & RASMUSSEN 2014). A recent replication study failed to replicate the effect of identification in the context of singular victims (MAJUMDER et al. 2022). Meta-analytic studies have confirmed the existence of the effect (BUTTS et al. 2019; see also LEE & FEELEY 2016), but estimate its overall magnitude as modest (LEE & FEELEY 2016), limiting the strength of the effect mostly to the radiant examples (e.g., a single victim of misfortune to which they did not contribute or that happened independently of them).

Other studies have explicitly analyzed the role of the identifiability effect in legal contexts. LEWINSOHN-ZAMIR et al. 2016, for example, presented a sample of law students with five tort law scenarios and two cases of violations punishable by fine and asked participants to take the

[4] See also ERLANDSSON et al. 2021; RAHAL & FIEDLER 2022 for a current discussion of the cognitive and affective background of prosocial behavior.

role of a policy-maker (a legislator) or decision-maker (the adjudicator). The cases in the latter condition included the names of the tortfeasor and the victim (in civil cases) or the perpetrator (in public law cases). In both types of cases, identifiability influenced the participants' decisions. BARAK-CORREN & LEWINSOHN-ZAMIR 2019 examined the role of identifiability in sexual harassment cases. Identified offenders were treated better than anonymous ones, and the reverse held for victims. A number of studies have also analyzed how legal outcomes might be determined by legally irrelevant factors, which could increase or decrease the judge's sympathy for a party (SPAMANN & KLÖHN 2016; WISTRICH et al. 2014).

Another line of research addressed identifiability strictly in court settings by changing the degree of case detail provided to the participants. In studies by STRUCHINER et al. 2020, one of the conditions identified a concrete individual (name and location) as a plaintiff as opposed to the other condition, which was also framed as a concrete case but did not include identifying information. As it turned out, identifying information made people evaluate judicial intervention more favorably.

The second example of ACEs can be found in the literature on the *construal-level theory of psychological distance* (EYAL & LIBERMAN 2012; TROPE & LIBERMAN 1996), which posits that, while reasoning about objects that transcend their direct experience and immediate situation, people form abstract mental construals of those objects. What is central to this process is *psychological distance*, that is, how far away from a given person the object *feels*, in terms of space, time, or the level of abstraction. Although the psychological distance corresponds to many such dimensions, the construal level theory (CLT) assumes that this subjective experience that something is close or far away will result in similar patterns of reasoning independently of which actual dimension is present in a given context.

Although the construal level theory has been applied in many areas of psychology (HO et al. 2015; LIBERMAN et al. 2007), moral reasoning remains one of the most visible examples of its application. In general, CLT predicts that when facing a moral dilemma, people are more likely to choose options that lead to overall better outcomes (even when it means violating some general principles or moral taboos) *when the object feels close*, but they are more likely to choose options consistent with general moral principles *when the object feels distant* (KÖRNER & VOLK 2014; PATIL et al. 2014). For example, BOSTYN et al. 2018 analyzed how the construal level can affect intuitive choices in the so-called trolley problem (in which the actor can choose to cause the death of one innocent person in order to save the lives of a larger group of people). The researchers asked participants to choose whether to kill one mouse by applying an electric shock to save the lives of five mice. However, one group of participants was presented with this scenario as a pure hypothetical, while the other group was made to believe that their decision would be immediately implemented to a group of mice they could see through a window. It turned out that people in the "real-life" condition (in which, arguably, the objects of their choice felt psychologically closer) were more likely to choose to kill the innocent mouse.

Legal decision-making appears to be an obvious field in which CLT can be expected to work similarly as in moral reasoning. After all, a judge typically makes decisions whose "objects," that is, parties to the proceedings, are psychologically close across many dimensions: They are concrete and located in the same place in time, and they are often also present in the same room. In contrast, the lawmaker's decisions will typically affect psychologically distant objects, as new rules normally apply only to future, hypothetical cases. A study by CAVIOLA et al. 2021 provided evidence that CLT correctly predicts a divergence in responses to the analytically same legal problem, depending on whether the participants are asked to decide on an existing one-shot problem or rather establish a rule for future cases. The authors asked a sample of lay participants to decide whether to prioritize saving the life of a socially useful individual over saving a socially less useful one in an emergency situation. The participants were systematically less inclined to give such priority when deciding on a rule rather than deciding on a specific, one-shot problem. Interestingly, this *generality effect* also

persisted in intermediate cases: When participants were asked to decide on a specific case but also to assume that their decision would form a precedent applicable to future cases, they would provide responses consistent with the abstract construal (i.e., refusing to prioritize).

The third example of ACEs, the *abstract-concrete paradox* (NAHMIAS et al. 2007; NICHOLS & KNOBE 2007; SINNOTT-ARMSTRONG 2008), refers to the situation in which the description of a problem in more abstract or more concrete terms results in people's diverging philosophical intuitions. The paradox was studied with respect to intuitions concerning free will, moral responsibility, knowledge, and rule-following. The most notable finding of the initial set of studies was that people tend to attribute moral responsibility to agents in a deterministic world when confronted with concrete cases of wrongdoing, while they find moral responsibility incompatible with determinism if asked the question in abstract terms. The concept of concreteness was operationalized by experimental philosophers in two major ways in some cases: as a reference to specific agents (SINNOTT-ARMSTRONG 2008) or as a detailed description of an action that was previously left undescribed (MANDELBAUM & RIPLEY 2012).

There are at least two kinds of theories that account for the persistence of the abstract-concrete paradox: affective and cognitive. According to the theory developed by NICHOLS & KNOBE 2007, the paradox is an effect of an affective performance error: A violation of moral norms in concrete cases creates affective responses, which in turn distort our reliance on the tacit theory of making judgments concerning responsibility, which tends not to elicit specific affective responses. This, in turn, explains why the attribution of responsibility is higher in concrete cases than in abstract ones. On the other hand, according to the *separate capacities hypothesis* developed by SINNOTT-ARMSTRONG 2008, the paradox results from the use of two different memory systems while judging different cases. This theory claims that in the case of the concrete level, we encode and represent it in episodic memory, which supports the attribution of responsibility. In contrast, in abstract cases, we encode them in semantic memory, which tends to produce decreased responsibility attributions (SINNOTT-ARMSTRONG 2008).

A notable study on judicial reasoning inspired by research on ACP was conducted by STRUCHINER et al. 2020. The authors demonstrated that legal reasoning can be affected by this phenomenon in at least two contexts. In the context of the application of vague legal principles, the authors hypothesized that when such a principle has at least two conflicting interpretations (e.g., the libertarian or paternalistic interpretation of the principle of the protection of human dignity), different interpretations might be associated with different levels of abstraction in which a given problem is presented. In the context of the *application of clear-cut legal rules*, on the other hand, Struchiner and colleagues hypothesized that people are more likely to apply the literal meaning of a rule in abstract but to decide against such a literal meaning in a concrete case. Struchiner and colleagues obtained substantial support for their hypotheses with both lay and professional (i.e., Brazilian judges) samples.

Furthermore, the participants in this study were asked to complete parts of the Interpersonal Reactivity Index questionnaire, thus capturing individual differences in cognitive (*perspective taking*) and affective (*empathic concern*) empathy. Echoing previous findings, affective but not cognitive empathy predicted the subjects' responses. In particular, highly empathic individuals were most susceptible to the manipulation of abstraction versus concreteness. While the original study was conducted using a between-subjects design (i.e., each participant decided cases on just one level of abstraction), in a follow-up study, Struchiner and colleagues presented professional judges with both abstract and concrete versions in a counterbalanced order (abstract-first or concrete-first), asking them to decide on each version. Strikingly, in such a setting, judges strived for consistency, with the effect present in reaction to the first version they saw but disappearing when they saw the second version.

One potential explanation of the pattern of results observed by Struchiner and colleagues is the fact that both lay and professional participants in their studies were recruited in Brazil, a country

whose legal culture is known for its particularism (i.e., the tendency to look for a just solution to a given case even at the expense of not following the relevant abstract rules). To address this potential challenge, BYSTRANOWSKI et al. (2022) replicated the effect on the population of judges from a country following the formalist approach to adjudication (Poland), who—as it can be hypothesized (SCHAUER 2006; see also SHERWIN 2006)—could have been expected to be consistent about the interpretation of legal rules, thus providing some evidence that the presence of ACEs in legal reasoning is not contingent on the peculiarities of a given legal system.

Just as the three analyzed strands of literature differ in terms of explanations provided for the postulated effect, so different are the normative perspectives from which they analyze whether the effect can be considered a bias and, if so, which level of reasoning, abstract or concrete, should be considered superior and preferable. Perhaps the most intense normative discussion has emerged around the literature on the identifiability effect. Most scholars contributing to this literature tended to argue that pure identifiability lacks moral significance (HOPE 2001), is irrelevant to a given target decision, and, as it is mostly driven by affective reactions, changing one's moral preferences on its basis is undesirable (GREENE 2008). Furthermore, to the extent that identification makes one focus on individual cases at the expense of maximizing some social good (which, in the public policy context, likely results in a suboptimal allocation of resources), decisions made at the abstract level seem preferable from the consequentialist point of view[5]. However, recently, a number of philosophers have started to argue that decisions affected by the identifiability effect can be defensible on at least some metaethical grounds, such as *ex ante* contractualism (ŻURADZKI 2018).

While the literature on CLT and the abstract-concrete paradox have been less involved in discussions on the relative advantages of reasoning in abstract or in concrete, we can see that at least some authors associate abstract judgment with more effortful and reflexive reasoning, which arguably implies that the resulting judgment is more reliable (SINNOTT-ARMSTRONG et al. 2010).

To recapitulate, in this section, we briefly discussed three examples of ACEs: psychological phenomena which affect people's intuitive judgment depending on the level of abstraction in which a problem is presented. Abstract-concrete effects appear to be present in many social contexts, including moral judgment and, as emerging evidence suggests, in legal decision-making, possibly including decisions made by professional legal officials. While some scholars see the effects as normatively problematic (and, specifically, consider abstract judgment superior in some important respects), the normative debate is still in its infancy. Thus, even if we were to assume that ACEs are pervasive in legal decision-making, the normative assessment of such a state of affairs would require some further theoretical work.

3. *The Rule of Law perspective*

In the previous section, we discussed the results of the empirical studies aimed at addressing the positive question of how people make normative decisions depending on the level of generality and the amount of, seemingly irrelevant, information they have about the persons whom the decision concerns. In this section, we move toward the normative question: How *should* officials make decisions in a system committed to the Rule of Law? As we demonstrate in what follows, even though the prescriptive dimensions of the Rule of Law might seem clear and coherent on paper, their implications for reacting to the widespread ACEs in normative reasoning are not straightforward.

[5] For an overview of the arguments for and against consequentialism in the context of the identifiability effect, see DANIELS 2012.

3.1. *The ideal of the Rule of Law*

The Rule of Law is one of the foundational concepts of liberal democratic legal orders, next to democracy, human rights, and, arguably, social justice and economic freedom (RAZ 1977; WALDRON 2020). It is both an ideal in political philosophy and, in some jurisdictions, a specific black letter doctrine (LENAERTS 2020). Given that the specific contents of such black letter doctrines differ between jurisdictions while remaining in dialogue with the philosophical ideal (SELLERS & TOMASZEWSKI 2010), in this article we refrain from engaging the specifics of the national doctrines and focus on the philosophical accounts.

Theorists of the Rule of Law distinguish its three aspects: (1) formal, (2) procedural, and (3) substantive (WALDRON 2020). The last one is the most contentious. Under some accounts (BINGHAM 2010), for a legal system to comply with the Rule of Law, it must protect basic human rights, that is, feature certain substantive guarantees. Other scholars (RAZ 1977; WALDRON 2020) have remarked that although human rights and substantive justice are clearly important ideals of a good legal system, for the sake of conceptual clarity, they should be distinguished from the Rule of Law in the strict sense. Without taking sides in this debate, for the sake of clarity, in this article, we focus on the formal and procedural aspects of the Rule of Law.

The formal understanding of the Rule of Law has been most famously, according to WALDRON 2008, theorized by Lon FULLER in his *Morality of Law* (1969). In his account, a legal system committed to the Rule of Law should feature norms that are general (addressed to all subjects and not specific persons), prospective (not retroactive), intelligible (understandable to their addressees), consistent (not contradictory), practicable (possible to realize), stable (not changing overnight), and congruent (applied by the officials exactly as they are promulgated). This last requirement that officials do what the law stated in advance has been highlighted by Jeremy Waldron as the central element of the Rule of Law:

> «The Rule of Law is a multi-faceted ideal. Most conceptions of this ideal, however, give central place to a requirement that people in positions of authority should exercise their power within a constraining framework of public norms, rather than on the basis of their own preferences, their own ideology, or their own individual sense of right and wrong (...) the Rule of Law is violated when the norms that are applied by officials do not correspond to the norms that have been made public to the citizens, or when officials act on the basis of their own discretion rather than norms laid down in advance» (WALDRON 2008, 6).

In this sense, the formal aspect of the Rule of Law emphasizes both certain features of legal norms (that they be general, prospective, clear, etc.) and the mode of their application by officials. Such a system allows individuals to plan their lives with full knowledge of what to expect from the state (RAZ 1977). It also, arguably, helps vindicate other ideals, like human rights or democracy, where, in liberal democratic systems, individuals know that their future fate depends on the commonly agreed-upon rules of the game, and not a personal preference of a particular judge or a clerk.

The procedural understanding of the Rule of Law complements the formal dimension by emphasizing that the application of the law by officials should occur with certain guarantees in place. Jeremy Waldron, summarizing the work of TASHIMA 2008, has postulated four procedural guarantees necessary for a system to qualify as compliant with the Rule of Law:

1. A hearing by an impartial and independent tribunal that is required to administer existing legal norms on the basis of the formal presentation of evidence and arguments.
2. A right to representation by counsel at such a hearing.
3. A right to be present, to confront and question witnesses, and to make legal arguments about the bearing of the evidence and the various legal norms relevant to the case.

4. A right to hear reasons from the tribunal when it reaches its decision that are responsive to the evidence and arguments presented before it (WALDRON 2020).

What matters from the point of view of the procedural aspect of the Rule of Law is not merely that the final decision is congruent with the existing law but primarily that all the features necessary (or, at least, increasing the probability) for such an outcome are present as well. Not just the content of the decision, but also the procedure for arriving at it, matters.

However, one can see how these two aspects of the Rule of Law ideal, formal, and procedural, can be in conflict. Waldron pointed this out:

> «For the most part, these two currents of thought sit comfortably together. They complement each other. Clear, general public norms are valueless if they are not properly administered, and fair procedures are no good if the applicable rules keep changing or are ignored altogether. But there are aspects of the procedural side of the Rule of Law that are in some tension with the ideal of formal predictability (...). Instead of the certainty that makes private freedom possible, the procedural aspects of the Rule of Law seem to value opportunities for active engagement in the administration of public affairs» (WALDRON 2008, 8).

Given our discussion of ACEs in the previous section, we postulate that this tension between the formal and procedural aspects of the Rule of Law is much more commonplace than Waldron admits. It is structurally present whenever an official must apply a general norm to a concrete case. Moreover, empirical research suggests that the decision might be *different* depending on how much, seemingly irrelevant, information the official has.

A clear case of the formal and procedural aspects of the Rule of Law being in conflict can be seen when discussing the phenomenon of judicial discretion. According to some scholars (DICEY 1915; HAYEK 2007), judicial discretion is, in principle, incompatible with the formal aspect of the ideal Rule of Law. If a judge "fills the gap" in the set of legal rules they need to apply, even when doing so based on high-level legal principles, they come closer to being a legislator than an adjudicator. A decision based on a rule generated in this way (or inferred depending on one's views) can hardly be deemed "prospective." Other scholars (DAVIS 1969), however, point out that the phenomenon of judicial discretion cannot be, as a matter of fact, and should not be, as a matter of justice, fully eliminated from the legal system. To uphold the Rule of Law, legislators should, instead, properly frame and authorize (WALDRON 2020) judicial discretion when necessary.

3.2. *Data show the tension might be much more common*

As we saw above, overall, legal scholars endorse values associated with both the formal and procedural aspects of the rule of law and see the two aspects as mutually reinforcing, with a fair procedure in which the law is applied being a necessary condition for the abstract formal Rule-of-Law values to be realized in life. If some of these scholars notice room for a conflict between the two aspects, it is primarily because of activist legal officials, who might abuse legal procedures and the scope of judicial discretion to enforce their private values at the expense of formal requirements of the Rule of Law. In other words, such a conflict would be a deviation from the norm, dependent on the (presumably conscious) choices of specific legal officials.

Against this more traditional picture, we postulate that the empirical research on ACEs implies that the tension between the two aspects is more typical and fundamental than has been acknowledged thus far. To the extent that deciding a concrete case might systematically change the way a legal decision-maker understands the relevant legal rule, legal procedure undermines, rather than reinforces, the Rule-of-Law virtues of general rules.

Let us revisit the empirical data most striking in this context. While the established body of research on ACEs aimed at demonstrating that the *moral* preferences of a given actor can shift

depending on the level of abstraction (CAVIOLA 2021), more recent studies in legal decision-making (BYSTRANOWSKI et al. 2022; STRUCHINER et al. 2020) demonstrate that the *interpretation* of existing law (including interpretation by professional judges) is affected by the level of abstraction. In other words, the same (textually) constitutional principle or statutory rule can be interpreted differently in the abstract and when applied to a particular case. Strikingly, STRUCHINER et al. 2020 observed that professional judges, if given an opportunity to interpret and apply the same rule both in abstract and concrete, managed to give consistent responses. This provides evidence that legal officials strive for consistency.

This observation (that legal decision-makers' tendency to interpret law under the influence of ACEs diminishes when they are asked to make choices both in abstract and concrete at the same time) can be seen as evidence that those effects constitute a bias, at least from the point of view of legal officials themselves (HSEE et al. 1999). However, from a normative perspective, more analysis is needed before we decide whether ACEs are indeed unwelcome in the domain of legal decision-making. So far, we have pointed out that their presence in this context amplifies the tension between the formal and procedural aspects of the Rule of Law. Whether, normatively, this tension should be accepted and welcome or resolved and eliminated is a question we address directly in the next section.

4. *The tension is problematic*

After reviewing empirical data providing evidence of the commonality of the assumed tension, one should emphasize why this tension may be problematic for legal scholars and lawmakers. We will do that by reviewing some theoretical insights and putting this issue against the ideal of the Rule of Law.

Having some understanding of a legal rule in the abstract, judges change their interpretations when facing the need to apply the same rule to a specific case. They change their interpretations after realizing that this rule will affect the lives of actual individuals, after familiarizing themselves with the details of the case and after listening to the parties' arguments. Are such changes necessarily something to be frowned upon? Certainly not, at least for some scholars who see the right to a fair trial as the main tenet of the Rule of Law—something that is not of a merely instrumental value (that is, worth practicing to the extent that it is conducive to achieving some more fundamental values, such as the formal Rule-of-Law values) but rather something of intrinsic value (HAREL 2014; HAREL & KAHANA 2010).

From a judge's point of view, if we assume that the purpose of a trial is not only to discover facts relevant to the case but also to hear the arguments of both parties, it appears very much welcome that the judge is open to revising their interpretation of the law in light of the details of the case. In addition, from the perspective of the parties, to the extent that the trial is a realization of their right to be heard, they should be able not only to present relevant evidence but also to argue, and perhaps convince the judge, that their situation and their claims are particular, possibly providing an exception to the abstractly understood rule.

Turning to the ideal of the Rule of Law, one might say that the tension's existence is problematic because it renders the mentioned ideal an empty postulate. To provide the full picture, we should move back to the formal aspects of the Rule of Law presented earlier, relying heavily on Fuller's work (FULLER 1969; WALDRON 2020). Bearing in mind the empirical data from the abovementioned studies on ACEs, those formal aspects (generality, intelligibility, etc.) might be analyzed in action (WALDRON 2016), and we might argue that those formal features might be especially vulnerable to the operation of ACEs, thus creating and strengthening the tension between the Rule of Law and judicial decision-making.

In *The Morality of Law*, Fuller listed possible ways in which features of the law, later interpreted as formal aspects of the Rule of Law, might be affected by bad legislation and a distorted process of

applying the law. Most of the formal features of the Rule of Law could be perceived as rules of creating legal provisions, and as such, one might say that in this interpretation, they will be safe from the operation of ACEs by simply not being applied to the application of the legal rules. However, one might say that if we treat the Rule of Law as an ideal that should be applied to the law in action as well, then suddenly features such as the generality or the prospective character of the legal rules might be threatened by the operation of ACEs. For example, if an interpretation of a rule made by a judge differs between abstract and concrete cases, one might say that the prospective character of this rule is violated. Another example might be one obvious feature of the legal system and one that is directly affected by the incoherent interpretation of the legal rules: congruence. This feature might be negatively affected: «mistaken interpretation, inaccessibility of the law, lack of insight into what is required to maintain the integrity of a legal system, bribery, prejudice, indifference, stupidity, and the drive toward personal power» (FULLER 1969, 81). Putting aside ways that require, obviously, criminal (or malicious) intent, one may ask to what extent ACEs may operate here. By briefly reintroducing the possibility of a different understanding of a given legal rule by the official applying the law, one might conclude that the idea of congruence, by mistaken interpretation or even prejudice, might be in danger.

The formal features of the Rule of Law might thus be affected by ACEs-related cases, effectively creating the situation in which the abovementioned tension should be perceived as problematic.

5. *Can the tension be resolved?*

From our discussion until now, it should be clear that, unlike with some other biases present in legal decision-making, the presence of ACEs, scrutinized from the point of view of the Rule of Law, there is no clear solution to this problem. It is hard to propose "solutions" to the problem—particular means, or concrete interventions—when the goals of the intervention, or even the very need for any intervention are disputed. In some ways, ACEs help to vindicate the procedural aspect of the Rule of Law; in others, they jeopardize its formal aspect. As there is no common understanding of what the ideal is, the proposed interventions could go in opposite directions.

For these reasons, we refrain from offering any "recommendations" on how judges should behave, or what lawmakers should do. We believe that the main function of this chapter is to raise awareness and illuminate, report empirical evidence, and point out normative problems, leaving the prescriptive work to future analyses that would start from stronger normative assumptions. Yet, as specific solutions usually seem more concrete to lawyers than the discussion of a problem in the abstract, in this section, we discuss a thought experiment playing with possible interventions and scrutinize them from the point of view of both the formal and procedural aspects of the Rule of Law. To the extent that ACEs in law result simply from applying general rules to specific cases, it seems that they will stay with us as long as it is human judges that are tasked with interpreting and applying the law. However, as these effects are magnified by some salient situational factors, such as the identification of parties, one could imagine what some potential interventions could look like.

Consider the possibility of holding "anonymous trials." We could imagine a world in which both parties, witnesses, and counsels are behind a veil of anonymity so that the decision-makers do not know their names, cannot see their faces, and potentially even hear their real voices. Moreover, anonymity could include the inability to share more information than is required by the legal norms being applied. Such a limited approach, addressing only some aspects of concretization, would render decision-making slightly more abstract. But would it be of any help from the point of view of the ideal Rule of Law?

Regarding the formal aspect, the answer could be affirmative. As rules are made in the abstract, their application in the abstract helps advance the ideals of predictability and congruence. However, from a procedural point of view, such a suggestion seems problematic. Arguably, the

Rule of Law requires that one's case is about oneself, and that includes the right to have one's face seen, one's voice heard, and being treated as a human being in the fullest possible sense, not just some anonymous number. Yes, a judge might be more lenient when, on top of rationally learning that a defendant has expressed remorse, they see the emotions on one's face and hear the tone of one's voice. However, is this always and necessarily bad? This is a hard call.

Such a simple thought experiment illuminates a more profound claim: from the mere fact that ACEs exist, even when studying them from the point of view of only one normative ideal, the Rule of Law, one cannot infer an obvious prescription regarding the goals of a lawmaker's intervention. Both the legislature and the decision-makers should be aware of the functioning and presence of ACEs. What to do about them, however, is a political choice made by the polity.

6. *Further research*

Thus far, we have discussed only the role that ACEs play in the tension between different conceptualizations of the Rule of Law ideal. The existence of these phenomena and their impact on legal decision-making open a number of issues to be explored, including general normative considerations in political and legal philosophy, empirical work, and some more direct implications for constitutional law, such as the role of judicial review. As these issues are complex, we will limit ourselves to highlighting them and pointing out further directions of inquiry.

Our considerations concern only the strain between the formal and procedural requirements of the Rule of Law. Further research should consider how the different Fullerian requirements should be structured in the legal systems, possibly balancing the use of concrete and abstract modes of decision-making and making the rationale of introducing these more explicit. Moreover, the substantive aspect of the Rule of Law may be affected by the tension as well.

Normative research should be supplemented by empirical work that aims to explore whether different models of decision-making already present in the legal system are actually prone to ACEs and, if so, whether the shifts in preferences, depending on the amount of details, are problematic from the viewpoint of cohesiveness of legally relevant decisions. For example, some aspects of the thought experiment presented above are already present in different jurisdictions: one may ask to what extent does the written or oral form of different proceedings tend to affect decisions made in practice?

As for the direct practical consequences for the implementation of the ideal of the Rule of Law into actual constitutional orders with ACEs in mind, the issue of judicial review stands out. Let us discuss this briefly. On one hand, the ability of the judiciary to strike down legislation and protect individual rights may be seen as a guarantee of political legitimacy and enforcing the Rule of Law principles (BERTOMEU 2011; FALLON JR. 2019; ZURN 2007), but on the other hand is challenged as *prima facie* undemocratic (see, e.g., TUSHNET 1999; TUSHNET 2020; WALDRON 2020).

From a theoretical perspective, the issue with arguments for and against judicial review as part of the Rule of Law and ACEs will be analogous to the problem of tension, as discussed above. However, an analysis of the influence of ACEs on legal reasoning seems crucial in implementing a particular model of judicial review. Different jurisdictions tend to employ very diverse approaches to controlling legislative and executive acts through the lens of constitutional provisions. Foremostly, we can distinguish between strong and weak forms of judicial review (WALDRON 2020). As described by Waldron, the former type refers to the systems in which courts, apart from controlling administrative and executive decisions, hold the power to either strike down provisions that they have found to be violating constitutional rights or do not bypass the application of the clearly relevant statute or modify the scope of its application with prospective binding power. A weak judicial review refers only to the courts' signaling power, which does not allow them to bypass the application of a potentially rights-violating rule. These forms can be described as prototype

categories of the strength of judicial control, while actual legal systems adopt a wide spectrum of these. The other perspective on judicial review is institutional, where we can distinguish between two main models: American and Kelsenian (FEREJOHN 2002). The main difference between the two refers to yielding the power to review to ordinary courts over their day-to-day adjudication in the American model, and creating a specialized body—constitutional court—what can assess the validity apart from a particular case in the European model (RAMOS 2007). The differentiation between various jurisdictions goes further when it comes to procedural and substantive relations between legislatures, ordinary courts, and constitutional courts (RAMOS 2007).

Three aspects of judicial review seem to be particularly relevant from the cognitive viewpoint of ACEs: the type of court that is allowed to review the legislation[6]; the mode of constitutional review (if it is performed in abstract, without a reference to a particular case, or requires a concrete case in which one's rights were infringed); and the outcome of the ruling (whether it establishes a precedent or only confirms the violation of rights in a given case).

Judicial review is not the only area of the legal system that seems particularly susceptible to ACEs. Apart from their impact on the judiciary and the level of application of law, the tension between concrete and abstract reasoning also exerts pressure on legislative bodies. Two recent studies (FANARRAGA 2020; SOCIA 2022) shed light on the practice of so-called "apostrophe laws," that is, naming statutes after specific cases (for example, using a victim's name). One might argue that these kinds of laws make use of identifiability and concrete reasoning in order to include some more radical solutions, which would be difficult to implement only when abstract reasoning is at play. Further research should carefully examine the potential role of ACEs in the political process accompanying such statutes.

7. *Conclusions*

In this chapter, we presented a cluster of cognitive phenomena that result in people judging a problem differently, depending on the level of abstraction described or the amount of potentially irrelevant details provided. The mechanisms behind these ACEs are complex, seem to include a number of cognitive and affective factors, and cannot be reduced simply to empathetic or negative reactions toward a given individual.

As we argue, research on ACEs in general, and on ACEs in legal contexts in particular, should not be overlooked by legal philosophers. While the discussion on when strict application of general rules should give way to recognizing the particularities of a specific case spans millennia, here we argue that the conflict between abstract and concrete legal decision-making might be much more common than traditionally acknowledged. The tendency of legal decision-makers to disregard general legal norms might actually not be limited to situations when a given case is exceptional in some possibly relevant sense, but instead simply results from the effect of different levels of abstraction on which decision-making takes place.

The discrepancy between these two modes of reasoning, abstract and concrete, amplifies the tension problem in the discussion of the formal aspects of the rule of law. In practice, choosing between abstract and concrete approaches to legal disputes remains a policy decision, and each solution bears its own problems (RACHLINSKI 2006). The impact of ACEs on judicial review and legislation should be carefully analyzed by legal philosophers, constitutional law scholars, and political theorists.

[6] The specialization and experience of judges may play a major role; see LEIBOVITCH 2017.

References

ALEXY R. 1985. *Theorie der Grundrechte*, Suhrkamp.

BARAK-CORREN N., LEWINSOHN-ZAMIR D. 2019. *What's in a Name? The Disparate Effects of Identifiability on Offenders and Victims of Sexual Harassment*, in «Journal of Empirical Legal Studies», 16, 4, 955 ff.

BASIL D.Z., RIDGWAY N.M., BASIL M.D. 2006. *Guilt Appeals: The Mediating Effect of Responsibility*, in «Psychology & Marketing», 23, 1035 ff. Available on: https://doi.org/10.1002/mar.20145.

BERTOMEU J.F.G. 2011. *Against the Core of the Case: Structuring the Evaluation of Judicial Review*, in «Legal Theory», 17, 2, 81 ff. Available on: https://doi.org/10.1017/S1352325211000073.

BILLI M., CALEGARI R., CONTISSA G., LAGIOIA F., PISANO G., SARTOR G., SARTOR G. 2021. *Argumentation and Defeasible Reasoning in the Law*, in «J», 4, 4, 897 ff. Available on: https://doi.org/10.3390/j4040061.

BINGHAM T. 2010. *The Rule of Law*, Penguin.

BOSTYN D.H., SEVENHANT S., ROETS A. 2018. *Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas*, in «Psychological Science», 29, 7, 1084 ff. Available on: https://doi.org/10.1177/0956797617752640.

BROŻEK B. 2004. *Defeasibility of legal reasoning*, Zakamycze.

BUTTS M.M., LUNT D.C., FRELING T.L., GABRIEL A.S. 2019. *Helping One or Helping Many? A Theoretical Integration and Meta-analytic Review of the Compassion Fade Literature*, in «Organizational Behavior and Human Decision Processes», 151, 16 ff.

BYSTRANOWSKI P., JANIK B., PRÓCHNICKI M., HANNIKAINEN I.R., DE ALMEIDA F.C.F., STRUCHINER N. 2022. *Do Formalist Judges Abide by Their Abstract Principles? A Two-Country Study in Adjudication*, in «International Journal for the Semiotics of Law-Revue Internationale de Sémiotique Juridique», 35, 5, 1903 ff.

CAVIOLA L., SCHUBERT S., MOGENSEN A. 2021. *Should You Save the More Useful? The Effect of Generality on Moral Judgments about Rescue and Indirect Effects*, in «Cognition», 206, 104501. Available on: https://doi.org/10.1016/j.cognition.2020.104501.

CHAIKEN S. 2003. *Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion*, in «Social Psychology: A General Reader», 461 ff.

COVA F., STRICKLAND B., ABATISTA A., ALLARD A., ANDOW J., ATTIE M., BEEBE J., BERNIŪNAS R., BOUDESSEUL J., COLOMBO M. 2018. *Estimating the Reproducibility of Experimental Philosophy*, in «Review of Philosophy and Psychology», 1 ff.

DAVIS K.C. 1969. *Discretionary Justice: A Preliminary Inquiry*, Louisiana State University Press. Available on: https://bac-lac.on.worldcat.org/oclc/422280062.

DICEY A.V. 1915. *Introduction to the Study of the Law of the Constitution* (8th Ed.), Macmillan and Co. Available on: http://link.library.utoronto.ca/eir/EIRdetail.cfm?Resources_ID=483687&T=F.

DUNCAN B. 2004. *A Theory of Impact Philanthropy*, in «Journal of Public Economics», 88, 9, 2159 ff. Available on: https://doi.org/10.1016/S0047-2727(03)00037-9.

ERLANDSSON A., MOCHE H., DICKERT S. 2021. *A New Typology of Psychological Mechanisms Underlying Prosocial Decisions*, manuscript.

EYAL T., LIBERMAN N. 2012. *Morality and Psychological Distance: A Construal Level Theory Perspective*, in MIKULINCER M., SHAVER P.R. (eds.), *The Social Psychology of Morality:*

Exploring the Causes of Good and Evil, American Psychological Association, 185 ff. Available on: https://doi.org/10.1037/13091-010.

Fallon jr R.H. 2019. *The Core of an Uneasy Case for Judicial Review*, in Id., *The Nature of Constitutional Rights: The Invention and Logic of Strict Judicial Scrutiny*, Cambridge University Press, 179 ff. Available on: https://doi.org/10.1017/9781108673549.008.

Fanarraga I. 2020. *What's in a Name? An Empirical Analysis of Apostrophe Laws*, in «Criminology, Criminal Justice, Law & Society», 21, 39 ff.

Ferejohn J.E. 2002. *Constitutional Review in the Global Context*, in «New York University Journal of Legislation and Public Policy», 6, 49.

Friedrich J., McGuire A. 2010. *Individual Differences in Reasoning Style as a Moderator of the Identifiable Victim Effect*, in «Social Influence», 5, 3, 182 ff.

Fuller L.L. 1969. *The Morality of Law*, Yale University Press.

Greene J.D. 2008. *The Secret Joke of Kant's Soul*, in Sinnott-Armstrong W. (ed.), *Moral Psychology. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, 35 ff.

Harel A. 2014. *Why Law Matters*, Oxford University Press.

Harel A., Kahana T. 2010. *The Easy Core Case for Judicial Review*, in «Journal of Legal Analysis», 2, 1, 227 ff. Available on: https://doi.org/10.1093/jla/2.1.227.

Hayek F. A. 2007. *The Road to Serfdom: Text and Documents. The Definitive Edition*, University of Chicago Press.

Ho C.K., Ke W., Liu H. 2015. *Choice Decision of E-learning System: Implications from Construal Level Theory*, in «Information & Management», 52, 2, 160 ff.

Hope T. 2001. *Rationing and Life-saving Treatments: Should Identifiable Patients Have Higher Priority?* in «Journal of Medical Ethics», 27, 3, 179 ff.

Hsee C.K., Loewenstein G.F., Blount S., Bazerman M.H. 1999. *Preference Reversals between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis*, in «Psychological Bulletin», 125, 576 ff. Available on: https://doi.org/10.1037/0033-2909.125.5.576.

Jenni K., Loewenstein G. 1997. *Explaining the Identifiable Victim Effect*, in «Journal of Risk and Uncertainty», 14, 3, 235 ff.

Kogut T. 2011a. *The Role of Perspective Taking and Emotions in Punishing Identified and Unidentified Wrongdoers*, in «Cognition & Emotion», 25, 8, 1491 ff.

Kogut T. 2011b. *Someone to Blame: When Identifying a Victim Decreases Helping*, in «Journal of Experimental Social Psychology», 47, 4, 748 ff.

Kogut T., Ritov I. 2005a. *The "Identified Victim" Effect: An Identified Group, or Just a Single Individual?*, in «Journal of Behavioral Decision Making», 18, 3, 157 ff.

Kogut T., Ritov I. 2005b. *The Singularity Effect of Identified Victims in Separate and Joint Evaluations*, in «Organizational Behavior and Human Decision Processes», 97, 2, 106 ff.

Kogut T., Ritov I. 2007. *"One of Us": Outstanding Willingness to Help Save a Single Identified Compatriot*, in «Organizational Behavior and Human Decision Processes», 104, 2, 150 ff.

Körner A., Volk S. 2014. *Concrete and Abstract Ways to Deontology: Cognitive Capacity Moderates Construal Level Effects on Moral Judgments*, in «Journal of Experimental Social Psychology», 55, 139 ff.

Lee S., Feeley T.H. 2016. *The Identifiable Victim Effect: A Meta-analytic Review*, in «Social Influence», 11, 3, 199 ff.

Leibovitch A. 2017. *Punishing on a Curve*, in «Northwestern University Law Review», 111, 5, 1205 ff.

LENAERTS K. 2020. *New Horizons for the Rule of Law Within the EU*, in «German Law Journal», 21, 1, 29 ff. Available on: https://doi.org/10.1017/glj.2019.91.

LESNER T.H., RASMUSSEN O. D. 2014. *The Identifiable Victim Effect in Charitable Giving: Evidence from a Natural Field Experiment*, in «Applied Economics», 46, 36, 4409 ff.

LEWINSOHN-ZAMIR D., RITOV I., KOGUT T. 2016. *Law and Identifiability*, in «Indiana Law Journal», 92, 505 ff.

LIBERMAN N., TROPE Y., WAKSLAK C. 2007. *Construal Level Theory and Consumer Behavior*, in «Journal of consumer psychology», 17, 2, 113 ff.

LOEWENSTEIN G., SMALL D., STRNAD J. 2005. *Statistical, Identifiable and Iconic Victims and Perpetrators*, Stanford Law School, working paper. Available on: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=678281.

MAJUMDER R., TAI Y. L., ZIANO I., FELDMAN G. 2022. *Revisiting the Impact of Singularity on the Identified Victim Effect: An Unsuccessful Replication and Extension of Kogut and Ritov (2005a) Study 2*, manuscript. Available on: https://doi.org/10.17605/OSF.IO/9QCPJ.

MANDELBAUM E., RIPLEY D. 2012. *Explaining the Abstract/Concrete Paradoxes in Moral Psychology: The NBAR hypothesis*, in «Review of Philosophy and Psychology», 3, 3, 351 ff.

MOCHE H., GORDON-HECKER T., KOGUT T., VÄSTFJÄLL D. 2022. *Thinking, Good and Bad? Deliberative Thinking and the Singularity Effect in Charitable Giving*, in «Judgment and Decision Making», 17, 1, 14 ff.

NAHMIAS E., COATES D.J., KVARAN T. 2007. *Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions*, in «Midwest Studies in Philosophy», 31, 1, 214 ff.

NICHOLS S., KNOBE J. 2007. *Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions*, in «Nous», 41, 4, 663 ff.

PATIL S. V., VIEIDER F., TETLOCK P.E. 2014. *Process Versus Outcome Accountability*, in BOVENS M., GOODIN R.E., SCHILLEMANS T. (eds.), *The Oxford Handbook of Public Accountability*, Oxford University Press, 69 ff.

RACHLINSKI J.J. 2006. *Bottom-up versus Top-down Lawmaking*, in «University of Chicago Law Review», 73, 3, 933 ff.

RAHAL R.M., FIEDLER S. 2022. *Cognitive and Affective Processes of Prosociality*, in «Current Opinion in Psychology», 44, 309 ff.

RAMOS F. 2007. *The Establishment of Constitutional Courts: A Study of 128 Democratic Constitutions*, in «Review of Law & Economics», 2, 6. Available on: https://doi.org/10.2202/1555-5879.1043.

RAZ J. 1977. *Rule of Law and its Virtue*, in «Law Quarterly Review», 93, 195 ff.

SABATO H., KOGUT T. 2021. *Happy to Help—If It's Not Too Sad: The Effect of Mood on Helping Identifiable and Unidentifiable Victims*, in «PloS One», 16, 6, e0252278.

SCHAUER F. 2006. *Do Cases Make Bad Law*, in «University of Chicago Law Review», 73, 883 ff.

SCHELLING T.C. 1968. *The Life You Save May Be Your Own*, in CHASE S.B. (ed.), *Problems in Public Expenditure Analysis. Studies of Government Finance*, The Brookings Institutions, 127 ff.

SCHNEIDER H.P. 1998. *Gesetzgebung und Einzelfallgerechtigkeit: Zum Verhältnis von Legislative und Judikative im sozialen Rechtsstaat*, in «Zeitschrift für Rechtspolitik», 31, 8, 323 ff.

SELLERS M., TOMASZEWSKI T. 2010. *The Rule of Law in Comparative Perspective. Volume 3*, Springer Science & Business Media.

SHERWIN E. 2006. *Judges as Rulemakers*, in «University of Chicago Law Review», 73, 919 ff.

SINNOTT-ARMSTRONG W. 2008. *Abstract + Concrete = Paradox*, In KNOBE J., NICHOLS S. (eds.), *Experimental Philosophy*, Oxford University Press, 209 ff.

SINNOTT-ARMSTRONG W., YOUNG L., CUSHMAN F. 2010. *Moral intuitions*, in DORIS J. (ed.) *The Moral Psychology Handbook*, Oxford University Press, 246 ff.

SMALL D.A., LOEWENSTEIN G. 2003. *Helping a Victim or Helping the Victim: Altruism and Identifiability*, in «Journal of Risk and Uncertainty», 26, 1, 5 ff.

SMALL D.A., LOEWENSTEIN G. 2005. *The Devil You Know: The Effects of Identifiability on Punishment*, in «Journal of Behavioral Decision Making», 18, 5, 311 ff.

SMALL D.A., LOEWENSTEIN G., SLOVIC P. 2007. *Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims*, in «Organizational Behavior and Human Decision Processes», 102, 2, 143 ff.

SMITH R. W., FARO D., BURSON K.A. 2013. *More for the Many: The Influence of Entitativity on Charitable Giving*, in «Journal of Consumer Research», 39, 5, 961 ff.

SOCIA K. M. 2022. *Driving Public Support: Support for a Law is Higher When the Law is Named After a Victim*, in «Justice Quarterly», 39, 7, 1449 ff. Available on: https://doi.org/10.1080/07418825.2022.2064329

SPAMANN H., KLÖHN L. 2016. *Justice Is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges*, in «The Journal of Legal Studies», 45, 2, 255 ff.

STRUCHINER N., ALMEIDA G., HANNIKAINEN I. 2020. *Legal Decision-Making and the Abstract/Concrete Paradox*, in «Cognition», 205, 1 ff.

STRUCHINER N., HANNIKAINENN I. 2016. *A insustentável leveza do ser: Sobre arremesso de anões e o significado do conceito de dignidade da pessoa humana a partir de uma perspectiva experimental*, in «Civilistica.com», 5, 1, 1 ff.

TASHIMA A.W. 2008. *The War of Terror and the Rule of Law*, in «Asian American Law Journal», 15, 245.

TROPE Y., LIBERMAN A. 1996. *Social Hypothesis Testing: Cognitive and Motivational Mechanisms*, in HIGGINS E.T, KRUGLANSKI A.W. (eds.), *Social Psychology: Handbook of basic principles*, The Guilford Press, 239 ff.

TUSHNET M. 1999. *Taking the Constitution Away from the Courts*, Princeton University Press. Available on: https://www.jstor.org/stable/j.ctt7spsn.

TUSHNET M. 2020. *Taking Back the Constitution: Activist Judges and the Next Age of American Law*, Yale University Press.

WALDRON J. 2008. *The Concept and the Rule of Law*, in «Georgia Law Review», 43, 1, 1 ff.

WALDRON J. 2020. *The Rule of Law*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). Available on: https://plato.stanford.edu/archives/sum2020/entries/rule-of-law/.

WISTRICH A. J., RACHLINSKI J.J., GUTHRIE C. 2014. *Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings*, in «Texas Law Review», 93, 855 ff.

ŻURADZKI T. 2018. *The Normative Significance of Identifiability*, in «Ethics and Information Technology», 21, 295 ff.

ZURN C.F. 2007. *Deliberative Democracy and the Institutions of Judicial Review*, Cambridge University Press.

Tunnel Vision in Decisions on Guilt: Preventing Wrongful Convictions

ENIDE MAEGHERMAN

1. *Introduction*

Although legal decisions are to a large extent controlled by the law, there are several aspects that the law cannot control. For instance, the law can regulate what evidence is admissible, but it rarely says how the judge should interpret it, or how much weight the judge should attach to it (ANDERSON et al. 2005). Legal decision-making therefore becomes a largely psychological or cognitive process, and should therefore also be studied from other perspectives. In this chapter, the focus lies on the decision on guilt that is made within criminal law. As the cognitive process of the decision on guilt is relevant within most legal systems, the issues discussed here are also thought to be relevant to most systems. There may however be some differences between the systems. For instance, a judge in an inquisitorial system is expected to be more active than a judge in an adversarial system, which is likely to influence some of the potential solutions suggested in this chapter. Nevertheless, it is expected that the topic of the chapter, namely the influence of bias on the decision of guilt, presents a risk in most systems. Legal processes are thought to not be immune to this influence, although the extent of the impact may differ between legal systems. Therefore, research and cases discussed here may come from several countries or systems, with the requirement that these focus on biases and the decision on guilt, unless otherwise specified. Throughout the chapter, the terms judge and jury or juror will be used interchangeably for readability reasons. The term legal decision-maker or trier of fact may also be used—it is believed that the cognitive processes discussed here are largely applicable to all these roles, as the decision to be made is also mostly comparable. There are a few elements of the decision on guilt that are important to the cognitive processes discussed in this chapter. This includes that the facts of the offence have to in some way be reconstructed in order for the judge to determine what happened, or what can be proven. Another cognitive factor is the requirement that the judge be convinced of the guilt of the suspect, in some systems "beyond a reasonable doubt".

Through applying knowledge gained in other disciplines, the cognitive processes that the legal system relies on could perhaps be improved. The need for such an improvement has, in recent years, become clear through several miscarriages of justice that have come to light. Perhaps due to these cases, the majority of the research on biases and preventing biases has taken place in the context of criminal law, which is also why this chapter will rely on criminal law cases. Most wrongful convictions have been brought to light by the Innocence Project in the US, but other countries have also had their fair share of, at least, questionable convictions (see for instance, the Schiedammer Park Murder in the Netherlands, the Parachute Murder in Belgium, the Amanda Knox case in Italy, and the Arnold Holst case in Germany). In some of these countries, it is extremely difficult to prove that a wrongful conviction has occurred. For example, in the Netherlands, a novum is required to reopen a case. A novum constitutes a fact that was not known to the judges at the time of the initial decision, and that would have led to a different decision if it had been known (FRANKEN 2021; NAN 2020). These requirements set the

bar quite high, thereby making it difficult to reopen the case. The US also applies the principle of finality, which renders it exceptional to overturn a verdict. Nevertheless, the Innocence Project has achieved great progress in that respect by overturning wrongful convictions, and identifying the possibilities to do so. Keeping in mind the Blackstone ratio, namely that it is better to have ten guilty people go free than to have one innocent person in prison, it becomes clear that the occurrence of wrongful convictions undermines the functions of the legal system.

Although acknowledging the fact that wrongful convictions can happen within any legal system is essential, this is not sufficient to also prevent them from happening. In order to do so, there also needs to be an improved understanding of how these wrongful convictions can occur. Often, when wrongful convictions come to light, there is a tendency to focus on a problematic piece of evidence. For instance, false confessions, mistaken eyewitness identifications, or contaminated DNA traces at the crime scene are often cited as causes of wrongful convictions. However, in addition to the problematic evidence that supported the suspect's guilt, there was likely also overlooked evidence that could have supported the suspect's innocence. For instance, in the case of Ronald Cotton, the victim mistakenly identified Ronald Cotton as her rapist. The case is therefore often used to demonstrate the fallibility of eyewitness testimony. However, Cotton also had an alibi, which was not properly taken into account when deciding on his guilt (THOMPSON-CANNINO et al. 2009). From this and many other examples, it becomes clear that deciding on guilt is not a simple matter of weighing evidence for and against guilt. The evidence has to somehow be integrated into an overarching decision, and different attention and weight can be given to different pieces of evidence, regardless of the evidential value it might actually have. That is where the human factor becomes decisive, and therefore the potential faults of the cognitive process should be understood.

In this chapter, several cognitive theories on legal decision-making will be discussed. Subsequently, biases and heuristics will be explained as errors that can occur while making decisions on guilt. In doing so, specific attention will be paid to confirmation bias, as this has been argued to be one of the most important biases to understand in the context of legal decision-making (FINDLEY & SCOTT 2006). The chapter will then discuss how biases and their influence can be prevented in order to reach an optimal decision.

2. *Cognitive theories of legal decision-making*

In the vast majority of cases, it is impossible for a judge to know exactly what happened, as they did not witness the event themselves. Yet, based on the available evidence, they have to come to a decision about what happened. These decisions also have high stakes, as the outcome can have a large impact on several individuals as well as society. An inherent leap is required from judges or jurors in order for them to become sufficiently convinced about what happened. How that leap is made has been the subject of many theories, but due to the internal and potentially subjective nature of the decision process, a lot remains unknown about how the leap is made.

According to BEX and colleagues (2010), a distinction can be made between argument-based and story-based approaches within theories on legal decision-making. One of the first theories that emphasised the role of persuasion in legal decision-making, thereby moving away from the question on admissibility, was Wigmore (TWINING 1985). Wigmore's theory was based on the idea that reasoning about judicial proof should be in line with the way in which people reason in everyday life. According to Wigmore, facts could be seen as propositions. In turn, evidence was the connection between a proposition that needed to be proven, and the proposition that supported the first proposition. That proposition was in turn supported by another proposition. Eventually, this results in a chain of propositions, or information, supporting each other, which is also often seen in judicial trials. A fact is accepted based on the evidence, which is supported

by other pieces of evidence or information. An example of how this would work in practice, for instance, in an international criminal trial was previously described by McDermott (2015). According to McDermott, the inferential proposition that several people took part in a criminal organisation could be supported by a meeting having taken place, and this could in turn be supported by direct evidence, such as witness statements.

It is important to acknowledge that the propositions in these models can be challenged at any point. For instance, an alternative explanation for the facts could be offered, or one of the facts could be negated by additional information. Such challenges could result in the inference between propositions losing value (Twining 1985; Bex et al. 2010). Although Wigmore's theory seems quite complex to use in practice (McDermott 2015), several elements of the theory were used in the further development of cognitive theories on legal decision-making.

For instance, Bennett and Feldman (2005 [1981]) also emphasised that the framework within which legal decisions can be understood should be in line with informal and commonplace social judgements. In that way, it should be possible for this to be understood even by those who do not have extensive training in the legal field. One way to achieve such a common understanding is the use of stories or narratives as a form of communication, which was also embraced by Pennington and Hastie in their story model (1991).

According to the story model, evidence within a legal context can be evaluated through the creation of a story. A decision is then made by seeing which verdict category the story fits with best. For example, if someone has killed another person, this could be both murder or manslaughter, dependent on whether premeditation was proven. Murder and man-slaughter would then represent two different categories. A third category could be that the perpetrator acted out of self-defense, which would in several jurisdictions cause the behaviour to be justified (Keiler & Roef 2016). The decision-maker has to determine which verdict category best matches the story that is created based on the evidence.

When the story model was tested experimentally, it was found that when evidence was presented in a story format, it was judged more favourably than when the evidence was presented according to legal issues (Pennington & Hastie 1992). Evidence that was congruent with the presented story was also remembered better by jurors than evidence that was not in line with the story (Pennington & Hastie 1992). The story is thought to be constructed based on the evidence, as well as by integrating evidence that is presented later. This occurs based on causal relations that are implied and on the general knowledge that the juror has (Pennington & Hastie 1986). Thus, the story model explains one possible theory of how different evidence can be integrated by the trier of fact, and is in line with the suggestion that legal decisions should be done in a similar way to everyday decisions.

The idea that the evidence is integrated into a causal story was further developed by the theory of Anchored Narratives (Wagenaar et al. 1993). According to the theory of anchored narratives, the initial story to be considered comes from the charge against the suspect. The charge is then judged on two aspects in order to reach a decision on the guilt of the suspect. The first question is the plausibility of the charge. The second question is whether the charge can be anchored in common-sense beliefs using the evidence available to the decision-maker (Wagenaar et al. 1993). Evidence can prove something if it is grounded in a general rule that can be considered valid most of the time. For instance, a witness statement can prove something if the accepted rule is that witnesses do not lie and do not make mistakes (Wagenaar et al. 1993). In this way, every piece of evidence that is used to prove the charge would have to be anchored into a generally accepted rule. The evidence can be seen as substories of the charge, and can be further divided into multiple substories based on the different elements that may be necessary. For instance, the intention to kill a victim may be divided into the substory of the suspect having told a witness that he wanted to kill the victim, and the substory of the suspect bringing a gun to his meeting with the victim. It is up to the court to consider when evidence is

sufficiently anchored (WAGENAAR et al. 1993). The anchoring of such evidence may be considered unsafe if it is for instance based on an incorrect belief, or if a large number of exceptions are possible for that anchor. If a charge is unsafely anchored, it could result in a wrongful decision on the suspect's guilt, for instance a decision based on information being accepted too easily by the judge. A typical belief on which such a conviction can be anchored is that people do not falsely confess (WAGENAAR et al. 1993).

Similarly to the story model, there is no need for all evidence that may exist in a case to be integrated into either the story or the anchors. For instance, evidence that contradicts the story as it has been constructed, may be disregarded by the decision-maker. This way of reasoning may be in line with the question of "beyond a reasonable doubt" or the requirement that the judge has to be convinced. The question then becomes whether the evidence that is not in line with the existing story can create sufficient doubt to undermine the conviction of the decision-maker. In wrongful convictions, it can be argued that such doubt should have existed in the mind of the decision-maker. Although the theories discussed so far provide insight into how a belief of guilt can be formed, they do not provide a clear answer as to why there was insufficient doubt created to undermine the conviction, despite the innocence of the defendant. The answer to that question can perhaps best be found in the fallibility of human reasoning.

3. *Biases*

Despite the legal regulations and rules that every country has, they still need to be applied and enforced by humans, thereby also making the process of legal decision-making prone to human error. As the theories above have demonstrated, the consideration, interpretation, and integration of evidence are to a large extent left up to the discretion of the trier of fact, and legal regulations can only control these to a limited extent. One of the most common human errors can be found in biases. Biases can be understood in the context of dual-system processing. According to dual-system processing, there are two systems that are used for cognition, including decision-making (KAHNEMAN 2011). Although differing opinions exist on to what extent the two systems are independent of each other, and whether and how tasks really are divided between the two systems (EVANS & STANOVICH 2013; STANOVICH & TOPLAK 2012), there is a general consensus regarding the features of these systems.

System 1 is thought to be an unconscious system that makes decisions quickly and automatically. The idea is that by allocating certain but limited cognitive processes to this system, more cognitive energy can be used for more difficult processes. For instance, if every decision—such as how to open the door—had to be consciously thought over, it would be extremely tiring. Decisions that are common and often made are therefore thought to be delegated to system 1. The more difficult processes are thought to be conducted by system 2. Contrary to system 1, system 2 requires more cognitive energy and time, is more controlled, and is less based on intuition. System 2, according to KAHNEMAN (2011), uses logical reasoning. It has also been argued that the default reasoning by system 1 can be overruled by the reflective thinking of system 2 (EVANS & STANOVICH 2013).

This latter argument is important as tasks that may initially be completed by system 2 can later be completed by system 1. That is likely to be true for most tasks which become automatic. For instance, when first learning to ride a bike or drive a car, every action requires conscious thought. Changing gears is considered to be cognitively taxing. However, as driving experience is gained, the handling and decision processes become more automatic. The cognitive processes that were completed by system 2 while learning have then been relegated to system 1. The same can be true for other tasks that might initially be quite demanding, but that become more automatic as more experience is gained. System 1 learns an automatic response through the decisions made by system 2, so that the task is eventually conducted by system 1's automatic response.

In order to be able to function efficiently, system 1 makes use of heuristics. These can be described as mental rules based on experience that allow for a quicker decision. A heuristic can therefore be seen as a rule of thumb. When a heuristic is applied, it can often quickly and efficiently lead to a correct outcome. However, a rule that has worked well in a previous situation does not necessarily apply to a new situation. As such, applying heuristics can also lead to mistakes, which can lead to biases. Biases can be described as discrepancies between the rational, normative behaviour, and the behaviour determined by the heuristic (GONZALEZ 2017). They can be intrinsic or learned and can have several consequences. More than 180 cognitive biases exist that interfere with how we process information, think, remember, and experience reality. As mentioned, the main focus in this chapter will be on what has been argued to be one of the most relevant biases in legal decision making.

3.1 *Confirmation bias*

Confirmation bias, which is one of the essential processes of what is known as tunnel vision, is a cognitive bias that has been described as one of the most influential cognitive errors (NICKERSON 1998). Furthermore, it has been argued to be of particular importance in the context of legal decision-making (FINDLEY & SCOTT 2006). Confirmation bias refers to the tendency to favour information that supports an existing belief or theory. That can happen both by paying disproportionate attention to information that supports an existing belief, as well as by interpreting information in such a way that it is in line with a prior theory (KASSIN et al. 2013; MENDEL et al. 2011; NICKERSON 1998). Tunnel vision has previously been described as a contributing factor to miscarriages of justice (RASSIN 2010), and could explain why exonerating evidence may be overlooked by the decision-maker.

Tunnel vision consists of several interrelated cognitive processes. At the start of these processes, there has to be a certain belief. For instance, the belief that a suspect is guilty could arise, based on an incriminatory case file, or even a particular fact about the defendant such as an earlier conviction. In essence, this belief could come from anything, and the judge is not necessarily aware of its existence or its source. Note though, that this belief, or the subsequent phenomenon of tunnel vision, is not necessarily a process that could only affect judges. An investigative team or prosecutor can also be hampered by it, which may result in a biased case file being presented to the trier of fact. The subsequent cognitive process that plays a role occurs when information that is contradictory to the belief arises. Holding contradicting information leads to an uncomfortable feeling, known as cognitive dissonance, a term introduced by FESTINGER (1957). He argued that when one experiences cognitive dissonance, one will try to reduce that uneasy feeling by attempting to achieve consonance. Cognitive dissonance can therefore be considered a precursor to behaviour aimed at achieving consonance, or aligning the beliefs held. According to CANCINO MONTECINOS (2020), there are several potential processes that can be used to reduce dissonance. These include trivialization, whereby the importance of either the original attitude or the dissonant information is reduced. Another strategy would be bolstering the initial opinion. The existence of cognitive dissonance has received ample support in the literature, despite the difficulty in testing it empirically (HARMON-JONES & HARMON-JONES 2007). Due to the need to reduce dissonance, it can have a considerable influence on further cognitive processes.

The process of achieving consonance in the context of legal decision-making has previously been studied by SIMON (2004), who argued that cognitive coherence is imposed on tasks that are complex, such as legal decision-making, in order to transform this process into a simpler one. In that way, an easier decision can be made with more confidence. SIMON (2004) also found the process of achieving consonance to be bidirectional—not only does the evaluation of the evidence influence the decision, but an already existing preference for a specific decision outcome also influences the evaluation of the evidence. One relatable way in which this could be understood is

perhaps the feeling one experiences when making a pros and cons list, only to find out they already had a preferred option. This need for consonance or consistency can also be related to the theories on legal decision-making described above. Namely, the extent to which the evidence is consistent with the story constructed by the decision-maker likely influences the manner in which it is integrated. An alternative possibility would be that the contradicting evidence could be trivialised in order to achieve consonance (Cancino Montecinos 2020).

Another way in which consonance can be achieved is through belief perseverance. Belief perseverance can be considered the next phase of tunnel vision. It refers to the tendency to adhere to one's belief, even when presented with contradictory information. Belief perseverance has been observed in practice—some prosecutors in wrongful conviction cases maintain a belief in the defendant's guilt, despite the evidence that led to their exoneration (Burke 2007).

In order to maintain a belief, confirmation bias can come into play. As previously described, confirmation bias is the tendency to disproportionally focus on, or favour, information that can confirm an existing belief compared to information that contradicts that belief (Kassin et al. 2013, Nickerson 1998). In that way, confirmation bias can maintain the previously held belief. There is a distinct difficulty in trying to study confirmation bias—as it is an internal and often unconscious process, finding the effect is dependent on the behaviour of the participants. Such behaviour, for example only following news outlets that align with political preference, can be considered the behavioural manifestation of tunnel vision. Confirmation bias is therefore a key element of tunnel vision, and the aspect or process that can most likely be observed. Subsequently, research most often focuses on measuring confirmation bias, including when studying tunnel vision.

There are several tasks that have been developed in an attempt to experimentally test confirmation bias. One task that is often used is Wason's card selection task (Wason & Johnson-Laird 1972). In this task, participants are given a rule to test and several options to investigate whether or not the rule is true. For instance, the rule could be "if one side has a vowel, the other side of the card must have an even number". Participants are then given 4 options that they can test in order to determine whether the rule is true. For the above rule, the cards could for instance show E, K, 4 and 7. Participants are asked to choose as few cards as possible to determine whether the rule is true. The E-card can here be used to verify the rule—if there is an even number on the back, that supports the rule. The K- and 4-card are uninformative, as the rule does not mention what must be on the back of a card with a consonant, nor does it say that an even number card must have a vowel on the other side. The 7-card can be used to falsify the rule, as a vowel on the other side would disprove the rule. The correct answer would therefore be to turn over the E-card and the 7-card (Cosmides 1989; Wason & Johnshon-Laird 1972). According to Rachlinski, typically, fewer than 10% of respondents give the correct answer to the selection task. The task has also been used in surveys answered by judges (Rachlinski 2012). Among Dutch judges, one study found that 32% of respondents gave the correct answer (Rachlinski 2012), whereas another found that 22% of judges gave the correct answer (Maegherman 2021). Therefore, it can be argued that, although judges may show less bias than the average population, only a minority showed no tendency to confirm their hypothesis when tested experimentally.

Confirmation bias has been shown to affect various key players in the judicial field such as police investigators and forensic scientists. For instance, Kassin and colleagues (2003) found that interviewers who had a stronger belief that the person they were interviewing was guilty tended to ask more questions confirming the guilt of the interviewee than those who had an expectation of the suspect being innocent. Moreover, when neutral observers then watched the interviews, they were also more likely to judge the former interviewees to be guilty than the latter. Dror and colleagues (2006) have previously demonstrated that a prior belief based on external information could influence the judgement of forensic fingerprint examiners. Liden and collegues (2018) also investigated confirmation bias in prosecutorial decision-making. In their study, prosecutors

showed a tendency towards confirming guilt once they had made the decision to press charges against the defendant. However, before that decision, they did not show this tendency, suggesting the commitment to a belief in guilt increased the confirmation bias, which can be understood in terms of the processes explained above. Although confirmation bias has been shown to have an effect on both the investigation and the prosecution, it could perhaps still be corrected at the trial stage. However, considering the final decision is also reliant on human cognition, it is also likely to be influenced by the same cognitive errors that other phases are affected by.

Based on findings from experimental research, it could indeed be argued that the decision on guilt or innocence is equally likely to be tainted by bias as the earlier decision or investigative processes. One study done by SCHÜNEMANN (1983, as cited in SCHÜNEMANN & BANDILLA 1989) seems to demonstrate the existence of belief perseverance in German trial judges. They compared a group of judges who had initially received mainly incriminating information prior to the trial to a group of judges who had received information that was less incriminating prior to the trial. Both groups subsequently read identical trial proceedings, and were unaware of the experiment being conducted. Of the group who received more incriminating information, 82% would convict the suspect, compared to only 53% in the group who received the less incriminating information. It therefore seems that the initial belief the judges had formed was maintained throughout reading the case file, and ultimately affected their decision on guilt. In other studies, using different populations, comparable results have been found. For instance, RASSIN (2010) used a sample of law students and presented them with a case file, after which participants had to make a judgment about the suspect's guilt. They were then asked to select further investigative measures, some of which were incriminating for the suspect, whereas others were exonerating. Based on the investigative measures that were chosen by participants, the further investigative measures that were chosen seem to have been influenced by their initial opinion on guilt. The selection of incriminating investigative measures was also associated with higher conviction rates. Thus, an initial belief tainted the further investigation and the subsequent decision. In practice, this may occur not only during the police investigation, but also in cases where the judge may have the opportunity to request further investigation. It should be mentioned that in Rassin's study, there was no general preference towards incriminating information—the preference that was observed was dependent on the initial belief of guilt. A similar result was observed by MARKSTEINER and colleagues (2011). They found that only police trainees with a prior belief in the suspect's innocence rated incriminating and exonerating evidence as equally reliable—those with a prior hypothesis of guilt found the incriminating evidence to be more reliable than the exonerating evidence.

4. *Combatting tunnel vision*

In this section, research on preventing biases, with a focus on tunnel vision, will be discussed. As explained above, tunnel vision can be seen as a combination of several cognitive process, including confirmation bias. In order to understand the difficulty of preventing tunnel vision, an explanation of the difficulty of fighting bias in general will first be discussed. Some of the research on bias, and bias reduction, are applicable to tunnel vision whereas other studies and methods have focused specifically on reducing tunnel vision. By looking at methods that have been researched to prevent bias in general, as well as those focused on tunnel vision, it becomes possible to identify remedies that continue to be promising.

4.1 *Difficulties in bias prevention*

Although biases are a widely accepted phenomenon, and have been found to have an influence in several areas of life, little is known about how biases, and their influences, can be prevented.

In general, it seems that training aimed at reducing biases has limited effect, which could be due, for instance, to the difficulty of voluntary suppression of an unconscious process, or to the fact that individuals may be less susceptible to learning during coerced training (WILLIAMSON & FOLEY 2018). Researchers have also found that the effects of training diminish over time (CLARKE et al. 2011). Nevertheless, as some trainings do seem to have an effect, although the longevity may be limited, it is still worth considering the aspects of these training.

One reason why training may have a limited effect, and why bias in general might be difficult to prevent, is thought to be the bias blind spot. The bias blind spot refers to the tendency to ignore one's own bias, despite acknowledging that other people or colleagues are impacted by biases (PRONIN et al. 2002). A study by KUKUCKA and colleagues (2017) showed a bias blind spot in forensic examiners. They surveyed forensic experts on the existence of bias in forensic science in general, and specifically in their field. Although the majority acknowledged the problem of bias in forensic science as a whole, fewer acknowledged it to be a problem in their discipline, and a distinct minority recognised that they themselves might also be impacted by bias. The bias blind spot in itself also seems hard to fight. WEST and colleagues (2012) found that cognitive ability was associated with a higher bias blind spot, suggesting it is something that affects everyone. They also found that not suffering from a bias blind spot does not necessarily lead to being less influenced by biases.

SCHMITTAT and ENGLICH (2016) investigated whether judicial experts are protected against biased reasoning. They included both several legal fields and several roles, such as judges, prosecutors, and defense lawyers. Their study used German practitioners. Based on their results, the practitioners showed a preference for information that supported their preliminary beliefs. They evaluated that information more positively than information that conflicted with their prior belief. However, there was a positive effect within specific domains. Namely, experts within their own domain showed less bias than general experts, who did not perform better than laypeople. Confirmatory reasoning could be reduced in general experts by inducing responsibility by emphasising the consequence of the decision for the defendants. However, the fact that expertise in itself does not protect against confirmatory reasoning also implies that judges may be affected by bias.

One could argue that, generally speaking, biases can be reduced by using system 2 instead of system 1, including when making important decisions as to whether somebody is guilty or not. As system 2 uses critical rather than automatic thinking, the use of system 2 could indeed lead to fewer cognitive errors. However, the activation of system 2 is not necessarily easily done, specifically in situations where there may be time pressure or a need to achieve cognitive closure (i.e., the urge to stop ambiguity and come to a clear conclusion). Moreover, when a task has been completed several times, it may also rely on system 1 (KAHNEMANN 2011). Therefore, structural ways in which critical thinking can be encouraged warrant attention. These can be focused on the individual, such as for example specific training, or they can be aimed at more structural changes, such as, for instance, reducing time pressure or increasing the requirements to explain a decision-making process, thereby activating system 2.

4.2 *Falsification and alternative scenarios*

When considering confirmation bias, one method that has been studied as an attempt to reduce bias is the use of falsification. Falsification is a concept that was used by Popper to explain the scientific method or a way of testing whether a hypothesis is true. Popper argued that a theory can only be considered true until contradictory evidence is found. Therefore, if one only looks for evidence that supports the theory, it may incorrectly be assumed to be true. According to Popper, a better way of determining whether or not a theory is true would be to look for evidence that can disconfirm the theory. If such evidence cannot be found after several serious

attempts, then that would again support that the theory is true. The process of looking for evidence that contradicts a scenario was termed falsification, and is thought to be equally as important as verification when trying to determine whether a theory or scenario is true (POPPER 2005 [1959]). Failed attempts at falsification increased the likelihood of the scenario being true (CROMBAG et al. 2006).

The classic example that is often used to explain the concept of falsification involves swans. If the hypothesis or theory is that all swans are white, it is likely that a lot of support can be found for that theory. However, it would only take one sighting of a swan of a different colour to prove that the theory is false. Subsequently, when trying to determine whether the theory is true, one should not only look for white swans, but should actively try to find swans of a different colour. If swans of a different colour are found, the theory can be considered false. If however, several serious attempts to find a swan of a different colour only result in finding more white swans, that gives support to the theory. The theory that all swans are white can still not be considered ultimately proven, as a sighting of a swan of a different colour could still occur and would disprove the theory.

Although the above example may be easily understood, it is also quite far removed from the practice of legal decision-making. However, similar reasoning can be applied in that context. There is a theory or a hypothesis, namely that the defendant is guilty of the crime they are charged with. Evidence confirming that theory can most likely be easily found in the case file. There is an inherent requirement for incriminating evidence to be present, as the case would otherwise not be brought before a judge. Therefore, verification of the theory that the defendant is guilty will in most cases be relatively easy. However, in order to determine whether the theory is true, falsification should also be applied. In the context of a decision on guilt, that would mean looking for exonerating evidence. For instance, determining whether the suspect might have an alibi, or whether there may be an alternative explanation for the DNA of the suspect found at the crime scene.

Exonerating evidence may not be explicitly present in the case file, and the judge or responsible party may be required to order additional investigative measures in order to obtain the exonerating evidence. Based on a survey and interviews conducted with Dutch judges, it became clear that exonerating evidence may not always be present in the case file. Moreover, in the documents that are prepared for the judge, the exonerating information may not be included. One example was that if a suspect was identified by a witness whereas another witness did not recognize the suspect, then it is possible that only the identification would be included in the preparation material (MAEGHERMAN 2021).

One method that is closely associated with the idea of falsification is the use of alternative scenarios. The scenario approach was explained in detail by VAN KOPPEN and MACKOR (2020). It is also related to the theory of anchored narratives discussed earlier, although more focus is placed on the consideration of multiple scenarios. According to VAN KOPPEN and MACKOR (2020), a scenario should consist of a chronological and causal account of an event, including a central action that can be understood in the context of the surrounding scene. Similarly to the story model by PENNINGTON and HASTIE (1992), in this approach, a scenario also relies on background knowledge, or scripts. Once several scenarios have been identified, the next component is what is known as the inference to the best explanation. The inference to the best explanation was explained by HARMAN (1965) as concluding that the hypothesis is true, based on the fact that the hypothesis is able to explain the evidence. Alternative hypotheses which could also explain the evidence should be rejected before making the inference to the best explanation. If one particular hypothesis can give a better explanation for the evidence than any other alternative hypothesis, it can be inferred to be true (HARMAN 1965). It should be remembered that the conclusion of the inference to the best explanation is not guaranteed to be

the truth—even though it might offer the best explanation possible, it is not necessarily true. The inference to the best explanation is nevertheless a key element of the scenario approach.

VAN KOPPEN and MACKOR (2020) explain several criteria that can be used to select and assess scenarios. These include the internal coherence of scenarios, the coherence of the scenarios with general background knowledge, and coherence of the scenarios with elements of the case that have been accepted to be true. Several of these requirements are also akin to the theory of anchored narratives, where the sub-scenarios of the main scenario should ultimately be anchored in commonly accepted knowledge. Furthermore, the scenario approach includes three ways in which scenarios relate to the evidence. The first is creation. A scenario is created during a criminal investigation on the basis of the evidence. The second way is accommodation. A scenario might have to be adapted in order to accommodate for contradicting evidence that is encountered. Alternatively, the scenario might have to be rejected when such evidence is encountered. Lastly, the third way in which a scenario can be used is to predict what evidence one would expect to find. For instance, if there is a scenario in which the suspect is thought to have touched the victim in several places, it could be expected that there would be traces of DNA. In this way, if evidence that is expected is not found, that would undermine how well the scenario can explain the evidence, and thus, the scenario might have to be rejected.

Falsification is also a key element of the scenario approach. As previously mentioned, multiple serious failed attempts at falsification increase the likelihood of the scenario being true (CROMBAG et al. 2006). According to VAN KOPPEN and MACKOR (2020), finding a good alternative scenario can also be considered part of the falsification process. The main scenario is presented by the prosecution and argues, based on a coherent selection of incriminating evidence, that the suspect is guilty. The defense can construct an alternative scenario based on another selection or interpretation of the evidence that is present in the case file. The alternative scenario could try to show that the suspect is not guilty, or that the scenario presented by the prosecution is not likely. Those scenarios should then be compared to one another in order to determine which scenario best explains the evidence.

When Dutch judges were asked about their use of alternative scenarios and falsification, it became clear that they understand the importance of considering alternative scenarios in order to avoid confirmation bias. However, their answers also suggested that their main focus was on trying to verify the main scenario or falsify the alternative scenario, which would not protect against confirmation bias (MAEGHERMAN 2021). On the contrary, what the judges described can be interpreted as looking for confirmation of the main scenario. In order for alternative scenarios to be used effectively, one should attempt to verify and falsify both the main and the alternative scenarios. In that way, the scenario that best explains the evidence can be identified.

In previous experimental research, it has also been found that the mere existence of alternative scenarios is unlikely to be enough to protect against bias. For instance, O'BRIEN (2009) presented participants with a mock case file. Some participants were asked to name a prime suspect after reading part of the case file, whereas others were not. Those who had expressed their belief in the prime suspect's guilt subsequently seemed to be more biased in their reasoning than those who had not. O'Brien then tested several ways in which this bias could be countered. One way this was tested was to have the participants think of explicit reasons why the scenario of guilt might be false. Another way was to have participants think of alternative scenarios, e.g. of another perpetrator having committed the crime. Whereas the first method did counter the effect of confirmation bias, the latter did not, which suggests the active assessment of scenarios might be necessary. RASSIN (2018) found similar results. Rassin presented participants with a case file including a scenario of guilt only, or also including an alternative scenario. Some of the participants were asked to use a pen-and-paper tool, whereas others were just presented with the case file. Those who used the pen-and-paper tool showed less confirmation bias than those who were simply presented with the alternative scenario, which again supports that the mere presence

of an alternative scenario would not necessarily protect against confirmation bias. Another study was conducted by TENNEY and colleagues (2009). They compared the effect of defense lawyers either refuting the incriminating evidence, or also arguing for an alternative scenario in which someone else committed the crime. Based on their finding, participants returned fewer guilty verdicts than when the defense lawyer only argued against the incriminating evidence. Based on these findings, it is not only important for the legal decision-maker to consider alternative scenarios, but also to carefully consider both the support for and evidence against each of the scenarios, in order to reach a valid conclusion about which scenario is most likely. One way in which this could be encouraged is perhaps through the use of training, or through specific instructions on explaining how the decision has been made.

4.3 *The analysis of competing hypotheses*

One type of training that was developed in intelligence analysis and which has previously found some success at reducing confirmation bias, is the analysis of competing hypotheses, hereinafter ACH (HEUER 1999). It is a structured analytic technique (CHANG et al. 2018). ACH involves carefully weighing alternative explanations again each other, which should prevent the decision-maker from settling on the first option that seems satisfactory. The analysis consists of eight steps, which will not all be described here; the most relevant aspects of the analysis will be explained. The first steps require the construction of potential hypotheses, and the listing of the available evidence. A matrix is then created where it can be indicated whether each piece of evidence is consistent, inconsistent or irrelevant to the hypothesis. The matrix helps the decision-maker with determining the diagnosticity of the evidence. For instance, if a piece of evidence is consistent with three out of four hypotheses in the matrix, its diagnostic value is low. This would be clear when using ACH, but might not be clear when the evidence is considered in light of only one hypothesis. Evidence that has no diagnostic value should be removed from the matrix. When looking at which hypothesis is most likely, ACH does not focus on the hypothesis with the most consistent evidence, but rather at the hypothesis with the least inconsistent evidence (HEUER 1999). It thereby also incorporates the principle of falsification. There have also been several criticisms of the ACH technique which have emerged in recent years. DHAMI and colleagues (2019) for instance found issues with the implementation: if the technique is not implemented properly, its effect on the decision-making process would be limited. Findings by MAEGHERMAN and colleagues (2021) also suggest that participants struggled to apply the methodology, as they did not seem to follow the steps of the technique, despite being explicitly instructed to do so. DHAMI and colleagues (2019) also criticized the technique for being too vague. Although the technique in its current form may not be suitable for application in legal decision-making, it may be further developed or adapted for use in the legal system, as the essential elements remain promising (MAEGHERMAN et al. 2021).

4.4 *Accountability*

The stimulation of critical thinking has previously been investigated through the concept of accountability. Accountability refers to requiring decision-makers to account for their decisions, or that one justifies one's views (LERNER & TETLOCK 2003). There are several types of accountability, each of which may have a different effect on the decision-making process and its consequences. For instance, a different effect has been observed depending on whether participants knew they had to account for a decision before making it, or whether they were asked to do so afterwards. Whereas prior accountability was observed to increase exploratory reasoning and to improve judgement, post-decisional accountability increased confirmatory reasoning and led to self-justifying behaviour (LERNER & TETLOCK 1999). Another distinction

that can be made is whether the decision-maker is asked to account for the decision itself, or whether they have to explain the decision-making process. This distinction is captured by contrasting outcome-accountability with process-accountability. Process-accountability is generally thought to lead to better decisions (LERNER & TETLOCK 2003), although some researchers have argued that this beneficial effect may be limited to elemental tasks that involve linear relations between the cues and the outcome (DE LANGHE et al. 2011).

A third factor that moderates the effect of accountability is the audience to whom the decision-maker has to account for the decision. For instance, PENNINGTON and SCHLENKER (1999) found that students who had to judge a cheating case gave harsher punishments when they expected to have to explain their decision to the professor who reported the cheating compared to if they had to explain their decision to the student who was accused of cheating. However, according to HALL and colleagues (2015), research is lacking on how having to account for the decision to multiple or varied audiences could influence the decision.

When placing these research findings in the context of legal decision-making, it seems certain elements are implemented well. For instance, in most jurisdictions, judges or jurors will be aware that they will need to explain their decision before making the decision. In the Netherlands this is required according to Art. 359 DCCP, but is fairly limited compared to the German requirement to motivate the decision (SIMMELINK 2001; MEVIS 2019). In addition, the explanation of the decision may serve different purposes to different audiences. For instance, for the victims or family members of the defendant, the explained decision can help them to understand why it was made. The decision may also be used by the court of appeal, or the Supreme Court, if the decision is appealed. Therefore, more research on the effect of varied audiences would be very valuable to understand the impact in the context of legal decision-making. Similarly, the focus on either the decision itself or on the decision-making process could also be improved within the context of legal decision-making, although this is likely to vary between different jurisdictions. In the Netherlands, for instance, the Supreme Court has previously ruled that the reasoned decision given by judges does not have to be a representation of what was considered, but should contain evidence that the decision could reasonably be based on. The reasoned decision is therefore not necessarily a valid representation of the decision-making process, but is instead focused on the decision itself (REIJNTJES & REIJNTJES-WENDENBURG, 2018). According to the existing research, this could increase the tendency for confirmatory reasoning and self-justification. On the other hand, the German code of criminal procedure seems to encourage a more critical perspective, as the judge has to account for their selection and evaluation of the evidence. Furthermore, they also have to pay attention to the facts that indicate an alternative version of events that was not accepted (MEVIS 2019; DREISSEN 2007). In the case of contradicting witness statements, the judge has to consider how both statements came about, and to consider the discrepancies between them. Due to the higher number of requirements attached to the written decision by the German judge, it can be expected that this would lead to more exploratory rather than confirmatory reasoning.

In an experimental study conducted by MAEGHERMAN and collegues (2021), participants were presented with a mock case file and asked to decide on the guilt of the suspect. They were asked to explain their decision according to instructions based on either the Dutch Code of Criminal Procedure, the German Code of Criminal Procedure, the principle of falsification, or minimal instructions in the control condition. They were also given these instructions prior to reading the case file, so as to mimic the prior accountability judges also experience. Furthermore, they were told that their decision would be reviewed by a panel of professional judges. It was found that those in the German condition used significantly more exonerating evidence in their written decision than those in the Dutch condition. Nevertheless, there was no difference in conviction rates between the two conditions, even though the amount of exonerating evidence in the decision was found to be a significant predictor of the final decision on guilt. There therefore

seems to be a discrepancy between the evidence that was included in the written decision and the actual consideration. Although the exact impact of the different accountability requirements were therefore hard to determine, there nevertheless seemed to be a positive effect of requiring a more critical explanation of the decision.

Changing or increasing the accountability requirements may also help to prevent the task of explaining a decision from becoming routine. When a reasoning task has been done many times, it can become routine. When this is combined with pressure, for instance due to time constraints, the reasoning process can become increasingly reliant on intuitive cognitive processes, which can in turn reduce the accuracy of the decision-making. Experience can contribute to faulty thinking in experts when feedback is not provided (KAHNEMANN 2011; TAY et al. 2016). In a review of physician's experience and the quality of the care they provide, CHOUDRY and colleagues (2005) found that those who have been in practice longer are actually likely to give lower-quality care. Interventions may help to maintain, or even improve the quality of care. The same is likely to be true for judges. Although the decisions they make cannot be considered simple, judges with several years of experience may have come across similar cases. Furthermore, judges are rarely provided with feedback on their work, except in the case of wrong decisions coming to light, or perhaps colleagues coming to a different decision. It would therefore also be beneficial for the decision of judges to be reviewed structurally and regularly, which could also aid the effect of the accountability requirements by adding a critical audience for the reasoned decision.

4.5 *Remedies requiring further investigation*

Of course, the remedies that have been discussed in this chapter are by no means the only measures that exist to protect against bias. In this final section, a few more will be discussed, which have, to my knowledge, not yet been sufficiently tested in the context of legal decision-making, or which may not be suitable for application in that context in their current form.

One such remedy that has been proposed to reduce the influence of bias in legal decision-making is to make use of Bayesian reasoning. Bayes theorem makes use of prior possibilities, which are updated with the evidence, in order to reach a posterior probability. When a decision-maker is given further evidence for a hypothesis, the probability of the hypothesis must be reconsidered given the evidence (DAHLMAN 2020). Although in theory, Bayesian reasoning sounds like a structured manner of reaching a decision that might be resistant to human cognitive biases, its application is less straightforward. Firstly, the method is quite difficult to understand and apply properly. It is also a method of reasoning that is likely to be very unfamiliar to those who have spent their life training in the law and are likely to be unfamiliar with mathematics (DAHLMAN 2020). Furthermore, DAHLMAN (2020) pointed out that the analysis is dependent on the subjective assumptions of the person charged with finding the information or making the decision. The need to estimate certain probabilities would result in there still being a significant amount of room for human error. The method also relies on likelihood estimations, which are unlikely to be available for all types of evidence. For example, using an identification made by a witness to determine a likelihood would be very difficult, as the validity of the identification could be affected by many factors, such as the visibility of the perpetrator by the eyewitness, or the identification procedure. Determining a mathematical value for the likelihood of all these factors is very difficult to do, and so a lot of estimation is still involved. Furthermore, there is a possibility of ignoring that pieces of evidence may not be independent of each other, among other things that the decision-maker may fail to take into account. For a more in-depth explanation of the Bayesian method of reasoning, and example of its application to a case, please see DAHLMAN (2020). For the scope of this chapter, it suffices to conclude that Bayesian reasoning is unlikely to offer an easily applicable solution to biases in

legal decision-making. According to ROBERTS (2020), it is far removed from the type of reasoning judges actually engage in.

Although a lot of the solutions that have traditionally been offered to counteract biases focus on the individual, an argument can also be made for changes to the system that could limit the opportunity for biases to arise. For instance, it is commonly accepted that being under time pressure can reduce the quality of reasoning. Therefore, it is important to make sure that judges and decision-makers have sufficient time available to study the case file and to carefully consider all the information that is available. That does not seem to be the case everywhere. For instance, in the Netherlands, there was a public letter sent by the judiciary that explained that the workload was too high, and that this would eventually lead to a reduction of the quality of the legal system (TEGENLICHT 2018).

Aside from the different legal systems, a distinction can also be made between the constellations of trier of facts. Whereas some courts, usually for lower-level offenses, require only one judge, other courts can consist of several judges. For instance, in the Netherlands, cases before the police judge will require a single judge, whereas more complicated cases will be assigned to three judges (VERBAAN 2016). In Belgium, a case can also be tried by one or multiple judges, or can even be tried by a jury (Assisen-court; TRAEST 2018). The difficulty that arises with having several decision-makers is that conformity can arise. Conformity can be defined as an intrinsic tendency to follow or agree with others, particularly in case of dominant opinions or socially desirable outcomes (PEOPLES et al. 2012). The tendency to conform can also be related to the concept of falsification—an alternative opinion may not be expressed due to not wanting to stand out from the group. For that reason, it is particularly problematic if a unanimous verdict has to be reached. WATERS and HANS (2009) conducted a study on juror decision-making in the US. They sent a questionnaire to 3500 jurors who had previously decided on felony cases. Out of those, one third had disagreed with the outcome of the jury deliberations, and would have decided differently if they had made the decision on their own.

In case of decisions made by multiple individuals, another potential remedy might be available. Closely linked to falsification, the idea of a devil's advocate has been used in several fields, such as for instance management and health practice. The devil's advocate is supposed to take a stance contrary to that of the group. That should in turn cause the other members of the group to consider the issue at hand from different perspectives, and to avoid making decisions without critically thinking about the possibilities (MACDOUGAL & BAUM 1997; BROHINSKY 2022). By expressing alternative opinions, the use of a devil's advocate could also be helpful against confirmation bias. The term devil's advocate is one of many that has been used for similar processes. Another way in which alternative views could be facilitated is through the use of dissenting opinions. For instance, the ECtHR publishes the dissenting opinions of the judges who did not agree with the final verdict. By giving more exposure to such differing opinions, alternative scenarios could also receive more attention and be considered more actively.

5. *Conclusion*

The current chapter has explained some of the existing theories that have been developed to explain the elusive process of bridging the gap between evidence and the decision on guilt. Furthermore, it has looked at the role of (confirmation) bias, and the problems it can cause when trying to make a correct decision on guilt. Although the problem of biases has been shown in several experimental studies, as well as case studies, a reliable way of counteracting biases is still lacking. This chapter has provided a short overview of several techniques that have been researched to try and avoid biases, and the mixed results these methods have found. Through an

overview of part of the literature, it has become clear that there is a lot left to learn about how to fight biases, and how to implement such protection within the legal system.

It is also important to acknowledge that the role of the judge may be different in different legal systems. For example, in an inquisitorial system, the judge has a much more active role than in an adversarial system. In the adversarial system, the judge does not decide in the same way as the judge in the inquisitorial system does (STRIER 1992; SPENCER 2016). It can be argued that several of the ways to fight biases which have been described here, could become part of the task of the active judge. However, at the moment, there seems to be no consensus on how to interpret this role, even within a single country (MAEGHERMAN 2021). The different legal systems may present varying difficulties when trying to prevent the influence of biases. Although some of the methods discussed here would most likely be applicable to a range of systems, and a range of decision-makers, more research should be done to find the optimal ways to prevent biases in each system. In order to do so, more cooperation between researchers and practitioners is needed. Research should be guided by the needs of practitioners, and practitioners should participate in research as much as possible. In that way, the remedies to protect against biases and reduce the risks of wrongful convictions can be further developed and optimized.

References

ANDERSON T., SCHUM D., TWINING W. 2005. *Analysis of Evidence*, Cambridge University.

BENNETT W.L., FELDMAN M.S. 2014. *Reconstructing Reality in the Courtroom: Justice and Judgement in American Culture*, Quid Pro Books. (Originally published in 1981)

BEX F.J., VAN KOPPEN P.J., PRAKKEN H., VERHEIJ B. 2010. *A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence*, in «Artificial Intelligence and Law», 18, 123 ff.

BROHINSKY J., SONNERT G., SADLER P. 2022. *The Devil's Advocate,* in «Science & Education», 31, 575 ff.

BURKE A. 2007. *Neutralizing Cognitive Bias: An Invitation to Prosecutors,* in «New York University Journal of Law & Liberty», 2, 512 ff.

CANCINO MONTECINOS S. 2020. *New Perspectives on Cognitive Dissonance Theory,* Doctoral dissertation, University of Stockholm.

CHANG W., BERDINI E., MANDEL D.R., TETLOCK P.E. 2017. *Restructuring Structured Analytic Techniques in Intelligence,* in «Intelligence and National Security», 33, 337 ff.

CHOUDHRY N.K., FLETCHER R. H., SOUMERAI S.B. 2005. *Systematic Review: The Relationship between Clinical Experience and Quality of Health Care,* in «Annals of Internal medicine», 142, 260 ff.

CLARKE C., MILNE R., BULL R. 2011. *Interviewing Suspects of Crime: The Impact of PEACE Training, Supervision and the Presence of a Legal Advisor*, in «Journal of Investigative Psychology and Offender Profiling», 8, 149 ff.

COSMIDES L. 1989. *The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task*, in «Cognition», 31, 187 ff.

CROMBAG H.F.M., VAN KOPPEN P.J., WAGENAAR W.A. 2006. *Dubieuze zaken: De psychologie van strafrechtelijk bewijs* [*Dubious Cases: The Psychology of Evidence in Criminal Law*], Olympus. (Originally published in 1992)

DAHLMAN C. 2020. *De-Biasing Legal Fact-Finders with Bayesian Thinking,* in «Topics in cognitive science», 12, 115 ff.

DE LANGHE B., VAN OSSELAER S.M.J., WIERENGA B. 2011. *The Effects of Process and Outcome Accountability on Judgment Process and Performance,* in «Organizational Behavior and Human Decision Processes», 115, 238 ff.

DHAMI M.K., BELTON I.K., MANDEL D.R. 2019. *The "Analysis of Competing Hypotheses" in Intelligence Analysis,* in «Applied Cognitive Psychology», 33, 1080 ff.

DREISSEN W.H.B. 2007. *Bewijsmotivering in strafzaken* [*Reasoned Decisions in Criminal Law Cases*], Boom Juridisch.

DROR I.E., CHARLTON D., PÉRON A.E. 2006. *Contextual information renders experts vulnerable to making erroneous identifications*, in «Forensic Science International», 156, 1, 74 ff. Available on: https://doi.org/10.1016/j.forsciint.2005.10.017.

EVANS J.S.B.T., STANOVICH KE. 2013. *Dual-Process Theories of Higher Cognition: Advancing the Debate*, in «Perspectives on Psychological Science», 8, 223 ff.

FESTINGER L. 1957. *A Theory of Cognitive Dissonance,* Stanford University Press.

FINDLEY K.A., SCOTT M.S. 2006. *Multiple Dimensions of Tunnel Vision in Criminal Cases*, in «Wisconsin Law Review», 2, 291 ff.

FRANKEN S. 2021. *Potentieel onveilige veroordelingen: afstand tot het novumvereiste*, in «Expertise en Recht», 1, 33 ff.

GONZALEZ C. 2017. *Decision-making: A Cognitive Science Perspective*, in CHIPMAN S. (ed.), *The Oxford Handbook of Cognitive Science*, Oxford University Press, 249 ff.

HALL A.T., FRINK D.D., BUCKLEY M.R. 2015. *An Accountability Account: A Review and Synthesis of the Theoretical and Empirical Research on Felt Accountability*, in «Journal of Organizational Behavior», 38, 204 ff.

HARMAN G. 1965. *The Inference to the Best Explanation*, in «Philosophical Review», 74, 88 ff.

HARMON-JONES E., HARMON-JONES C. 2007. *Cognitive Dissonance Theory after 50 Years of Development,* in «Zeitschrift für Sozialpsychologie», 38, 7 ff.

HEUER R.J. 1999. *Analysis of Competing Hypotheses*, in «Psychology of Intelligence Analysis», 95.

KAHNEMANN D. 2011. *Thinking, Fast and Slow,* Penguin.

KASSIN S.M., GOLDSTEIN C.C., SAVITSKY K. 2003. *Behavioral confirmation in the interrogation room: On the dangers of presuming guilt*, in «Law and Human Behavior», 27, 187 ff.

KASSIN S.M., DROR I.E., KUKUCKA J. 2013. *The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions,* in «Journal of Applied Research in Memory and Cognition», 2, 42 ff.

KEILER J., ROEF D. (eds.). (2016). *Comparative Concepts of Criminal Law*, Intersentia.

KUKUCKA J., KASSIN S.M., ZAPF P.A., DROR I.E. 2017. *Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners,* in «Journal of Applied Research in Memory and Cognition», 6, 452 ff.

LERNER J. S., TETLOCK P. E. 1999. *Accounting for the Effects of Accountability,* in «Psychological Bulletin», 125, 255 ff.

LERNER J.S., TETLOCK P.E. 2003. *Bridging Individual, Interpersonal, and Institutional Approaches to Judgment and Decision Making: The Impact of Accountability on Cognitive Bias*, in SCHNEIDER S., SHANTEAU J. (eds.), *Emerging Perspectives on Judgment and Decision Research*, Cambridge University Press, 431 ff.

LIDÉN M., GRAENS M., JUSLIN P. 2018. *From Devil's Advocate to Crime Fighter: Confirmation Bias and Debiasing Techniques in Prosecutorial Decision Making*, in «Psychology Crime and Law», 25, 494 ff.

MACDOUGALL C., BAUM F. 1997. *The Devil's Advocate: A Strategy to Avoid Groupthink and Stimulate Discussion in Focus Groups*, in «Qualitative Health Research», 7, 532 ff.

MAEGHERMAN E., ASK K., HORSELENBERG R., VAN KOPPEN P. J. 2021. *Test of the Analysis of Competing Hypotheses in Legal Decision-Making,* in «Applied Cognitive Psychology», 35, 62 ff.

MAEGHERMAN E. 2021 *Facilitating Falsification in Legal Decision-Making: Problems in Practice and Potential Solutions*, Doctoral Dissertation, Maastricht University.

MARKSTEINER T., ASK, K., REINHARD M.A., GRANHAG P. A. 2011. *Asymmetrical Scepticism towards Criminal Evidence: The role of Goal- and Belief-Consistency,* in «Applied Cognitive Psychology», 25, 541 ff.

MCDERMOTT Y. 2015. *Inferential Reasoning and Proof in International Criminal Trials: The Potentials of Wigmorean Analysis*, in «Journal of International Criminal Justice», 13, 507 ff.

MENDEL R., TRAUT-MATTAUSCH E., JONAS E., LEUCHT S., KANE J.M., MAINO K., KISSLING W., HAMAN J. 2011. *Confirmation Bias: Why Psychiatrists Stick to Wrong Preliminary Diagnosis*, in «Psychological Medicine», 41, 12, 2651 ff.

MEVIS P.A.M. 2019. *Modernisering van het strafprocesrecht op z'n Duits* [*Modernisation of criminal procedural law in the German way*], in «Delikt & Delinkwent», 7, 40 ff.

NAN J. 2020. *Herziening ten voordele van de gewezen verdachte als buitengewoon rechtsmiddel*, in «Nederlands Tijdschrift voor Strafrecht», 1, 11 ff.

NICKERSON R.S. 1998. *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises,* in «Review of General Psychology», 2, 175 ff.

O'BRIEN B. 2009. *Prime suspect: An Examination of Factors That Aggravate and Counteract Confirmation Bias in Criminal Investigations*, in «Psychology, Public Policy, and Law», 15, 315 ff.

PENNINGTON J., SCHLENKER B.R. 1999. *Accountability for Consequential Decisions: Justifying Ethical Judgments to Audiences,* in «Personality and Social Psychology Bulletin», 25, 1067 ff.

PENNINGTON N., HASTIE R. 1986. *Evidence Evaluation in Complex Decision Making,* in «Journal of Personality and Social Psychology», 51, 242 ff.

PENNINGTON N., HASTIE R. 1991. *Cognitive Theory of Juror Mecision making: The Story Model,* in «Cardozo Law Review», 13, 519 ff.

PENNINGTON N., HASTIE R. 1992. *Explaining the evidence: Tests of the Story Model for Juror Decision Making,* in «Journal of Personality and Social Psychology», 62, 2, 189 ff.

PEOPLES C.D., SIGILLO A.E., GREEN M., MILLER M.K. 2012. *Friendship and Conformity in Group Opinions: Juror Verdict Change in Mock Juries,* in «Sociological Spectrum», 32, 2, 178 ff.

POPPER K. 2005. *The Logic of Scientific Discovery*, Routledge. (Originally published in 1959)

PRONIN E., LIN D.Y., ROSS L. 2002. *The Bias Blind Spot: Perceptions of Bias in Self Versus Others,* in «Personality and Social Psychology Bulletin», 28, 3, 369 ff.

RACHLINSKI J.J. 2012. *Judicial Psychology,* in «Rechtstreeks», 2, 15 ff.

RASSIN E. 2010. *Blindness to Alternative Scenarios in Evidence Evaluation,* in «Journal of Investigative Psychology and Offender Profiling», 7, 153 ff.

RASSIN E. 2018. *Reducing Tunnel Vision with a Pen-and-Paper Tool for the Weighting of Criminal Evidence*, in «Journal of Investigative Psychology and Offender Profiling», 15, 227 ff.

REIJNTJES J.M., REIJNJES-WENDENBURG C. 2018. *De Bewijsconstructie* [*The Evidence Construction*], in «Handboek Strafzaken», 34.

ROBERTS P. 2020. *Scenarios, Probability, and Evidence Scholarship, Old and New*, in «Topics in Cognitive Science», 12, 1213 ff.

SCHMITTAT S.M., ENGLICH B. 2016. *If You Judge, Investigate! Responsibility Reduces Confirmatory Information Processing in Legal Experts*, in «Psychology, Public Policy, and Law», 22, 386 ff.

SCHÜNEMANN B. 1983. *Experimentelle Untersuchungen zur Reform der Hauptverhandlung in Strafsachen*, in KERNER H.I., KURY H., SESSAR K. (eds.), *Deutsche Forschungen zur Kriminalitatsentstehung und Kriminalitatskontrolle*, Heymanns, 1109 ff.

SCHÜNEMANN B., BANDILLA W. 1989. *Perseverance in Courtroom Decisions*, in WEGENER H., LÖSEL F., HAISCH J. (eds.), *Criminal Behavior and the Justice System: Psychological Perspectives*, Springer, 181 ff.

SIMMELINK J.B.H.M. 2001. *Bewijsrecht en bewijsmotivering [Evidence Law and Reasoned Decisions]*, In GROENHUIJSEN M.S., KNIGGE G., (ed.), *Het onderzoek ter zitting,* Rijksuniversiteit Groningen, 397 ff.

SIMON D. 2004. *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, in «The University of Chicago Law Review», 71, 511 ff.

SPENCER J. R. 2016. *Adversarial vs Inquisitorial Systems: Is There Still Such a Difference?*, in «The International Journal of Human Rights», 20, 601 ff.

STANOVICH K.E., TOPLAK M.E. 2012. *Defining Features Versus Incidental Correlates of Type 1 and Type 2 Processing*, in «Mind & Society», 11, 3 ff.

STRIER F. 1992. *What Can the American Adversary System Learn from an Inquisitional System of Justice*, in «Judicature», 76, 109 ff.

TAY S.W., RYAN P., RYAN C.A. 2016. *Systems 1 and 2 Thinking Processes and Cognitive Reflection Testing in Medical Students*, in «Canadian Medical Education Journal», 7, 97 ff.

TEGENLICHT 2018. *Brandbrief rechters: wij vrezen voor de toekomst van de rechtspraak*, in «Nieuwsuur», 8 November 2018. Available on: https://nos.nl/nieuwsuur/artikel/2258390-brandbrief-rechters-wij-vrezen-voor-de-toekomst-van-de-rechtspraak.html.

TENNEY E.R., CLEARY H.M., SPELLMAN B.A. 2009. *Unpacking the Doubt in "Beyond a Reasonable Doubt": Plausible Alternative Stories Increase Not Guilty Verdicts*, in «Basic and Applied Social Psychology», 31, 1, 1 ff. Available on: https://doi.org/10.1080/01973530802659687.

THOMPSON-CANNINO J., COTTON R., TORNEO E. 2009. *Picking Cotton*, St. Martin's Press.

TRAEST P. 2018. *België*, in VERREST. P.A.M., MEVIS P.A.M., (eds.), *Rechtsvergelijkende inzichten voor de modernisering van het Wetboek van Strafvordering*, Boom Juridisch, 19 ff.

TWINING W. 1985. *Theories of Evidence: Bentham & Wigmore,* Stanford University Press.

TWINING W. 1985. *Theories of Evidence: Bentham & Wigmore,* Stanford University Press.

VAN KOPPEN P.J., MACKOR A.R. 2020. *A Scenario-Approach to the Simonshaven Case,* in «Topics in Cognitive Science», 12, 1132 ff.

VERBAAN J.H.J. 2016. *Straf(proces)recht begrepen* [*Criminal (Procedural) Law Understood*], Boom Juridisch.

WAGENAAR W.A., VAN KOPPEN P.J., CROMBAG H.F.M. 1993. *Anchored Narratives: The Psychology of Criminal Evidence,* Harvester Wheatsheaf.

WASON P.C., JOHNSON-LAIRD P.N. 1972. *Psychology of Reasoning: Structure and Content,* Harvard University Press

WATERS N.L., HANS V.P. 2009, *A Jury of One: Opinion Formation, Conformity, and Dissent on Juries,* in «Journal of Empirical Legal Studies», 6, 513 ff.

WEST R.F., MESERVE R.J., STANOVICH K.E. 2012. *Cognitive Sophistication Does Not Attenuate the Bias Blind Spot*, in «Journal of Personality and Social Psychology», 103, 506 ff.

WILLIAMSON S., FOLEY M. 2018. *Unconscious Bias Training: The 'Silver Bullet' for Gender Equity?,* in «Australian Journal of Public Administration», 77, 355 ff.

Legal Rules as a Bias-Counteracting Device

NOAM GUR

In his essay *Of the Origin of Government*[1], David Hume brings to light an intriguing link between the justificatory basis of government and the phenomenon nowadays called by social psychologists *systematic biases*[2]. According to Hume, although people have an interest in «the maintenance of order in society», a natural propensity that springs up in their «circumstances and situation» often «hinders [them] from seeing distinctly» that interest and inclines them to act against it[3]. He argues further that, since the «circumstances and situation» of government do not prompt the above propensity, people can resort to government as an «expedient» against that propensity[4]. While Hume's primary focus is an actor's deliberation in view of his or her interest, the pattern observed by Hume is extendable to an actor's deliberation on how he or she ought to act from the perspective of equity and morality—an extension that, if valid, lends further significance to the observed pattern.

Since Hume's essay, however, systematic biases have generally not occupied a salient place in political and legal theory discourse about the justification, role, and normative significance of political and legal authority[5]. While I can surmise what has caused this omission, tracing its causes is not my aim here. My aim, instead, is to help rectify the omission. In doing so, I will focus particularly on the context of law and legal rules; I will seek to benefit from modern empirical work in psychology regarding biases; I will attempt to broaden the range of biases taken into account; and I will highlight some of the implications of the observed link between law and biases. I should add, without wishing to detract from my opening tribute to Hume, that notwithstanding my affinity with some of his intuitions, my argument hereunder is a self-standing argument that is put forward for independent consideration.

Stated more specifically, my purpose is to illustrate that one of the key functions of legal guidance of conduct (in its appropriate use and legitimate instantiations) is to serve as a corrective device against several systematic biases present in the settings of activity that law typically regulates[6]. I will begin with a few clarificatory comments essential for understanding my claim (Section 1). I will then bring into focus several relevant biases and explain how law—

* I am grateful to Marco Brigaglia and to an anonymous referee for helpful suggestions and comments. I also wish to thank Tatiana Catsiapis, Lorenzo Estero, and Catherine Ievers for their help as research assistants at different stages of my work on this paper.

[1] HUME 2007 [1740], 324-345 [SB534-SB539].

[2] Instead of the term *bias*, Hume uses terms such as *propensity*, *propension*, or *inclination* (HUME 2007 [1740], 343-344 [SB535-SB537]).

[3] HUME 2007 [1740], 344 [SB538].

[4] HUME 2007 [1740], 344 [SB537].

[5] That is not to suggest that no work has been carried out on the link between law and biases. Some notable work has been done, especially by scholars of regulation and behavioural economic analysis of law. See, e.g., FARNSWORTH 2003; JOLLS & SUNSTEIN 2006; ZAMIR 2015; THALER & SUNSTEIN 2021.

[6] The word "serve" in the accompanying body text might evoke Joseph Raz's "service conception" of authority (RAZ 1986, 36-80). I have discussed Raz's service conception of authority and explained how my approach differs from it elsewhere (see, e.g., GUR 2018, 21-70, 127-130, 155-156, 213-218).

and, particularly, legal rules—is structurally suited to counteract some of their common instantiations in social life (Section 2). I will discuss possible doubts arising from several factors, such as the corrupting capacity of power, social injustice facilitated by law, law-generated errors, and the prospect of debiasing oneself of one's own accord (Subsection 2.2. and Section 3). I will conclude with some remarks about the implications of my claim (Section 4).

1. *Preliminary comments*

I should first clarify that my claim will *not* be that law can counteract certain biases because law-making officials tend to be persons less amenable to them (which is, of course, not the case). Instead, law's comparative advantage in this regard—though dependent for its realization on the reasonable personal competence of the law-makers involved—will be attributed to certain *structural* characteristics of legal norms and of the settings and mode of decision-making in which law-makers typically operate. I will explain shortly what these structural characteristics are. At this point, I only wish to emphasize that my claim does not revolve around a personal comparative advantage and that my analysis will recognize, and take into account, the fallibility of law-making officials.

The above comment must be complemented by a further important caveat. Nothing I will say in this paper should be read as a claim that all instances of law, or all legal systems, fulfil a bias-counteracting function[7]. Nor will I claim that legal officials or systems can fulfil this function without meeting certain prerequisites in addition to their being legal. Although the said bias-counteracting function owes its existence primarily to structural attributes of law, this function is not likely to be fulfilled if the relevant law-making officials fail to satisfy some prerequisites of competence and moral decency, or if the relevant legal system fails to pass a threshold of reasonable quality. Thus, for example, if the relevant officials have no regard for the interests of the law's subjects, these officials are not likely to make good use of law's potentially beneficial structural attributes. Nor are they likely to do so if they (or, as the case may be, the advisors on whom they rely) are not reasonably judicious and reasonably informed about the regulated domain of activity. These examples are not intended as an exhaustive list of the relevant prerequisites. They are merely intended as illustrations of the general caveat just entered. Before turning to the next comment, the conjunction of the last two points can be summed up thus: the bias-counteracting capacity highlighted herein is not *rooted* in system-specific or personal characteristics, but it does *depend for its realization* on the reasonable quality of a legal system and on the possession of certain personal attributes (e.g. reasonable degrees of knowledge, understanding, and judiciousness) by those involved in law-making.

A third clarificatory comment is about the conditions that warrant recourse to the bias-counteracting function of law (assuming that the legal officials and system involved meet the prerequisites noted in the previous paragraph). None of my arguments hereunder is intended to imply that the presence or influence of the relevant biases is *sufficient* to warrant the law's intervention. Invoking the rather heavy machinery of law is a course of action that often implicates significant costs and adverse effects, such as the restriction of individual liberty[8], the potential emergence of a legalistic culture[9], and enforcement costs—all of which must be taken

[7] "Legal" in the accompanying body text is meant to imply, *inter alia*, a requisite degree of compliance with the precepts of legality expounded by Lon Fuller, also associated with the label the *rule of law* (FULLER 1969, 33-38, 46-91). According to Fuller, a total failure to comply with any one of those precepts results in something that is not a legal system properly so called (FULLER 1969, 39).

[8] Or, at least, what Isaiah Berlin called negative freedom (BERLIN 1969, 118-172). Relevantly, Berlin refers there to Hobbes's characterization of law as a "fetter" (see HOBBES 1991a [1642], 274).

[9] In this connection, see, e.g., Leslie Green's discussion of what he calls the vices of legalism (GREEN 2008, 1058).

into account when legal intervention is considered in response to a problem, including the problem of bias-induced errors. Thus, to justify recourse to law as a bias-counteracting device in a given context of activity, it must be shown, not only that human decision-making in that context is influenced by bias, but also that the resulting errors are consequential or common enough (or both) to outweigh the negative effects of deploying the legal apparatus. Now, there is a further dimension to this calculus. Counter-bias regulatory action need not always take the form of rules that dictate the behaviour of biased actors. There are softer or less direct forms of regulatory intervention in response to bounded rationality problems[10]. Thus, an adequate assessment of regulatory intervention is not simply binary—i.e. intervention vs. non-intervention—but one that takes into account alternative types of regulatory action and their associated pros and cons. The balance of pros and cons may vary between alternative types of regulatory action because they may differ, for example, in terms of how costly, intrusive, restrictive, and effective they are. I will revert to this issue in due course.

A fourth comment concerns the non-exclusivity of the said function of law. As indicated by my wording at the outset (i.e. "one of the key functions")[11], by no means do I intend to suggest that counteracting biases is the only, or the only important, function of law. Nor is it my claim that the problems addressed by law when it counteracts biases are *sheer* problems of bias. I readily acknowledge that law has other important functions, such as correcting for informational deficiencies, facilitating social coordination, extricating actors from prisoner's dilemma-type situations, upholding schemes of social cooperation, and generating common normative frameworks in the face of ideological and moral disagreements. And I also acknowledge that such functions are often performed in conjunction with the bias-counteracting function. In other words, the problems that law addresses when it counteracts biases are, for the most part, problems wherein other difficulties (e.g. difficulties of coordination, cooperation, and deficient information) are *entwined with* systematic biases. This entwinement will be made visible through some of my illustrations in Section 2.

Given the cognitive-science focus of this volume, a final preliminary word should be devoted to the relationship between biases and cognition. Are biases a cognitive phenomenon? They are a cognitive, but not purely cognitive, phenomenon. At least some biases (including some of those I will discuss here) are linked with both cognitive and motivational factors[12]. They have cognitive effects, such as their influence on how we perceive and interpret relevant factual information and reasons for or against an action. But the sources and triggering conditions of these biases are partly motivational—as is the case, for example, when self-interest motivations manifest themselves in the form of a self-serving bias. Thus, biases can be classified as a partly cognitive and partly motivational phenomenon. The dual character of biases has significant implications, inter alia, for law's normativity—implications that I will highlight in Section 4 below and have discussed at greater length elsewhere[13].

[10] See, e.g., JOLLS & SUNSTEIN 2006 (distinguishing between, on the one hand, debiasing strategies that insulate outcomes from the effects of bounded rationality and, on the other hand, strategies that alter the bias-triggering situation or environment in an attempt to steer people in more rational directions, which they call *debiasing through law*); THALER & SUNSTEIN 2021 (advancing the notion of *nudge*, namely the design of choice environments in (not restrictive or incentive-based) ways that facilitate better decisions).

[11] In the third paragraph.

[12] See, e.g., PYSZCZYNSKI & GREENBERG 1987; KUNDA 1990; DITTO & LOPEZ 1992; DUNNING 2001; DAWSON et al. 2002; BALCETIS & DUNNING 2006; DUNNING 2015.

[13] GUR 2018, 121-130, 163, 165, 215-216.

2. *Systematic biases and legal guidance*

Before turning to the relevant biases, notice should be taken of two distinctions that will inform the discussion. These are, more specifically, two structural differences between the typical environment and mode of decision-making of subjects in everyday activity on the one hand, and of law-makers operating in their capacity as such on the other. The qualifier "typical" is important because the differences I am about to highlight mark no more than approximate tendencies, ones that shade into each other in more than one way. But they are tendencies that nonetheless have significant implications for our discussion. The two distinctions are as follows. (1) In the course of everyday activity as ordinary citizens, many (though not all) of our decisions are, in the operative sense, decisions about *our* actions. Law-makers operating in their capacity as such, on the other hand, typically decide about actions of a general class of people. There is, if you like, a directional difference between these modes of decision-making: the former often consists in a decision whose operative content is first-personal, whereas the latter's operative content (e.g. a rule drafted or voted on) is paradigmatically addressed to a general class comprised almost entirely of other actors[14]. (2) Private decision-making, especially in the course of everyday activity, is often (though, again, not always) *particularistic* in character. Decisions made by a citizen in the practical settings of day-to-day life are often decisions about how to act here and now. The private decision-maker in these settings is therefore placed in relatively close proximity to situational stimuli. On the other hand, the type of decision-making in which legislative and regulatory authorities are engaged is primarily *non-particularistic*. Decisions made by legislators and regulators are characteristically general (not only in the aforementioned sense of their class of addressees, but also regarding the class of circumstances encompassed)[15] and prospective in application—they are intended to apply to future action[16]. Such public officials are placed in decisional settings that are relatively distant from the specificities of each and every case that falls within the purview of their decisions. These distinctions provide relevant background for the argument below, as they will make it easier to appreciate how, in certain contexts of activity where individual decision-making tends to be affected by certain types of bias, legal modes of regulation offer a suitable remedy.

2.1. *Self-enhancement bias*

The directional difference highlighted above between modes of decision-making—i.e. decisions directed at self/others—brings into play two types of bias with special pertinence to this inquiry. They will be discussed in turn in this and the following subsection. The first type, known to psychologists as the *self-enhancement bias*[17], tends to emerge when people evaluate their

[14] There is an obvious sense in which a law-maker's decision is also self-personal, in that it can be framed in terms such as: "should I (i.e. the law-maker) put forward this or that rule?" or "how should I vote in a parliamentary proceeding?" However—and this is the sense in which it is not self-personal—the rule he or she is putting forward or voting on is addressed to a general class comprised almost entirely of other actors. This latter fact has significant implications for our purpose.

[15] See, e.g., ARISTOTLE 1984, 1795-1796 [1137b]; HART 2012, 21; FULLER 1969, 33-34, 46-49. The case of judicial decisions is more complicated in this respect. Unlike private decisions, they are exclusively addressed to others; but unlike typical legislative and regulatory acts, they are particularistic in that they are primarily addressed to specific, named individuals (or corporations) and concern specific disputes. However, to the extent that judicial decisions have binding force as precedents for future cases, their function bears some resemblance to that of legislative or regulatory acts. Another relevant difference between judicial and private decision-making is that the former is paradigmatically about past events. In this sense, it takes place in conditions more conducive to careful and reflective decision-making than are the normal conditions in which daily decisions on immediate actions are made.

[16] See, e.g., FULLER 1969, 35, 51-52.

[17] See, e.g., BAUMHART 1968, 20-25; LARWOOD & WHITTAKER 1977; SVENSON 1981; BROWN 1986; KRUGER & DUNNING 1999; ZHAO 2021. See also CROSS 1977, 8-12.

own performance and skills in comparison with those of others. In this context, more often than not, judgement seems to be compromised by inner motives for upholding one's self-esteem[18], resulting in a pattern documented in several empirical studies under the label the *better-than-average effect*: most people rate their performance and skills as better than those of the average person, with a disproportionately large percentage placing themselves towards the top end of the comparative scale[19]. Thus, for instance, experimental evidence suggests that the majority of drivers consider themselves to be more skilful and less dangerous than most other drivers[20]. And a similar pattern of self-appraisal inflation has been identified through surveys pertaining to other personal traits and skills, and different domains of activity, such as self-assessment of ethical conduct in business[21], managerial abilities[22], academic teaching performance[23], and logical reasoning[24].

One important implication of these findings is that they help explain why legal systems need to regulate some of the matters they commonly regulate. The fact that most drivers tend to overestimate their driving skills and underestimate their dangerousness as drivers, for example, is part of what makes it sensible to have the road traffic safety laws that we find in so many jurisdictions. If no legal speed limit, overtaking restrictions, drink-drive limitations, rules about multitasking while driving, tailgating prohibition, and other similar rules were in force, and drivers were to individually decide on these matters, they would frequently fail to take into account the full extent of their shortcomings and the actual limits of their abilities as drivers. In such a state of affairs, there would be too many wrong decisions on the part of too many drivers[25]. This problem can be overcome by recourse to legal regulation, partly because the judgement exercised by law-makers will turn on how *they* estimate the driving skills of *other people* (i.e. the ordinary driver), rather than on how any given driver estimates his or her own skills. This is not to suggest that the self-enhancement bias—or biases more generally—fully explain the recourse to traffic regulation or that other factors are insignificant. The explanation features other important considerations, including, notably, the need for coordination between road-users. This illustrates my preliminary comment that the biases countered by law are typically entwined with other difficulties. Now, reverting to the self-enhancement bias, its role in the traffic rules example can be generalized. It is applicable to other contexts of regulation wherein significant stakes are attached to the manner in which people perform an activity and their the manner of performance is thought to depend on characteristics such as skills, knowledge, or understanding—including, for example, regulations concerning safety in work environments, product safety, construction standards, hygiene standards for food businesses, the positioning of outdoor electricity lines and cellular base stations, domestic use of substances such as pesticides and herbicides, storage and disposal of hazardous materials, and so on.

[18] Along with other, cognitive and situational factors—see DUNNING et al. 1989; ALICKE et al. 1995; KRUGER & DUNNING 1999; KRUGER 1999; DUNNING et al. 2003.

[19] See citations in fn. 17.

[20] SVENSON 1981. See also PRESTON & HARRIS 1965; DELHOMME 1991; HORSWILL et al. 2004; WAYLEN et al. 2004; DOGAN et al. 2012; STEPHENS & OHTSUKA 2014.

[21] BAUMHART 1968.

[22] LARWOOD & WHITTAKER 1977.

[23] CROSS 1977.

[24] KRUGER & DUNNING 1999, 1124-1125. This may evoke Hobbes's remark that «such is the nature of men, that howsoever they may acknowledge many others to be more witty, or more eloquent, or more learned; Yet they will hardly believe there be many so wise as themselves: For they see their own wit at hand, and other mens [*sic*] at a distance» (HOBBES 1991b [1651], 87).

[25] On the influence of the better-than-average effect/self-enhancement bias on driving decisions and behaviour, see, e.g., MĂIREAN & HAVÂRNEANU 2018.

2.2. *Self-serving bias*

The second type of relevant bias—self-serving bias—tends to emerge when people exercise judgement about questions of equity that bear on their own interest. In these situations, more often than not, people perceive relevant facts and principles in a manner somewhat beneficial to themselves[26]. Thus, even if the decision stems from a process of reasoning that is not wittingly egoistic, excessive weight is nonetheless likely to be assigned to those considerations that coincide with the deliberating actor's needs and wants at the expense of other relevant considerations. As Albert Venn Dicey succinctly described it: «[M]en come easily to believe that arrangements agreeable to themselves are beneficial to others. A man's interest gives a bias to his judgment far oftener than it corrupts his heart»[27].

Consider, as relevant evidence, the results of an experiment in which students were asked to make fair judgements about the remuneration that should be given to them and to others for different amounts of working hours as exam readers[28]. The questionnaires administered to one group of participants premised that they had worked seven hours, whereas another student had worked ten hours (Self=7, Other=10). The questionnaires administered to another group of participants premised the reverse, namely that they had worked ten hours, whereas another student had worked seven hours (Self=10, Other=7). The results revealed a clear bias towards overpayment to self. For instance, there was a significantly higher rate of participants in the former (Self=7, Other=10) group than in the latter (Self=10, Other=7) group who consistently replied that fairness requires that equal payments be made to themselves and to the other student (despite the difference in working hours)[29]. Moreover, among participants who did not consistently opt for an equal payments outcome as the fairest, the gap between payment to self and payment to other was significantly larger in the latter (Self=10, Other=7) group than in the former (Self=7, Other=10) group[30].

The tendency reflected in such findings ties in with some of the central roles law fulfils in social life. Consider, for example, the regulation of conduct pertaining to common resources or goods such as fresh water, pasturage, parks, beaches, rivers, and other parts of the environment. Or think of laws levying taxes to finance the provision of public or common goods such as national defence, sanitation, roads, and railways. If such matters were left unregulated and each citizen were to decide what he or she ought to do in the way of protecting those goods or resources, and how much to contribute to their sustenance, there would arise, among various other problems[31], the problem of self-serving bias[32]. That is, many citizens, even those who would not act in a deliberately selfish way, would tend to perceive the facts and to balance reasons for actions in a manner somewhat overly sensitive to their individual circumstances and needs, and would thus tend to exempt themselves too often from sharing collective burdens[33].

[26] MESSICK & SENTIS 1979; MESSICK 1985, 94-100; THOMPSON & LOEWENSTEIN 1992; BABCOCK & LOEWENSTEIN 1997; DAWSON et al. 2002. See further POGARSKY & BABCOCK 2001.

[27] DICEY 1914, 15.

[28] MESSICK & SENTIS 1979.

[29] MESSICK & SENTIS 1979, 428, 432.

[30] MESSICK & SENTIS 1979, 428-29, 432.

[31] For example, coordination difficulties, lack of information about collective needs, and deliberate egoism.

[32] There is, of course, an extent to which self-serving bias may operate even after regulation has been introduced, namely as a cause of non-compliant behaviour. But, notwithstanding this limitation, the role played by law in diminishing and attenuating the operation of self-serving bias remains significant and crucial. In Section 4, I will refer to a conception of legal normativity that I consider to be integral to law's ability to fulfil this role.

[33] Cf. John Locke's comment that in the state of nature «though the law of nature be plain and intelligible to all rational creatures; yet men being biased by their interest . . . are not apt to allow of it as a law binding to them in the application of it to their particular cases» (LOCKE 2003 [1689], 325 [ch. IX, para. 124]).

Legal regulation suggests itself as a suitable measure of solving, or significantly attenuating, this problem[34], not least because of its typically generic character—that is, because the paradigmatic mode of decision associated with it is decision directed at the public at large couched in general rules, rather than a decision about this or that person[35].

I should add, parenthetically, that the use of common resources may also involve what Garrett Hardin famously labelled the *tragedy of the commons*[36], namely situations where unrestricted access to such resources is destined to result in their depletion or ruin. However, the range of situations that elicit the self-serving bias is wider than the tragedy of the commons-type situations[37]. Moreover, these two problems are associated with different modes of deliberation: the tragedy of the commons, as characterized by Hardin, focuses on self-interested rational deliberation, whereas self-serving bias occurs in deliberation that is not wittingly selfish, such as judgement of morality or equity[38].

A possible challenge to the above claim should be mentioned and addressed at this point. Law-makers and other legal officials, it may be argued, are themselves susceptible to certain forms of bias associated with their position and power. As Lord Acton famously noted, «power tends to corrupt»[39], and, as various critical legal theorists argued, there are patterns of legal thought and practice that facilitate social injustice through their tendency to preserve or reinforce the domination of traditionally powerful groups in society (e.g. in terms of economic class[40], gender[41], or race[42]). Now, I do not deny that such ill patterns exist, to a varying extent, in legal and political systems[43], and can interfere with law's proper functioning against individual self-serving bias. But they do not invalidate my foregoing argument about law's ability to correct for this bias, for the following combination of reasons.

To begin with, it should be borne in mind that a respectable legal and political system will have in place several safeguards that mitigate the fallibilities associated with legal and political power. And in my preliminary comment at the outset regarding requisite conditions for the law's ability to counter biases, I envisaged, as one of the relevant prerequisites, the presence of such safeguards[44]. These include, for example, constitutional protection of civil and political rights (including rights that facilitate social and political change, such as freedom of speech and assembly), institutional checks and balances between separate branches of government, a

[34] I use the more qualified wording "significantly attenuating", because, as noted in fn. 32 above, law cannot entirely avert all relevant manifestations of self-serving bias. In this connection, it should also be noted that the use of law must be accompanied by other means such as the cultivation of civic awareness and conscientious attitudes among members of society.

[35] This is not to deny that, in some contexts, "the invisible hand" of the market can effectively operate so that self-interested individual behaviour will incidentally promote collective welfare. The above discussion, however, focuses on other types of situation, in which a regulatory vacuum is not likely to produce collectively desirable results.

[36] HARDIN 1968. Among other resources, Hardin notably draws on LLOYD 1883. For a relevant literature review, see MESSICK, BREWER 1983. On the operation of self-serving bias in this context, see, e.g., WADE-BENZONI et al. 1996. And on additional biases in this context, see THOMPSON 2000.

[37] The former, unlike the latter, is not limited to situations wherein the potential risk is of depletion or ruin of a common resource.

[38] Hardin does consider and rules out the possibility of averting the tragedy of the commons through «an appeal to conscience» (HARDIN 1968, 1246-1247). But this is a merely subsidiary element of his argument, and it is discussed with a rather narrow focus on evolutionary considerations.

[39] In his letter to Bishop Mandell Creighton, 5 April 1887 (DALBERG-ACTON 1949, 364). But cf. DECELLES 2012.

[40] See, e.g., HORWITZ 1977; TUSHNET 1978; ABEL 1998.

[41] See, e.g., TAUB & SCHNEIDER 1998; MACKINNON 1983; MACKINNON 1987; MOSSMAN 1987. Cf. HUNTER 2012.

[42] See, e.g., BELL 2008; CRENSHAW et al, 1995; DELGADO & STEFANCIC 2013.

[43] Though it should also be kept in mind that law can play an important role in combating such patterns through equality laws, both directly and by means of its expressive effect. For a recent relevant study, see ALBISTON & CORRELL 2023.

[44] Though I was only partly explicit about it (in fn. 7 above).

reasonable degree of adherence by officials to "the rule of law" constraints on the exercise of power, and a reasonable level of compliance with rules of due process, transparency, and accountability. However, such safeguards, and rights in particular, have themselves been a target of critique by several CLS scholars[45], in the light of which it becomes especially important to delineate my modest claim in this regard: it is readily conceded that such safeguards are effective only up to a point, and may even have certain negative effects (e.g. insofar as their associated rhetoric can work to conceal or blind us to existing patterns of social injustice). It is only claimed here that, on balance, such safeguards make a beneficial contribution towards constraining and reducing the possibility for power abuse. Or, to put the point differently, if I had to choose between living with these safeguards or living without them, I would emphatically choose to live *with* them, despite their shortcomings.

What about the remaining extent of inequity that even worthwhile legal and political systems (with the above safeguards in place) may harbour? There is a risk that this measure of inequity will work its way into law's envisaged operation against self-serving bias. But this risk does not *negate* law's ability to correct for self-serving bias. Nor does it mean that society should, or reasonably can, dispense with this function of law. For a state of affairs wherein matters such as the use of common goods, the allocation of collective burdens, and the trade-off between mutually incompatible individual liberties would all be left unregulated by law hardly appears a palatable option—and is, at any rate, worse than having such matters regulated by a reasonably well-designed legal system that attains a significant, albeit deficient, level of social justice. To try to address the shortfall in social justice by renouncing law's regulatory role would be, as the idiom goes, to throw the baby out with the bathwater. A better way forward is to advance the law towards a higher attainment of social justice through a combination of political, legal, and civil action, employing either conventional tools or, insofar as necessary, tools that lie «at the outer edge»[46] of fidelity to law, such as (non-violent) civil dissent[47].

2.3. *The availability heuristic and its biasing effect*

Another pertinent type of bias emanates from a cognitive mechanism labelled by Amos Tversky and Daniel Kahneman the *availability heuristic*[48]. These two renowned psychologists have shown that people's intuitive estimations of the probability of events are influenced by availability, that is, the ease with which instances or associations of the relevant event come to one's mind[49]. Availability is indicative of the probability of events, since frequently occurring events have, *ceteris paribus*, better chances to be noticed than infrequent ones, and, to this extent, they are likely to be better recalled and easier to imagine[50]. Availability, therefore, can function as a heuristic device that reduces complex tasks of probability computation to simpler mental operations[51]. However, despite its general utility, the availability heuristic is subject to significant limitations[52]. First, there is an obvious sense in which the connection between frequency of occurrence and availability is merely probabilistic and less than conclusive: while a frequently occurring event is likely to be discerned by people and register in their minds, it is always possible that some individuals have never, or rarely, or not recently, experienced or witnessed it, which means that it

[45] See, e.g., TUSHNET 1993.
[46] RAWLS 1999, 322.
[47] RAWLS 1999, 319-323, 326-331.
[48] TVERSKY & KAHNEMAN 1973; TVERSKY & KAHNEMAN 1974.
[49] TVERSKY & KAHNEMAN 1973, 207-208.
[50] TVERSKY & KAHNEMAN 1973, 208; TVERSKY & KAHNEMAN 1974, 1127.
[51] TVERSKY & KAHNEMAN 1974, 1124. This may be especially useful in everyday dealings where people only have a limited amount of time and energy for investigation prior to action and may not have relevant statistical data on hand.
[52] TVERSKY & KAHNEMAN 1974, 1124.

may have low availability from their perspective[53]. For example, young people may have relatively little exposure to the occurrence of diseases linked with age[54]. If availability exerts powerful influence on probability judgements, it may render them, as it were, prisoners of their own experience who underestimate the likelihood of developing such diseases in the future[55]. Second, availability is affected by additional factors not related to frequency of occurrence, such as how vivid and salient the relevant event is[56]. For example, terrorist attacks are more dramatic and tend to make more of a news item than do road accidents, and thus may well have increased cognitive availability[57]. By allowing such attributes to impinge on how we perceive the likelihood of events, the availability heuristic leads to systematic deviations from correct statistical assessment[58]. It has, in other words, a biasing effect.

That these biases permeate people's evaluations of probability has been demonstrated in several experimental studies. In one experiment, for example, participants were presented with one of the letters K, L, N, R, V in each trial, and were asked whether randomly selected words from a standard English text are likelier to be words beginning with that letter or words wherein that letter appears third[59]. Each of the above letters is, in fact, likelier to appear as the third letter in a word. However, it is easier to think instantly of words that begin with these letters, which is to say that they have greater cognitive availability. The results attested to a biasing influence of this factor: of 152 participants, 105 thought that the majority of the above-listed letters are likelier to appear at the beginning of a word[60]. In another experimental study, in which participants were asked to estimate the frequency of different fatalities[61], significant over- and underestimations, many of which seem to mirror a biasing effect of availability, were made by the participants[62]. Overestimations often pertained to fatalities by dramatic and vivid causes—for example, floods, tornadoes, and venomous bites or stings—whereas underestimations often pertained to "quiet killers" that hardly attract public attention—for example, death from leukaemia, emphysema, diabetes, and heart disease[63].

The notion of availability-related bias can offer further insight into the type of practical difficulties law aims to overcome in various domains of regulation. We often rely in daily reasoning on subjective assumptions or intuitive estimations regarding the likelihood of desirable outcomes, potential ramifications, and possible harms associated with alternative courses of

[53] There are, of course, also indirect means of communication by which impressions can be conveyed, e.g., television and social media. But it is not clear how much these do to correct the potentially distorting influence of availability, and it is possible that they sometimes even reinforce it. For one thing, frequency of occurrence does not seem to be one of the leading criteria by which television broadcasters select their reported items (see body text accompanying fn. 57). And it is also conceivable that some of the phenomena observed in the context of social media, such as the so-called echo chamber effect, have availability-limiting/distorting effects.

[54] LICHTENSTEIN et al. 1978, 575.

[55] LICHTENSTEIN et al. 1978, 575; TVERSKY, KAHNEMAN 1973, 230; SUNSTEIN 2002, 33. The availability heuristic can, to some extent, affect even the clinical judgement of physicians (see, e.g., LY 2021). See further on availability and the perception of disease likelihood, SHERMAN et al. 1985.

[56] TVERSKY & KAHNEMAN 1973, 228-229; TVERSKY & KAHNEMAN 1974, 1127.

[57] SUNSTEIN 2002, 50-52. See more generally SUNSTEIN 2002, 33-35, 78-98.

[58] TVERSKY & KAHNEMAN 1973, 209; TVERSKY & KAHNEMAN 1974, 1127.

[59] TVERSKY & KAHNEMAN 1973, 211-212.

[60] TVERSKY & KAHNEMAN 1973, 212. Furthermore, each of these letters was thought by a majority of the participants to be likelier to appear at the beginning of a word than as the third letter in it. The study included additional experiments denoting a similar tendency.

[61] LICHTENSTEIN et al. 1978. See discussion in SLOVIC 2000, 106-107; SUNSTEIN 2002, 34.

[62] LICHTENSTEIN et al. 1978, 562-571; SLOVIC 2000, 106-107.

[63] LICHTENSTEIN et al. 1978, 562-567; SLOVIC 2000, 107. A corresponding tendency appeared when participants were given pairs of lethal events and were asked which is the more frequent event in each pair: the bulk of them, for example, incorrectly judged road accidents to be a more frequent cause of death than stroke, and homicide to be more frequent than suicide (LICHTENSTEIN et al. 1978, 553-559).

action. So it should be clear in light of the above discussion that bias linked with the availability heuristic can readily find its way into practical decision-making and lead us to sub-optimal action. And when the subject of our decision or the setting in which we make it is highly susceptible to such bias, this may significantly strengthen the case for regulatory intervention[64]. Thus, for example, a miner, builder, technician, or factory worker may be continually exposed in the course of their work to noise, dust, radiation, or other unhealthy agents that, although not causing immediately noticeable harm, have detrimental effects in the long term. Given their gradual and unspectacular modus operandi, the risks associated with such physical or chemical agents may have low cognitive availability from the perspective of the worker and their employer[65]. So, if left to their own devices to freely determine what precautions to take or whether to purchase a related insurance policy, the worker and their employer are prone to overlook or underweight the above types of risk, and thus fail to take the appropriate measures[66]. Legal regulation often provides a suitable way to avert such failures, partly because the normal design of law-making systems and procedures allows scope for less visceral probability evaluations, which are thus less vulnerable to the biasing influence of the availability heuristic. Law-making officials who contemplate occupational safety and health requirements, or compulsory insurance schemes for certain work environments, will be expected, and will normally have the time and resources, to initiate a methodical risk assessment that relies on statistical evidence[67].

2.4. *Intertemporal choice and hyperbolic discounting*

The fourth type of bias worth considering is the tendency to overvalue imminent rewards at the expense of long-term rewards, also known in behavioural science as *myopic* or *hyperbolic discounting*[68]. Daily life manifestations of this tendency are ubiquitous. «Imagine», as Richard Herrnstein writes, «that we could always select meals for tomorrow, rather than for right now. Would we not all eat better than we do? We may find it possible to forgo tomorrow's chocolate cake or second helping of pasta or third martini»[69]. When it comes to the meal at hand, however, this becomes harder, and often the temptation of such instant gratifications prevails over concerns for our health and figures. As Hume put it:

> «In reflecting on any action, which I am to perform a twelve-month hence, I always resolve to prefer the greater good ... But on my nearer approach ... [a] new inclination to the present good springs up, and makes it difficult for me to adhere inflexibly to my first purpose and resolution. This natural infirmity I may very much regret, and I may endeavour, by all possible means, to free myself from it»[70].

[64] This is not to suggest that people could or should generally dispose of the availability heuristic. The point, instead, is that this heuristic should not be settled for in specific situations where (1) it predictably produces errors that are consequential or common enough (or both), and (2) more accurate assessment methods can be relied upon with comparatively small costs and minimal adverse effects.

[65] Of course, there are additional sources of difficulty in the foregoing situation. Importantly, the employee is often less able than is the employer to finance measures such as safety adjustments to the work environment and an adequate insurance coverage, or to insist that such measures be taken. And the employer's interest in taking such measures is often not equal to that of the employee.

[66] There are also cases in which availability produces the opposite effect. For example, a recent experience of an accident, or overly dramatic news reports about a certain peril, may lead to exaggerated fears and unduly deter people from engaging in a normal activity (SUNSTEIN 2002, 33-35, 50-52, 78-98).

[67] The availability heuristic is not the only source of errors in people's intuitive perception of risks and probabilities. For a discussion of related error-producing cognitive and emotional phenomena, see SUNSTEIN 2002, 28-49. See also GUTTEL & HAREL 2005.

[68] See, e.g., AINSLIE 1975; HERRNSTEIN 1990; WINSTON & WOODBURY 1991; KIRBY & HERRNSTEIN 1995; KIRBY 1997.

[69] HERRNSTEIN 1990, 359.

[70] HUME 2007 [1740], 343-344 [SB536].

This tendency has been tested empirically with quantifiable rewards, such as monetary payments. In this context, the mere observation that people discount value from delayed monetary payments is not regarded as evidence of irrational tendencies[71]. For this behaviour is sometimes justifiable by factors like the risk of future frustration or default (due to the debtor's death, bankruptcy, forgetfulness, etc.) and the prospect of interest gains on capital in one's possession[72]. However, the fashion in which people discount confirms that irrational tendencies are also at work here. Experiments have shown that the fashion of discounting usually approximates hyperbolic curves[73], which is to say that the rate of discounting diminishes with time[74]. By way of illustration, hyperbolic discounters situated at a given point in time (t) discount more for the time interval between t and t + 2 than for the time interval between t + 2 and t + 4, and more for the time interval between t + 2 and t + 4 than for the time interval between t + 4 and t + 6, although all these intervals are of equal length. This manner of discounting results in incidents of choice inconsistency such as this: people who, at time point t, make a rational choice between future rewards—e.g. opting for some (sufficiently) larger reward payable at t + 6 instead of a smaller reward payable at t + 4—would often fail to make such a choice when they are at, say, t + 3[75]. Such experimental findings help explain how, when we express a preference or form an intention as to a future course of action, temptations pulling in an opposite direction may become harder to resist as the relevant occasion approaches, and we sometimes find ourselves departing from what we earlier held to be, and what may indeed be, the optimal course of action[76].

Hyperbolic discounting tendencies, and the resulting phenomenon of preference reversal, have key pertinence for law's bias-counteracting role. Private decision-making in the course of everyday activity, as was noted earlier, is to a large extent particularistic, in the sense that the actor often decides about how to act here and now. Decisions of legislative or regulatory authorities are structurally different in that here it is not the immediate actor but someone else who makes the decision, and who does so prospectively. These differences render the latter decisions structurally less prone than the former to hyperbolic discounting. And, in turn, they help explain the role played by rules of law, with their relative persistence through time, in countering this bias. This role is performed, in different ways and to different degrees, by a large variety of legal rules, but, for illustrative purposes, one domain in which it is particularly evident can be mentioned: pension and social security policy. Human patterns of hyperbolic discounting suggest that many young adults tend to underestimate the importance of saving money for old age, as it comes at the expense of presently available payoffs. Policymakers are placed at a vantage point that allows them more easily to appreciate the full picture of people's changing earning capacities through a lifetime and their old age needs, and to formulate adequate responses, this being part of the reason why in many different jurisdictions pension matters are not left wholly unregulated and certain funds for retirement are insured by statutory social security schemes.

In this and the previous subsections, the operation of four types of bias—i.e. self-enhancement, self-serving, availability, and myopic discounting bias—has been discussed separately. It should be noted, however, that law's bias-countering function is often at work in situations where more than one of these biases operate simultaneously in a mutually reinforcing manner. Take, for

[71] HERRNSTEIN 1990, 358; KIRBY 1997, 54-55 and fn. 2.
[72] KIRBY 1997, 55 fn. 2.
[73] KIRBY 1997, 59-68; AINSLIE 1975; HERRNSTEIN 1990, 359; KIRBY & HERRNSTEIN 1995.
[74] HERRNSTEIN 1990, 360; KIRBY & HERRNSTEIN 1995, 83-84; KIRBY 1997, 54-55.
[75] See KIRBY & HERRNSTEIN 1995 (reporting a series of experiments in which individuals were presented with choices of the above form) and HERRNSTEIN 1990, 358 (discussing a similar example). See also SOLNICK et al. 1980; AINSLIE & HAENDEL 1983; MILLAR & NAVARICK 1984, 213-217.
[76] KIRBY 1997, 54-55; KIRBY & HERRNSTEIN 1995, 83.

example, France's recent anti-waste laws that limit the commercial use of plastic packaging and single-use tableware. The practical decisions to which these laws apply—i.e. decisions of whether to stop using packaging and tableware that are quite possibly more convenient and cheaper in the short term, but are less environmentally sustainable—are prone to be affected by both self-serving bias and myopic bias[77]. In addition, since the effects of climate change (e.g. global warming) occur incrementally and are not *equally* felt throughout the world, availability-related bias is also likely to be at work here[78]. Examples of this sort can be multiplied.

3. *Responding to further doubts: law-generated errors; cognitive self-debiasing*

Earlier on I considered relevant doubts arising from the corrupting capacity of power and from the extent of social injustice that legal and political systems harbour[79]. I now wish to consider two further sources of doubt. The first is the possibility of incidental errors generated by some of the same structural attributes that I associated with law's bias-counteracting capacity. I am referring here especially to the fact that legal rules tend to be comprised of relatively coarse-grained categories with limited revisability in response to particularistic features of individual cases. For ease of reference, this might be called the generic character of legal rules[80]. Their generic character renders them a somewhat blunt instrument. It means, in other words, that legal rules are to some degree both over- and under-inclusive in relation to their underlying justifications—which is to say that, on some occasions of their application, legal rules yield suboptimal outcomes[81]. Now, if the generic character of legal rules can, on the one hand, work to avert errors in the ways highlighted in Section 2, but, on the other hand, generates suboptimalities in the way just noted, one might wonder if its overall effect is positive.

I will point out three factors that, I believe, address this query. The first is that rules produced by reasonably competent legislators and draftspersons are likely to be couched in a manner that moderates their *degree* of over- and under-inclusiveness. This can be attained by several techniques, such as linguistic choices that introduce *some* degree of granularity into the rule; considered selection between terminological alternatives that vary in scope, vagueness, and elasticity; the provision of legal definitions for terms used in the rule; the incorporation of exceptions; or a combination of several such techniques[82]. Second, if a given aspect of human activity simply does not lend itself to prescriptive generalizations that approximately match underlying justifications—not even with the aid of the above draftsmanship techniques—then this could well be a compelling reason to avoid using legal rules *to that extent*. And the third and final factor is judicial interpretation. Judicial interpretation can bring further refinement to statutory provisions, and thereby limit their over- and under-inclusiveness—with an effect that, insofar as the court decision binds as precedent, extends to the future[83]. The foregoing three

[77] On hyperbolic discounting in the environmental context, see, e.g., VISCUSI et al. 2008; HARDISTY, WEBER, 2009; MEYER 2013; KAPLAN et al. 2014; BERRY 2017; SARGISSON & SCHÖNER 2020.

[78] Furthermore, there is evidence that the better-than-average effect is also involved in assessment of environmental behaviour. See, e.g., LEVISTON & UREN 2020. And finally, error-inducing factors other than biases, such sheer deficiency of knowledge, may also be involved in this context.

[79] In Section 2.2.

[80] These attributes may be evocative of Frederick Schauer's characterization of rules as *entrenched generalizations* (SCHAUER 1991, 47-52).

[81] See in this regard, e.g., SCHAUER 1991, 31-34; ALEXANDER & SHERWIN 2001, 34-36.

[82] My reference to reasonable competence here reinvokes an earlier-mentioned prerequisite for the fulfilment of law's bias-counteracting function (see Section 1 above).

[83] This point is compatible with my argument in GUR 2018, 57-67—for my argument there was not that the scope of legal rules cannot be delineated through interpretation, but rather that in some of the situations considered there

factors—i.e. reasonably competent draftsmanship, attention to the generalizability of the matter, and judicial interpretation of statutory enactments—do not eliminate the phenomenon of over- and under-inclusiveness altogether. However, the limited degree to which over- and under-inclusiveness persist when these factors are in play is likely to be a price worth paying for the potential benefits of law's generic character[84].

A second type of doubt about my argument might arise from the possibility of self-debiasing. If self-debiasing is possible widely and to a full or nearly full degree, one might wonder if it does not render the bias-counteracting function of law redundant. To place this issue in a somewhat sharper focus, suppose a society-wide educational scheme about biases that would render at least the bulk of people adequately informed about the foregoing biases. And, for the moment, let us set aside questions about the cost and feasibility of such a scheme[85]. Cannot people with such information detect and correct for the above biases in their own practical reasoning simply by means of their cognitive capacities, namely through self-reflection? The answer to this is a *qualified* "yes", and the attendant qualification has key significance for present purposes. Let me explain. It is possible that information about the above biases can improve people's ability to recognize their own biases in retrospect or even reduce their effect in the future. However, this possible improvement is, first, not easy to attain or sustain even assuming adequately informed individuals and, second, likely to be limited—namely, it is doubtful that people generally can eliminate those biases altogether, or to anywhere near that extent, by means of knowledge and reasoning alone[86]. Why so? Mainly because of the following factors.

First, recall the partly-cognitive-and-partly-motivational character of some of the biases involved[87]. As described at the outset[88], this dual character means that, while these biases manifest themselves in the form of systematic cognitive error, their precursors are partly motivational. Thus, for example, part of what underlies hyperbolic discounting tendencies is the motivational lure that some gratifications obtain through their imminence and our frequently felt motivational difficulty to make present or proximate sacrifices in pursuit of remote goods. Such motivational factors have a persistent manner of regaining access to our reasoning in the form of biases—which is part of what makes biases hard to defend against by cognitive means alone.

Second, in some of the situations considered in the previous section, the *same* type of bias that tends to influence the actor's judgement of the action on its merits is liable to compromise his or her assessment of whether, and to what extent, he or she is biased. Take, for example, an actor's judgement of what and how much he or she ought to do in the way of upholding certain common goods, in terms of restricting his or her own action or shouldering positive burdens. The actor is supposed not simply and only to judge the matter by reference to substantive

it cannot be delineated by means that are consistent with Raz's pre-emption thesis, i.e. without recourse to the balance of first-order reasons.

[84] I should, perhaps, add another qualification to the above proposition. It is likelier to hold true insofar as we treat the normative force of law as *overridable* by highly compelling contrary reasons that might crop up in particular contingencies. The overridability contained in this attitude to law—which I have advocated elsewhere (GUR 2018, chs. 7-9) and will touch upon in Section 4 below—is a type of safety valve, so to speak, apt to avert the harshest potential consequences of law's relative bluntness. On overridability, see also SCHAUER 1991, 113-118, 196-206. For a relevant discussion that foregrounds the psychological aspects of rule-application and reconsideration, see BRIGAGLIA & CELANO 2017.

[85] I am not certain as to whether setting up and operating a scheme of this type and scale is a realistic possibility.

[86] Similarly, though somewhat more categorically, Hume writes about an attempt to free oneself from one's «inclination to the present»: «I may have recourse to study and reflection within myself; to the advice of friends; to frequent meditation, and repeated resolution: And having experience'd how ineffectual all these are, I may embrace with pleasure any other expedient, by which I may impose a restraint upon myself, and guard against this weakness» (HUME 2007 [1740], 344 [SB536-SB537]).

[87] See citations in fn. 12.

[88] Body text accompanying fn. 12.

factors that bear directly on its merits, but also to take into account his or her own epistemic limitations, including the possibility and extent of his or her self-serving bias. But the actor's assessment of such epistemic factors is itself an assessment that bears on his or her own interest in this case, because it may tip the balance in his or her deliberation for or against adherence to the contemplated restrictions and burdens—so it, too, may be affected by a self-serving bias.

Third, and more generally, it is part of the nature of biases that they tend to operate at an unconscious or not fully conscious level. They tend to work, in other words, below our cognitive radar—colouring our perception of things while leaving us under the impression that we see things as they are. And, thus, in situations where individuals are affected by biases, they tend not to recognize the fact that, or the degree to which, they are thus affected. This much is not only implied by the very notion of bias; it is also a pattern that finds empirical support in relevant experimental studies, where the tendency for people to overlook their own biases or underestimate their extent has been borne out and given the name the *bias blind spot*[89].

4. *Implications*

Finally, what are the implications of my claims heretofore? Rather than enumerating all the implications, I will only flag up some of the principal ones, which I divide into two types: implications for law-making and implications for legal normativity.

4.1. *Implications for law-making*

The first type of implication is internal to a legal system, in the sense that it pertains directly to the work of law-makers and their auxiliaries (e.g. policy advisors and draftspersons). That is, the notion that law is structurally suited to counteract certain common biases importantly informs (even if it does not conclusively or exclusively determine) the answer to questions such as what sort of social problems law is apt to solve and, in the light of this, when it is desirable to introduce legal regulation. The answer to these questions obviously has direct bearing on the content of law. This general point, however, should be supplemented by a couple of comments. First, some notes of caution. When seeking to use law as a bias-counteracting device, law-makers should exercise vigilance and restraint. They should refrain from overbearingly extensive or indiscriminate recourse to regulatory action. In particular, they should bear in mind that the relevant structural advantage of law pertains to *specific* biases and that only *some* instantiations thereof ultimately merit legal intervention, not least because, for all its potential benefits, recourse to the legal apparatus also carries notable risks and adverse effects[90].

Furthermore, law-makers should be conscious of the *variety* of counter-bias regulatory tools at their disposal. Rather than thinking exclusively of rules that mandate outcomes and limit choice, they should be alive also to other regulatory means which are often less intrusive or restrictive and are sometimes more cost-effective. Notably, these include what might be described as light-touch regulatory tools expounded by scholars such as Cass Sunstein, Christine Jolls, and Richard Thaler—which, instead of removing bias-induced outcomes from the range of legally permissible options, redesign the decision-making environment in (not

[89] PRONIN et al. 2002; PRONIN et al. 2004; EHRLINGER et al. 2005; MCPHERSON-FRANTZ 2006. See also FRIEDRICH 1996. Note that the observation is not that people never acknowledge their biases or always underestimate the extent of those biases. Rather, what has been observed is a general tendency: a pattern that characterizes most people's self-perception in most of the cases where they show biases.

[90] As noted in one of my preliminary comments in Section 1.

incentive-based) ways that contribute to better decisions[91]. While in some key instances of law's counter-bias operation the latter tools cannot substitute for the more traditional forms of regulation, they are nonetheless an important part of the regulatory toolkit.

Another, related comment concerns legal paternalism[92]. Does law's bias-counteracting function necessarily involve paternalism? The answer is negative. The exact definition offered for paternalism varies between theorists, but for present purposes paternalism can be broadly described as interference with a person, against their will, that is motivated or defended by «a claim that the person interfered with will be better off or protected from harm»[93]. Now, there *can* be instances of law's bias-counteracting operation that correspond with the above description. The use of law in such instances should be assessed, inter alia, through the prism of normative arguments for and against paternalism[94]. But many instances of law's bias-counteracting operation are *not* subsumable under the above description. To see this, it suffices to identify the intended class of persons protected by the relevant laws. This class includes persons other than the actor whose behaviour is constrained by a given application of these laws. This is the case, for instance, regarding many traffic laws, environmental laws, child welfare laws, building safety regulations, financial regulations, and product safety laws, to list but a few examples. Some of these instances featured in my illustrations in Section 2, which need not be reproduced at this point. To conclude the present comment, then, law's bias-counteracting function should not be seen as bound up with paternalism.

4.2. *Implications for legal normativity*

The second type of implication relates to the interface between law and its subjects[95]. It pertains, more specifically, to two—interconnected and partly overlapping—questions. First, what mode of practical reasoning should legal subjects use when faced with legal requirements? And, second, what kind of reasons can law constitute or generate? I will comment on the relation between these two questions shortly. But let me first point out what implications my observations have for the first question. The fact that one of law's key functions is to counteract biases strongly militates against subjects' recourse to a particularistic mode of reasoning whereby they act simply on weight assessments of the reasons for and against compliance with the law as applicable to the case at hand. For, as I have explicated at greater length elsewhere[96], this mode of reasoning is fully exposed and highly susceptible to the biases highlighted in Section 2—biases that are often part of what made it necessary and justified to invoke legal forms of regulation in the first place. Since action through this mode of reasoning requires assessments by the actor of whether, and to what extent, the benefits of the rule apply to the particularities of each situation of his or her daily activity, this mode of reasoning clearly runs the risk of eliciting, instead of restraining, human propensities such as short-sightedness and self-favouritism. So this mode of reasoning with legal rules is, in effect, likely to hinder the actor's prospect of making optimal use of the potential benefits of legal conduct-guidance. And if we zoom out from the mode of decision-making of an individual agent and consider the dynamic of interacting agents, the possibility of an even worse

[91] JOLLS & SUNSTEIN 2006 (advocating some such irrationality offsetting strategies under the label *debiasing through law*); THALER & SUNSTEIN 2021 (advancing choice environment architecture methods of this type under the label *nudge*).

[92] I thank an anonymous referee for bringing to my attention the need to clarify how my argument relates to paternalism.

[93] DWORKIN 2020.

[94] A discussion of which is not possible within the scope of this paper. For relevant surveys, see, e.g., DWORKIN 2020; HUSAK 2003. See also DWORKIN 1972; SARTORIUS 1983; FEINBERG 1989; GOODIN 2002; ENOCH 2016.

[95] By the word "subjects" in the above sentence, I mean non-legal actors to whom the law is addressed.

[96] GUR 2018, 121-127, 131, 163.

state of affairs comes into view. For each agent's recognition that other agents are less than disposed to comply with legal requirements may itself operate as an impetus to non-compliant behaviour, which would, in turn, further deepen and reinforce a disinclination to comply[97]. It thus seems highly questionable whether orderly social life and the essential goods that hinge on it would be attainable if the above mode of reasoning were to dominate the interface between law and its addressees[98].

Elsewhere I have set out what I believe to be a desirable mode of reasoning to be employed by subjects of a (reasonably just and well-functioning)[99] legal system[100]. Their mode of reasoning, I have argued, should be shaped by a relatively settled (but overridable) disposition to comply with the system's requirements. I have elaborated there on the properties of the envisaged disposition and how its associated mode of reasoning differs from the particularistic mode of reasoning disapproved above. Here I will only briefly mention one set of properties that makes the disposition auspicious to law's bias-counteracting function. The disposition I am referring to has a degree of deep-seatedness in the actor's attitudinal profile. As a relatively embedded disposition, it enjoys, in turn, motivational persistence or, in slightly more figurative terms, motivational stickiness or grip. This means that—although it is overridable by other motivational forces—its *activation* is not conditional on an assessment of whether the reasons for its adoption extend to the case at hand or on an assessment of the merits of the legally prescribed action. In other words, even if the disposition does not always "win" in the sense of determining the agent's eventual action, it makes its motivational force felt in a manner not conditional on reasons for action applicable to a particular situation or directive, and on how the agent assesses them. That is why a disposition of this sort has limited susceptibility to the biases discussed above, and deliberation shaped by this disposition is apt to facilitate law's bias-counteracting function.

As noted above, the preceding question—concerning modes of practical reasoning—intimately connects to, and partly overlaps with, a second question—namely, the question of what kind of reasons law can constitute or generate. What is the connection and overlap between these two questions? The answer lies in the following fact. The range of practically relevant reasons generated by law includes, apart from reasons for or against an action, reasons that pertain directly to the choice between alternative modes of reasoning: if you like, reasons about how to reason[101]. Thus, for instance, suppose that certain attributes of law mean that, when a legal system passes a certain quality threshold, reasoning with its directives through mode of reasoning X tends to be more conducive to correct decisions than using alternative modes of reasoning in response to its directives. If so, the existence of such a legal system is a reason for its subjects to employ mode of reasoning X, rather than alternative modes of reasoning. A similar inference can be made from my foregoing argument about the relationship between law's bias-counteracting function and a disposition to comply with the law. The inferential move, in a very rough and abridged form, is this: since a legal system's capacity to fulfil its bias-counteracting function (and thereby optimize our action) depends for its realization on there being a disposition to comply with the system's requirements, the existence

[97] In a somewhat similar vein, Hume notes: «You are ... naturally carried to commit acts of injustice as well as me. Your example both pushes me forward in this way by imitation, and also affords me a new reason for any breach of equity, by shewing me, that I should be the cully of my integrity, if I alone shou'd impose on myself a severe restraint amidst the licentiousness of others» (HUME 2007 [1740], 343 [SB535]).

[98] In this context, I have also considered and rejected the thought that legal sanctions and law-independent moral dispositions could suffice to adequately prevent or solve the problem envisaged above (GUR 2018, 100, 170-178).

[99] I explain more specifically what is meant by the above parenthetical prerequisite in GUR 2018, 135-136, 138, and 179.

[100] That mode of reasoning features in what I have called the *dispositional model* (GUR 2018, chs. 7-9), but it is not synonymous with the dispositional model and does not fully encompass the set of claims put forward in this model.

[101] Or reasons that have a necessary entailment for the question of how to reason.

of a legal system that meets all other prerequisites for bias-counteracting[102] is a reason to adopt a disposition to comply with its requirements. The latter proposition—concerning a reason to adopt a disposition to comply with the law—is a central element of a normative model I have advocated elsewhere under the label the *dispositional model*[103].

As a final note, it is worth stressing exactly why the normative implication of biases, as I see it, is *distinctive*. The explanation lies partly with the previously indicated fact that biases are characteristically hard to self-detect and to correct for through cognitive means alone. As was pointed out, biases tend to operate below our cognitive radar and often exercise a type of persistence in finding their way into our judgement. These characteristics of biases reflect, in turn, on the set of tools needed to effectively solve problems of bias. They mean that the required set must include, inter alia, tools from outside the cognitive toolbox. What is needed, in other words, is a mental mechanism that is not confined to knowledge and reasoning alone, and that enjoys a type of inherent persistence capable of countering the tenacity of biases. Before this line of thought is continued, it should be linked to a more specific aspect of biases. As highlighted earlier, some of the relevant biases have a partly-cognitive-and-partly-motivational character—in the sense that they have cognitive effects but partly motivational roots (as clearly exemplified by the self-interest motivational roots of a self-serving bias). The partly motivational roots of such biases call for a means of response that itself holds some motivational purchase. In other words, since the *problem* is partly motivational, its *solution* must be one that incorporates a motivationally resistant mechanism.

The foregoing specifications can all be found in a law-abiding attitude that I have characterized elsewhere[104], and that consists, inter alia, of a (relatively settled, but overridable) disposition to comply with the law, as partially described above. And this, as I have further argued, leads to the idea of reasons vis-à-vis attitudes and, more specifically, reasons to adopt a disposition to comply with the law. Here then lies part of the distinctive normative significance of law's structural ability to counter biases: it points to the above-stated nexus between reasons and attitudes. By so doing, it supports an approach to legal normativity that might be dubbed *attitudinal rationalism* and that I consider to be a helpful alternative to a purely rationalist normative framework.

[102] As pointed out in Section 1 above, although law's bias-counteracting function owes its existence primarily to structural attributes of law, this function is not likely to be fulfilled when the system in question fails to pass a threshold of reasonable quality.

[103] GUR 2018, chs. 7-9.

[104] GUR 2018, chs. 7-9.

References

ABEL R.L. 1998. *Torts*, in KAIRYS D. (ed.), *The Politics of Law: A Progressive Critique* (3rd ed.), Basic Books, 445 ff.

AINSLIE G. 1975. *Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control*, in «Psychological Bulletin», 82, 463 ff.

AINSLIE G., HAENDEL V. 1983. *The Motives of the Will*, in GOTTHEIL E., DRULEY K.A., SKOLADA T.E., WAXMAN H.M. (eds.), *Etiologic Aspects of Alcohol and Drug Abuse*, Charles C Thomas, 119 ff.

ALBISTON, C., CORRELL, S. 2023. *Law's Normative Influence on Gender Schemas: An Experimental Study on Counteracting Workplace Bias against Mothers and Caregivers*, in «Law & Social Inquiry», 1 ff. Available for early online view at: doi:10.1017/lsi.2022.102 (accessed 31 July 2023).

ALEXANDER L., SHERWIN E. 2001. *The Rule of Rules: Morality, Rules, and the Dilemmas of Law*, Duke University Press.

ALICKE M.D., KLOTZ M.L., BREITENBECHER D.L., YURAK T.J., VREDENBURG D.S. 1995. *Personal Contact, Individuation, and the Better-Than-Average Effect*, in «Journal of Personality and Social Psychology», 68, 804 ff.

ARISTOTLE. 1984. *Nicomachean Ethics*, in BARNES J. (ed.), *The Complete Works of Aristotle*, Princeton University Press, 1729 ff.

BABCOCK L., LOEWENSTEIN G. 1997. *Explaining Bargaining Impasse: The Role of Self-Serving Biases*, in «Journal of Economic Perspectives», 11, 109 ff.

BALCETIS E., DUNNING D. 2006. *See What You Want to See: The Impact of Motivational States on Visual Perception*, in «Journal of Personality and Social Psychology», 91, 612 ff.

BAUMHART R. 1968. *An Honest Profit: What Businessmen Say About Ethics in Business*, Holt, Rinehart and Winston.

BELL D.A. 2008. *Race, Racism and American Law* (6th ed.), Aspen Publishers.

BERLIN I. 1969. *Four Essays on Liberty*, Oxford University Press.

BERRY M.S., NICKERSON N.P., ODUM A.L. 2017. *Delay Discounting as an Index of Sustainable Behaviour: Devaluation of Future Air Quality and Implications for Public Health,* in «International Journal of Environmental Research and Public Health», 14, 997 ff.

BRIGAGLIA M., CELANO B. 2017. *Reasons, Rules, Exceptions: Towards a Psychological Account*, in «Analisi e Diritto», 14, 131 ff.

BROWN J.D. 1986. *Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments,* in «Social Cognition», 4, 353 ff.

CRENSHAW K., GOTANDA N., PELLER G., THOMAS K. (eds.) 1995. *Critical Race Theory: The Key Writings That Formed the Movement*, The New Press.

CROSS K.P. 1977. *Not Can, But Will College Teaching Be Improved,* in «New Directions for Higher Education», 17, 1 ff.

DALBERG-ACTON, J.E.E. 1949. *Essays on Freedom and Power*, The Free Press (Selected, and with an introduction by G. Himmelfarb).

DAWSON E., GILOVICH T., REGAN D.T. 2002. *Motivated Reasoning and Performance in the Wason Selection Task,* in «Personality and Social Psychology Bulletin», 28, 1379 ff.

DeCelles K.A., DeRue D.S., Margolis J.D., Ceranic T.L. 2012. *Does Power Corrupt or Enable? When and Why Power facilitates Self-Interested Behavior*, in «Journal of Applied Psychology», 97, 681 ff.

DELGADO R., STEFANCIC J. (eds.) 2013. *Critical Race Theory: The Cutting Edge* (3rd ed.), Temple University Press.

DELHOMME P. 1991. *Comparing One's Driving with Others: Assessment of Abilities and Frequency of Offences. Evidence for a Superior Conformity of Self-Bias?*, in «Accident Analysis and Prevention», 23, 493 ff.

DICEY A.V. 1914. *Lectures on the Relation Between Law and Public Opinion in England During the Nineteenth Century* (2nd ed.), Macmillan.

DITTO P.H., LOPEZ D.F. 1992. *Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions*, in «Journal of Personality and Social Psychology», 63, 568 ff.

DOGAN E., STEG L., DELHOMME P., ROTHENGATTER T. 2012. *The Effects of Non-Evaluative Feedback on Drivers' Self-Evaluation and Performance*, in «Accident Analysis and Prevention», 45, 522 ff.

DUNNING D. 2001. *On the Motives Underlying Social Cognition*, in SCHWARZ N., TESSER A. (eds.), *Blackwell Handbook of Social Psychology*, Vol. 1, Blackwell, 348 ff.

DUNNING D. 2015. *Motivated Cognition in Self and Social Thought*, in MIKULINCER M., SHAVER P.R. (eds.), *APA Handbook of Personality and Social Psychology*, Vol. 1, American Psychological Association, 777 ff.

DUNNING D., JOHNSON K., EHRLINGER J., KRUGER J. 2003. *Why People Fail to Recognize Their Own Incompetence,* in «Current Directions in Psychological Science», 12, 83 ff.

DUNNING D., MEYEROWITZ J.A., HOLZBERG A.D. 1989. *Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability*, in «Journal of Personality and Social Psychology», 57, 1082 ff.

DWORKIN G. 1972. *Paternalism*, in «The Monist», 56, 64 ff.

DWORKIN G. 2020. *Paternalism*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). Available at: https://plato.stanford.edu/entries/paternalism/.

EHRLINGER J, GILOVICH T, ROSS L. 2005. *Peering Into the Bias Blind Spot: People's Assessment of Bias in Themselves and Others*, in «Personality and Social Psychology Bulletin», 31, 680 ff.

ENOCH D. 2016. *What's Wrong with Paternalism: Autonomy, Belief, and Action*, in «Proceedings of the Aristotelian Society», 116, 21 ff.

FARNSWORTH W. 2003, *The Legal Regulation of Self-Serving Bias*, in «UC Davis Law Review», 37, 567 ff.

FEINBERG J. 1989. *The Moral Limits of the Criminal Law: Vol. 3: Harm to Self*, Oxford University Press.

FRIEDRICH J. 1996. *On Seeing Oneself as Less Self-Serving Than Others: The Ultimate Self-Serving Bias?*, in «Teaching of Psychology», 23, 107 ff.

FULLER L.L. 1969. *The Morality of Law* (2nd ed.), Yale University Press.

GOODIN R.E. 2002. *Permissible Paternalism: Saving Smokers from Themselves* (2nd ed.), in LAFOLLETTE H. (ed.), *Ethics in Practice: An Anthology*, Blackwell, 307 ff.

GREEN L. 2008. *Positivism and the Inseparability of Law and Morals*, in «New York University Law Review», 83, 1035 ff.

GUR N. 2018. *Legal Directives and Practical Reasons*, Oxford University Press.

GUTTEL E., HAREL A. 2005. *Matching Probabilities: The Behavioral Law and Economics of Repeated Behavior*, in «University of Chicago Law Review», 72, 1197 ff.

HARDIN G. 1968. *The Tragedy of the Commons*, in «Science», 162, 1243 ff.

HARDISTY D.J., WEBER E.U. 2009. *Discounting Future Green: Money Versus the Environment*, in «Journal of Experimental Psychology: General», 138, 329 ff.

HART H.L.A. 2012. *The Concept of Law* (3rd ed.), Oxford University Press, (ed. by L. Green, first published 1961).

HERRNSTEIN R.J. 1990. *Rational Choice Theory: Necessary but Not Sufficient*, in «American Psychologist», 45, 356 ff.

HOBBES T. 1991a. *De Cive (The Citizen)*, in GERT B. (ed.), *Thomas Hobbes, Man and Citizen*, Hackett Publishing Company, 87 ff. (first published 1642).

HOBBES T. 1991b. *Leviathan*, Cambridge University Press (ed. by R. Tuck, first published 1651).

HORSWILL M.S., WAYLEN A.E., TOFIELD M.I. 2004. *Drivers' Ratings of Different Components of Their Own Driving Skill: A Greater Illusion of Superiority for Skills that Relate to Accident Involvement*, in «Journal of Applied Social Psychology», 34, 177 ff.

HORWITZ M.J. 1977. *The Transformation of American Law, 1780-1860*, Harvard University Press.

Hume D. 2007. *A Treatise of Human Nature*, Clarendon Press (ed. by F.T. Norton and M.J. Norton, first published 1740).

HUNTER R. 2012. *The Power of Feminist Judgments?*, in «Feminist Legal Studies», 20, 135 ff.

HUSAK D. 2003. *Legal Paternalism*, in LAFOLLETTE H. (ed.), *The Oxford Handbook of Practical Ethics*, Oxford University Press, 387 ff.

JOLLS C., SUNSTEIN C.R. 2006. *Debiasing through Law*, in «The Journal of Legal Studies», 35, 199 ff.

KAPLAN B.A., REED D.D., MCKERCHAR T.L. 2014. *Using a Visual Analogue Scale to Assess Delay, Social, and Probability Discounting of an Environmental Loss*, in «The Psychological Record», 64, 261 ff.

KIRBY K.N. 1997. *Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards*, in «Journal of Experimental Psychology: General», 126, 54 ff.

KIRBY K.N., HERRNSTEIN R.J. 1995. *Preference Reversals due to Myopic Discounting of Delayed Reward*, in «Psychological Science», 6, 83 ff.

KRUGER J. 1999. *Lake Wobegon Be Gone! The "Below-Average Effect" and the Egocentric Nature of Comparative Ability Judgments*, in «Journal of Personality and Social Psychology», 77, 221 ff.

KRUGER J., DUNNING D. 1999. *Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessment*, in «Journal of Personality and Social Psychology», 77, 1121 ff.

KUNDA Z. 1990. *The Case for Motivated Reasoning*, in «Psychological Bulletin», 108, 480 ff.

LARWOOD L., WHITTAKER W. 1977. *Managerial Myopia: Self-Serving Biases in Organizational Planning*, in «Journal of Applied Psychology», 62, 194 ff.

LEVISTON Z., UREN H.V. 2020. *Overestimating One's "Green" Behavior: Better-Than-Average Bias May Function to Reduce Perceived Personal Threat from Climate Change*, in «Journal of Social Issues», 76, 70 ff.

LICHTENSTEIN S., SLOVIC P., FISCHHOFF B., LAYMAN M., COMBS B. 1978. *Judged Frequency of Lethal Events*, in «Journal of Experimental Psychology: Human Learning and Memory», 4, 551 ff.

LLOYD, W.F. 1883. *Two Lectures on the Checks to Population*, Oxford University Press (available also in *W. F. Lloyd on the Checks to Population*, in «Population and Development Review», 6, 473 ff).

LOCKE J. 2003. *The Second Treatise of Civil Government*, in WOOTTON D. (ed.), *John Locke, Political Writings*, Hackett Publishing Company, 261 ff. (first published 1689).

Ly D.P. 2021. *The Influence of the Availability Heuristic on Physicians in the Emergency Department*, in «Annals of Emergency Medicine», 78, 650 ff.

Mackinnon C. 1983. *Feminism, Marxism, Method and the State: Toward Feminist Jurisprudence*, in «Signs», 8, 635 ff.

Mackinnon C. 1987. *Feminism Unmodified: Discourses on Life and Law*, Harvard University Press.

Măirean C., Havârneanu C.E. 2018. *The Relationship Between Drivers' Illusion of Superiority, Aggressive Driving, and Self-Reported Risky Driving Behaviors*, in «Transportation Research Part F: Traffic Psychology and Behaviour», 55, 167 ff.

Mcpherson-frantz C. 2006. *I AM Being Fair: The Bias Blind Spot as a Stumbling Block to Seeing Both Sides*, in «Basic and Applied Social Psychology», 28, 157 ff.

Messick D.M., Brewer M.B. 1983. *Solving Social Dilemmas: A Review*, in Wheeler L., Shaver P. (eds.), *Review of Personality and Social Psychology*, Vol. 4, Sage Publications, 11 ff.

Messick D.M. 1985. *Social Interdependence and Decision Making*, in Wright G. (ed.), *Behavioral Decision Making*, Plenum, 87 ff.

Messick D.M., Sentis K.P. 1979. *Fairness and Preference*, in «Journal of Experimental Social Psychology», 15, 418 ff.

Meyer A. 2013. *Intertemporal Valuation of River Restoration*, in «Environmental and Resource Economics», 54, 41 ff.

Millar A., Navarick D.J. 1984. *Self-Control and Choice in Humans: Effects of Video Game Playing as a Positive Reinforcer*, in «Learning and Motivation», 15, 203 ff.

Mossman M.J. 1987. *Feminism and Legal Method: The Difference it Makes*, in «Wisconsin Women Law Journal», 3, 147 ff.

Pogarsky G., Babcock L. 2001. *Damage Caps, Motivated Anchoring, and Bargaining Impasse*, in «Journal of Legal Studies», 30, 143 ff.

Preston C.E., Harris S. 1965. *Psychology of Drivers in Traffic Accidents*, in «Journal of Applied Psychology», 49, 284 ff.

Pronin E., Gilovich T., Ross L. 2004. *Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others*, in «Psychological Review», 111, 781 ff.

Pronin E., Lin D.Y., Ross L. 2002. *The Bias Blind Spot: Perception of Bias in Self Versus Others*, in «Personality and Social Psychology Bulletin», 28, 369 ff.

Pyszczynski T., Greenberg J. 1987. *Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-Testing Model*, in Berkowitz L. (ed.), *Advances in Experimental Social Psychology*, Vol. 20, Academic Press, 297 ff.

Rawls J. 1999. *A Theory of Justice* (rev. ed.), Harvard University Press.

Raz J. 1986. *The Morality of Freedom*, Clarendon Press.

Sargisson R.J., Schöner B.V. 2020. *Hyperbolic Discounting with Environmental Outcomes across Time, Space, and Probability*, in «The Psychological Record», 70, 515 ff.

Sartorius R. (ed.) 1983. *Paternalism*, University of Minnesota Press.

Schauer F. 1991. *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Clarendon Press.

Sherman S.J., Cialdini R.B., Schwartzman D.F., Reynolds K.D. 1985. *Imagining Can Heighten or Lower the Perceived Likelihood of Contracting a Disease: The Mediating Effect of Ease of Imagery*, in «Personality and Social Psychology Bulletin», 11, 118 ff.

SLOVIC P. 2000. *The Perception of Risk*, Earthscan Publications.

SOLNICK J.V., KANNENBERG C.H., ECKERMAN D.A., WALLER M.B. 1980. *An Experimental Analysis of Impulsivity and Impulse Control in Humans*, in «Learning and Motivation», 11, 61 ff.

STEPHENS A.N., OHTSUKA K. 2014. *Cognitive Biases in Aggressive Drivers: Does Illusion of Control Drive us off the Road?*, in «Personality and Individual Differences», 68, 124 ff.

SUNSTEIN C.R. 2002. *Risk and Reason: Safety, Law and the Environment,* Cambridge University Press.

SVENSON O. 1981. *Are We All Less Risky and More Skillful Than Our Fellow Drivers?*, in «Acta Psychologica», 47, 143 ff.

TAUB N., SCHNEIDER E.M. 1998. *Women's Subordination and the Role of Law*, in KAIRYS D. (ed.), *The Politics of Law: A Progressive Critique* (3rd ed.), Basic Books, 328 ff.

THALER R.H., SUNSTEIN C.R. 2021. *Nudge: The Final Edition,* Yale University Press (original ed. 2008).

THOMPSON B.H., JR. 2000. *Tragically Difficult: The Obstacles to Governing the Commons*, in «Environmental Law», 30, 241 ff.

THOMPSON L., LOEWENSTEIN G. 1992. *Egocentric Interpretations of Fairness and Interpersonal Conflict*, in «Organizational Behavior and Human Decision Processes», 51, 176 ff.

TUSHNET M. 1978. *A Marxist Analysis of American Law*, in «Marxist Perspectives», 1, 96 ff.

TUSHNET M. 1993. *The Critique of Rights*, in «SMU Law Review», 47, 23 ff.

TVERSKY A., KAHNEMAN D. 1973. *Availability: A Heuristic for Judging Frequency and Probability*, in «Cognitive Psychology», 5, 207 ff.

TVERSKY A., KAHNEMAN D. 1974. *Judgment under Uncertainty: Heuristics and Biases*, in «Science», 185, 1124 ff.

VISCUSI W.K., HUBER J., BELL J. 2008. *Estimating Discount Rates for Environmental Quality from Utility-Based Choice Experiments*, in «Journal of Risk and Uncertainty», 37, 199 ff.

WADE-BENZONI K.A., TENBRUNSEL A.E., BAZERMAN M.H. 1996. *Egocentric Interpretations of Fairness in Asymmetric, Environmental Social Dilemmas: Explaining Harvesting Behavior and the Role of Communication*, in «Organizational Behavior and Human Decision Processes», 67, 111 ff.

WAYLEN A.E., HORSWILL M.S., ALEXANDER J.L., MCKENNA F.P. 2004. *Do Expert Drivers Have a Reduced Illusion of Superiority?*, in «Transportation Research Part F: Traffic Psychology and Behaviour», 7, 323 ff.

WINSTON G.C., WOODBURY R.G. 1991. *Myopic Discounting: Empirical Evidence,* in KAISH S., GILAD B. (eds.), *Handbook of Behavioral Economics*, Vol. 2B, JAI Press, 325 ff.

ZAMIR E. 2015. *Law, Psychology, and Morality: The Role of Loss Aversion*, Oxford University Press.

ZHAO Y., CHEN R., YABE M., HAN B., LIU P. 2021. *I Am Better Than Others: Waste Management Policies and Self-Enhancement Bias*, in «Sustainability», 13, 13257.

PART V.

Law and Emotions

Emotion in Criminal Law

MIHA HAFNER

1. *Introduction*

In the past decades, a growing body of scholarly work has boldly attempted to explore the intersection of law and emotion. While these explorations are manifold, an important strand of research inquiries into the role of emotion in legal reasoning. It has gained a strong impetus after cognitive (and other) sciences have convincingly overthrown the previously persistent dichotomy between reason and emotion, and established that the gap between them was largely illusionary. This importantly impacted the legal thought. The notion of the realm of law as entirely rational, reason-based discipline that needs to be guarded from the impure and distorting influence of uncontrollable, unpredictable, and overwhelming emotions, has been slowly transformed by the recognition that emotions are in fact part and parcel of both law and legal reasoning[1]. Scholars have acknowledged that «the law is [...] imbued with emotion» (Bandes 1999, 2), that legal reasoning cannot be fully comprehended without considering the emotion component, and that emotion and objectivity are in no way incompatible in law (Grossi 2019).

In criminal law, however, the assumption that emotions do not and should not have any citizens' rights therein was hard to defend even before the described turn. In fact, «criminal law is one of the few areas of doctrine in which an examination or assessment of emotions [...] has been a standard feature of the doctrinal and adjudicative landscape» (Abrams & Keren 2010). Nonetheless, recent scholarly work has both deepened and broadened the understanding of the multifaceted role of emotion in criminal law. This chapter briefly outlines some of the most salient research trajectories in this field. Any attempt of being exhaustive in this pursuit would turn out futile as research in the area of emotions and (criminal) law is extremely interdisciplinary and fast expanding. Moreover, many challenges remain yet unresolved. However, despite being unable to provide definite answers to these questions, it is crucial to spell them out.

The chapter begins with the problem of conceptualising emotions and further focuses on the family of moral emotions as particularly relevant for criminal law. Section 3 deals with the role of emotions in the general and legal decision making. It sketchily summarises some findings on emotions and everyday decision making and introduces more specific research on the impact of emotions in legal decision making. It also identifies and speculates on some crucial but yet not determined problems relating particularly to professional decision makers in criminal law. Section 4 revolves around the ambiguous notion of empathy and its potential when employed by the criminal law protagonists. Section 5 ventures beyond the strictly cognitive context of emotion discourse and explores the many normative traces of emotions in the valid criminal legislation. It provides many examples of emotions being basic normative elements of a multitude of criminal law norms or at least laying implicit foundations for them. The chapter concludes with drawing attention to the burgeoning field of the socio-legal research of emotion

[1] For overviews of the development in the field, see e.g., Abrams & Keren 2010; Maroney 2016.

in the criminal justice. It provides examples how scrutinising emotion from the sociological perspective can enrich this discourse with fresh and equally indispensable insights.

2. *The notion of emotion*

Delving into the intersection of criminal law and emotion may be quite daunting as different researchers approach this field presupposing different notions of emotion. As a vantage point, authors often assume psychological definitions of emotion[2], but they may also consider emotions from sociological, philosophical, neurological (see GENDRON 2021) or some other perspective. These various concepts overlap significantly, and their divergence may be simply a consequence of different theoretical, methodological, or research approaches. Moreover, the abundance of emotion-related terms circulating in literature further muddles these discussions. Expression that are sometimes used as synonyms and other times as distinct but related concepts commonly include feeling, sentiment, affect, mood, passion, temperament, affective state etc.[3] However, despite the fact that there is no generally accepted definition of emotion within a particular discipline, let alone between the fields (MULLIGAN & SCHERER 2012), there is still a strong consensus on what causes emotions, what are their elements, and what are their effects (SCARANTINO 2016, 5).

Emotions are usually elicited by an event or stimulus that is of relevance to a person. Such event may be external (e.g., a sudden loud noise) or internal (e.g., a memory or a mental representation of an event). Thus, a necessary component of an emotion is a (usually swift) appraisal process evaluating the stimulus as important to the person. Such appraisal informs the individual of the nature of the event and prepares her organism for adaptive behaviour (action tendencies). This coordinated response of the organism results in felt and often displayed physiological changes (e.g., increased heartrate, modified facial expression). Typically, emotions disturb our set behavioural course and plans by creating new goals. Therefore, emotions can have a strong impact on our interaction with the social environment (SCHERER 2005, 700-702).

As subjective experiences, emotions may seem distinctly idiosyncratic phenomena. However, psychology has established that emotions are evoked, operate, and produce effects under universal principles. Frijda calls these principles the laws of emotion (FRIJDA 2007, 1-22). They represent general rules, such as that emotions arise as a reaction to events that the individual appraises as important to her; or that in ambiguous circumstances, people tend to interpret the situation in a way that minimises negative emotional load.

Besides psychological and philosophical takes on the role of emotion in law, sociological approaches also feature strongly in the expanding literature on law and emotion. Sociological research of emotion focuses on emotions' social aspects. Sociological notion of emotion «also assumes that the emotional arousal, cognitive appraisals, expressions, and language that compose emotional experience are constrained by both culture and structure» (LIVELY & WEED 2016, 67). An important concept in the sociology of emotion are the so-called feeling rules (HOCHSCHILD 2012). These are historically and culturally dependent social norms that regulate which emotions should be felt and how they should be expressed in a particular social environment.

Both these commonalities in notions of emotion as well as the divergences among disciplines tackling this subject matter should be considered when illuminating emotions and criminal law from different perspectives.

[2] It should be noted that there is furthermore a broad diversity between different strands of psychology researching emotions. For example, neuropsychology, evolutionary psychology, social psychology, and clinical psychology, to name just a few, each focus on different aspects of these phenomena.

[3] In this chapter, the term emotion will be predominately used and definitional differences between related expressions will not be further discussed. However, when imported from other literature, other terms will be used as well.

2.1. *Moral emotions*

The entire legal realm is strongly tied to moral emotions[4], while the domain of criminal law provides a particular arable ground for this family of emotions. Thus, moral emotions deserve a brief introduction. While emotions in general are usually evoked by events and stimuli that directly affect one's self (FRIJDA & MESQUITA 1994), the orientation of moral emotions is less self-centred and more prosocial. Haidt defines moral emotions as «emotions that are linked to interests or welfare either of society as a whole or at least of persons other than the judge or agent» (HAIDT 2003, 853), whereas Prinz delineates them more generally as «emotions that arise in the context of morally relevant conduct» and that they «promote or detect conduct that violates or conforms to a moral rule» (PRINZ 2007, 68). Among different taxonomies of moral emotions stands out Haidt's categorisation to four families of prototypical moral emotions: (1) other-condemning (contempt, righteous anger, disgust), (2) self-conscious (shame, embarrassment, guilt), (3) other-suffering (compassion), and (4) other-praising (gratitude, elevation) (HAIDT 2003; see also TANGNEY et al. 2007 and PRINZ 2007).

The example of righteous anger (sometimes also referred to as indignation or moral outrage) illuminates the relevance of moral emotions for criminal law and reveals emotional and cognitive parallels to many criminal law concepts. Research has thus shown that righteous anger implies blame (QUIGLEY & TEDESCHI 1996) and provokes direct punitive response (HUTCHERSON & GROSS 2011), sometimes termed as altruistic punishment (BOYD et al. 2003; FEHR & GÄCHTER 2002). In fact, the role of moral outrage has been researched specifically in criminal law context (ASK & PINA 2011; BASTIAN et al. 2013; FEIGENSON 2016; GOLDBERG et al. 1999). These studies have confirmed that moral outrage did influence the legal decisions of the decision makers. Moreover, this research clearly illustrates how moral emotions play a central role in discussions on cognitive mechanisms underpinning legal decision making and reasoning in (criminal) law. In cognitive terms, ultimately, these mechanisms are integral to moral decision making.

3. *Emotions experienced by the decision maker*

3.1. *Dilemmas on the role of emotion in legal decision making*

The illusion of a judge as a cold and emotionally detached mouthpiece of the law has been long debunked. It comes as no surprise that jurists are as susceptible to emotions in their decision making as any other human beings. However, acknowledging the presence of emotions in the process of legal decision making opens a plethora of other more complicated questions. They relate to the overarching theme of how do, and how should, emotion integrate into legal decision making. The first set of dilemmas pertains to the relation between emotion and the so-called rational reasoning. Can—and should—these two components be detached? Does emotion inevitably obscure rationality in legal reasoning? Are there some (types of) emotions that are more desirable than the others in this process or is it the intensity or another quality of experienced emotions that is of higher relevance. Perhaps the instigator of the emotion should be also taken into account? Emotions may be more or less preferable whether incited by the alleged crime, its outcome, the perpetrator, or other stakeholders involved in the criminal trial. Finally, there are many different decisions being taken in the context of a criminal trial. Emotions may not have (equal) relevance for all of them. For example, righteous anger may

[4] See for example SAJÓ 2016 on the relationship between moral emotions on the one hand and constitutionalism and fundamental human rights on the other.

influence a sentencing decision but not the decision on criminal responsibility of the defendant. It appears that none of these questions can be resolved with a general answer. As Bandes aptly concludes «the appropriateness of particular emotions cannot be discussed apart from the context in which they appear» (BANDES 1996, 372).

In this chapter, we will often assume a figure of a judge as a prototypical legal decision maker in the criminal law context. However, *mutatis mutandis* the addressed dilemmas are applicable to other professionals in the criminal justice system, such as prosecutors, law enforcement officers, defence attorneys and other legal representatives, members of parole panels and other similar bodies. Moreover, regarding the emotional impact on legal decisions, justifiability of the demarcation between legal professionals and lay decision makers in criminal law should be considered. Do emotions shape decisions of jurors and (as is the case in many jurisdictions) lay judges in mixed panels differently than those of professional judges? This question seems particularly pertinent in light of a scarce research on professional jurists in that area, compared to more prevalent studies on mock juries.

3.2. *Emotion and general decision making*

Before venturing into the inquiries on the role of emotion in legal decision making, the place of emotion in our every-day judgments should be briefly illuminated. A useful starting point in this discussion is Kahneman's division of human cognition to System 1 and System 2 (KAHNEMAN 2011; see also KAHNEMAN & FREDRICK 2002). System 1 is quick, intuitive, effortless, and draws from emotion while System 2 is slow, analytical, determined, and effortful. System 1 is much more prevalent and more efficient in our every-day decision making. However, it is also more prone to mistakes. This is why the role of System 2 is to either re-evaluate (and correct) particular judgments of System 1, or to independently resolve more complex cognitive tasks. Therefore, in this model, emotion first acts as a powerful indicator to System 1, facilitating its quick and efficient decisions. However, emotion also indirectly proposes a decision to System 2, which may through more thorough deliberation either confirm or reject it.

Much research has built upon this model and further elaborated how emotion impacts our judgment. Various experimental work has also differentiated between modes of emotions and their temporal dimensions in the process of general decision making. In a succinct overview of this field, VÄSTFJÄLL and SLOVIC (2013, 258-266) explain that emotion may first be experienced as predecisional affect influencing decision prior to its being made. Predecisional affect may come as current mood, anticipatory emotions, or anticipated emotions. Anticipatory emotions are emotional reactions experienced at present by thinking of the future outcome. Anticipated emotions, conversely, are not felt at present, but are cognitively anticipated to occur after the decision. On the other hand, postdecisional affect represents emotion experienced when the decision has already been made.

Further distinction between emotions involved in the decision-making process concerns the cause of the experienced emotion. Thus, incidental affect is unconnected to the decision task (e.g., sadness due to received bad news), while integral affect is related to the very decision (e.g., anxiety whether one will choose the best option). This chapter does not allow for explanation on how these various types of emotions integrate into the decision-making process. These mechanisms are complex and differ according to the outlined taxonomy of affective or cognitive state, a particular experienced emotion and a particular type of decision taken. Suffice it to say that plethora of experimental evidence confirms the assertion that emotions indeed shape our judgments and decisions in our everyday life[5]. The provided framework also sets a theoretical foundation for exploring the impact of emotion in legal decision making.

[5] For a comprehensive overview see VÄSTFJÄLL & SLOVIC 2013.

3.3. *Moral emotions and moral decision making*

Recent decades have seen an increased attention in research of moral judgment and the relevance of moral emotion therein. We zoom in to this area of decision making as decisions in criminal law are inevitably intertwined with moral judgments (WEST 2020). Evading complex discussion on the relationship between morality and (criminal) law that has intrigued many legal scholars, this chapter will undertake the assumption that many criminal law norms overlap with moral norms and thus moral decisions inevitably underpin crucial criminal law decisions.

Moral judgment usually pertains to «either the moral value of an action—its being good/bad or right/wrong—or whether one should/should not or ought/ought not [to] perform it» (MAIBOM 2010, 1001). Similarly as in any decision making, the dual-process models of cognition (KAHNEMAN & FREDRICK 2002) can be also applied in the realm of moral judgment. HAIDT (2007, 998) thus contrasts moral intuition and moral reasoning:

> «Moral intuition refers to fast, automatic, and (usually) affect-laden processes in which an evaluative feeling of good-bad or like-dislike (about the actions or character of a person) appears in consciousness without any awareness of having gone through steps of search, weighing evidence, or inferring a conclusion. Moral reasoning, in contrast, is a controlled and "cooler" (less affective) process; it is conscious mental activity that consists of transforming information about people and their actions in order to reach a moral judgment or decision».

HAIDT (2007) further claims that the role of moral emotions, however, is not reduced only to moral intuition. In fact, moral emotions in his view, have a pivotal impact on moral reasoning as well by instantly suggesting the outcome of a moral judgment. In the subsequent process of moral reasoning, typically only those rational arguments are sought that buttress the initial intuitive moral judgement signalled by the affect. While in the process of rational deliberation the intuitive, affect-laden moral judgment may be corrected or changed by reasoning, this requires much cognitive effort and does not occur often.

This model of moral decision making is sometimes termed as the new sentimentalism. The term implies a sharp departure from the opposing and older paradigm of moral rationalism, which emphasises the prevalence of rational reasoning in morality (MAIBOM 2010). While the new sentimentalism is not unanimously accepted in literature on moral decision making, many scholars in the field tend to embrace one form of sentimentalism (see e.g., GREENE 2013; GREENE et al. 2008; HAIDT 2003; NICHOLS 2002; PRINZ 2007; SLOTE 2014). Moreover, even authors who do not subscribe to sentimentalist explanations of moral judgment, acknowledge that the role of emotions in human morality was underestimated in previous rationalistic approaches (BLOOM 2013; CRAIGIE 2011; GREENSPAN 2011; HUEBNER 2013; MAIBOM 2010; NUCCI 2001).

This shift to emotionalism is substantiated by research findings from many different fields. Neuroscience provides one such important area. Brain imaging experiments have revealed that brain areas generating and regulating emotions are also involved in moral reasoning (YOUNG & KOENIGS 2007). Moreover, clinical studies with patients who sustained damage to these brain structures indicate that these individuals also have impaired moral judgment (e.g., DAMASIO 2005; KOENIGS et al. 2007; MARTINS et al. 2012). Research from developmental psychology also seems congruent with sentimentalism. It indicates that very young children make intuitive moral judgments even before they develop more sophisticated cognitive mechanisms. When justifying why particular behaviour is good or bad, children draw reasons from other people's emotions (e.g., this would make them sad, angry or upset). Moreover, when very young children themselves break moral rules, they already display moral emotions (guilt, shame) (NUCCI 2001). Further support for the claim that emotions are indeed pivotal in moral judgment comes from different strands of experimental psychology. Research has thus shown that individuals' moral judgments

changed when a particular emotion was invoked in them. For example, by induced disgust, test subjects' moral judgments became harsher (HORBERG et al. 2009; WHEATLEY & HAIDT 2005), whereas with invoked mirth their moral judgments became more utilitarian (STROHMINGER et al. 2011; VALDESOLO & DESTENO 2006). The observation that people heavily rely on their moral sentiments in moral judgments is further supported by the discovered moral dumbfounding phenomenon. HAIDT (2001) coined this term for situations when people make intuitive moral stance on a certain matter (e.g., moral disgust on incest) and try to justify it with rational arguments. When their arguments are rebutted, however, they still stick to their intuitive judgment despite admitting to the fact that they cannot rationally defend their decision.

3.4. *The role of (moral) emotions in criminal law decision making*

Recently, the body of literature exploring the effect of emotion on decision making in (criminal) law has been steadily growing. This research usually builds on the previous findings on the role of emotion in general and in moral decision making, particularly on research on how people attribute blame and/or develop propensity to punish (in a non-legal meaning of the terms) (e.g., ASK & PINA 2011). However, by applying these concepts to a legal context, researchers face many hurdles specific to the domain of legal decision making. They first pertain to the fact that legal decision making is governed by a complex set of substantive and procedural legal rules, legal principles, and rules of interpretation, which legal theory usually denotes as legal reasoning. Secondly, professional legal decision makers are (through education, training, and professional socialisation) typically much aware of these rules and hence of the fact that legal reasoning is not and ought not to be done in the same fashion as every-day judgments. It would be reasonable to hypothesise that such motivated cognition, put in psychological terms, mitigates the effect of emotion in legal decision making to at least a certain degree[6]. Methodologically, however, this hypothesis is difficult to test; not at the least because in the process of legal decision making, many emotions might be experienced and many different decisions are taken[7]. Researchers thus typically focus only on selected emotions and on particular legal decisions (e.g., the role of anger on the criminal responsibility assessment). However, in such a complex system, it is methodologically perilous to draw causal conclusions from one single variable to the other (internal validity). Another critical obstacle concerns the test subjects. As neither actual jury members nor professional jurists are readily accessible for this type of experiments, researchers most often resort to mock jurors[8] as test subjects, which further obfuscates extrapolation of results to both real juries and even more so to professional decision makers (ecological validity) (see also PHALEN et al. 2021, 288-290). With these caveats in mind, one should be very careful in drawing broad conclusions from this area of research. Notwithstanding that, this research still offers intriguing insights into the relationship between emotion and legal decision making.

FEIGENSON and PARK (2006) offer a useful model on how emotion might influence legal decision making, which may help us navigate through particular findings of various studies in this field. They propose three ways in which emotion impacts the process of legal judgment. First, emotions can influence individual's strategies for processing information. Thus for example, some emotions such as anger, disgust, and happiness typically pair with a sense of a higher certainty in a decision maker, which may in turn reduce her analytical processing of legally-relevant information and increase susceptibility for heuristics-based solution (TIEDENS & LINTON 2001). Interestingly,

[6] See, e.g., LERNER & TETLOCK 1999 and DESTENO et al. 2000 on evidence that motivation may correct emotion-related biases.

[7] Cf. FEIGENSON 2016 on possible effects of multiple emotions on legal decisions.

[8] Mock juries in this kind of experiments do not consist of actual jury members, but of selected individuals (most often students) who are assigned this role in deciding a hypothetical or actual case.

this effect has been confirmed in a criminal law context with professional decision makers, namely the Swedish crime investigators. When assessing reliability of a witness, angry as opposed to sad investigators more readily employed heuristic processing of information (Ask & Granhag 2007; see also Semmler & Brewer 2002 for similar results with mock jurors).

Second, emotions may produce mood-congruency effect in legal decision makers[9]. As summarised by Feigenson and Park: «People in positive moods tend to make more positive evaluations of ambiguous information; people in negative moods tend to interpret the same information more negatively» (Feigenson & Park 2006, 148). Thus, evaluating inconclusive evidence, a prosecutor in a negative mood might perceive, interpret, and memorise more details unfavourably for the defendant compared to a prosecutor in a more cheerful mood.

Finally, emotions may affect the way people make particular decisions by providing informational cues. These mechanisms operate in variety of very distinctive and sometimes complex ways, in which emotions may affect directly or indirectly and operate as incidental or integral factors (see Feigenson & Park 2006 for an overview). In a criminal-law context, the impact of integral emotions (affects evoked by some features of the criminal case) seems of particular relevance to our discussion. A study by Bright and Goodman-Delahunty (2006), for example, showed that conviction rate in a group of mock jurors with greater anger (induced by gruesome photographic evidence) was significantly higher than in a group that was not shown any evidence. Feigenson and Park (2006, 152) assume that decision-makers in such situations use their emotional state as an informational cue concerning the judgment target (e.g., defendant's blameworthiness).

After accounting for various mechanisms through which emotions may impact legal decision-making, other variables should also be included to the picture. They mostly pertain to integral emotions and concern the questions of who or what invokes the emotion and towards whom the emotion is directed to (Salerno 2021). Thus, for example a judge may feel sympathy for the victim of the offence, and anger towards the defendant. However, she may also be compassionate towards the defendant and disappointed by the victim. A prosecutor might be repulsed by the offence itself but not necessarily by the mentally challenged defendant. These distinct factors should also be carefully considered when discussing the impact of emotions on legal decisions.

By sketching the background of variability of the mechanisms and modes through which emotions impact legal decision making in criminal law, we take an example of anger and briefly present some research looking into the effects of this emotion. Anger is usually experienced when a person deems that a responsible other has caused an event that one appraises as personally relevant, but incongruent with one's goals (Tangney et al. 2007, 361), particularly when such events are perceived as unjust or unfair (Mikula et al. 1998). Psychologists, however, distinguish anger on the one hand from righteous anger or moral outrage on the other. The latter is more common with decision makers in criminal justice. Moral outrage is typically caused not by perceived harm to the person experiencing this affect but by harm caused to someone else or by a breach of moral norms (Russell & Giner-Sorolla 2011).

Research indicates that anger in decision makers directly or indirectly increases their punitiveness towards the defendant. An interesting experiment by Ask and Pina (2011) investigated how anger impacts mock jurors' assessment of criminal intent in an embezzlement case. They found that angry jurors attributed more criminal intent to the defendant compared to neutral and sad ones[10]. Consequently, this led to jurors' greater punitiveness. Georges et al. (2013) measured how mock jurors' anger reflected in their sentencing decisions in a capital case trial. They found that the angrier the jurors were, the more likely it was for them to decide for a death sentence, as angrier jurors estimated mitigating circumstances of the case presented by the

[9] As opposed to emotion, mood is usually characterised by lower intensity and longer duration (Scherer 2005, 702).

[10] The effect of anger increasing the intentionality judgment has been also confirmed in a non-legal setting (e.g., Subra 2021).

defence as weaker. Another study utilising a murder case, where anger was imposed by a victim impact statement, also revealed that angry jurors were more likely to opt for death sentence compared to both neutral and sad jurors (NUÑEZ et al. 2015; see also NUÑEZ et al. 2017). Similar effect of anger on punitiveness of decision makers in a criminal trial scenario was also found in experiments by MATSUO and ITOH (2017) and LAURENT et al. (2014).

As pointed out, we should not make any direct inferences from studies on mock juries to judgments of professional legal decision makers. In fact, experiments investigating the impact of emotion on decisions of judges and other criminal justice professionals are extremely scarce. One such intriguing study with 53 Norwegian judges revealed that emotions displayed by the witness-victim during her testimony did not influence judges' witness credibility assessment nor their decisions on defendant's guilt (WESSEL et al. 2006). Conversely, an identical scenario tested on lay decision makers revealed that lay people in their assessments relied heavily on victim's displayed emotion, rather than on the content of her testimony (KAUFMANN et al. 2003). Discordant with these findings are results of a study with the Swedish police investigators. It revealed that angry (as opposed to sad) professional criminal investigators evaluating witness credibility did not pay attention to the consistency of the witness statement with the main investigation hypothesis (ASK & GRANHAG 2007). This indicates that anger indeed influenced investigators' information processing strategy and consequently their professional judgment.

The few cited studies do not allow for any general conclusions on the impact of emotion on the decision making of the criminal justice professionals. However, they allow for the assumption that the significance of the experienced emotion on legal judgments does not only differ between lay persons and legal professionals but might also vary between different groups of professionals within the criminal justice.

4. *Empathy and legal decision making*

Much discussion at the intersection of law and emotion revolves around the desirability of empathy in legal decision making. However, every mentioning of the empathy should come with a caveat of a considerable terminological and definitional confusion about the term in literature (CUFF et al. 2016). Most scholars, nonetheless, agree that empathy is not an emotion (as it is sometimes presented)[11] but rather a capacity to feel or understand other people's emotions, as well as their thoughts, perceptions, feelings and other cognitive states. Here, a distinction should be drawn between affective and cognitive empathy, which is not always clearly made in discussions on legal contextualisation of empathy.

> «[Affective] empathy refers to situations in which the subject has a similar emotional state to an object as a result of perceiving the object's situation. [...] Cognitive empathy refers to situations when the subject arrives at an understanding of the object's state through cognitive processes. It implies that the subject has used cognitive perspective taking to project him or herself into the position of the subject» (PRESTON & DE WAAL 2002, 2).

When exploring the role of empathy in legal decision making it is therefore useful to clarify which type of empathic capacity one has in mind. Perhaps also due to this definitional vagueness, the debates on whether it is beneficial for judges to utilize empathy in their work entail variety of arguments pro et contra. BANDES (2009) emphasises that empathy is an

[11] Some authors tend to conflate empathy with the emotions of compassion and/or sympathy as well as with the phenomenon of emotional contagion. Nevertheless, empathy may incite and can be compatible with these emotions. For distinctions see e.g., PRESTON & DE WAAL 2002; WISPÉ 1986.

essential human capacity which, through understanding other people's affective and cognitive states, enables us to function in a social world. Hence, it is also an essential prerequisite in judging, as it enables judges to understand other people's conduct and gives them basis for their moral reasoning. Similarly, HENDERSON (1987) recognises a valuable source of (human) knowledge in empathy (see also WEST 2020). She argues that empathy benefits a legal decision maker both in the process of discovering a conclusion as well as in the process of justifying this conclusion, in a manner that is unreachable to disembodied reason. Providing an example from criminal law, Bandes posits: «In a criminal case, the effort to understand the defendant's perspective can yield information valuable for both the guilt and sentencing phases of the trial» (BANDES 1996, 379).

On the other pole, some authors raise concerns over (improperly) utilising empathy in judging. These range from claims that empathic imagination has no normative significance in judging (POSNER 1995) to fear that empathy may lead to partiality and bias and, therefore, might be inconsistent with objectivity and the rule of law (ROACH ANLEU & MACK 2021). In that respect, BANDES (2009) warns against selective empathy; when a judge more easily empathises with a party having a background familiar to the judge compared to a person with dissimilar life experience. The criminal justice system seems particularly sensitive setting for selective empathy where defendants are often marked by disadvantaged socioeconomic, family and educational backgrounds or come from otherwise marginalised social groups. In contrast, this is typically not the case with judges, prosecutors, and other legal professionals. Judge's empathy might thus be more readily accessible for some victims or perhaps defendants with more relatable personal profiles to the decision maker herself (e.g., white-collar crime defendants). In fact, research indicates that in sentencing decisions, white male jurors in United States are more likely to show bias against defendants from other racial and demographic backgrounds. As an explanation for this bias, researchers propose the flip side of the selective empathy mechanism, which they term as empathic divide (HANEY 2003; LYNCH & HANEY 2011).

An interesting study by WETTERGREN and BERGMAN BLIX (2016) reveals how, apart from judges, empathy is employed by other criminal law professionals, namely the prosecutors in Sweden. This study convincingly shows how empathy is used as a valuable legal reasoning tool in various important prosecutorial decisions. In one presented case in the study, a prosecutor was in a dilemma whether to press charges for the offence of aggravated unlawful threat, as the suspect's conduct caught on CCTV cameras was somewhat ambiguous. By thoroughly analysing the suspect's behaviour, the prosecutor has concluded that the suspect panicked and reacted under fear rather than with the intention to threaten. Interestingly, in substantiating to the researcher her decision not to prosecute, the prosecutor buttressed her conclusion by employing empathy. She explained that the emotions of fear and panic is what she would have felt in the suspect's position (WETTERGREN & BERGMAN BLIX 2016).

A study conducted among Australian judicial officers, indeed, reveals that a majority of them believe that empathy is essential or very important in their day-to-day work. Moreover, several «describe their judicial practice as entailing impartiality and empathy, almost as complementary forces requiring careful and persistent monitoring of their boundaries» (ROACH ANLEU & MACK 2021, 74). This reflection perhaps best encapsulates the ambiguous nature of empathy in legal decision making. It is an indispensable human ability allowing a legal decision maker to fully comprehend different perspectives of the stakeholders involved in a case—along with the legally relevant emotional aspects. On the other hand, its potential selective application together with its ability to invoke emotions in the decision maker herself, may go contrary to the postulates of impartiality and objectivity.

5. *Emotions as normative elements of the criminal law norms*

In the first part of this chapter, emotions were tackled from the vantage point of extra-legal phenomena influencing legal decision making. In that role, emotions were recognised as both an indispensable element in the process as well as potentially detrimental factor obscuring and unduly biasing rational process of legal reasoning. Therefore, it is easy to overlook that on the other hand, emotions in criminal law often play a prominent legal role of the very normative elements in legal norms. Performing in this role, they pose a duty (and often a challenge) to jurists to furnish them with legal definitions, to recognise, interpret, and prove them. Take for example the following provision on excessive self-defence from the Slovenian Criminal Code:

> «In the event the perpetrator has acted beyond the limits of justifiable self-defence, he or she may receive a more lenient sentence; when he or she acts due to *severe irritation* or *great fear* caused by attack, his or her punishment may be remitted» (emphasis added)[12].

When the court is applying this provision and is considering whether the defendant exceeding self-defence acted due to severe irritation or great fear caused by the attack, it will first need to interpret the terms "severe irritation" and "great fear". The defence (that has been shifted the burden of proof in this case) needs to establish that the perpetrator was indeed in an emotional state that matches one of these definitions. Despite these being legal terms, they derive their meaning from and need to correspond to actual experienced emotions. Hence, in establishing these facts the court needs to find the way to unravel (*ex post facto*) deeply subjective experiences of affects in the perpetrator. This may pose difficult evidentiary challenges, which in practice usually require assistance of an expert (e.g., psychiatrist or psychologist).

The selected example pertains to the rules on sentencing (at least in the provided legal order). However, emotions are weaved into legal norms of many other eminent criminal law subject matters of both substantive and procedural nature[13]. One such important area are rules on culpability and excuses, more specifically on criminal insanity and (substantially) diminished capacity[14]. While typically the application of these rules requires underlying mental condition or sometimes intoxication with psychoactive substances, many legislations also allow for the application of the rules on diminished capacity and insanity even due to extreme emotional excitation. This might be the case with offenders acting under uncontrollable rage, severe anxiety, shock, or in a panic attack (Blomsma & Roef 2016).

Moreover, emotions may stand as normative elements of particular (modes of) criminal offences. Let us take for example a provision on manslaughter from the Swiss Criminal Code: «Where the offender acts in a state of extreme emotion that is excusable in the circumstances, or in a state of profound psychological stress, the penalty is a custodial sentence from one to ten years»[15]. In this provision, the perpetrator's excusable state of extreme emotion or psychological stress is the normative element constituting a mitigated form of homicide. A comparable common law definition can be found in voluntary manslaughter (Kahan & Nussbaum 1996).

Even more often, however, emotions appear in criminal offences as their implicit presuppositions. Hate crimes present one such example. What typically qualifies these acts as hate crimes as opposed

[12] Criminal Code of the Republic of Slovenia, Article 22, Paragraph 2 (unofficial English translation) (Official Gazette of the Republic of Slovenia, no. 50/12—official consolidated version, 6/16—cor., 54/15, 38/16, 27/17, 23/20, 91/20, 95/21, 186/21 and 105/22—ZZNŠPP).

[13] Needless to point that provided examples may be regulated differently in different legal orders.

[14] It must be noted that different jurisdictions use different legal terms for this concept, as well as different normative levels of diminished capacity.

[15] Swiss Criminal Code of 21 December 1937, Article 113 (unofficial English translation).

to other forms of (violent) offences is that they are motivated by the emotions of hatred (or perhaps moral disgust[16], contempt, or similar affect) rooted in prejudice or bias against certain social group[17]. Furthermore, many jurisdictions have recently criminalised offences generally termed as revenge pornography. Although both theoretical and legal conceptions vary, narrower definitions distinguish this type of offences from similar unlawful sharing of another's intimate content, in that it is incited out of revenge[18] by the victim's (ex sexual) partner (WALKER & SLEATH 2017).

Finally, some criminal offences can only be committed if the perpetrator's conduct causes a certain emotion in the victim or if such a conduct is typically capable of causing a certain feeling. In many jurisdictions this is the case with rather common crimes, including threat, extortion, or stalking, which might require that the perpetrator's conduct is capable of instilling fright or other emotional distress in the victim.

Another area where emotions play an important and often controversial normative role in criminal law are sentencing decisions. Legal orders vary in how they prescribe emotions of defendants and victims to be taken into account in deciding upon an appropriate sentence. Some legal orders explicitly lay down particular emotions that decision makers need to consider, for example remorse (ROACH ANLEU & MACK 2021, 41). Some legal orders only exemplify typical emotions that can be considered in sentencing, whereas other jurisdictions leave a wider discretion relating to the factors the courts may or may not take into account. In practice, emotions in the perpetrator are often considered as aggravating or mitigating factors when they are weighed as motives for an offence (e.g., HESSICK 2006). Vengeance, jealousy, envy, and hatred are examples of such aggravating factors. On the other hand, a judge or jury may consider motives in offences committed out of compassion, pity, love, or provoked rage as mitigating factors. An even more controversial matter in sentencing are emotions displayed by the offender after committing a crime, such as shame[19], regret, or guilt. In this context, remorse has gained the most theoretical attention (e.g., BANDES 2016; BENNETT 2016; PROEVE & TUDOR 2016; SARAT 1999). The displayed remorse by the defendant is normally considered a mitigating factor leading to a more lenient sentence. Interestingly, however, the court sometimes even expects the defendant to show this particular emotion during a trial. Hence, the lack of displayed remorse or remorse that is feigned, can be used as an aggravating factor when a court imposes a criminal sanction (ROSSMANITH et al. 2018).

Victim's emotional distress caused by the committed offence can be a similarly deciding factor in sentencing. While regarding victim's suffering as a relevant circumstance in applying sentence has not been disputed, more controversy has been stirred by the victim impact statements allowed in many common law jurisdictions. With a victim impact statement, the victim obtains an opportunity to present to the court how the offence has affected her life, but also to propose a sentencing recommendation to the court (BOOTH 2016). Such statements, particularly in the capital punishment trials in the United States, have become increasingly emotional; sometimes with an included video material underlaid with evocative music, they resemble short documentary films about the victim's life (WINOGRAD 2008). This led to the dilemma, dealt even by the Supreme Court of the United States, whether particular victim

[16] In this vein, KAHAN 1998, 1634 «suggest[s] that the "hate crimes" debate is better understood as a "disgust crimes" debate».

[17] It should be emphasised that there is no universal definition of hate crime and that conceptualisations of this phenomenon vary both in theory and in legal regulation between countries (see e.g., SCHWEPPE 2021). Many authors agree with SULLAWAY 2004, 253 that «the presence or absence of the emotion of hate is a poor criterion by which to define hate crimes».

[18] As succinctly explained by MCDERMOTT et al. 2017, 71: «[R]evenge is not motivated by the rational expectation of future deterrence. It is instead driven by the intrinsic pleasure that one expects to experience upon striking back».

[19] Critically on the effect of shame pursued by the criminal law, see MASSARO 1999.

impact statements were overly emotional[20]. Many feared that emotionally charged content might unduly bias the jury or judge and thus render sentencing decision unfair (BANDES 1996). The presented dilemma provides a good example on the complex and metamorphic role that emotion plays in the criminal law decision making. A substantive criminal law question (emotional distress by the victim to be considered in applying criminal sanction) invokes a procedural (evidentiary) challenge of establishing this fact. This, in turn, implies concern whether provoked emotions in decision makers (as an extra-legal factor) will meddle with and bias legal reasoning, which finally results in a broader procedural issue, whether these emotional factors undermine fair trial and due process rights in a criminal trial.

Finally, focusing on solely procedural criminal norms, we find that emotions sometimes take a central stage in that area as well. In this context, especially emotions of victims and other witnesses are of concern. On the one hand, procedural regulations should strive to prevent (additional) emotional harm (secondary victimisation) that the criminal trial might cause to these vulnerable participants. On the other hand, however, such protective measures should not overly impose on the defendants' fair trial rights. Pursuing the first goal, many contemporary criminal procedures include sets of regulations aiming at preventing intimidation, humiliation, and fear in victims and other witnesses. In fact, in the European Union, Directive 2012/29/EU[21] imposed an obligation on all member states to adopt procedural measures in their domestic laws that would recognise and prevent potential emotional distress in victims during pre-trial and trial proceedings. On the other hand, however, the jurisprudence of the European Court of Human Rights (ECtHR) cautions that the concern for the (emotional) wellbeing of witnesses should not breach the defendant's conventional fair trial rights, namely the right to examine witnesses against him. Interestingly, tackling the relevance of witnesses' fear, ECtHR calls for a closer investigation of this affect.

> «A distinction must be drawn between two types of fear: fear which is attributable to threats or other actions of the defendant or those acting on his or her behalf and fear which is attributable to a more general fear of what will happen if the witness gives evidence at trial»[22].

Without further examining the quoted argument, that ECtHR develops in the cited and other decisions, it is evident that emotions as the interest of procedural criminal law may sometimes take a central stage in legal fora.

Moreover, affects as more distant procedural factors should be also considered in other procedural undertakings of various criminal trial participants. Let us take, for example, false admission of guilt made by the defendant under threat, or a witness's perjury motivated by revenge. Such acts, which are procedurally invalid, are directly motivated by emotions. Perhaps more indirectly, but no less importantly, emotions act in the background of the traditional instruments of procedural law—oaths[23]. Historically, the effectiveness of oath stems from the fear of deity's wreath and punishment in case the person taking the oath breaches it (WHITE 1903). Notwithstanding its archaic roots, the oath or affirmation taken by various procedural actors remains an inevitable component in almost any contemporary criminal procedure. It seems that an important factor for its effectiveness still nowadays lies in the psychological,

[20] See e.g., *Payne* v. *Tennessee*, 501 U. S. 808 (1991) and *Kelly* v. *California*, 07-11073 (2008).

[21] Directive 2012/29/EU of the European Parliament and of the Council of 25 October 2012 establishing minimum standards on the rights, support and protection of victims of crime, and replacing Council Framework Decision 2001/220/JHA.

[22] ECtHR, Al-Khawaja and Tahery v. the United Kingdom, 15 December 2011, no. 26766/05 and 22228/06, Para. 122.

[23] Interestingly, oaths in some jurisdictions directly invoke on emotions. For example, Australian judicial officers swear that they will dispense justice «without fear or favour, affection or ill-will» (CAMPBELL 1999, 146).

particularly emotional weight that the act of swearing imposes[24]. The person taking the oath is not only motivated by the fear of a pending legal sanction if she breaks it, and the expected emotions of guilt or shame, but also by moral pride in keeping to the oath[25].

Furthermore, all criminal proceedings usually give authority to the presiding judge(s) to prevent and sanction undue acts of insult, humiliation, or other emotional harm to any participants of the proceedings in order to protect the dignity of the persons involved and the authority of the court itself.

Finally, attention should be also brought to restorative justice mechanisms that increasingly complement regular criminal proceedings in contemporary jurisdictions[26]. While the advantages of these alternatives to traditional criminal proceedings are manifold, their emotional aspect should not be underestimated. A crucial advantage of successful restorative justice programmes is the alleviation of negative emotions, not only those of the victim and the offender but also of other stakeholders; sometimes even the wider community affected by the offence (ROSSNER 2013; STRANG 2002). This emotional "discharging" prevents new criminogenic situations among those involved in the incident and in their immediate social environment. Decision makers in the police, prosecuting authorities, and courts selecting suitable cases to be diverted to restorative justice mechanisms are therefore often instructed by procedural norms or guidelines to consider these emotional aspects of cases as criteria.

Lastly, we should also touch upon the acts of clemency. In various jurisdictions pardons may be granted by different authorities; often by the head of state, government or some other body. In effect, clemency absolves a convicted person (or defendant before conviction) of all or some consequences of the criminal conviction[27]. Reasons for a granted pardon vary. Nevertheless, this institution is often understood (not without controversy) as a correction of a particular (unjust) criminal justice system outcome (NOVAK 2015). While it would be naïve to claim that nowadays clemency is based on actual emotions felt by those in charge of this decision, the etymology of the term clearly reveals its emotional background at least in its symbolic dimension. Thus, an act of clemency often reflects a public—rather than initially sovereign's—sentiment about a particular criminal case or a particular convicted person[28]. The public might feel sympathy, compassion or pity towards the convicted person that the public believes did not deserve a conviction or (particularly harsh) punishment. The public might also exhibit forgiveness or mercy for someone who is believed to have already paid their debt to the society. Such public sentiments may legitimise discretionary clemency decisions and thus reduce concerns over arbitrariness[29].

The provided examples are not exhaustive in any way but nevertheless support the assertion that emotion is indeed part and parcel of the very normative structure of both substantive and procedural criminal law.

6. *Emotions from the socio-legal perspective*

Criminal trials and related criminal proceedings are highly formalised and uniform processes if viewed through normative lenses. However, criminal law proceedings can be observed also as

[24] For empirical evidence in a non-criminal law context on how oath-taking markedly improves truth telling, see e.g., JACQUEMET et al. 2019.

[25] On guilt, shame, and moral pride as instigators of moral behaviour, see TANGNEY et al. 2007.

[26] For a comprehensive comparative overview see DÜNKEL et al. 2015.

[27] Similar can be claimed for amnesty, which in comparison to pardon, is usually passed by a legislative body and might be motivated by different, e.g., political or pragmatic, reasons.

[28] Cf. SAJÓ 2016, on the influence of public sentiment in the formation of constitutional norms.

[29] «[A]ll countries must wrestle with clemency's underlying tensions between unchecked discretion and law; between individualization and arbitrariness; and between mercy and justice» (NOVAK 2015, 820).

social phenomena, as dynamic interactions among the involved stakeholders embedded in a specific cultural environment and institutional setting. Moreover, the participants (e.g., judges, attorneys, defendants, witnesses) are of different social status and hold various amounts of social power. Socio-legal research, that explores these aspects of court proceedings, has been increasingly paying attention to the function of emotions therein. This strand of research observes how emotions are displayed, managed, interpreted, or instrumentally employed by various procedural actors for various purposes (e.g., BERGMAN BLIX & WETTERGREN 2016).

Socio-legal analysis of emotion in judicial context builds upon concepts developed in sociology of emotion and applies them to legal settings. One such important concept is emotional labour[30]. It denotes adapting (suppressing or inducing) emotions to fit a particular social or professional setting according to the so-called feeling rules (HOCHSCHILD 2012). Feeling rules in judicial (criminal law) context entail expectations concerning which emotions and how ought or ought not to be displayed in a particular court situation, for example during witness interrogation[31].

The concept of emotional labour can be easily linked to the notion of self-regulation that MARONEY (2020) discusses as a crucial component of a desirable judicial temperament—a set of personal traits ideally possessed by a judge. Maroney argues that besides positive and negative emotionality, judge's self-regulation is pivotal. It entails not only the ability to control which emotions and how are exhibited by a judge in a particular judicial setting, but also which emotions and how strongly are felt by her.

However, ROACH ANLEU and MACK (2021, 11) note that emotion management in court also entails regulating experienced and expressed emotions of other trial participants:

> «the judiciary can use their own feelings and emotion displays to accomplish their daily tasks and professional goals. Judicial officers can adopt a certain demeanour to evoke particular emotions as a way to foster trust, or as an attempt to induce certain feeling states among participants in their workplaces».

It should be pointed out that properly applied emotion management can be a highly beneficial and powerful tool in criminal proceedings. It can be used to strengthen decision making in criminal law through various strategies. For example, a judge, who uses warm, calm, and compassionate communication mode in addressing a frightened or nervous witness may acquire from the witness more, and more accurate information which will foster the truth finding process in the trial. Conversely, using cold and authoritative tone to remind a defence attorney that he is abusing his procedural rights might be an effective way for a judge to keep an adversarial balance between the parties and to ensure orderly progress of a procedure.

On the other hand, a legal decision maker may also employ emotion management to regulate her own emotions. By doing so, a judge may again pursue different objectives. Perhaps a judge might want to exhibit dispassion during an emotionally heavily charged testimony to convey the appearance of neutrality, impartiality, and procedural fairness to the public (MACK & ROACH ANLEU 2010). However, a judge might also want to regulate her experienced (not just displayed) emotion, in order to avoid unwanted bias in her decision making (see above, 3.4). In an insightfully study of Australian judicial officers, ROACH ANLEU and MACK (2021, 104-111) reveal different methods and techniques that judges use to regulate their own affects. These include self-talk, an internal discourse with oneself as conscious self-reminder of one's formal

[30] In literature, the notion of emotional labour partially or completely overlaps with some related terms, such as emotion regulation, emotion management, emotion work, or affective practice. For an overview of these concepts see ROACH ANLEU & MACK 2021, 8-12.

[31] As noted, feeling rules also vary culturally. For example, as acknowledged by BERGMAN BLIX & WETTERGREN 2016, 32: «If anger is expressed, it is likely to be more subtle in Sweden than in the US, due to the societal emotional regime».

function. Adjournment of proceedings—often resorted to when a judge feels anger—is an efficient and immediate way for a judge to distance herself from the emotion-triggering situation, to reflect upon it and her feelings, and to regain composure. Furthermore, judges make use of debriefings with colleagues and peers as a method to unburden emotion-related stress, whereas sometimes they also take advantage of humour for this purpose. Notwithstanding our focus on judges' emotion management, it should be pointed out that this practice is equally available and utilised by other criminal justice participants. They use it to pursue their own specific goals and to perform their own specific roles. For example, FLOWER (2021) reports on emotional performance of defence attorneys whose primary concern is conveying loyalty to defendants they represent, while WETTERGREN and BERGMAN BLIX (2016) document on prosecutors' use of empathy in emotion management.

However, the flip side of the emotion management is the danger of its instrumental use that can be an equally powerful weapon working against procedural fairness and legitimacy of criminal justice. Thus criminal justice participants may take advantage of emotions to manipulate other stakeholders[32], to exert their social power or to reaffirm their social status in the courtroom in manners that lack legitimacy (BERGMAN BLIX & WETTERGREN 2016; MACK & ANLEU 2010).

7. *Conclusion*

The many recent endeavours of scholars to tackle the role of emotions in criminal law have opened fascinating new perspectives on the decision making in the criminal law. Most importantly, they have allowed us to better understand the complexities and multidimensionality of this process. Despite still many unresolved challenges, these new insights should not be neglected. This is equally in the interest of the legal theory and of legal practitioners, who deal with emotions and legal decisions on daily bases; but ultimately, it is in the interest of all the criminal justice participants. Embracing a thoroughly multidisciplinary research on emotion in criminal justice does in no way threaten the central criminal law postulates as we have nurtured and developed through centuries. On the contrary, by better understanding the element of emotion that had previously been either intentionally ignored or unintentionally overlooked by the criminal law doctrine, we may make a better use of the fundamental criminal law principles and develop them further with the aspiration of a just application of the criminal law.

[32] Cf. Pillsbury's discussion on deliberate or unconscious "emotional deception" in written judicial opinions (PILLSBURY 1999, 341).

References

Abrams K., Keren H. 2010. *Who's Afraid of Law and the Emotions?,* in «Minnesota Law Review», 94, 1997 ff.

Ask K., Granhag P.A. 2007. *Hot Cognition in Investigative Judgments: The Differential Influence of Anger and Sadness,* in «Law and Human Behavior», 31, 537 ff.

Ask K., Pina A. 2011. *On Being Angry and Punitive: How Anger Alters Perception of Criminal Intent,* in «Social Psychological and Personality Science», 2, 494 ff.

Bandes S.A. 1996. *Empathy, Narrative, and Victim Impact Statements*, in «The University of Chicago Law Review», 63, 361 ff.

Bandes S.A. 1999. *Introduction*, in Id., *The Passions of Law,* New York University Press, 1 ff.

Bandes S.A. 2009. *Empathetic Judging and the Rule of Law*, in «Cardozo Law Review De Novo», 133 ff.

Bandes S.A. 2016. *Remorse and Criminal Justice*, in «Emotion Review», 8, 14 ff.

Bastian B., Denson T.F., Haslam, N. 2013. *The Roles of Dehumanization and Moral Outrage in Retributive Justice,* in «PLOS ONE», 8, 1 ff.

Bennett C. 2016. *The Role of Remorse in Criminal Justice*, in Tonry M. (ed.), *Oxford Handbook Online in Criminology and Criminal Justice*, Oxford University Press.

Bergman Blix S., Wettergren Å. 2016. *A Sociological Perspective on Emotions in the Judiciary*, in «Emotion Review», 8, 32 ff.

Blomsma J., Roef D. 2016. *Justifications and Excuses*, in Keiler J., Roef D. (eds.), *Comparative Concepts of Criminal Law,* Intersentia, 157 ff.

Bloom P. 2013. *Just Babies: The Origins of Good and Evil*, Crown Publishers.

Booth T. 2016. *Accommodating Justice: Victim Impact Statements in the Sentencing Process*, in «University of Technology Sydney Law Research Series», 33, 430 ff.

Boyd R., Gintis H., Bowles S., Richerson P.J. 2003. *The Evolution of Altruistic Punishment,* in «Proceedings of the National Academy of Sciences», 100, 3531 ff.

Bright D.A., Goodman-Delahunty J. 2006. *Gruesome Evidence and Emotion: Anger, Blame, and Jury Decision-Making*, in «Law and Human Behavior», 30, 183 ff.

Campbell E.M. 1999. *Oaths and Affirmations of Public Office,* in «Monash University Law Review», 25, 132 ff.

Craigie J. 2011. *Thinking and Feeling: Moral Deliberation in a Dual-process Framework,* in «Philosophical Psychology», 24, 53 ff.

Cuff B.M.P., Brown S.J., Taylor L., Howat D.J. 2016. *Empathy: A Review of the Concept,* in «Emotion Review», 8, 144 ff.

Damasio A. 2005. *Descartes' Error: Emotion, Reason, and the Human Brain,* Penguin Books.

DeSteno D., Petty R.E., Wegener D.T., Rucker D.D. 2000. *Beyond Valence in the Perception of Likelihood: The Role of Emotion Specificity,* in «Journal of Personality and Social Psychology», 78, 397 ff.

Dünkel F., Grzywa-Holten J., Horsfield P. 2015. *Restorative Justice and Mediation in Penal Matters: A Stock-taking of Legal Issues, Implementation Strategies and Outcomes in 36 European Countries,* Forum Verlag Godesberg.

Fehr E., Gächter S. 2002. *Altruistic Punishment in Humans*, in «Nature», 415, 137 ff.

Feigenson N. 2016. *Jurors' Emotions and Judgments of Legal Responsibility and Blame: What Does the Experimental Research Tell Us?,* in «Emotion Review», 8, 26 ff.

Feigenson N., Park J. 2006. *Emotions and Attributions of Legal Responsibility and Blame: A Research Review,* in «Law and Human Behavior», 30, 143 ff.

Flower L. 2021. *The Loyal Defence Lawyer,* in Bandes S.A., Madeira J.L., Temple K.D, Kidd White E. (eds.), *Research Handbook on Law and Emotion,* Edward Elgar Publishing, 165 ff.

Frijda N. H. 2007. *The Laws of Emotion,* Lawrence Erlbaum Associates.

Frijda N. H., Mesquita B. 1994. *The Social Roles and Functions of Emotions*, in Kitayama S., Markus H.R. (eds.), *Emotion and Culture: Empirical Studies of Mutual Influence,* American Psychological Association, 51 ff.

Gendron M. 2021. *The Evolving Neuroscience of Emotion: Challenges and Opportunities for Integration with the Law,* in Bandes S.A., Madeira J.L., Temple K.D, Kidd White E. (eds.), *Research Handbook on Law and Emotion,* Edward Elgar Publishing, 27 ff.

Georges L.C., Wiener R.L., Keller S.R. 2013. *The Angry Juror: Sentencing Decisions in First-Degree Murder*, in «Applied Cognitive Psychology», 27, 156 ff.

Goldberg J.H., Lerner J.S., Tetlock P.E. 1999. *Rage and Reason: The Psychology of the Intuitive Prosecutor*, in «European Journal of Social Psychology», 29, 781 ff.

Greene J.D. 2013. *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*, Atlantic Books.

Greene J.D., Morelli S.A., Lowenberg K., Nystrom L.E., Cohen J.D. 2008. *Cognitive Load Selectively Interferes with Utilitarian Moral Judgment,* in «Cognition» 107, 1144 ff.

Greenspan P. 2011. *Craving the Right*, in Bagnoli C. (ed.), *Morality and the Emotions*, Oxford University Press, 38 ff.

Grossi R. 2019. *Law, Emotion and the Objectivity Debate,* in «Griffith Law Review», 28, 2019, 23 ff.

Haidt J. 2001. *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment,* in «Psychological Review», 108, 814 ff.

Haidt J. 2003. *The Moral Emotions,* in Davidson R.J., Scherer K.R., Goldsmith H.H. (eds.), *Handbook of Affective Sciences*, Oxford University Press, 852 ff.

Haidt J. (2007). *The New Synthesis in Moral Psychology*, in «Science», 316, 998 ff.

Haney C. 2003. *Condemning the Other in Death Penalty Trials: Biographical Racism, Structural Mitigation, and the Empathic Divide Symposium: Race to Execution*, in «DePaul Law Review», 53, 1557 ff.

Henderson L.N. 1987. *Legality and Empathy,* in «Michigan Law Review», 85, 1574 ff.

Hessick C.B. 2006. *Motive's Role in Criminal Punishment,* in «Southern California Law Review», 80, 89 ff.

Hochschild A.R. 2012. *The Managed Heart: Commercialization of Human Feeling*, University of California Press.

Horberg E.J., Oveis C., Keltner D., Cohen A.B. 2009. *Disgust and the Moralization of Purity,* in «Journal of Personality and Social Psychology», 97, 963 ff.

Huebner B. 2013. *Do Emotions Play a Constitutive Role in Moral Cognition?,* in «Topoi», 34, 427 ff.

Hutcherson C.A., Gross J.J. 2011. *The Moral Emotions: A Social-functionalist Account of Anger, Disgust, and Contempt,* in «Journal of Personality and Social Psychology», 100, 719 ff.

Jacquemet N., Luchini S., Rosaz J., Shogren J.F. 2019. *Truth Telling Under Oath,* in «Management Science», 65, 426 ff.

Kahan D.M., Nussbaum M.C. 1996. *Two Conceptions of Emotion in Criminal Law,* in «Columbia Law Review», 96, 269 ff.

Kahneman D. 2011. *Thinking, Fast and Slow,* Lane.

Kahneman D., Fredrick S. 2002. *Representativeness Revisited: Attribute Substitution in Intuitive Judgment,* in Gilovich T., Griffin D.W., Kahneman D. (eds.), *Heuristics and Biases: The psychology of Intuitive Judgement*, Cambridge University Press, 49 ff.

Kaufmann G., Drevland G.C.B., Wessel E., Overskeid G., Magnussen S. 2003. *The Importance of Being Earnest: Displayed Emotions and Witness Credibility,* in «Applied Cognitive Psychology», 17, 21 ff.

Koenigs M., Young L., Adolphs R., Tranel D., Cushman F., Hauser M., Damasio A. 2007. *Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements,* in «Nature», 446, 908 ff.

Laurent S.M., Clark B.A.M., Walker S., Wiseman K.D. 2014. *Punishing Hypocrisy: The Roles of Hypocrisy and Moral Emotions in Deciding Culpability and Punishment of Criminal and Civil Moral Transgressors,* in «Cognition & Emotion», 28, 59 ff.

Lerner J.S., Tetlock P.E. 1999. *Accounting for the Effects of Accountability,* in «Psychological Bulletin», 125, 255 ff.

Lively K.J., Weed E.A. 2016. *The Sociology of Emotions,* in Barrett L.F., Lewis M., Haviland-Jones J.M. (eds.), *Handbook of Emotions*, Guilford Press, 66 ff.

Lynch M., Haney C. 2011. *Mapping the Racial Bias of the White Male Capital Juror: Jury Composition and the "Empathic Divide",* in «Law & Society Review», 45, 69 ff.

Mack K., Anleu S.R. 2010. *Performing Impartiality: Judicial Demeanor and Legitimacy,* in «Law & Social Inquiry», 35, 137 ff..

Maibom H. 2010. *What Experimental Evidence Shows Us about the Role of Emotions in Moral Judgement,* in «Philosophy Compass», 5, 999 ff.

Maroney T. 2016. *A Field Evolves: Introduction to the Special Section on Law and Emotion*, in «Emotion Review», 8, 3 ff.

Maroney T. 2020. *(What We Talk About When We Talk About) Judicial Temperament,* in «Boston College Law Review», 61, 2085 ff.

Martins A.T., Faísca L.M., Esteves F., Muresan A., Reis A. 2012. *Atypical Moral Judgment Following Traumatic Brain Injury,* in «Judgment and Decision Making», 7, 478 ff.

Massaro T.M. 1999. *Show (Some) Emotion*, in Bandes S.A. (ed.), *The Passions of Law,* New York University Press, 80 ff.

Matsuo K., Itoh Y. 2017. *The Effects of Limiting Instructions about Emotional Evidence Depend on Need for Cognition,* in «Psychiatry, Psychology, and Law», 24, 516 ff.

McDermott R., Lopez A.C., Hatemi P.K. 2017. *'Blunt Not the Heart, Enrage It': The Psychology of Revenge and Deterrence,* in «Texas National Security Review», 1, 68 ff.

Mikula G., Scherer K.R., Athenstaedt U. 1998. *The Role of Injustice in the Elicitation of Differential Emotional Reactions,* in «Personality and Social Psychology Bulletin», 24, 769 ff.

Mulligan K., Scherer K.R. 2012. *Toward a Working Definition of Emotion*, in «Emotion Review», 4, 345 ff.

Nichols S. 2002. *Norms with Feeling: Towards a Psychological Account of Moral Judgment,* in «Cognition», 84, 221 ff.

Novak A. 2015. *Transparency and Comparative Executive Clemency: Global Lessons for Pardon Reform in the United States,* in «University of Michigan Journal of Law Reform», 49, 817 ff.

Nucci L.P. 2001. *Education in the Moral Domain,* Cambridge University Press.

Nuñez N., Myers B., Wilkowski B.M., Schweitzer K. 2017. *The Impact of Angry Versus Sad*

Victim Impact Statements on Mock Jurors' Sentencing Decisions in a Capital Trial, in «Criminal Justice and Behavior», 44, 862 ff.

NUÑEZ N., SCHWEITZER K., CHAI C.A., MYERS B. 2015. *Negative Emotions Felt During Trial: The Effect of Fear, Anger, and Sadness on Juror Decision Making*, in «Applied Cognitive Psychology», 29, 200 ff.

PHALEN H.J., SALERNO J.M., NADLER J. 2021. *Emotional Evidence in Court,* in BANDES S.A., MADEIRA J.L., TEMPLE K.D, KIDD WHITE E. (eds.), *Research Handbook on Law and Emotion,* Edward Elgar Publishing, 288 ff.

PILLSBURY S.H. 1999. *Harlan, Holmes, and the Passions of Justice,* in BANDES S.A. (ed.), *The Passions of Law,* New York University Press, 330 ff.

POSNER R.A. 1995. *Overcoming Law*, Harvard University Press.

PRESTON S.D., DE WAAL F.B.M. 2002. *Empathy: Its Ultimate and Proximate Bases*, in «Behavioral and Brain Sciences», 25, 1 ff.

PRINZ J.J. 2007. *The Emotional Construction of Morals,* Oxford University Press.

PROEVE M., TUDOR S. 2016. *Remorse: Psychological and Jurisprudential Perspectives,* Routledge.

QUIGLEY B.M., TEDESCHI J.T. 1996. *Mediating Effects of Blame Attributions on Feelings of Anger,* in «Personality and Social Psychology Bulletin», 22, 1280 ff.

ROACH ANLEU S., MACK K. 2021. *Judging and Emotion: A Socio-Legal Analysis*, Routledge, Taylor & Francis Group.

ROSSMANITH K., TUDOR S., PROEVE M. 2018. *Courtroom Contrition: How Do Judges Know?*, in «Griffith Law Review», 27, 366 ff.

ROSSNER M. 2013. *Just Emotions: Rituals of Restorative Justice*, Oxford University Press.

RUSSELL P.S., GINER-SOROLLA R. 2011. *Moral Anger, but Not Moral Disgust, Responds to Intentionality,* in «Emotion», 11, 233 ff.

SAJÓ A. 2016. *Emotions in Constitutional Institutions,* in «Emotion Review», 8, 44 ff.

SALERNO J.M. 2021. *The Impact of Experienced and Expressed Emotion on Legal Factfinding,* in «Annual Review of Law and Social Science», 17, 181 ff.

SARAT A. 1999. *Remorse, Responsibility, and Criminal Punishment,* in BANDES S.A. (ed.), *The Passions of Law,* New York University Press, 168 ff.

SCARANTINO A. 2016. *The Philosophy of Emotions and Its Impact on Affective Science*, in BARRETT L.F., LEWIS M., HAVILAND-JONES J.M. (eds.), *Handbook of Emotions*, Guilford Press, 1 ff.

SCHERER K.R. 2005. *What are Emotions? And How Can They Be Measured?,* in «Social Science Information», 44, 695 ff.

SCHWEPPE J. 2021. *What Is a Hate Crime?,* in «Cogent Social Sciences», 7, 1 ff.

SEMMLER C., BREWER N. 2002. *Effects of Mood and Emotion on Juror Processing and Judgments,* in «Behavioral Sciences & the Law», 20, 423 ff.

SLOTE M. 2014. *A Sentimentalist Theory of the Mind,* Oxford University Press.

STRANG H. 2002. *Repair or Revenge: Victims and Restorative Justice*, Oxford University Press.

STROHMINGER N., LEWIS R.L., MEYER D.E. 2011. *Divergent Effects of Different Positive Emotions on Moral Judgment,* in «Cognition», 119, 295 ff.

SUBRA B. 2021. *The Effect of Anger on Intentionality Bias*, in «Aggressive Behavior», 47, 464 ff.

SULLAWAY M. 2004. *Psychological Perspectives on Hate Crime Laws,* in «Psychology, Public Policy, and Law», 10, 250 ff.

TANGNEY J.P., STUEWIG J., MASHEK D.J. 2007. *Moral Emotions and Moral Behavior,* in «Annual Review of Psychology», 58, 345 ff.

TIEDENS L.Z., LINTON S. 2001. *Judgment under Emotional Certainty and Uncertainty: The Effects of Specific Emotions on Information Processing,* in «Journal of Personality and Social Psychology», 81, 973 ff.

VALDESOLO P., DESTENO D. 2006. *Manipulations of Emotional Context Shape Moral Judgment*, in «Psychological Science», 17, 476 ff.

VÄSTFJÄLL D., SLOVIC P. 2013. *Cognition and Emotion in Judgment and Decision Making*, in ROBINSON M.D., WATKINS E., HARMON-JONES E. (eds.), *Handbook of Cognition and Emotion*, Guilford Press, 252 ff.

WALKER K., SLEATH E. 2017. *A Systematic Review of the Current Knowledge Regarding Revenge Pornography and Non-consensual Sharing of Sexually Explicit Media*, in «Aggression and Violent Behavior», 36, 9 ff.

WESSEL E., DREVLAND G.C.B., EILERTSEN D.E., MAGNUSSEN S. 2006. *Credibility of the Emotional Witness: A Study of Ratings by Court Judges,* in «Law and Human Behavior», 30, 221 ff.

WEST R. 2020. *The Anti-Empathic Turn*, in FLEMING J.E. (ed.), *Passions and Emotions*, New York University Press, 243 ff.

WETTERGREN Å., BERGMAN BLIX S. 2016. *Empathy and Objectivity in the Legal Procedure: The Case of Swedish Prosecutors,* in «Journal of Scandinavian Studies in Criminology and Crime Prevention», 17, 19 ff.

WHEATLEY T., HAIDT J. 2005. *Hypnotic Disgust Makes Moral Judgments More Severe*, in «Psychological Science», 16, 780 ff.

WHITE T. 1903. *Oaths in Judicial Proceedings and Their Effect upon the Competency of Witnesses,* University of Pennsylvania Law Review, 51, 373 ff.

WINOGRAD, B. 2008. *Petition Preview: Enya, the Death Penalty, and Video Victim Impact Evidence,* in «SCOTUSblog». Available at: https://www.scotusblog.com/2008/08/petition-preview-enya-the-death-penalty-and-video-victim-impact-evidence/.

WISPÉ L. 1986. *The Distinction between Sympathy and Empathy: To Call forth a Concept, a Word Is Needed,* in «Journal of Personality and Social Psychology», 50, 314 ff.

YOUNG L., KOENIGS M. 2007. *Investigating Emotion in Moral Cognition: A Review of Evidence from Functional Neuroimaging and Neuropsychology*, in «British Medical Bulletin», 84, 69 ff.

Prescriptive Descriptions: Reason-Emotion Binary through Feminist Critique

KRISTINA ČUFAR

1. *Introduction: Reason-emotion binary and social inequalities*

Critical thought in general and feminist critique in particular have long mistrusted the narrative that dualisms are but simple descriptions of objective reality. Pure description, like pure rationality, is an elusive and deceiving ideal, as the dualisms employed to describe the world take part in creating and organizing it. In so doing, the dualisms inscribe difference and hierarchically arrange the opposite poles in terms of the favored and the devaluated one. Reason-emotion, male-female, strong-weak, active-passive, culture-nature, (hu)man-animal, white-black, West-the rest, good-evil, and many others reflect the social structures and the imbalances of power within our society. The reason-emotion binary has long served as a tool to exclude women, people of color, the colonized, uneducated/poor, and others from the "reasonable man" mold. Subordinate groups were (and still are) commonly presented as the Other: emotional, animalistic, closer to nature, and consequently denied education, opportunities, and full membership in the political community.

The exclusion of subordinated groups from the political community is reflected in the very concept of (hu)man in law and philosophy. Some men, i.e., those endowed with high social status, material prosperity, and white skin, have long felt warm and comfortable in the law's empire that grants them rights and protects their interests. They are the original (hu)man invoked by Article 1 of the *Universal Declaration of Human Rights*: «All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood». The Other, excluded from the brotherhood of reasonable men, often experience the law as Kafka's trial. Article 2 of the *Declaration* neatly expresses some of the markers of oppression that have severe consequences for perceived humanity, daily lives, and interactions with the law on the part of the Other: «Everyone is entitled to all the rights and freedoms set forth in this *Declaration*, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status».

Articles 1 and 2 of the *Declaration* offer a neat illustration of the reason-emotion binary's legacy in Western (legal) tradition. The reason is perceived as the defining faculty of the human being, the bearer of rights and property owner, capable of meaningful judgment. Everyone and everything else is defined precisely through their lack of reason. Those determined by their difference vis-à-vis the reasonable man were long perceived as both mysterious and defective. Only the proper legal subject, the reasonable free autonomous individual, can speak and judge, while the Other's concerns are easily dismissed as emotional prattle. Rather than being seen as proper subjects of law, women, people of color, and other excluded groups were long perceived as objects of law and occasionally even reduced to property of those in the position of power (Montoya 2016).

Unsurprisingly, women, the poor, and people of color gained the right to vote much later than rich white men. Access to higher legal education and public office was basically impossible

for women before the 20th century, while the role of the judge remained reserved for men even longer (SCHULTZ & SHAW 2013). The struggle of indigenous populations of colonialized territories and various minorities across the Western world is also far from complete, as the office of state judge long remained reserved for white men of a specific background (SCHULTZ & SHAW 2013). Arguments against allowing women to serve on the bench stressed that women are too emotional, irrational, disorderly, and in need of protection from the harsh realities of the courtroom (SOMMERLAD 2013). It is worth examining how this subordination is established and maintained through a genealogy of the reason-emotion binary.

In this vein, this chapter briefly frames the complicated relationship between law and emotion. It then proceeds with a rough overview of the reason-emotion binary in the history of Western thought. This overview adopts the point of view of the feminist critique of the reason-emotion binary to illuminate the impact of theoretical concepts on lived realities and the legal rights of individuals. It also highlights that feminist theory has long been interested in the role of emotions in law and politics. Paying attention to feminist critique illustrates that the recent 'discovery' of emotions by the law and emotion scholarship builds upon a preexisting critique. Instead of focusing on legal scholarship alone, this overview draws on broader cultural and political feminist critique to demonstrate the law's embeddedness in different regimes of knowledge.

2. *Law, reason, and emotions*

In the Western tradition, the law is often portrayed as the domain of reason. The reason is perceived as central to legal theory and practice, as the guarantor of objectivity, neutrality, and the rule of law. Nevertheless, the law is a very emotional business: legislative, administrative, and judicial procedures encompass and provoke intense feelings of spectators and those directly involved. Regardless, the ideal image of a dispassionate judge presiding over the emotional drama, rationally applying legal norms to the factual mess, remains persistent (MARONEY 2011). Emotions are viewed with suspicion as potential contaminants threatening impartiality and objectivity of judgment (BANDES 1999a; UMPHREY et al. 2003; GROSSI 2019). In recent decades, a diverse body of law and emotions scholarship seeks to challenge this view and introduce nuance to this binary representation (e.g., MINOW & SPELMAN 1988; MILLER 1998; BANDES 1999b; ABRAMS & KEREN 2010; NUSSBAUM 2006; MARONEY & GROSS 2014; FRIEDLAND 2019; COTTERRELL 2018; ROACH ANLEU & MACK 2021).

Since emotions are present in animals and infants, they were conceptuality separated from reason and perceived as intuitive involuntary forces (POSNER 1999). Nevertheless, emotions develop and mature with age, can be trained, and are integral to moral, aesthetic, and reasoning in general (WALLACE 1993). Current trends in law and emotion scholarship focus on understanding and theorizing how emotions inform legal reasoning, identifying the benefits and pitfalls of emotional responses of the decision-makers, distinguishing between emotions that supposedly enhance or distort legal reasoning, and various other thought-provoking issues. The findings of cognitive sciences demonstrate that pure reason is neither realistic nor desirable, as emotions play a crucial part in what we understand as the process of reasoning.[1] These insights notwithstanding, the idea that emotions are an irrational force that comes over

[1] Damasio demonstrated that patients who suffer a brain injury hindering their ability to emote experience extreme difficulty in reasoning and decision-making. It seems that emotions are a crucial part of the reasoning. See generally DAMASIO 2005. Though this is not always the case, several studies suggest a beneficial impact of emotion on logical reasoning and a correlation between intense emotional response to a situation and the capacity to logically evaluate it. See: BLANCHETTE & CAPAROS 2013; For importance of emotional granularity for legal decision-making, see GENDRON & FELDMAN BARRETT 2019.

us and must be controlled and tamed by reason seems to persist and conserve the hierarchical structure of the reason-emotions divide.

Despite academic interest in law and emotion, the legal domain, in general, still approaches emotions based on the beliefs of folk psychology, even if these often conflict with contemporary scientific narratives that consider the oppositional perception of reason and emotions as outdated (FELDMAN BARRETT 2017, 219-251; NUSSBAUM 1995, 53-78). Nowadays, emotions are understood as a complex interaction of human embodiment residing in the body/brain and the sociocultural context that affects how emotions are perceived, expressed, and experienced (SCARANTINO 2016). It is important to stress that we still do not know much about emotions, reasoning, mind, and brain. Various cognitive processes, including emotions' exact nature and role, remain open to diverse interpretations.[2] Emotions have long incited the interest of biology, psychology, sociology, philosophy, cognitive and neurosciences, yet their definition remains open-ended (see, e.g., FELDMAN BARRETT et al. 2016; AHMED 2014, 1-19). In line with this diversity, this chapter does not intend to offer an all-encompassing definition of emotions. Instead, the chapter zooms in on how perceptions of emotions developed through time and contributed to social inequalities, marginalization, and exclusion of certain groups of people. A rather rudimentary overview of the history of the reason-emotion binary and its consequences does not pretend to be exhaustive; it instead provides a bricolage of diverse examples illustrating how this binary influenced Western thought, society, and political and legal systems.

3. *Feminist critiques of the reason-emotion divide*

The reason-emotion dualism and its consequences for the lived experience of people has long excited feminist critique. This section provides a brief overview of feminist engagement with emotions and reason through time and across different strands of feminisms. While sharing the core idea that women are not naturally inferior to men and the desire for a happier and more just society, feminisms cover a vast range of theoretical ideas and activisms with diverse epistemological and ideological points of view. The examples provided below are not exhaustive but intended to illustrate some of the strands of feminisms and feminist scholars who recognized the political thrust of binary conceptualization of reason and emotions. This section departs from the feminist critique of the classical thinkers that shaped Western political philosophy from antiquity to modernity. Illustration of the reason-emotion dualism's development and solidification is followed by a rough exploration of diverse feminist debates on emotions and reasoning in the 20th and 21st centuries.

3.1. *Antiquity*

Ancient Greece and its philosophy are widely perceived as the cradle of Western civilization. This founding myth fetishizes ancient Greek democracy, which was actually a rule of a small group of wealthy male citizens. This commonly unaddressed bias at the heart of the idealized origin is reflected in ancient Greek philosophy representing Western thought's base. For example, Plato's famous allegory of the charioteer qua reason in control of wild horses qua emotions hints at distrust towards unruly emotions. And yet, the allegory stresses interdependence between

[2] Various issues ranging from small and unrepresentative sample sizes, different theoretical ideas about whether emotions are contextual, biological, or both, and a wide variety of interpretative methods for brain scans contribute to widely different conclusions about what emotions are and how they function. When it comes to emotion and gender, for instance, it seems that researchers find what they anticipated to find, namely differences in specific mental processes of different genders or the absence of such differences. See: BRODY et al. 2016; FINE 2013.

reason and emotion that is absent in the modern version of the dualism. As for women, Plato described them as reincarnations of wicked irrational men, inferior to men in both reason and virtue, even though he proposed a (near) abolition of the sexual difference in his ideal republic (OKIN 1979, 15-50). Despite his sexist rhetoric, Plato's idea of philosopher queens was revolutionary in the context of ancient Greek society.

Aristotle, on the other hand, was a defender of the status quo. He saw women as deformities and defined them according to their function (for men) in the reproduction of *man*kind. In his view, what distinguishes a human from an animal is his reason. Still, not all humans are equal in reason: menial workers, women, and slaves (excluded from humanity altogether) are denied full rationality and thus excluded from Aristoteles's best state, prompting Genevieve Lloyd to interpret his philosophy as one of the cornerstones of male reason and its dominance (LLOYD 1993, 1-30). Nevertheless, Aristotle's view of emotions was more nuanced than Plato's. The Aristotelian reason offers itself to re-appropriations by feminist thinkers who produced more inclusive interpretations, opening avenues for the cooperation of reason and emotion, contributing to happy lives and relationships of a broad(er) specter of human beings (HOMIAK 2018).

The dichotomies between reason and emotion, mind and body, masculine and feminine, can be traced to ancient Greek philosophy and its medieval interpretations (LYONS 1999). Nevertheless, these pairs' sharp polarizations were not established until the 17th century (KELLER 1985, 44). Even the idea that male and female bodies are radically different was not present in science until the 18th century.[3] Instead, antique and medieval thinkers perceived the human body according to the "one-sex model," treating the female embodiment as an imperfect version of the male (LAQUEUR 1992). The idea that emotions belong to the feminine sphere, while man is marked by reason, is intimately connected with the age of Enlightenment.

3.2. *Enlightenment*

The age of Enlightenment is commonly perceived as the era that reinforced the mind-body divide and thus contributed to the creation of different spheres for women and men.[4] This is reflected in the early capitalist public-private dualism that confined women in their roles as wives and mothers and men as wage earners and political actors (LITTLE 1995; FISCHER 2016). Descartes' "think therefore I am" elevated reason as superior to the body and pawed the way for interpreting emotions as the unwelcome Other (DAMASIO 2005). Women and people of color were aligned with nature and body, while the white male body conspicuously disappeared as the house of decentered, objective and universal reason (AHMED 1995).

Nevertheless, the 17th century Europe saw an unprecedented number of women expressing their ideas in print. Thinkers like Mary Astell and Damaris Lady Masham embraced Descartes' formulation of reason and did not perceive it as exclusively male (ATHERTON 2018). The notion of rational equality of men and women was, if not widely accepted, put forward and echoed across Europe (PERRY 2005; PERUGA 2005; STUURMAN 2005). François Poullain de la Barre famously argued that the mind has no sex and women, depending on their class and geographical location, were always actively participating in science (SCHIEBINGER 1991). It would thus be erroneous to flatten down the Enlightenment narratives as homogenous and the

[3] The idea that biological traits demark male and female bodies is largely accepted but not scientifically plausible. Consequently, Western societies resorted to the "normalization" of intersex people who exhibit sexual characteristics of "both" sexes to uphold the male-female dualism. This normalization involves surgeries at an early age and long-term hormonal treatments, exemplifying the normative import of "descriptive" dualisms: "nature" has to be modified to be made consistent with the "rational" order. See, e.g., DEVOR 1989; FAUSTO-STERLING 2000; GEERTZ 1975.

[4] I refer to Enlightenment as a historical epoch, a current of thought developed in Europe and spread around the globe in the 17th and 18th centuries, and a philosophical concept.

transition to modernity as a simple linear progression of ever-stricter separation of genders and gender roles (Robertson 2005; Carey & Festa 2009; Tricoire 2017; Israel 2002). Regardless, ideas that women might be capable of reason were largely marginalized. Arguments asserting women's biological incapability of reasoning dominated, and women's achievements were often appropriated by men, forgotten, or simply ignored. Idealizations of Enlightenment as the epoch of reason, the overcoming of superstition, and the dawn of democracy and human rights are mostly blind to this movement's internal contradictions and complexities.

The uncomfortable fact that the glorious epoch of reason also produced an array of sexist and racist myths is difficult to ignore from the perspective of critical scholarship (see, e.g., Spivak 1999). Mind-body and other hierarchical pairings allowed thinkers like Rousseau to preach absolute equality and freedom on the one hand and argue for subordination and exclusion of women as naturally inferior and in need of male dominance on the other (Okin 1979, 99-198). Views of women as deficient in reason, destined to serve men, and naturally belonging to the private sphere are not foreign to Locke, Kant, Comte, Hegel, and other giants in philosophy (Kristeva 1996 [1979]; Le Doeuff 1991 [1977]; Herman 2018; Kleingeld 2019). These powerful narratives were never simple descriptions of reality; they played a part in (re)constructing social hierarchies, (re)structuring the place of different individuals, and informed popular beliefs about the different natures of men and women, which partially persist to this day.

It is worth stressing that it was not only women who were perceived as inferior and less reasonable; such labels were also attached to the men belonging to lower social classes and those enslaved during the European colonialization of the globe. Men might have been the sovereigns of their households (the private sphere), yet their political participation (the public sphere of politics) was often severely limited. French and American revolutions represent celebrated steps toward greater equality of men, yet most of the population remained excluded from the category of free and equal citizen. While largely barred from entering into public discussions, women and colonialized subjects have nevertheless responded to the Enlightenment's narratives of progress and universal reason.

3.3. *Enlightened revolutions and the Other*

Mary Wollstonecraft, one of the most famous early modern advocates of women's rights, was critical not only of the position of women but of social hierarchies in general, challenging the hereditary privilege governing the English society and the monarchy itself. As the events of the French Revolution shook the old social order, Wollstonecraft was contemplating its promise of equality and its internal contradictions (Wollstonecraft 1995 [1790 and 1792]). She internalized the Enlightenment's emphasis on reason and education. Still, she did not accept that women are inferior in reason and should be educated differently, as many, including Rousseau, suggested (Rousseau 1979 [1762]). On the contrary, Wollstonecraft correlated the unequal position of women with their socialization and education into false excessive sensibility and submission.

Olympe de Gouges, an advocate of women's rights and the abolition of slavery, pointed out the lack of concern for women's equality in the *Declaration of the Rights of Man and the Citizen* with her 1791 *Declaration of the Rights of Woman and the Female Citizen* (De Gouges 1979 [1791]). The alternative text of the *Declaration* urges a reconsideration of the status of women in French society. It also stresses that women participated in the revolutionary struggle, which their male companions quickly forgot once their goal was achieved. In her critique, she appeals to reason: «Woman, wake up; the tocsin of reason is being heard throughout the whole universe; discover your rights» (De Gouges 1979 [1791]). These ideas eventually led to de Gouges' decapitation.

The revolutionary demand for equal rights excluded not only women but also the populations of the colonized territories. Purposely ignoring that the *Declaration of the Rights of Man and the Citizen* explicitly allowed social distinctions based on "public utility," the Haitian

Revolution of 1791 represents a uniquely successful slave uprising that led to Haitian independence in 1804 (MARTEL 2017, 62-74). The first modern "black state" abolished slavery, if not forced labor, and was curiously erased from historiography for over two centuries. Such an erasure of a successful slave revolution is rooted in its disruption of the dominant narrative that the Other are incapable of agency and institution-building, reinforced by many enlightened philosophers, including Kant (TROUILLOT 2015).

American Revolution and its republicanism similarly overlooked large portions of the population. American *Declaration of Independence* was thus challenged by the *Declaration of Sentiments* written by Elizabeth Cady Stanton in 1848 (WEISS 2009, 100-123). Signed by both women and men, it strategically utilized the identification of women with emotion in its title. *Declaration of Sentiments* amended the phrase "all men are created equal" by adding "and women," decried illegitimacy of patriarchal organization of society, exposed grievances of oppression of women in private and public spheres, and urged for not only a genuinely democratic state but a truly democratic and equal society. Like Haitian Revolution and de Gouges' *Declaration*, the *Declaration of Sentiments* and its critique remains mainly overlooked in both historiography and legal and political theory. What is remembered and celebrated and what is erased and forgotten reflects the power imbalances underpinning descriptions of past events.

The line dividing public and private spheres, reasonable and emotional beings, was firmly entrenched by the end of the 18th century; in the 19th century, it seemed natural and immutable. The idea that women are led by emotions and naturally incapable of reasoning was widely supported by influential liberal intellectuals. Among modern liberals, John Stuart Mill represents a notable exception. He loudly questioned the narrative that women are emotional and irrational creatures (MILL 2017 [1869]). He pointed out that femininity is differently defined across different societies and could hardly be considered a natural given. Instead of explaining women's supposed lack of intellectual abilities with recourse to nature, he attributed the perceived feminine qualities to the content and lack of education available to women and girls. Indeed, women's education was one of the key concerns at the time, as the prophets of emotion-reason distinction often warned that education and intellectual development would hinder women's emotional capacity and even cause the atrophy of their reproductive organs (ROSENBERG 1982, XII-25). In contrast, Mill proposed that greater happiness of women achieved through their access to education and personal freedom would translate into greater happiness of men and thus benefit society as a whole.

3.4. *Towards abolition and universal suffrage*

Happiness was a central concept of women's attempts to theorize society in the late 19th and early 20th centuries. Aware of their subject position, early feminists did not accept the dominant idea of social sciences as objective and generalizing. Critically assessing the politics of gender and knowledge, they assumed a more grounded and activist standpoint (LENGERMANN & NIEBRUGGE-BRANTLEY 1997). The perspective of classical feminist social theory, engaging with material and emotional dimensions of pain, was often dismissed by male counterparts as too emotional, activist, and lacking objectivity. The reason-emotion epistemic binary continues to serve as an avenue to discredit critical and feminist thought (JAMES 1997). Classical feminist theorists like Harriet Martineau, Jane Adams, Anna Julia Cooper, Marianne Weber, and others are largely marginalized in mainstream social sciences and their records. For our purposes, it is interesting that instead of shutting emotions out, these authors called attention to the lived experience. Not trying to mask their emotional responses, these authors achieved vigilant observations on how markers of gender, class, ethnicity, and race limit the opportunities of individuals. Their approach was explicitly political, aiming to change society, the economic system, the state, and the legal order understood as the guardian of the status quo.

In the 19th century, the status quo was problematized on several fronts. The abolitionist movement was growing, questioning the institution of slavery as grossly immoral. Women were essential actors in this emancipatory movement closely related to the emergence of the feminist suffrage movement. While the two movements intersected, cooperated, and involved white and women of color as prominent members, their relationship is somewhat complicated. Not all abolitionists supported equal suffrage for women, nor did all white feminists advocate universal suffrage for people of color, fearing that black men would get the right to vote before women, as it actually happened (McDaneld 2013). Black women, on the other hand, were mistrustful of middle-class feminists and their tendency to describe the status of white women as slaves, thus erasing the legal status and reality of black enslaved women and their plight.[5] Racism was a feature of the early feminist movement in the USA, and many of its colored proponents were marginalized in its scholarly reconstructions. The actual subject positions and ascribed features of white and women of color were widely different (as they still are). White women are commonly conceived as fragile, meek, and in need of protection – a stark contrast to the cultural interpretation of black women as strong, resilient, hypersexual, and unruly. Formerly enslaved abolitionist and feminist activist Sojourner Truth is one of the most iconic voices of black women who pointed out that the experience of white women is not universal (Truth 2020 [1863]).

In general, women did share an expectation of remaining silent in public spaces. Public spaces and political discourse were perceived as a male domain unfit for women belonging to the private sphere, destined to serve their spouses and children. Those who decided to speak out were portrayed as unfeminine monsters and ridiculed in numerous ways, including cartoons depicting women activists as grotesque abominations (Rodríguez Durán 2015; Krueger 1992, 3-10) and sexist, classist and racist depictions of "new women" in anti-suffrage plays (Dassori 2005). Some women abolitionist activists resorted to emotional speeches based on personal experience, as their employment of rational argumentation was often met with hostility (Lamb-Books 2016, 123-155). Emotional portrayals of the suffering of slaves, particularly the pitiful exploitation of female slaves that male speakers tended to omit, resonated more favorably among their audience, as emotional speeches corresponded to the stereotypical image of a woman as empathic and caring. The arguments against women's right to participate in political processes or to speak in public or to do so in a manly, that is, abstract and rational manner, exhibit the still-existing double bind in which women tend to find themselves. On the one hand, women are (to be) excluded from decision-making because they are too emotional; on the other hand, those who do not conform to their role as highly emotional are judged for not being feminine and discredited on this basis (Crozier-De Rosa 2014; Brody et al. 2016).

The struggle for the right to vote was one of the most visible early feminist endowers. Not only women and people of color but also the colonialized men were excluded from voting based on being too childlike, irrational, erratic, and emotional to be recognized as true men and thus entitled to (full) political rights (Rosa 2021). Suffrage activists sometimes worked hard to avoid emotional public addresses to avoid the stereotype of hysterical women and chose instead to overcome sentiment with reason and republican arguments (Linkugel 1963). Yet, the struggle for suffrage was an extremely emotional process for those involved, as their experience of anger, injustice, joy, and comradeship motivated the movement (Florin 2009). The different subject position of women, especially married women under the sovereignty of their husbands who could not hold property or enter into contracts, was established and perpetuated by the law. Women's right to vote exemplified the complete legal subjectivity only available to white men.

[5] For more on African American women that engaged in political speech on the issues of slavery, civil rights and women's rights, see, e.g., Logan, 1999.

As such, the demand for women's suffrage is inseparable from the demand for equality in economic rights and equal excess to education and jobs.

This stage of feminist effort is often referred to as the "first-wave feminism," "liberal feminism," or "equality stage of feminism." However, such classification entails oversimplifications, threatens to gloss over the differences within the movement, and contributes to the erasure of some strands of feminism, especially socialist feminisms. Socialist feminists focused on the injustices of the capitalist organization of society and the plight of (working-class) women (DISCH & HAWKESWORTH 2016). Feminism was never a movement; it was a multiplicity of movements from the start. Classification is nevertheless somewhat helpful in understanding how feminist thought and demands developed and dealt with the emotion-reason divide in different historical contexts. The first-wave of feminist engagement was focused mainly on asserting the artificiality of the supposedly natural differences between women and men, building upon the idea of equality of human beings. The argument of a primordial difference between emotional women and rational men was refuted as a product of education and socialization. Formal equality before the law was mostly achieved by the 1960s and 70s, which allowed for a new round of appraisal of different standards, lived experiences, and legal treatments of men and women, as well as the differences among women themselves.

3.5. *Personal is political*

The equality stage consumed itself and was succeeded by the so-called "diversity stage of feminism" or "second-wave" feminisms. Reevaluating the outcomes of the struggle for formal equality, this stage of feminist thought focused on the differences between women and men and the differences among women, veiled over by the legal proclamation of equality. This era of feminist ideas is most famously expressed in the slogan "personal is political," denoting that formal equality before the law did not result in actual equality. Second-wave feminist movements were intertwined with the progressive social movements of the 1960s and 70s, like May 68 confronting the injustice of the capitalist system in Europe and the Civil Rights Movement that challenged racial inequalities in the US, demanding the end of racial segregation, greater economic and social justice, access to education, employment, housing, and other basic provisions.

As we have seen, the reason-emotion divide casts a long shadow over anyone who is incompatible with the ideal model of a reasonable white, educated, heterosexual, wealthy man. The basic premise of first-wave feminist activists was to minimize and refute the difference between genders as an artificial product of upbringing, tradition, and morality. Achieving this formal equality revealed that differences between genders exist and that glossing over them often disadvantages women. How to approach the differences, whether to refute or vindicate them, is an evergreen topic inscribing diversity into feminisms. Simone de BEAUVOIR's (2010 [1949]) *Second Sex* and Betty FRIEDAN's (2013 [1963]) *The Feminine Mystique* were influential works sparking the flame of the second-wave feminisms, as they highlighted the different experiences and limitations that defined the lives of women and inspired a new surge of feminist organizing. Feminist legal theory was also emergent within the Critical Legal Studies (CLS) movement in the USA but was marginalized as a niche topic, which led "Fem-Crits" to distance from the original CLS (WEISBERG 1993).

Topics like reproductive rights, gendered violence, legal (mis)treatment of sexual harassment and rape, discrimination, economic disparities between men and women, unpaid reproductive work women perform at home, and many other issues were highlighted on both sides of the Atlantic. While second-wave feminisms were whitewashed in media and scholarship and are still often presented as a movement of white bourgeois women, women of color and women belonging to lower social classes engaged in feminist organizing as well (ROTH 2003). Various movements shared the core issues and sometimes cooperated – though racial and class dominance patterns

often hindered collaboration. Simultaneously challenging several factors of oppression was never easy: for example, black women often felt forced to choose between the struggles for racial and gender equality, while white leftist women were often ridiculed and exploited by their male counterparts and felt the need to organize an alternative, feminist, leftist agenda.

Sex, traditionally understood as a private and emotional matter, stood out as another territory of inequality whatever its setting, from sexual dynamics in marriage to the false promise of liberation after the sexual revolution, from sexual harassment on the street and at work to reproductive rights. Radical feminism explicitly exposed sex and sexual relations as power relations for the first time in Western history. Furthermore, they pointed out that crimes like rape are not grounded in incontrollable male sexual desire but are instead an expression of power and domination and, thus a tool of patriarchal oppression (BROWNMILLER 2013). Women were often excluded from the debates about contraception and abortion in favor of male (objective) experts, and homosexuality was taboo. This prompted many women to speak out and demand concrete societal and legal transformations (SHULMAN 1980). The victim-blaming culture, for example, was called out. While aggressive male sexuality has long been normalized and a man's sexual history did not necessarily reflect on his reputation, women were (are) judged by different standards. Women construed as hysterical or hypersexual in psychiatric discourse were designated as liars in legal procedures through most of the 20th century (LUNBECK 2003). This intersection of legal and medical discourses in rape trials produced interest in the victim's sexual history, personal reputation, and "unfeminine habits" to protect innocent men from malicious accusations. Framing victims of sexual assaults as inciting and provoking was (is) ingrained in the legal system, police procedures, as well as judicial procedures. Focusing on the victim of gendered violence was (is) often accompanied by disturbing emotional identification of (male) judges with the perpetrators and a tendency to invent explanations and rationalizations for their actions (NEDELSKY 2002 [1997]; LEES 1996; HOWE & ALAATTINOĞLU 2018).

Regarding inequality as perpetuated and created by the law, legal scholar Catharine MacKinnon, one of the most recognizable voices of radical feminism, argued that law is constructed and operates from a male point of view and represents the institutionalized domination of men over women (MACKINNON 1987). She criticized the notion of equality as *equality with men* and rejected such equality as an avenue of enshrining the male perspective and norms as the yardstick for everyone. Others joined her in recognizing law as the symbol of patriarchy and observed how capitalist legal systems contributed to the exclusion of women from public life (RIFKIN 1993). While the law's role in maintaining male dominance made many radical feminists suspicious of its liberatory potential, some legal transformations did occur in the wake of their critique. Sexual harassment and discrimination, for example, found their way into the body of law as actual concepts, while legal treatment of rape underwent a reform.

In this period, the feminist legal theory was developing as a discipline with multiple approaches and threads. It attacked not only the law but also legal method and reasoning. Feminist analysis of judicial decisions revealed that neutrality of legal reasoning, supposedly limited to the relevant facts and the law, often serves as a convenient mask for what judges are actually doing – solidifying commonly accepted stereotypes about women and their place in society, family, and politics (MOSSMAN 1987). Strategic use of commonly known "facts," religious inspirations, and precedents was routinely implemented to protect the status quo from a position of objective authority. Behind the veil of rationality, however, decision-makers' personal experiences, convictions, and emotions often stood out as the deciding factor. This issue continues to be problematized and addressed by feminist legal scholars today. The *Feminist Judgment Project*, for example, aims to point out that judicial decisions are not mathematical equations with one correct solution. The *Feminist Judgment Project* is a collaborative initiative in which legal scholars and practitioners rewrite significant legal decisions from a feminist

perspective. The project deals with cases that were decided against women's interests by writing alternative judgments employing the same legal method, facts, and laws to produce different, feminist decisions (e.g., HUNTER et al. 2010). The *Feminist Judgment Project* was started by the legal scholar Rosemary Hunter in the 2000s, but it is an ongoing project involving many scholars and experts from around the globe. The re-written judgments are published and used as a teaching and research tool and are also intended to influence the development of the law. A common lament of *Feminist Judgment Project* participants nevertheless echoes concerns put forward in the 1980s: that the method and style of argumentation developed for centuries exclusively by and for men are restrictive for women who desire an alternative way of legal reasoning.

The question of difference stands out again. Women identified with the body and their reproductive function were not only perceived as less rational but also more unstable and less moral than men (SMART 1989, 80-113). Male researchers investigating the development of moral reasoning have been using all-male samples to construct a scale of moral development. When they evaluated women utilizing this scale, they concluded that women's moral development is hindered and inferior. Carol Gilligan proposed a different explanation: the all-male samples only investigate the moral reasoning of men without wondering if women's moral reasoning might be different without being inferior (GILLIGAN 1982). While male moral development is identified with autonomy, individual rights, and application of general rules to specific situations, women tend to focus on relationships and the possibility of compromise. Women's supposed lack of detachment and depersonalization was habitually interpreted as a lack of individual agency. Gilligan termed the standardized (male) moral reasoning as the "justice perspective" and traced a strong connection between the justice perspective and the presuppositions of contractualism centered on the autonomous individual with his rights. She termed the alternative style of reasoning the "care perspective" or "ethics of care." She conducted studies using mixed samples – the outcome demonstrated that both men and women use the care perspective in their moral reasoning but, when pressed, tend to replace it with the justice perspective. Most men resort to the justice perspective when prompted, while some women retain the care perspective. She concluded that all-male samples thus erase the care perspective and contribute to its demonization, calling attention to the fact that women's psychology was always considered a mysterious defective version of the male.

The question of whether women reason differently than men was translated into other empirical studies. Previous investigations measured the representation of women in state institutions in numbers; new studies began to ask whether such representation also makes a substantial difference (e.g., MENKEL-MEADOW 2002 [1985]; THOMAS 1995). Affirming the difference between men and women is risky, as it might provide grist for the mill of opponents of women's equality. The difference between "male" and "female" ethics and moral reasoning also finds diverse interpretations within feminisms. The idea that the ethics of care might represent a valid and welcome alternative to the egocentric approach of liberal tradition significantly influenced cultural feminisms.

Cultural feminists celebrated the difference between men and women by asserting that supposed feminine trades like being nurturing, caring, and empathetic are positive, significant, and ought to be celebrated and encouraged rather than disparaged as irrational (KITTAY & MEYERS 1991). Weaker versions of cultural feminisms like Gilligan's allow for the possibility that men might adopt and adjust to some of the women's virtues and thus contribute to a better society. Strong cultural feminisms, in contrast, elevate the difference to the point of demanding women's segregation from men and the creation of their own cultural world. Such ideas find expression in, for example, Adrianne Rich's "lesbian continuum" and her critique of the institution of "compulsory heterosexuality" (RICH 1980). On her account, women could only reclaim their identities and be freed from patriarchal control if they embraced lesbianism as not only sexual orientation but a true bond of love and friendship.

For some, the ideas of strong cultural feminisms that connect different moral reasoning with biological sex or human anatomy come uncomfortably close to the 19th century conceptions of women and femininity (WILLIAMS 1989; FRUG 1993). The idea that care perspective is somehow innate to women and marks them as different from men was criticized as failing to recognize that the difference in moral reasoning is a product of subordination and not merely gender (TRONTO 1993). Gilligan's sampling was charged for mainly including white middle-class subjects. Subsequent studies with more diverse samples revealed that men of color and lower social classes also tend to rely more on the care perspective than privileged white men (TRONTO 1987; HARDING 1991). It is worth noting that Gilligan's studies refuted biological determinism and pointed out that the moral reasoning of most people operates along the lines of both care and justice perspectives. Her work largely departed from developmental psychologist Nancy Chodorow who traced the differences between boys' and girls' development through their different inclusion in child-care rather than through biological differences between the sexes (CHODOROW 1974).

Again, education and upbringing seem to play an important role, as personal development is a part of the intersubjective process through which human beings learn their place and the (ir)relevance of their voices, minds, and ideas in relation to others. The messages women get from early childhood and throughout their lives thus profoundly influence their sense of self and contribute to the "different voice" in which they might speak and reason (BELENKY et al. 1986). French historian Elisabeth Badinter, for example, challenged the idea that women-mothers are by definition nourishing and selflessness (BADINTER 1982). Through her exploration of the history of maternal indifference, she illustrated many mothers in the past centuries saw their children as a nuisance and did not seem to care greatly about their faith. Motherly love and portrayal of women as naturally loving and caring thus appears to be a theoretical construct without a universal history.

An alternative account of difference can be found in the works of European feminists, with French feminism standing out as one of its more notable instances. Sexual difference was also at the forefront of their discussions, yet French feminists attempted a deconstruction of the subject as such, even the celebrated female subject. While second-wave feminist activism was very influential in continental Europe and the USA, French feminisms or poststructuralist feminisms are considered more theoretical compared to the 1980s USA feminist scholarship. Primarily growing out of a critique of Freudian and Lacanian psychoanalysis and French poststructuralist philosophy, French feminisms rejected the essentialization of the Women as the Other, either erased from historiography or theorized as an eternal victim.

The proponents of French feminisms adopted diverse theoretical approaches but can be linked through their emphasis on the role of language and its binaries in constructing subjectivities. Rather than advocating separatism or searching for the mythical femininity, theorists like Hélène CIXOUS (1976), Luce IRIGARAY (1991 [1984]), and Julia KRISTEVA (2019) searched for a (re)construction of ethics that would allow men and women to live together beyond the artificial dualisms. Such binaries are, after all, forever haunted by the residue that does not fit into either pole. The reason-emotion pair in the history of Western thought is complicated, as reason is often used to legitimize the dismissal of women, savages, and other marginalized groups. Ironically, such logic is usually based on the emotional inclinations of great male thinkers, established as the privileged subjects of knowledge, from Plato to Freud (IRIGARAY 1985). The idea of alternative ethics demands a move away from celebrating either of the poles of a given binary and points towards the possibility of love and a different reality.

3.6. *Emotions, experiences, epistemologies, knowledges*

As hinted above, gender is only one of the markers of oppression that profoundly marks an individual's place in society. Despite the presence of women of color in emancipatory movements of the 20th century, they were often marginalized: feminism was busy constructing

its essentialist category of the Woman on the template of white, middle-class, heterosexual women, while racial equality movements focused on the grievances of men of color, sidelining the feminist concerns. While the fact that race, class, sexual orientation, and other circumstances profoundly impact a woman's experience was consistently recognized by some feminist thinkers, coherent scholarship on the topic was lacking. A powerful critique of second-wave feminisms was beginning to flourish in the works of intersectional, decolonialist, and postmodernist feminists.

Kimberlé Crenshaw famously elaborated on the legal implications of intersectional discrimination in the late 1980s (CRENSHAW 1989). She problematized how the law treats gender and race as separate eventualities. The interplay of these factors, which is far more complicated than a simple sum of two types of discrimination, is thus overlooked, leaving women of color without legal protection in cases where bias is rooted precisely in their overlapping identities. Women of color were unable to prove that they were discriminated against *as women of color* since the legal system only recognized either racial or gender discrimination. Sensitivity to the complexity of the "matrix of domination" (HILL COLLINS 1998 [1990]) as a web of interlocking hierarchies posed a severe challenge to white liberal feminism. Intersectionality as a methodological framework quickly grew and became widely applied in critical and feminist legal studies and practice (MACDONALD et al. 2005). The struggle of intersectional feminisms in the legal domain is far from complete, as marginalized groups of women continue to fight to be heard instead of spoken for by the more powerful social actors (e.g., SHAUKAT 2020).

Beyond the law, the complexity of identity in terms of gender, race, class, age, sexual orientation, nationality, postcoloniality, ability, etc. challenged the monolith representation of women in liberal feminist tradition (SGRABHAM et al. 2008). Instead of the reformative stance of liberal feminists, intersectional feminists like Bell HOOKS (2014 [1981]), Patricia HILL COLLINS (1998 [1990]), Audre LORDE (2020 [1984]), Gloria ANZALDÚA (2012 [1987]), Angela DAVIS (1983), and many others adopted a more revolutionary stance. Their important works confronting the complexity of personal experience are often grounded in emotions, the experience, and defiance of women facing marginalization on several fronts. Intersectional feminism announced the collapse of the category of Woman as the core of feminist engagement. It exposed the fluidity and individuality of personal identities, baring the diversity silenced in dualistic thinking, which reconstructs the abundance in two opposing poles.

Shamelessly admitting that one's emotions and experience shape one's perspective and inform one's knowledge is blasphemous in modern positivist science. Yet, as we have seen, feminist thinkers have long been suspicious of the objective reason narrative. Elisabeth Spelman, for example, refuted the positivist "view from nowhere" (referring to NAGEL 1989) and the related distrust of emotion as the "Dumb View" that ought to be replaced by different epistemology (SPELMAN 1990 [1988]). Such transformation should not only involve a displacement of male exclusion of women but should also involve the concerns of race and class excluded by white middle-class feminisms and seriously consider the pain women are inflicting on each other. According to Spelman, emotions are integral to politics, as she demonstrated through a survey of appropriations of suffering (of others) in philosophy, art, and feminist activism and theory (SPELMAN 1998). She proposed that suffering is not just a negative emotion but an open-ended potentiality that could trigger political transformations.

When an individual's emotions do not correspond to their assigned social role, the order of things is threatened. Alison Jaggar concentrated on the epistemic value of such "outlaw emotions" (JAGGAR 1989). Pointing out that positivist accounts tend to identify and conflate emotion with feeling – an involuntary physical sensation – she proposed that emotion and reason are hopelessly entangled. Rather than banishing emotions, Western science tends to suppress them and remains blind to their contribution to its allegedly objective concepts.

Embracing race, gender, and class distinctions as important aspects of emoting-reasoning, Jaggar focused on outlaw emotions, exemplified by, for example, the fear of a woman experiences when a man exposes her to sexual harassment intended as a compliment. The expected emotion in the patriarchal cultural script would be happiness and gratitude expressed by the woman's smile, while fear and anger are seen as the wrong responses. Such outlaw emotions of women, people of color, and other marginalized groups are constitutive of both critical theory and practice. In this context, it is worth mentioning how women and/or people of color often feel crazy because their emotional reactions seem so out of line with the regime of expected normalcy (e.g., SHULMAN 1980; HOOKS 1995).

Marxist feminist theory is another important thread in intersectional and standpoint feminisms. While Marx and Engels assumed women's reproductive work in the household as a given, their analysis of class consciousness is attractive to many feminists who set to inform Marxism's gender-blind spot. Christine Delphy, for example, theorized sexual division of labor and framed women as a class (DELPHY 1993; see also BARRETT & MCINTOSH 1979). Analyzing the structural differences between men's and women's lives – with close attention to the fact that neither group is homogenous in terms of race, class, and other markers of oppression – Nancy Hartsock elucidated women's "double day" (HARTSOCK 2004 [1983]). Women's days are composed of a job outside the home, emotional labor consisting of managing and negotiating the feelings of others and performing certain emotions (like empathy, nurturing, loving, etc.), and domestic work. Much like the proletariat, women are not simply passive victims but active participants and creators of their oppression in patriarchal capitalism.

Intersectional and standpoint feminisms envelop a radical epistemological claim that the oppressed groups know differently and that what they know is not meaningless and devoid of reason but silenced knowledge. The idea that knowledge is a product of social location or standpoint points towards the need for epistemological pluralism that makes Western science uneasy. According to reason-emotion, mind-body, and subject-object divides, scientific knowledge is produced by the subject (the knower) investigating the object of cognition (the known). The knower must disassociate from their personal experience, feelings, and emotions to build objective knowledge representing but a neutral description of the known. Thus, knowers are presupposed to function as generic fungible subjects adopting a universal perspective and speaking in the name of the universal Truth. As we have seen, judicial decision-making is ideally imagined to work similarly. As elaborated so far, the original subject-knower implies a (white, affluent, etc.) man, the claim to his universality notwithstanding. While many women work(ed) hard to comply with this model to prove they are capable of rational thought, others reject such epistemic domination.

Challenging the seemingly neutral epistemology of modern science is thus one of the most important endowers of feminist critique. Through this critique, ideological implications of positivist objectivity are revealed. The professed neutral description emerges as highly normative, creating the world and social roles we play in it. Unsurprisingly, feminists have long been suspicious of liberal ideology such as Rawls' "veil of ignorance," a mental experiment in which a person must choose a society to live in without knowledge of the identities with which they will have to navigate it. Rawls' speculation is based on a presupposition of egoistic self-interest as a given. His consequent affirmation of Western capitalist liberal society as the best possible world indeed implies that the person hiding behind the veil might be Rawls himself, as proposed by Mari MATSUDA (1986). The disembodied universal subject thus emerges, time and again, as ideological construction in service of perpetuating the preeminence of a specific point of view.

Standpoint theorizing departs from a reminder that the positivist objective knower always-already had an agenda and a standpoint of his own. The presuppositions supporting the supposedly objective knowledge were never thoroughly examined as they maintain the dominance of privileged groups. Rethinking these presuppositions is everyone's task. Ironically,

the objectivity of science might be strengthened if the standpoint of the knower was seriously considered (HARTSOCK 2004 [1983]). Donna Haraway developed one possible expression of such objectivity through a refusal of the universal subject with an infinite vision (HARAWAY 1988). Instead of the customary model of objectivity, she proposed feminist objectivity, which does not pretend to be universal and all-knowing. Instead of promising transcendence, feminist objectivity is a bricolage of partial, localized, and diverse knowledges, less prone to exclusions and oppositional thought, and grounded in embodied objectivities. The break with mind-body ideology promises to close the gap between the knower and the known – the subject and the object – as two distinct entities where the former is active and the latter passive. Situated knowledges are thus marked by a passionate detachment and a hope for transformation.

Standpoint theorizing significantly contributed to the revaluation of epistemology as a value-neutral practice of a disembodied individual knower. Nevertheless, it is only one mode of feminist theorizing. Helen Longino criticized the lure of standpoint theorizing's excessive relativism and its inability to determine which of the incompatible standpoints ought to prevail, suggesting the need for democratic consensus on what knowledge should be considered valid (LONGINO 1992). Nonwestern postcolonial feminists might be reluctant to give up entirely on positivism and the related concept of individual rights, forced as they are to navigate both the colonialist attitudes of the Westerns and the oppression of women within their communities. Chandra Mohanty traced racist discourses of some Western feminists that other Nonwestern "third world women" according to the binary logic of Western as progressive and Nonwestern as uncivilized (MOHANTY 1988). Uma Narayan warned against romanticizing the "epistemic privilege" of the oppressed, as such narratives risk overlooking the complexities and darker sides of inhibiting marginalized perspectives (NARAYAN 2004 [1989]). While most feminists agree that transnational feminism is necessary, universalization of the Western patriarchal structures is widely criticized as counterproductive.

3.7. *Gender trouble*

The influences of French feminisms, poststructuralist philosophy, and intersectional feminisms are all at work in Judith Butler's theory of gender performativity (BUTLER 2008 [1990]; BUTLER 2011 [1993]). Departing from de Beauvoir's famous proclamation that one is not born but instead becomes a woman, Butler took issue with the sex-gender distinction in feminist thought. Attempting to both represent women and escape the narrative that women qua women are defined by their biology, the concept of gender as a cultural interpretation of sexed bodies became prominent in feminist theory. Yet, the sex-gender binary is no less problematic than other hierarchical oppositions, as it presupposes biological sex as an immutable passive given on which gender is inscribed. Butler revolutionized feminist and queer theory by arguing that sex, too, is a discursive phenomenon and that material sexed bodies are created through numerous repetitions of sex/gender norms. Eliminating the idea that there is an identity behind an individual's expression of gender, she focused on gender as performance. Gender performativity does not imply that gender is not real or is a free choice, but rather that gender is power materializing. According to Butler, gender is a set of highly regulated practices that create intelligible bodies conforming to the heterosexual matrix of men/masculine and women/feminine. All bodies resist this process – hence the need for the repetitive reassertion of "the law of sex" – yet some fail completely by not corresponding to the demanded coherence between sex, gender, desire, and sexual practice. Rather than spinning in the vicious circle of binary sex-gender imaginary, Butler proposed relinquishing the essential woman as the totalizing subject of feminisms and engaging with diverse "marginal genders" that might offer a line of escape, a subversive confusion of the fixed categories.

As we have seen, feminist thought developed as a resistance to hierarchical oppositions like male-female, mind-body, reason-emotion, and public-private. It is hardly surprising that resistance

to oppressive dualisms remained entrenched in oppositional thinking and took categories like men-women, equality-difference, and sex-gender as stable references. Dualisms were constantly questioned, yet their logic was oft repeated. Diverse strands of critique of second-wave feminist theorizing like intersectional, decolonial, and poststructuralist feminisms exposed the limitations of this approach: oppression does not function on a single axis nor is it composed of several parallel axes. Instead, markers of oppression interact and intertwine. The Other, excluded from the concept of full humanity, cannot be articulated as a homogenous entity. Butler's thesis that the body is not a given object inscribed by cultural interpretations of race, gender, disability, etc. exposes the presuppositions of feminist thought that limit feminist theory and politics.

Abandoning the search for the quintessential woman allows feminist thought to respond to a broader range of oppressions more inclusively. In postmodernity, the world and gender relations have undergone various transformations and feminisms had to respond to multiple issues from (de)colonialization to neoliberal capitalism, ecology to queer liberation, and new technologies to increased cultural diversity (GILLIS et al. 2007; BUDGEON 2011; DAVIES 2018). Feminism and gender equality became a part of political and legal jargon in international and national legal systems. Feminist jurisprudence has developed into a discipline encompassing many strains. Liberal feminist jurisprudence focused on individual rights is still the most visible and influential, yet in constant dialog and friction with radical, cultural, intersectional, queer, decolonial, and postmodern feminist critique of the law's complicity in (re)creating inequalities and injustices.

Nevertheless, emotions and emotional responses are still racialized and gendered in legal proceedings. While men of color have long been understood as emotional, (white) men were (are) socialized to repress and hide their emotions (FREVERT 2013, 87-147; DE BOISE & HEARN 2017; LEE 2003). Some of the few emotions that white men could freely express are anger and jealousy, which are often rationalized in legal proceedings. Murder can be rearticulated as voluntary manslaughter if the man argues that he was enraged by, for example, his wife's infidelity (FELDMAN BARRETT 2017, 225-228). Outwardly expressed anger of a (black) woman or men of color, on the other hand, is treated as deviant and problematic (FELDMAN BARRETT 2017). If women are expected to be emotional and caring – whether this is used to present them as inferior or to build them up as ethically superior – (white) men are expected to be cold, rational beings. It is important to note that despite the landmark transformations of legal systems, women, people of color, queer and transgender people, disabled, elderly, Nonwestern, indigenous, poor, and many others still struggle to feel genuinely protected and heard in legal procedures. At the same time, climate change, rampant wars, and technological developments demand adequate political and legal responses. Feminist jurisprudence thus remains a strand of critical legal theory and retains its explicitly normative outlook.

Some designate the shifts in feminist theorizing since the 1990s as "third-wave feminisms." Marked by fragmentations and diversity, this wave, like all attempts at classifying feminisms, is a contested concept. Indeed, feminist critique is always multiple and diverse, and the periodization in waves does violence to its varieties and (dis)continuities. Liberal feminists are still around and did not disappear with the first-wave feminism. Women of color have long understood that their experience differs from that of white women and did not suddenly realize this only as a response to the second-wave. Queer feminists had sought the lines of flight within the system of gender relations in heteronormative societies long before homosexuality and transgenderism were (largely) decriminalized, demedicalized, and allowed to appear in public. Most feminist thinkers are not enthusiastic about adopting labels of different genres or waves of feminism ascribed to them. The impossibility of clearly defining the borders and scopes of different waves of feminism is illustrative of the impossibility of delimitations and closure so often exposed by feminist thinkers. When it comes to the feminist theorizing of the past decades, new materialist approaches and the affectual turn are the most interesting when considering the reason-emotion divide.

3.8. *Affectual turn*

Postmodernism, poststructuralism, and deconstruction dominated critical and feminist thought in the late 20th century. These critical readings of subjectivity, representation, knowledge, discourse, epistemology, and culture are sometimes labeled as the "linguistic" or "textual turn" in social sciences and humanities. Linguistic turn has been superseded in recent decades by the "affectual turn" (Greco & Stenner 2013; Clough & Halley 2007). Affectual turn is responding to the wider "emotionalization of society" with a renewed interest in affects, feelings, and emotions (Pedwell & Whitehead 2012). Affectual turn problematizes the treatment of emotions and affects as discursive phenomena and is intimately connected with the rise of new materialist scholarship. New materialisms focus on non-human agency and the affect, aim to challenge dualisms like nature-culture, human-nonhuman, mind-matter, and some such, and are grounded in critical engagement with empirical natural sciences (see, e.g., Barad 2007; Bennett 2010; Coole & Frost 2010; Braidotti 2006). Renewed interest in affect and emotions is thus accompanied by a strong emphasis on the material, the bodily, and becoming, as well as by a shift in methodological and onto-epistemological orientation in critical scholarship. In social sciences and humanities, affective turn is often presented as a revolutionary shift of focus, as the discovery of emotions by law and emotion scholarship exemplifies.

While affectual turn announces a novel approach to affect and emotion, interest in the bodily and the emotive as political forces itself is far from new. Scholars like Elizabeth Grosz (2004) and Eve Sedgwick (Sedgwick & Frank 2003) stressed the exciting new horizon beyond dualistic thought opened up by new materialisms and suggested that previous (feminist) scholarship neglected biology, body, and nature. Many others, like Clare Hemmings (2005) and Sara Ahmed (2008), warned against such a flattening down of inherited feminist narratives to present the study of affect as new and groundbreaking. Whether or not feminist thinkers associated with the affectual turn construct their theories with or against their predecessors, their work is, in a way, a continuation of the destabilization of the separation of reason and emotion traced in this chapter. In feminist scholarship, emotions and affects are perceived as crucial for critical revaluation and reconstruction of politics and ethics in the gap between the public and private.

Affect, the key concept of the affectual turn, has no fixed definition and is mainly perceived as distinctive from emotion. Some authors clearly distinguish between pre-subjective bodily affect and culturally mediated emotions, while others perceive affect precisely as an entanglement of biology and culture (Liljesröm 2016). Thus, affect can be understood as an assemblage of pre-individual physical and life forces not limited to human beings, a relational quality of affecting and being affected, an excess that escapes the rational over-coding and thus as a free and potentially transformative force. While potentially transformative, affect also plays a part in the (re)production of social hierarchies and oppressions, as our emotional attachments to social norms ensure their durability (Butler 1997). Specific emotions and feelings are often the points of departure for a feminist philosophical critique.

Queer and feminist scholar Ann Cvetkovich investigated the political effects of affective expression and sensational representations on the example of Victorian sensation novels to challenge the idea that the expression of feelings is a path to liberation (Cvetkovich 1992). She disputed the medical and clinical discourses on trauma as based on a strict separation of the public and private, erasing the experiences of women and queer people (Cvetkovich 2003). She continues this line of argument in her work on depression as a cultural, social, and political phenomenon rooted in capitalist exploitation and systemic racism and sexism rather than an individual biochemical imbalance (Cvetkovich 2012; Cvetkovich & Michalski 2021). Sianne Ngai investigated the interlocking between (lack of) agency and emotions on the examples of "ugly feelings" like irritation, paranoia, envy, and disgust to flesh out the racialized and gendered implications of cultural artifacts (Ngai 2005). Ranjana Khanna focused on melancholia and

psychoanalysis, and proposed the notion of affect as an interface in cultural production (KHANNA 2012). Sara Ahmed also approaches emotions as cultural and political practices rather than individual states of being (AHMED 2004). In her work on happiness, she reconstructed the intellectual history of this presumably positive emotion to demonstrate its multifacetedness (AHMED 2010). The imperative to be happy or to make others happy influences people's choices and lives. Happiness can be used to justify oppression, and the revolt against oppression might cause unhappiness. Adding to explorations of complexities of supposedly positive emotions and affects, Lauren Berlant's engagement with cruel optimism theorizes how people cope and survive amid the crisis of the neoliberalist economy and shuddering personal relationships (BERLANT 2011). She theorized cruel optimism as the attachments that sustain the fantasy of the "good life" and simultaneously cause pain and injury. The desire for a good life itself is thus an obstacle to personal flourishing: hard work no longer guarantees financial stability, for example. The complexity of affects, feelings, emotions, sentiments, and their role in society thus continues to be scrutinized with renewed ardor and diverse and innovative methodological approaches.

4. *Conclusions*

This chapter outlined the long and diverse engagement with the reason-emotion binary in feminist theory. This epistemic binary was consolidated in the age of the Enlightenment, an era of paradoxes. While Enlightenment is celebrated for advances in science and political liberalism with values of equality, democracy, and the rights of men, Enlightenment was also an era of European colonialization, racialized slavery, and exclusion of women from the public sphere. While emancipatory movements achieved formal legal equality for all people – regardless of gender, race, and other attributes, the legacy of oppositional thinking remains entrenched in our collective imaginations. In feminist critiques of the reason-emotion binary, the hierarchical relationship between the two poles is usually highlighted. Reason is traditionally perceived as superior to uncontrollable emotions and reason is identified with maleness and whiteness. Such hierarchical definitions are not neutral descriptive tools and cause effects far beyond theoretical discussions. While theories of emotions are becoming more complex and reject the emotion-reason dualism, it is essential to remember that this dualism is not just a naïve epistemological relic of the past but a political tool that played an instrumental role in constructing our lived realities. Moreover, this dualism, especially if not scrutinized, continues to haunt scientific discussions and affect countless lives.

In the legal and political domain, the reason-emotion binary was employed to justify slavery, colonialism, and other social hierarchies, severely limiting the legal rights of women, people of color, and other marginalized groups. Access to education and public office was thus long precluded for devaluated groups. This chapter traced resistance to the reason-emotion binary in feminist thought to highlight the effects of this dualism and the struggle for its destabilization in the spheres of political critique, ethics, law, society, epistemology, and science. The multiplicity of feminisms, their internal controversies and contradictions also emerged through the chapter, illustrating the complexity of human organization of society and knowledge, as well as the interrelations of the two. While feminisms are diverse in their approaches and scopes, they share an ideal of equality (however imagined) and a suspicion toward epistemic binaries. Feminist critique is a reminder that professed neutrality and objectivity often gloss over problematic presuppositions, and that declared scientific or juridical disengagement and objectivity repeatedly import underlying stereotypes. These lessons are fundamental to (interdisciplinary) legal research, stressing the importance of carefully examining scientific and theoretical narratives. Understanding the extent and implications of the reason-emotion binary thus contributes toward critical scrutiny of received knowledge and seemingly self-evident facts in legal theory and practice.

References

Abrams K., Hila K. 2010. *Who's Afraid of Law and the Emotions?*, in «Minnesota Law Review», 94, 6, 1997 ff.

Ahmed S. 1995. *Deconstruction and Law's Other: Towards a Feminist Theory of Embodied Legal Rights*, in «Social & Legal Studies», 4, 1, 55 ff.

Ahmed S. 2004. *The Cultural Politics of Emotion*, Edinburgh University Press.

Ahmed S. 2008. *Open Forum Imaginary Prohibitions: Some Preliminary Remarks on the Founding Gestures of the 'New Materialism'*, in «European Journal of Women's Studies», 15, 1, 23 ff.

Ahmed S. 2010. *The Promise of Happiness*, Duke University Press.

Ahmed S. 2014. *Cultural Politics of Emotion*, Edinburgh University Press.

Anzaldúa G. 2012. *Borderlands / La Frontera: The New Mestiza* (4th Ed.), Aunt Lute Books. (Originally published 1987)

Atherton M. 2018. *Cartesian Reason and Gendered Reason*, in Louise A., Witt C. (eds.), *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, Routledge, 21 ff.

Badinter E. 1982. *The Myth of Motherhood: An Historical View of the Maternal Instinct*, Souvenir Press Ltd.

Bandes S. (ed.) 1999b. *The Passions of Law*, New York University Press.

Bandes S. 1999a. Introduction. In Id. (ed.) *The Passions of Law*, New York University Press, 1ff.

Barad K. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*, Duke University Press Books.

Barrett M., McIntosh M. 1979. *Christine Delphy: Towards a Materialist Feminism?*, *in* «Feminist Review», 1, 95 ff.

Belenky M.F., Mcvicker Clinchy B., Rule Goldberger N., Mattuck Tarule J. 1986. *Women's Ways of Knowing: The Development of Self, Voice, and Mind*, Basic Books.

Bennett J. 2010. *Vibrant Matter: A Political Ecology of Things*, Duke University Press Books.

Berlant L. 2011. *Cruel Optimism*, Duke University Press.

Blanchette I., Caparos S. 2013. *When Emotions Improve Reasoning: The Possible Roles of Relevance and Utility*, «Thinking & Reasoning», 19, 3-4, 399 ff.

Braidotti R. 2006. *Transpositions: On Nomadic Ethics*, Polity.

Brody L.R., Hall J.A., Stokes L.R. 2016. *Gender and Emotion: Theory, Findings, and Content*, in Feldman Barrett L., Lewis M., Haviland-Jones J.M. (eds.) *Handbook of Emotions, Fourth Edition*, Guilford Publications, 369 ff.

Brownmiller S. 2013. *Against Our Will: Men, Women and Rape*, Open Road Media. (Originally published in 1975)

Budgeon S. 2011. *Third-Wave Feminism and the Politics of Gender in Late Modernity*, Palgrave Macmillan.

Butler J. 1997. *The Psychic Life of Power: Theories in Subjection*, Stanford University Press.

Butler J. 2008. *Gender Trouble: Feminism and the Subversion of Identity*, Routledge. (Originally published in 1990)

Butler J. 2011. *Bodies That Matter: On the Discursive Limits of Sex*, Routledge. (Originally published in 1993)

CAREY D., FESTA L. 2009. *Some Answers to the Question: 'What is Postcolonial Enlightenment?'*, in ID., *The Postcolonial Enlightenment: Eighteenth-century Colonialism and Postcolonial Theory*, Oxford University Press, 1 ff.

CHAMALLAS M. 2013. *Introduction to Feminist Legal Theory*, Wolters Kluwer Law & Business.

CHODOROW N. 1974. *Family Structure and Feminine Personality*, in ZIMBALIST ROSALDO M., LAMPHERE L., BAMBERGER J. (eds.), *Women, Culture and Society*, Stanford University Press, 43 ff.

CIXOUS H. 1976. *The Laugh of the Medusa*, in «Signs: Journal of Women in Culture and Society», 1, 4, 875 ff.

CLOUGH P.T., HALLEY J. (eds.) 2007. *The Affective Turn: Theorizing the Social*, Duke University Press Books.

COOLE D., FROST S. (eds.) 2010. *New Materialisms: Ontology, Agency, and Politics*, Duke University Press Books.

COTTERRELL R. 2018. *Law, Emotion and Affective Community*, «SSRN Scholarly Paper». Available on: https://papers.ssrn.com/abstract=3212860.

CRENSHAW K. 1989. *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics,* in «University of Chicago Legal Forum», 140, 139 ff.

CROZIER-DE ROSA S. 2021. *Emotions and Empire in Suffrage and Anti-suffrage Politics: Britain, Ireland and Australia in the Early Twentieth Century*, in HUGHES-JOHNSON A., JENKINS L. (eds.), *The Politics of Women's Suffrage*, University of London Press, 309 ff.

CROZIER-DE ROSA S. 2014. *Shame and the Anti-Suffragist in Britain and Ireland: Drawing Women back into the Fold?,* in « Australian Journal of Politics & History», 60, 3, 346 ff.

CVETKOVICH A, MICHALSKI K. 2021. *The Alphabet of Feeling Bad Now*, in HOLLENBACH J., ALEX MCDONALD R. (eds.), *Re/Imagining Depression: Creative Approaches to "Feeling Bad"*, Springer International Publishing.

CVETKOVICH A. 1992. *Mixed Feelings: Feminism, Mass Culture, and Victorian Sensationalism*, Rutgers University Press.

CVETKOVICH A. 2003. *An Archive of Feelings: Trauma, Sexuality, and Lesbian Public Cultures*, Duke University Press.

CVETKOVICH A. 2012. *Depression: A Public Feeling*, Duke University Press.

DAMASIO A. 2005. *Descartes' Error: Emotion, Reason, and the Human Brain*, Penguin.

DASSORI, E. 2005. *Performing the Woman Question: The Emergence of Anti-suffrage Drama*, in «ATQ: 19th Century American Literature and Culture», 19, 4, 301 ff.

DAVIES E.B. 2018. *Third Wave Feminism and Transgender: Strength through Diversity*, Routledge.

DAVIS A.Y. 1983. *Women, Race & Class*, Knopf Doubleday Publishing Group.

DE BEAUVOIR S. 1953. *The Second Sex*, Vintage Books. (Originally published in 1949)

DE BOISE S., HEARN J. 2017. *Are Men Getting More Emotional? Critical Sociological Perspectives on Men, Masculinities and Emotions*, in «The Sociological Review», 65, 4, 779 ff.

DE GOUGES O. 1979. *Declaration of the Rights of Woman*, in GAV LEVY D., BRANSEN APPLEWHITE H.B., DURHAM JOHNSON M. (eds.), *The French Revolution and Human Rights: A Brief Documentary History*, University of Illinois Press, 87 ff.

DELPHY C. 1993. *Rethinking Sex and Gender*, « Women's Studies International Forum », 16, 1, 1 ff.

DEVOR A. 1989. *Gender Blending; Confronting the Limits of Duality*, Indiana University Press.

DISCH L., HAWKESWORTH M. 2016. *Feminist Theory: Transforming the Known World*, in ID., *The Oxford Handbook of Feminist Theory*, Oxford University Press, 1 ff.

FAUSTO-STERLING A. 2000. *Sexing the Body: Gender Politics and the Construction of Sexuality*, Basic Books.

FELDMAN BARRETT L. 2017. *How Emotions are Made: The Secret Life of the Brain*, HarperCollins.

FELDMAN BARRETT L., LEWIS M., HAVILAND-JONES J.M. (eds.) 2016. *Handbook of Emotions* (4th Ed.), Guilford Publications.

FINE C. 2013. *Is There Neurosexism in Functional Neuroimaging Investigations of Sex Differences?*, in «Neuroethics», 6, 2, 369 ff.

FISCHER C. 2016. *Feminist Philosophy, Pragmatism, and the "Turn to Affect": A Genealogical Critique*, in «Hypatia», 31, 4, 810 ff.

FLORIN C. 2009. *Heightened Feelings! Emotions as 'Capital' in the Swedish Suffrage Movement*, in «Women's History Review», 18, 2, 181 ff.

FREVERT U. 2013. *Emotions in History – Lost and Found*, Central European University Press.

FRIEDAN B. 2013. *The Feminine Mystique*, W. W. Norton & Company. (Originally published in 1963)

FRIEDLAND S. 2019. *Fire and Ice: Reframing Emotion and Cognition in the Law*, in « SSRN Scholarly Paper». Available on: https://papers.ssrn.com/abstract=3386184.

FRUG M.J. 1993. *Progressive Feminist Legal Scholarship: Can We Claim 'A Different Voice'?*, in ID., *Postmodern Legal Feminism*, Routledge.

GEERTZ C. 1975. *Common Sense as a Cultural System*, in «The Antioch Review» 33, 1, 5 ff.

GENDRON M., FELDMAN BARRETT L. 2019. *A Role for Emotional Granularity in Judging*, in « Oñati Socio-Legal Series», 9, 5, 557 ff.

GILLIGAN C. 1982. *In a Different Voice: Psychological Theory and Women's Development*, Harvard University Press.

GILLIS S., HOWIE G., MUNFORD R. (eds.) 2007. *Third Wave Feminism: A Critical Exploration*, Palgrave Macmillan.

GRABHAM E., COOPER D., KRISHNADAS J., HERMAN D. (eds.) 2008. *Intersectionality and Beyond: Law, Power and the Politics of Location*, Routledge-Cavendish.

GRECO M., STENNER P. (eds.) 2013. *Emotions: A Social Science Reader*, Routledge.

GROSSI R. 2019. *Law, Emotion and the Objectivity Debate*, in «Griffith Law Review», 28, 1, 23 ff.

GROSZ E. 2004. *The Nick of Time: Politics, Evolution, and the Untimely*, Duke University Press.

HARAWAY D. 1988. *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, in «Feminist Studies», 14, 3, 575 ff.

HARDING S. 1991. *The Curious Coincidence of Feminine and African Moralities: Challenges for Feminist Theory*, in FEDER KITTAY E., MEYERS D.T. (eds.), *Women and Moral Theory*, Rowman & Littlefield Publishers, 296 ff.

HARTSOCK N. 2004. *The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism*, in HARDING S.G. (ed.), *The Feminist Standpoint Theory Reader: Intellectual and Political Controversies*, Psychology Press, 35 ff. (Originally published in 1983)

HEMMINGS C. 2005. *Invoking Affect*, in «Cultural Studies», 19, 5, 548 ff.

HERMAN B. 2018. *Could It Be Worth Thinking About Kant on Sex and Marriage?*, in ANTONY L., WITT C. (eds.), *A Mind Of One's Own: Feminist Essays On Reason And Objectivity*, Routledge, 53 ff.

Hill Collins P. 1998. *Fighting Words: Black Women and the Search for Justice*, University of Minnesota Press. (Originally published in 1990)

Homiak M.L. 2018. *Feminism and Aristotle's Rational Ideal*, in Antony L., Witt C. (eds.), *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, Routledge, 3 ff.

Hooks b. 1995. *Feminism: Crying for Our Souls*, in «Women & Therapy», 17, 1-2, 265 ff.

Hooks b. 2014. *Ain't I a Woman: Black Women and Feminism*, Routledge. (Originally published in 1981)

Howe A., Alaattinoğlu D. 2018. *Contesting Femicide: Feminism and the Power of Law Revisited*, Routledge.

Hunter R., McGlynn C., Rackley E. (eds.) 2010. *Feminist Judgments: From Theory to Practice*, Hart Publishing.

Irigaray L. 1985. *Speculum of the Other Woman*, Cornell University Press.

Irigaray L. 1991. *Sexual Difference*, in Moi T. (ed.), *French Feminist Thought: A Reader*, Wiley-Blackwell, 118 ff.

Israel J.I. 2002. *Radical Enlightenment: Philosophy and the Making of Modernity 1650-1750*, Oxford University Press.

Jaggar A.M. 1989. *Love and Knowledge: Emotion in Feminist Epistemology*, in «Inquiry», 32, 2, 151 ff.

James C.A. 1997. *Feminism and Masculinity: Reconceptualizing the Dichotomy of Reason and Emotion*, in «International Journal of Sociology and Social Policy», 17, 1/2, 129 ff.

Keller Fox E. 1985. *Reflections on Gender and Science*, Yale University Press.

Khanna R. 2012. *Touching, Unbelonging, and the Absence of Affect*, in «Feminist Theory», 13, 2, 213 ff.

Kittay E.F., Meyers D.T. (eds.) 1991. *Women and Moral Theory*, Rowman & Littlefield Publishers.

Kleingeld P. 2019. *On Dealing with Kant's Sexism and Racism*, in « SGIR Review», 2, 2, 3 ff.

Kristeva J. 1996. *Women's Time*, in Garry A., Pearsall M. (eds.), *Women, Knowledge, and Reality*, Routledge, 61 ff. (Originally published in 1979)

Kristeva J. 2019. *Passions of Our Time*, Columbia University Press.

Krueger, C.L. 1992. *The Reader's Repentance: Women Preachers, Women Writers, and Nineteenth-Century Social Discourse*, University of Chicago Press.

Lamb-Books B. 2016. *Angry Abolitionists and the Rhetoric of Slavery: Moral Emotions in Social Movements*, Springer International Publishing.

Laqueur T. 1992. *Making Sex: Body and Gender from the Greeks to Freud*, Harvard University Press.

Le Doeuff M. 1991. *Women and Philosophy*, in MOI T. (ed.), *French Feminist Thought: A Reader*, Wiley-Blackwell, 181 ff. (Originally published in 1977)

Lee C. 2003. *Murder and the Reasonable Man: Passion and Fear in the Criminal Courtroom*, New York University Press.

Lees S. 1996. *Ruling Passions: Sexual Violence, Reputation and the Law*, Open University Press.

Lengermann P.M., Niebrugge-Brantley J. 1997. *The Women Founders: Sociology and Social Theory, 1830-1930, A Text with Readings*, McGraw-Hill Humanities/Social Sciences/Languages.

Liljesröm M. 2016. *Affect*, in Disch L., Hawkesworth K. (eds.), *The Oxford Handbook of Feminist Theory*, Oxford University Press, 16 ff.

LINKUGEL W.A. 1963. *The Woman Suffrage Argument of Anna Howard Shaw*, in «Quarterly Journal of Speech», 49, 2, 165 ff.

LITTLE MO 1995. *Seeing and Caring: The Role of Affect in Feminist Moral Epistemology*, in «Hypatia», 10, 3, 117 ff.

LLOYD G. 1993. *The Man of Reason: 'Male' and 'Female' in Western Philosophy*, Routledge.

LOGAN, S. W. 1999. *We are Coming: The Persuasive Discourse of Nineteenth-century Black Women*, SIU Press.

LONGINO H.E. 1992. *Subjects, Power, and Knowledge: Description and Perscription in Feminist Philosophies of Science*, in ALCOFF L., POTTER E. (eds.), *Feminist Epistemologies*, Routledge, 101 ff.

LORDE A. 2020. *Sister Outsider: Essays and Speeches*, Penguin Publishing Group. (Originally published in 1984)

LUNBECK E. 2003. *Narrating Nymphomania between Psychiatry and the Law*, in UMPHREY M., DOUGLAS L., SARAT A. (eds.), *Law's Madness*, University of Michigan Press, 49 ff.

LYONS W. 1999. *The Philosophy of Cognition and Emotion*, in Dalgleish T., Power M. (eds.), *Handbook of Cognition and Emotion*, John Wiley & Sons, 21 ff.

MACDONALD G.M., OSBORNE R.L., SMITH C. 2005. *Feminism, Law, Inclusion: Intersectionality in Action*, Sumach Press.

MACKINNON C.A. 1987. *Feminism Unmodified: Discourses on Life and Law*, Harvard University Press.

MARONEY T.A. 2011. *The Persistent Cultural Script of Judicial Dispassion*, in «California Law Review», 99, 2, 629 ff.

MARONEY T.A., GROSS J.J. 2014. *The Ideal of the Dispassionate Judge: An Emotion Regulation Perspective*, «Emotion Review», 6, 2, 142 ff.

MARTEL J.R. 2017. *The Misinterpellated Subject*, Duke University Press Books.

MATSUDA M. 1986. *Liberal Jurisprudence and Abstracted Visions of Human Nature: A Feminist Critique of Rawls' Theory of Justice*, «New Mexico Law Review», 16, 3, 613 ff.

MCDANELD J. 2013. *White Suffragist Dis/Entitlement: The Revolution and the Rhetoric of Racism*, «Legacy», 30, 2, 243 ff.

MENKEL-MEADOW C. 2002. *Portia in a Different Voice: Speculations on a Women's Lawyering Process*, in NAFFINE N. (ed.), *Gender and Justice*, Routledge, 37 ff. (Originally published in 1985)

MILL J.S. 2017. *The Subjection of Women*, Routledge. (Originally published in 1869)

MILLER, W.I. 1998. *The Anatomy of Disgust*, Harvard University Press.

MINOW M.L., SPELMAN E.V. 1988. *Passion for Justice*, in «Cardozo Law Review», 10, 37 ff.

MOHANTY C. 1988. *Under Western Eyes: Feminist Scholarship and Colonial Discourses*, in «Feminist Review», 30, 1, 61 ff.

MONTOYA C. 2016. *Institutions*, in DISCH L., HAWKESWORTH K. (eds.), *The Oxford Handbook of Feminist Theory*, Oxford University Press, 367 ff.

MOSSMAN M. 1987. *Feminism and Legal Method: The Difference it Makes*, in «Wisconsin Women's Law Journal», 3, 147 ff.

NAGEL T. 1989. *The View from Nowhere*, Oxford University Press.

NARAYAN U. 2004. *The Project of Feminist Epistemology: Perspectives from a Nonwestern Feminist* in HARDING S.G. (ed.), *The Feminist Standpoint Theory Reader: Intellectual and Political Controversies*, Psychology Press, 213 ff. (Originally published in 1989)

NEDELSKY J. 2002. *Embodied Diversity and the Challenges to Law* in NAFFINE N. (ed.), *Gender and Justice*, Routledge. (Originally published in 1997)

NGAI S. 2005. *Ugly Feelings*, Harvard University Press.

NUSSBAUM M.C. 1995. *Poetic Justice: The Literary Imagination and Public Life*, Beacon Press.

NUSSBAUM M.C. 2006. *Hiding from Humanity*, Princeton University Press.

OKIN MOLLER S. 1979. *Women in Western Political Thought*, Princeton University Press.

PEDWELL C., WHITEHEAD A. 2012. *Affecting Feminism: Questions of Feeling in Feminist Theory*, in «Feminist Theory», 13, 2, 115 ff.

PERRY R. 2005. *Mary Astell and Enlightenment*, in KNOTT S., TAYLOR B. (eds.), *Women, Gender and Enlightenment*, Palgrave Macmillan, 357 ff.

PERUGA BOLUFER M. 2005. *'Neither Male, Nor Female': Rational Equality in the Early Spanish Enlightenment*, in KNOTT S., TAYLOR B. (eds.), *Women, Gender and Enlightenment*, Palgrave Macmillan, 389 ff.

POSNER R. 1999. *Emotion Versus Emotionalism in Law*, in BANDES S. (ed.), *The Passions of Law*, New York University Press.

RICH A. 1980. *Compulsory Heterosexuality and Lesbian Existence*, in «Signs: Journal of Women in Culture and Society», 5, 4, 631 ff.

RIFKIN J. 1993. *Toward a Theory of Law and Patriarchy*, in Weisberg K.D. (ed.), *Feminist Legal Theory: Foundations*, Temple University Press, 412 ff.

ROACH ANLEU SH., MACK K. 2021. *Judging and Emotion: A Socio-Legal Analysis*, Routledge.

ROBERTSON J. 2005. *Women and Enlightenment: A Historiographical Conclusion*, in KNOTT S., TAYLOR B. (eds.), *Women, Gender and Enlightenment*, Palgrave Macmillan, 692 ff.

RODRÍGUEZ DURÁN, J. 2015. *An Introduction to Anti-women-suffrage Propaganda*, in «Independent Journal of Interdisciplinary Arts», 1, 19 ff.

ROSENBERG R. 1982. *Beyond Separate Spheres: Intellectual Roots of Modern Feminism*, Yale University Press.

ROTH B. 2003. *Separate Roads to Feminism: Black, Chicana, and White Feminist Movements in America's Second Wave*, Cambridge University Press.

ROUSSEAU J-J. 1979. *Emile: Or on Education*, Basic Books. (Originally published in 1762)

SCARANTINO A. 2016. *The Philosophy of Emotions and Its Impact on Affective Science*, in FELDMAN BARRETT L. 2017. *How Emotions are Made: The Secret Life of the Brain*, HarperCollins.

SCHIEBINGER L. 1991. *The Mind Has No Sex?: Women in the Origins of Modern Science*, Harvard University Press.

SCHULTZ U., SHAW G. 2013. *Introduction: Gender and Judging*, in ID., *Gender and Judging*, Bloomsbury Publishing, 1 ff.

SEDGWICK KOSOFSKY E., FRANK A. 2003. *Touching Feeling: Affect, Pedagogy, Performativity*, Duke University Press.

SHAUKAT M. 2020. *American Muslim Women: Who We Are and What We Demand from Feminist Jurisprudence*, in «Hastings Women's Law Journal», 31, 155 ff.

SHULMAN KATES A. 1980. *Sex and Power: Sexual Bases of Radical Feminism*, in «Signs: Journal of Women in Culture and Society», 5, 4, 590 ff.

SMART C. 1989. *Feminism and the Power of Law*, Routledge.

SOMMERLAD H. 2013. *Let History Judge? Gender, Race, Class and Performative Identity*, in SCHULTZ U., SHAW G. (eds.), *Gender and Judging*, Bloomsbury Publishing, 355 ff.

SPELMAN E.V. 1990. *Inessential Woman*, Beacon Press. (Originally published in 1988)

SPELMAN E.V. 1998. *Fruits of Sorrow: Framing Our Attention to Suffering*, Beacon Press.

SPIVAK C.G. 1999. *A Critique of Postcolonial Reason: Toward a History of the Vanishing Present*, Harvard University Press.

STUURMAN S. 2005. *The Deconstruction of Gender: Seventeenth-Century Feminism and Modern Equality* in KNOTT S., TAYLOR B. (eds.), *Women, Gender and Enlightenment*, Palgrave Macmillan, 371 ff.

THOMAS S. 1995. *How Women Legislate*, OUP USA.

TRICOIRE D. 2017. *Introduction*, in ID., *Enlightened Colonialism: Civilization Narratives and Imperial Politics in the Age of Reason*, Springer International Publishing, 1 ff.

TRONTO JC 1987. *Beyond Gender Difference to a Theory of Care*, in «Signs: Journal of Women in Culture and Society», 12, 4, 644 ff.

TRONTO JC 1993. *Moral Boundaries: A Political Argument for an Ethic of Care*, Psychology Press.

TROUILLOT M-R. 2015. *Silencing the Past: Power and the Production of History*, Beacon Press.

TRUTH, S. 2020. *Ain't I a Woman?*, Penguin. (Originally published in 1863)

UMPHREY MA., DOUGLAS L., SARAT A. 2003. *Madness and Law: An Introduction*, in ID., *Law's Madness*, University of Michigan Press, 1 ff.

WALLACE K. 1993. *Reconstructing Judgment: Emotion and Moral Judgment*, in «Hypatia», 8, 3, 61 ff.

WEISBERG K.D. 1993. *Introduction*, in ID., *Feminist Legal Theory: Foundations*, Temple University Press.

WEISS P.A. 2009. *Canon Fodder: Historical Women Political Thinkers*, Penn State Press.

WEST R., BOWMAN C.G. (eds.) 2019. *Research Handbook on Feminist Jurisprudence*, Edward Elgar Publishing.

WILLIAMS J. 1989. *Deconstructing Gender*, in «Michigan Law Review», 87, 4, 797 ff.

WOLLSTONECRAFT M. 1995. *A Vindication of the Rights of Men and A Vindication of the Rights of Woman*, Cambridge University Press. (Originally published in 1790 and 1792)

PART VI.

Defeasibility and Legal Cognition

Defeasibility and Balancing

MANUEL ATIENZA

1. *Introduction. New names for traditional concepts*

Defeasibility and balancing are more or less new names for phenomena that are not new; they could not be, because they are closely related to basic features of legal systems and legal practice.

Let us start with "defeasibility". The expression (defeasibility) was introduced into legal theory at the end of the 1940s by Herbert Hart in one of his first writings: "The Ascription of Responsibility and Rights" (HART 1948). It is a work that Hart did not want to publish again later, but for reasons that do not seem to have had anything to do with this notion, but rather with that of ascription or, more precisely, with an excessively wide conception of ascriptivism, of the weight assigned to the ascriptive use of language, which entailed (Hart reached this conclusion as a consequence of various criticisms that were directed against his writing) a risk of incurring in reductionism (see LACEY 2006, 146)[1]. In fact, it seems that Hart was «unusually proud throughout his life» of having found something that showed the importance of paying attention to the legal use of language in order to develop notions of general philosophical interest (LACEY 2006, 144).

Hart's "discovery" is relatively simple, and he explains it with the clarity and elegance that always characterised him. It is that certain legal concepts, such as 'contract' or 'trespass', and, more generally, many of the most typical ones in criminal law, cannot be completely understood (defined) in terms of necessary and sufficient conditions, but rather that it is indispensable to include in their characterisation an "unless" clause:

> «In consequence, it is usually not possible to define a legal concept such as 'trespass' or 'contract' by specifying the necessary and sufficient conditions for its application. For any set of conditions may be adequate in some cases but not in others and such concepts can only be explained with the aid of a list of exceptions or negative examples showing where the concept may not be applied or may only be applied in a weakened form» (HART 1948, 174).

[*] Translated from Spanish by Mª Carmen Martínez Gómez.

[1] Anna Pintore, in a 1990 book (PINTORE 1990), considers that work of Hart to represent an initial and "deviant" stage from a path that leads (fundamentally in *The Concept of Law*) to «a conception of law and legal concepts that is commonly considered to be closed and belonging to legal positivism» (9). According to Pintore, the defeasibility of legal concepts that Hart defends here (and which would be something different from conceptual vagueness) takes us to an image of the law «as an open system, with no boundaries» (15). Hart, again according to Pintore, would have abandoned, in his mature stage, that idea of law «not as a system, and even less as a closed system of rules and concepts» (18) which, however, would have been assumed by someone like Neil MacCormick, who would represent (it is important to remember that Pintore writes in 1990), a "third way" between Hartian positivism and Dworkinian principialism; and, to carry out that operation, MacCormick would be based precisely in the defeasible character of legal concepts (PINTORE 1990, 183 ff.; all translations from Italian are mine). Anyway, the development of that notion in MacCormick's work is found in MACCORMICK 1995 (which later was part of MACCORMICK 2005).

In a later essay (see CHIASSONI 2019, 233) the conditions that would go behind the "unless" clause are classified by Hart into two categories: excusing conditions or invalidating conditions. But what is perhaps more interesting to highlight here is that Hart thought that there was no word in ordinary English to account for this feature, and his choice of "defeat" or "defeasible" was, in fact, a consequence of his familiarity with legal practice (of his experience as a lawyer), and also shows what has already been pointed out: that the careful analysis of legal language can have a more general scope:

> «This characteristic of legal concepts [needing the 'unless' clause] is one for which no word exists in ordinary English. The words 'conditional' and 'negative' have the wrong implications, but the law has a word which with some hesitation I borrow and extend: this is the word '*defeasible*' used of a legal interest in property which is subject to termination or '*defeat*' in a number of different contingencies but remains intact if no such contingencies mature. In this sense then, contract is a defeasible concept» (HART 1948, 175).

About a decade later, Stephen Toulmin, in a book that is often considered as the beginning of studies on "informal logic", *The uses of argument* (TOULMIN 1958), introduces the same idea to account for a typical feature of argumentation, as he understands it[2].

In short, what Toulmin proposes there is an approach to argumentation seen as a social interaction, which takes place between a proponent and an opponent (the classic scheme of dialectics). At the beginning of the argumentation, the proponent holds a thesis (*claim*: for example, "Harry is a British subject"), which can be objected to by the opponent; otherwise, there would be no need to argue. If so, if it is objected, then the proponent has to give reasons (*data* or *ground*) in favour of his initial claim, which are at the same time relevant and sufficient (for example: "Harry was born in Bermuda"). The opponent may now dispute those reasons, those facts, but even if he accepts them, he can require the proponent to justify the step from the *data* to the *claim*. The general statements that authorise said step constitute the *warrant*, that is, a statement that is not descriptive, and that Toulmin explains by making an analogy with the role that a recipe has in the baking of a cake, and once all the ingredients are in place (for example: "A man born in Bermuda will generally be a British subject"). Finally, it is sometimes necessary to show that the guarantee is valid, relevant and of enough weight, which constitutes the *backing* of the argument (in our example: "On account of the following statutes and other legal provisions: ..."). Those elements are enough to account for when we have a valid or correct argument. But the *strength* of an argument depends on two other factors that, when added to the previous ones, allow us to obtain a general model of argumentation: the *qualifiers* that graduate the strength with which the *data*, the *warrant* and the *backing* provide support for the *claim* ("most certainly", "presumably", "most likely"...); and the *rebuttals*, that is, the support provided for the claim may stop existing or weaken when certain extraordinary circumstances or certain exceptions occur (for example: "unless both his parents were aliens, or he has become a naturalised American").

[2] It is worth clarifying here that Toulmin's way of understanding argumentation is not that of classic logic, of formal deductive logic. His model, as I will now explain, is that of traditional dialectics, which consists of seeing argumentation as an interaction, as an activity. Juan Carlos Bayón has questioned the idea that legal reasoning is defeasible and, with it, also the need or pertinence of building a type of non-classic (non-monotonic) logic to account for justificatory judicial reasoning. But he understands argumentation, the justifying judicial reasoning, in the sense of classic logic, that is, as «la inferencia con la que se justifica una determinada conclusión acerca del derecho aplicable a un caso individual» (BAYÓN 2001, 50). He is right, but Toulmin's idea of defeasibility (of refutability) refers to something different, namely, to the process of argumentation, to argumentation seen from a pragmatic perspective.

Toulmin, by the way, points out that this last element coincides with what Hart had called "defeasibility" in his work. At the same time, he underlines that Hart had shown that this phenomenon had relevance not only in the field of law, but also in the field of philosophy (regarding notions such as freedom of will or responsibility), and suggests what could have been the cause of Hart's discovery: «It is probably no accident that he reached these results while working in the borderland between jurisprudence and philosophy» (TOULMIN 1958, 142).

As a precursor of this notion, in the field of ethics, Toulmin also refers to the thesis defended by David Ross in his influential book, of 1930, *The Right and the Good*, according to which it is necessary to recognise that all moral norms have exceptions. As it is well known, Ross introduced there the distinction between *prima facie* duties and real or absolute duties, in order to account for the (according to him—that is, according to the distinction he introduces—only apparent) conflicts between moral duties. So, for example, the duty to tell the truth or to keep a promise may have an exception in certain circumstances, for instance, in a case in which acting in accordance with these duties would cause a person unjustified harm:

> «If, as almost all moralists except Kant are agreed, and as most plain men think, it is sometimes right to tell a lie or to break a promise, it must be maintained that there is a difference between *prima facie* duty and actual or absolute duty. When we think ourselves justified in breaking, and indeed morally obliged to break, a promise in order to relieve some one's distress, we do not for a moment cease to recognize a *prima facie* duty to keep our promise, and this leads us to feel, not indeed shame or repentance, but certainly compunction, for behaving as we do» (ROSS 1930, 28).

I believe it is important to highlight here some features that Ross underlines in relation to ethics, which contrast what happens in other fields of experience and which would explain the need to introduce the distinction in question. One is that Ross considers that the opinions of the majority of people or of the wise people play a very important role in ethics, and would constitute something like a starting point of the ethical method[3], which could not be said, of course, of the physical sciences, which construct theories and hypotheses that seem to move further, and increasingly further away, from our intuitions about how the physical world is and how it works. Another one is that mathematical notions, such as that of the isosceles triangle, differ from those of an ethical nature, for example: that of correctness, because the former could be defined—we could say—by a set of necessary and sufficient properties: thus, a triangle that has two equal angles is isosceles, independently of any other feature it possesses; but this does not happen in relation to the rightness of acts. And the third characteristic (a consequence of the previous one) is that the (moral) rightness of a particular act (as opposed to its *prima facie* rightness) depends on a set of circumstances[4] or, in other words, the act in question falls under various moral standards, so that according to one (for example, "no lying") it could be wrong, but, according to another, it could be right ("no causing unjustified harm").

Furthermore, what we understand today as defeasibility (that rules contain implicit exceptions) has such remote antecedents that they could be placed in the very emergence of philosophy; at least, of practical philosophy. In a way, it is what lies behind Plato's distrust of

[3] What Ross defends as a method of ethics, both in that book and in a later book, *Foundations of Ethics* (ROSS 1939), is nothing but a version of the "reflective equilibrium".

[4] «But no act is ever, in virtue of falling under some general description, necessarily actually right; its rightness depends on its whole nature and not on any element in it. The reason is that no mathematical object (no figure, for instance, or angle) ever has two characteristics that tend to give it opposite resultant characteristics, while moral acts often (as everyone knows) and indeed always (we must admit after reflecting) have different characteristics that tend to make them at the same time *prima facie* right and *prima facie* wrong; there is probably no act, for instance, which does good to anyone without doing harm to someone else, and *vice versa*» (ROSS 1930, 33 f.).

legislation, of the government of men by means of general rules, as it emerges from dialogues such as *The Republic* (PLATO 1997a) or *The Statesman* (PLATO 1997b). In the latter, government by laws (and customs) appears as a kind of rationality of the second best, since «the best thing» says the Stranger (who in the dialogue represents the role usually played by Socrates), «is not that the laws should prevail, but rather the kingly man who possesses wisdom» that is, the wise and good man: the philosopher. And the reason for this would be that

> «the law could never accurately embrace what is best and most just for all at the same time, and so prescribe what is best. For the dissimilarities between human beings and their actions, and the fact that practically nothing in human affairs remains stable, prevent any sort of expertise whatsoever from making any simple decision in any sphere that covers all cases and will last for all time» (PLATO 1997a, 294a).

And the idea of defeasibility is also one of those underlying Aristotle's presentation of the concept of equity, in one of the most brilliant pages, in my opinion, in the entire history of philosophy of law. Aristotle defends the need to deviate in certain cases from the literal meaning of the law, that is, to introduce an exception, in order to account for the singularities of the specific case, which the legislator could not foresee, due to the «nature [...] of practical affairs». The text deserves to be quoted at some length:

> «the equitable is just, but not the legally just but a correction of legal justice. The reason is that all law is universal but about some things it is not possible to make a universal statement which will be correct. In those cases, then, in which it is necessary to speak universally, but not possible to do so correctly, the law takes the usual case, though it is not ignorant of the possibility of error. And it is none the less correct; for the error is not in the law nor in the legislator but in the nature of the thing, since the matter of practical affairs is of this kind from the start. When the law speaks universally, then, and a case arises on it which is not covered by the universal statement, then it is right, when the legislator fails us and has erred by over-simplicity, to correct the omission—to say what the legislator himself would have said had he been present, and would have put into his law if he had known. Hence the equitable is just, and better than one kind of justice—not better than absolute justice but better than the error that arises from the absoluteness of the statement. And this is the nature of the equitable, a correction of law where it is defective owing to its universality. In fact this is the reason why all things are not determined by law, viz. that about some things it is impossible to lay down a law, so that a decree is needed. For when the thing is indefinite the rule also is indefinite, like the lead rule used in making the Lesbian moulding; the rule adapts itself to the shape of the stone and is not rigid, and so too the decree is adapted to the facts» (ARISTOTLE 1984a, book V, sect. 10, 1137b-1138a)[5].

[5] References to these classical texts can also be found in SCHAUER (2012), who rightly recalls the importance of courts of equity in the development of law (including, of course, common law). Curiously enough, the way of understanding defeasibility in law proposed by Alchourrón, what he calls "dispositional approach", is precisely the same as Aristotle regarding equity. According to Alchourrón, the circumstance C can be considered as an implicit exception from the moment of the enactment of a law, even if the legislator did not consider it at the moment, but as long as there are reasons to think that, if he had considered it, he would have introduced it. Alchourrón thinks that many of the conditional sentences in our everyday language (and that is also for legal language) are defeasible: we formulate our sentences for normal circumstances, knowing that in certain situations our sentences will be defeated. And that because «las construcciones condicionales de la forma 'Si A entonces B' son frecuentemente usadas de un modo tal que no se pretende con ellas afirmar que el antecedente A es una condición suficiente del consecuente B, sino sólo que el antecedente, sumado a un conjunto de presupuestos aceptados en el contexto de emisión del condicional, es condición suficiente del consecuente B» (ALCHOURRÓN 2000, 23-26).

With regard to the other term, "balancing", something very similar could be said, precisely because, in reality, defeasibility and balancing are different aspects of the same reality, instruments, one could say, with which one tries to achieve the same purpose (speaking in abstract terms): to avoid excessive rigidity in the law and to contribute to bringing the law closer to justice.

In recent times, the person who seems to have contributed most to spreading the idea of balancing in legal theory—mainly, in the Latin world—has been Robert Alexy. This notion (the German expression is "Abwägung"), by the way, does not appear in the German author's first work, from 1978, dedicated to legal argumentation (ALEXY 1989), but instead, years later, when he deals with fundamental rights (ALEXY 2002)[6] and introduces the distinction (essentially inspired by Dworkin) between rules and principles. Fundamental rights, for Alexy, are essentially principles. Unlike rules, which would be norms that order something definitely, principles would be characterised as "optimisation commands", that is, norms that order something to be achieved to the highest possible degree, according to the existing factual and legal possibilities. Well, while the application of rules requires subsumptive reasoning, in the case of principles the type of argumentation to be resorted to would be balancing. I will not go now into other details about the way in which Alexy understands balancing (I will say more about this later), but I am interested in highlighting these two points.

The first is that Alexy's conception of balancing has not undergone any change that can be considered essential throughout all these years (about 40, during which it has been discussed *ad nauseam*), but it has undergone some additions and adjustments. One of them consists precisely of the following. In his recent polemic with POSCHER (2022), the latter reproaches him, among other things, that principles cannot be conceived as "optimisation commands", simply because an optimisation requirement, following Alexy's definitions, would be a rule: it orders something to be done (whatever the optimisation consists of, that is, the achievement of something "to the highest possible degree") in a definitive manner. Well, to face this criticism (which had already been made by AARNIO 1990 and by SIECKMAN 1990), Alexy establishes a distinction between an "optimisation command" and a "command to be optimised" (which is what principles would be), and for that he relies precisely on Ross' differentiation between two types of duties, that was previously mentioned. Therefore, in short, what Alexy holds is that the key distinction to understanding balancing is the one that can be established between two types of duties: ideal duties, *prima facie* or *pro tanto* (fixed in principles), and real duties, definitive or considering all the circumstances of the case (fixed in the rules resulting from the balancing of principles). And the other point I want to make here is that Alexy's elaboration of the method of balancing does not pretend to be anything other than a rationalisation of the way in which the German Constitutional Court and other European courts proceed when solving problems that involve conflicts between rights (between principles): balancing is, one might say, a way of solving those conflicts by moving from the principles to the rule, from ideal duties (which conflict with each other) to the duty considering all the circumstances of the case.

The idea of balancing, under this or another name, has always been present both in the practice of law and in its theorisation, in what has traditionally been called legal methodology. Precisely, one of the most influential methodological directions—not only in Germany, but in all civil law countries—in the 20th century has been the so-called "Jurisprudence of interests", headed by Philip Heck and inspired—inevitably—by the work of the second Jhering. The basic idea (as happens with all anti-formalist directions) is that conflicting, hard cases can arise in law (cases of legal gaps, contradiction, etc.), which cannot be solved simply by applying the legal rules, in accordance with their literal or textual meaning, but instead, to solve them it is

[6] The first edition of his *Theory of Constitutional Rights* is from 1986.

necessary to do a "balancing" of the interests at stake; and, in turn, the law itself would be nothing else, for Heck, than what results from an opposition of forces, of interests, which pull in different directions[7].

Moreover, the usual assertion that the balancing method is preferred by those who promote a finalist interpretation of the norms (the anti-formalists) and who are, therefore, opposed to those in favour of a strict, literal, interpretation of the law (the formalists), seems to me to be questionable or, at least, in need of some nuance. And not only because of the usual imprecision with which these terms are usually used ("formalism" and "anti-formalism"), but also because, at least very often, those who are supposed to—those who say they do—take their decisions strictly bound by the law (the formalists or legalists), do not fail to also really consider the interests, the purposes, that are at stake when interpreting a rule and arriving to a decision; in other words, they do not fail to balance. A typical example of this can be found in the famous *Lochner* case, decided by the Supreme Court of the United States in 1905, and which is usually considered (the majority's decision—and its justification—which was opposed—as is well known—by Holmes' dissenting vote—which was not the only one) as the epitome of legal formalism. Well, what was at issue there, as is well known, was whether a New York State law limiting work in bakeries to 10 hours a day and six days a week should be considered constitutional or not. And what I find interesting to remark here is that both the anti-formalist Holmes (who defended the constitutionality of the law) and the majority of the Court (who overturned the law because they considered it unconstitutional) resorted to a ponderative type of scheme, which, by the way, does not imply at all an abandonment of formal logic. As far as the majority is concerned, the ruling is based on the observation that, on the one hand, there is the freedom of contract established in the 14th Amendment of the US Constitution, and, on the other hand, the "police powers" that grant each State of the Union the competence to legislate (and limit freedom of contract) for reasons of health, safety, etc. And what had to be determined then was «which shall prevail—the right of the individual to labor for such time as he may choose or the right of the State to prevent the individual from laboring or from entering into any contract to labor beyond a certain time prescribed by the State»; for reasons that are not to be noted now (and neither whether or not they were justified), the Court opted for the former. And that same balancing scheme (which, I insist, does not imply any distancing from deductive logic, despite some of Holmes' misguided expressions in that respect[8]) is the one used by the dissenting judge, but with an opposite result to that of the majority, since he made the second of the rights prevail or, rather, the reasons in favour of recognising a State the competence to establish those limits to freedom of contract[9].

Finally, as happened in the case of defeasibility, the notion of balancing, of pondering, of weighing the interests, the reasons, of opposite signs and which may be present in certain cases requiring a decision to be taken, is so rooted in the very idea of law that, as is well known, the scales are part of the usual symbolism of the administration of justice: in the deliberation that must take place in conflicting, hard cases, the two sides of the scales represent the places where the arguments, the reasons, for and against, should be placed in order to reach a "balanced"

[7] This is an analogical use of the "parallelogram of forces" method, which shows the result of applying two forces to a (physical) object. In *La jurisprudencia de intereses de Philipp Heck*, the author, María José García Salgado, concludes that «puede verse la Jurisprudencia de intereses como una teoría normativa de la ponderación de intereses, cuya finalidad es proporcionar al juez pautas que le permitan proteger, en caso de conflicto, el interés preferido por el legislador» (GARCÍA SALGADO 2010, 242). And in a later work she connects these ideas directly with the contemporary discussion on balancing (GARCÍA SALGADO 2019).

[8] Particularly in *The Path of the Law* (HOLMES 1897).

[9] Hart was right when, commenting on this case, he pointed out that what here «is stigmatized as 'mechanical' and 'automatic' is a determined choice made indeed in the light of a social aim, but of a conservative social aim» (HART, 1958, 611).

decision. But it is not only that, but also that the scales, the "scales of reason", have been the image that has dominated conceptions of rationality in the West. Marcelo Dascal has studied this metaphor of the scales of reason which, according to him, allows, at least, two interpretations: a "metric" or "algorithmic" one, that leads to a "hard" conception of reason; and another of a "dialectical" nature and which leads to a "soft" conception of rationality. In his opinion, both are complementary, but the second is the one that should be used fundamentally in contingent matters and in matters linked to the notions of "burden of proof" and "presumption". And he illustrates this with a statement by Leibniz (in whose work both senses, both conceptions, of reason would be present), according to which «no one has as yet pointed out the scales [for weighing and evaluating considerations that go against each other until a decision is reached], though no one has come closer to doing so and offered more help than the jurists» (DASCAL 1996, nt. 24).

2. *Defeasibility, balancing and conceptions of law*

At the beginning I said that the abundance in law of references to the notions of defeasibility and balancing were related to basic—intrinsic—characteristics of legal systems and legal practice[10]. The examples could be multiplied. Thus, the classic—structural—theory of crime in the criminal dogmatics of continental law could very well be considered as a scheme of defeasibility: a typical action is unlawful unless... and if it is typical and unlawful, then it is guilty unless... Presumptions, the burden of proof, maxims of experience or rules of evidence are constructions that presume something like a legal institutionalisation of defeasibility: if the circumstances X and Y are present, then it is understood that event H has occurred, unless... The same could be said of courts of equity, whose function would be precisely to avoid the bad consequences that the application without exceptions of general rules could have (but without going against the principle of universality—generality is not the same as universality). "Atypical torts" (such as abuse of law, legal fraud or deviation of power) are also examples of the defeasibility of rules and of the use of a balanced reasoning (see ATIENZA & RUIZ MANERO 2000). The procedure for deviating, in general, from a merely literal interpretation of a rule involves a balancing judgement (in order to be able to create an exception). Also the "judgement of proportionality" to which jurists very often resort is nothing other than a balancing exercise. The resolution of conflicts between rights—a central problem in the law of the Constitutional State—inevitably involves resorting to balancing. Et cetera, et cetera[11].

But, at the same time, all those statements may be more or less obvious, depending on one's conception of law. And the way of understanding those notions and of assigning them a role of greater or lesser significance in the theory and practice of law is also dependent on that—on how one conceives the law. Moreover, I have the impression that much of the (very abundant) literature on defeasibility and balancing that exists today is at risk of focusing on rather irrelevant issues or, in any case, of little interest, simply because many of the authors of all those texts do not seem to be aware (or are not aware to an adequate extent) of the main conclusion that is drawn from what I pointed out in the previous section. It is that law is, above all, a social practice, an activity, aimed at the satisfaction of certain ends and values. And practical questions (in the sense of traditional practical reason) cannot be solved in the same

[10] According to Guastini, the notion of defeasibility (and of the axiological gap) does not belong to the theory of legal systems, but to that of interpretation (GUASTINI 2008, 149). But this can only be understood if it is connected with a certain conception of law—the one that he holds—and to which I will later refer, in critical terms.

[11] Schauer gives many examples of defeasibility in law, some of them characteristic of common law. See SCHAUER 2012, 79.

way as would be appropriate for problems posed in the empirical sciences or in the formal sciences; which does not mean, beyond that, that empirical or formal knowledge can be disregarded in the resolution of practical problems. But what seems fundamental is to realise that law, morality or politics are "rational enterprises" (to use Toulmin's expression) with their own peculiarities, and hence the importance of paying attention to the way in which we argue within those practices. And, when this is done, the result is that the concepts involved cannot always be defined by a set of necessary and sufficient properties, the correct answer to a moral (or legal) case requires carrying out an analysis that takes into account what Ross called toti-resultant attributes and not parti-resultant attributes (see ROSS 1930, 28, and nt. 5), because—to use the poetic expression of the Platonic dialogue— «nothing in human affairs remains stable», but instead "the nature of practical affairs" means that not all the circumstances of future cases can be foreseen. Hence, the task of governing human behaviour by means of rules cannot be done by resorting exclusively to classificatory (subsumptive) operations, but instead, it is sometimes necessary to deliberate, to use balancing; in other words, to generate new rules in a coherent way, respecting the established system, but including in that system the reasons underlying the rules, that is, the purposes and values that underlie them. To put it extremely synthetically, the phenomena of defeasibility and balancing can only be properly understood if law is fundamentally considered as a social practice, as an activity, and not exclusively as an object, that is, as a type of reality consisting simply of a set of statements, a normative system. And it is not that the normative system is not part of law, but rather, that it is a necessary, but not sufficient, component. Law is not only a (coercive and dynamic) system of norms but, above all—to put it in Jhering's terms—means to an end; norms (and coercion) constitute (indispensable) organisational means for the achievement of that end, for the satisfaction of certain social needs[12].

The latter (the post-positivist conception) constitutes, in my opinion, the most appropriate way of understanding law, especially if what is pursued is to account for the rights of the Constitutional State and the era of globalisation. But, of course, it is not the only existing one, and not even the dominant one.

It is, for example, very different from the one held by Niklas Luhmann, which seems to continue being a considerable influence on sociologists (and theorists) of law. Although I do not believe that Luhmann's schemes have ever served to satisfactorily explain legal phenomena, it could nevertheless be accepted that they capture some features of law in the age of legal positivism that have nevertheless become, so to speak, obsolete. For example, the process of positivisation of law which has been taking place (in some European countries or countries of European influence) since the beginning of the 19th century, implied, according to him, that the legitimisation of law would no longer depended on any material element, but exclusively on procedure; but that—I would say—has been clearly refuted in recent times with the introduction in Constitutions of declarations of fundamental rights (and the institutionalisation of constitutional courts) which precisely set a limit to the very idea of positivisation: the law is not established and valid simply, or in all cases, by virtue of a decision that can be transformed at any time (see LUHMANN 1977 and LUHMANN 1990, spec. 115 ff.): the law cannot have any content. And the same could be said of his thesis of the progressive autonomisation of law and its configuration as an autopoietic system, which is self-regulating and self-reproducing regardless of the other social subsystems and guided solely by the idea of reducing complexity; on the contrary, the evolution of our legal systems goes towards making more and more permeable the boundaries between law and politics, morality, economy... All of which explains,

[12] Recall Jhering's definition of law: «Law is the sum of the conditions of social life in the widest sense of the term, as secured by the power of the State through the means of external compulsion» (JHERING 1913, 380).

in my opinion, that even though the phenomenon of defeasibility and the use of balancing have always been an important aspect of legal practice, it could be said that nowadays their weight has increased considerably. Therefore, a conception such as Luhmann's, which is "obsessed" with the value of security, which leaves little room for "openness" to ideas of justice, and which sees—we could say—law almost exclusively in terms of rules, does not seem to be functional in relation to the legal systems of our time[13].

But post-positivism is also not the dominant conception in contemporary legal theory. In particular, it is not so in the Latin world where, on the other hand, there has been much discussion in recent times about defeasibility and about balancing, and also about whether the vindication (or recognition) of these phenomena implies or not the abandonment of legal positivism. Thus, Riccardo Guastini (par excellence representative of the "realist" positivism of the Genoese school) considers that defeasibility of legal norms has nothing to do with legal positivism, despite what the following reasoning seems to suggest:

> «Legal positivism claims that law can be identified independently of any moral evaluation. But if legal norms are defeasible, their content cannot be identified without moral evaluations. Then the scientific project of legal positivism is doomed to failure: in order to identify the law, moral evaluations must be assumed».

However, this reasoning is not valid, according to him, among other things, because it leads to the following confusion:

> «It is one thing to identify something—specifically a normative text—as law, it is quite another thing to determine its normative content: what is commanded (permitted, prohibited), to whom, under what circumstances. Legal positivism simply says that the first of these two things can be done without evaluation; it says nothing about the second. Methodological legal positivism is not, and does not include, a theory of interpretation» (GUASTINI 2008, 155; my translation).

But if this is so, that is, if legal positivism means only that, then the only comment that can be added is that such a poor conception of law simply lacks interest, regardless of whether its theses are true or not[14].

Going back to an earlier idea. The greater importance (and visibility) of the phenomena of defeasibility and balancing in recent times perhaps allows us to explain (and solve) a certain controversy that can be detected among researchers of defeasibility: while some, such as RODRÍGUEZ & SUCAR (1998) or POGGI (2021), are in favour of abandoning the notion, as it would be nothing but a new label for designating things that are well known, others, such as Chiassoni, think that this would be a mistake, because the turn towards defeasibility in contemporary legal thought points towards central problems of law that could be clarified if this notion is carefully analysed (CHIASSONI 2019, 229).

This last author, precisely, has distinguished up to 11 different notions of defeasibility, that is, there would be—according to him—11 types of entities, of objects, to which philosophers of Law attribute this feature[15]; and it is possible that a similar analysis could be made with regard to

[13] Although perhaps that cannot be said of the last Luhmann, according to whom the legal form just as we know it would have been a "European anomaly" linked to the Nation-state, but which would stop being functional in relation to the law of the global society. On this see CAMPOS 2023, ch. 1.

[14] A critique of Guastini's conception can be found in ATIENZA 2018.

[15] They are the following: «(1) defeasible *facts*; (2) defeasible *beliefs*; (3) defeasible legal *concepts*; (4) defeasible legal *provisions* or legal *texts*; (5) defeasible legal *interpretations*, or defeasible *meaning*, of legal provisions; (6) defeasible legal *norms*, rules, principles, standards, etc. (norm defeasibility); (7) defeasible legal *reasoning*; (8) defeasible legal

balancing: many things can be balanced and the activity of balancing can also be seen from very different points of view. But Chiassoni himself concludes that many of those uses are parasitic and that the truly relevant and interesting notion is that of defeasible rule[16]. Well, although I personally consider that Chiassoni's analysis of defeasibility (and of the indeterminacy of law) clarifies some things, it seems to me that the most fruitful (I would also say the most "natural") way of proceeding to analyse this notion (and also that of balancing) consists of starting from the two main instances that can be distinguished in legal practice: the activity of establishing general rules (I leave out contracts, wills and other legal transactions, although here too both balancing and defeasibility play a role) and that of interpreting and applying them in the solution of cases. In both instances it is about ensuring that the law can satisfy the characteristic aims and values of practice, and that is what explains, as I said, why those two notions—and others to which I have already referred in part—have acquired a singular importance in contemporary legal theory. Let us see.

3. *Defeasibility and balancing in the process of legislation. Rules and principles*

Although when we speak of balancing we usually refer to the balancing carried out by judges, the bodies that apply the law, it should not be forgotten that the establishment of general rules, of laws (or of other types of measures that may not have a general scope) is fundamentally governed by the idea of balancing, of deliberation. This is why, for example, the rhetorical tradition called the type of (persuasive) discourse that took place in the assembly "deliberative genre", whose time horizon was the future (as opposed to the judicial genre, which looked at the past) and which included what we would call today legislative argumentation: to establish laws. Aristotle pointed out in his *Rhetoric* that we only deliberate about matters which are contingent (not about what must necessarily happen), and which are also under our control ("which we have it in our power to set going") (ARISTOTLE 1984b, book I, sect. 4, 1359[b]). The ultimate goal of deliberation, in general terms, would be, for him, happiness (*eudaimonia*), which consists of different parts, of different goods, although what is actually deliberated upon—let us say, the most immediate goal—would be constituted by the means, by the actions that are convenient to achieve those ends (ARISTOTLE 1984b, book I, sect. 6, 1362[a]).

Well, what could be called the "internal justification" of legislative argumentation could then be seen as a type of balancing, not of subsumption: each of the normative provisions of a legal text would be the fruit of a deliberation in which the "balance of reason" would have given a certain statement as a result (the resultant of the parallelogram of forces in Heck's metaphor). But it is a balancing that is very different from the reasoning to which, sometimes, judges have to resort to and which is called by that name. The fundamental difference is that legislative argumentation is much more open than judicial argumentation, the reasons to which a legislator can (must) resort are not authoritatively determined or, to put it differently, those limits are much wider, so that, in short, it is about a more complex rationality which does not admit, for example, its reduction to a binary scheme: it is not a matter of choosing between the

positions, jural relations, legal entitlements, etc. (status defeasibility); (9) defeasible legal *arrangements*, like contracts, wills, etc. (legal arrangements defeasibility); (10) defeasible legal *claims;* (11) defeasible legal *conclusions*)» (CHIASSONI 2019, 231).

[16] The definition he gives is this: «*Defeasible norm:* a norm is defeasible, if and only if, the normative consequence it states is liable (i.e., may be subject) to a set of negative conditions of application ('exceptions', 'defeaters', 'defeating conditions')» (CHIASSONI 2019, 249). And then he establishes more specific notions, depending on whether they are explicitly or implicitly defeasible norms, and whether the norms are closed-defeasible (of different types) or "open-defeasible". In total there would be seven more specific notions of "defeasible norm".

constitutionality or unconstitutionality of a law or between the conviction or acquittal of the accused, but of choosing a text from a plurality, almost an infinity, of possibilities.

In order to carry out this task[17], the legislator needs to mobilise scientific and technical knowledge of many different kinds; as well as starting on the basis of a moral and political philosophy. In other words, the ends to be achieved through legislative intervention must be morally justified or, at least, they must not contradict constitutional values and principles; the established statements—the rules—must be drafted with sufficient clarity; they must fit harmoniously into the previously existing legal system (so as not to generate gaps or contradictions); the appropriate subjective incentives (sanctions in the broad sense) and objective means (financial, institutional...) must be established so that the addressees comply with the requirements of the rules (to make the transition from law in texts to law in action); and it is also necessary to ensure that compliance with the provisions of the law leads to the achievement of the pursued goals (the transition from effectiveness to social effectiveness); but all of this must also be done in a reasonable (efficient) way. Well, within this extremely complex task, one aspect of considerable importance is the choice of the types of legal statements (I am referring, then, to the formal aspect, not to the contents) that are most suitable for achieving all those purposes.

Here it is worth starting by recalling that legislative statements do not only express norms[18]. There are also definitions—theoretical statements—, practical statements that express normative acts (for example, that of repealing a law) or evaluative statements. And, within norms, we should make a distinction between those of a deontic or regulative nature (they establish that, given certain conditions, the performance of an action or the achievement of a state of affairs is deontically modulated as obligatory, prohibited or permitted) and constitutive norms (if certain conditions are met, then a certain normative result is produced—constituted—: a legal event or a legal action). All these statements differ in terms of their structure, but also with regard to the role they play within legal reasoning and in relation to the social system (inasmuch as they articulate in a certain way the social and individual powers and interests).

In order to deal with the problem of defeasibility, I will focus on regulative norms, because this is where the distinction between rules and principles is situated, which, as will be seen, is of particular significance. However, this does not mean that defeasibility only has a place here; for example, when it comes to establishing the conditions of validity of a contract (one of Hart's examples) we would be in the context of constitutive rules: those conditions of validity, at least on many occasions, cannot be established—as he told us—by pointing out a set of necessary and sufficient conditions, but instead, the list would have to be followed by the famous "unless" clause. The same could be said, of course, of legislative definitions. And, in any case, the classifications that can be made of legal statements must always be understood in an open, functional, and—so to speak—contextual sense: it is not only that there may be penumbral cases (statements that do not fully fit into any of these categories), but also that each one of those statements can only be properly understood if we take into consideration its relation to other statements of the other types: what functions as a unit is the set of statements, articulated in a certain way, that makes legislatively created law (or a fragment of it) capable of fulfilling its purpose.

Well, principles and rules (which—I insist—are characteristic types of legal statements, but are not the only pieces of law) differ from each other, as I said, from diverse perspectives. Thus, both rules and principles have a conditional structure, but the difference would be that the antecedent (the conditions of application) in the case of principles have an "open" character, while in rules it is "closed"; which could also be expressed, following von Wright's

[17] I present here a summary of different works on the theory and technique of legislation, now collected in Atienza 2019.

[18] I take the classification of legal sentences that can be found in Atienza & Ruiz Manero 1996.

terminology[19], by saying that principles are categorical norms, that is, their conditions of application do not contain other properties than those derived from the content of the norm itself, while in rules there are additional conditions of application. To illustrate this with an example: "it is forbidden to discriminate on the basis of sex (whenever there is an opportunity to perform such an act)" is a principle; "it is forbidden to pay a woman a lower wage than a man, if both do the same work", is a rule. From the point of view of how they operate in legal reasoning, rules work as peremptory or exclusionary reasons, so that, if the fixed conditions of application are met, then what is established in the rule must be done, without entering into any type of deliberation, whereas principles provide only non-peremptory reasons (thus, weaker reasons, with less force, but with a wider scope)[20], that must be weighed against other reasons (to return to the example, reverse discrimination or affirmative action may be justified in some cases). And, finally, principles limit the pursuit of individual and social interests (which is a way of saying that they establish rights) and promote the satisfaction of social interests; and rules also play this role, but by imposing positive and negative duties and thus generating reciprocal restrictions (without the need for balancing) or by granting a power of discretionality (rules of end[21]) that would affect only the means.

Those differences can also be seen in terms of defeasibility, in the following way. Principles are conditional statements (norms) that are presented as intrinsically defeasible: they are non-peremptory reasons; that is, *prima facie* reasons to carry out a certain conduct, but that, when balanced against others, can be defeated, all circumstances considered. This is what we saw in Ross's classic book (or in Alexy): they presuppose the existence of a distinction between two types of duties: ideal and real. Whereas the vocation of rules, we could say, is to not be defeated (to operate as peremptory reasons), although we cannot discard that exceptionally they may be, that is, that they include implicit exceptions. And we have already seen why: human affairs cannot stand still[22] and it is impossible that the legislator, who necessarily has to express himself in general and future-referring terms, has taken into account all the elements that are relevant regarding the reasons underlying the rules, that is, the aims and values they seek to achieve[23].

[19] See on this AGUILÓ 2000, 135 f.; VON WRIGHT 1963.

[20] The way of drawing the distinction between rules and principles (ATIENZA & RUIZ MANERO 1996) is very similar to that found in HAGE & PECZENIK 2000. They speak of decisive reasons and contributive reasons, but the meaning is the same as the one we outlined between peremptory or exclusionary reasons and non-peremptory reasons. One difference with our analysis, however, is that they assume Alexy's conception of principles: principles «only generate (as opposed to rules) reasons that plead for actions that contribute as much as possible to goal states» (HAGE & PECZENIK 2000, 306). And I do not see clearly the point of constructing two different kinds of logical functors—of conditionals—to symbolise a rule or a principle.

[21] In our scheme, the distinction between rules and principles is combined with the other distinction we made between action rules and end rules (see ATIENZA & RUIZ MANERO 1996).

[22] So defeasibility is not simply due to certain features of natural language, but rather to certain features of law. On this, see SCHAUER 2012, 77.

[23] There is a clarification to be made here. Authors such as Guastini (in general, the members of the Genoese school) start from a basic distinction between provision and norm, that is, one thing is the text, the statement, and another thing is what it means, the norm; so that norms only exist when statements are interpreted; Guastini insists, for example, that it is a mistake to confuse a statement with its literal interpretation. As a consequence of all this, he affirms that defeasibility can only be a feature of norms, not of provisions (see CHIASSONI 2019, 249, who—following Guastini's thesis—thinks that legal provisions would only be defeasible in a metonymic sense); or, in other words, defeasibility does not exist prior to the interpretation, but instead it depends on the interpretation. And hence the statement I referred to earlier, according to which defeasibility would not belong to the theory of normative systems (norms understood here as mere dispositions), but to that of interpretation. In my opinion, it is a way of speaking that does not contribute much to clarifying things, for the following reasons. I believe that, sometimes, the distinction in question is indeed relevant, but not always. Frequently, a jurist will refer to such and such an article of a law, and by this he may (usually) be alluding both to the text and to something like its basic meaning; no one (or almost no one), I believe, speaks of a legal system by referring exclusively to a set of

In relation to the above, there are a few things to be clarified. To begin with—and I return to something I said earlier—this difference between rules and principles must be seen in relative terms; to put it differently, it is a distinction within a continuum, in the sense that the "open" or "closed" character of the conditions of application is an obviously gradable element: between very specific guidelines for conduct (indubitable rules) and very abstract principles there is a very wide intermediate zone; and the same could be said of the more or less peremptory character of a reason. This also translates into a greater or lesser tendency for rules to have exceptions, to be defeasible. It is sometimes said that, if all norms are defeasible, then the very distinction between rules and principles collapses or, at the very least, that it could not be seen as a qualitative, strong distinction. Well, I believe that this distinction is of great importance (indispensable to understand many aspects of our legal systems), but it certainly cannot be interpreted in essentialist terms, but in the functional and dynamic way I suggested before. I am not so sure that it can be said that *all* legal rules (like all conditionals) are defeasible[24], but, certainly, most of them are (they can be defeated in some occasion), even in very extraordinary circumstances[25]. And this difference between what happens usually or extraordinarily is what allows us to maintain the distinction in question: principles usually function (whether they are principles explicitly fixed by the legislator or by the constituent, or—implicit—principles

statements, and excluding any idea of what the statements mean. But, in addition, there is a certain ambiguity in the use of the expression "interpretation" which, it seems to me, Guastini does not take into account in his work. Because "interpretative statement" can be understood as a statement of the form "T means S" (GUASTINI 2008, 152), but such utterances are only relevant in case there is any doubt about T. So one thing is interpretation in the noetic sense (as a mere act of apprehension of a meaning) and another in the dianoetic sense (when it is a matter of solving a doubt and a discursive activity is carried out). On this, see LIFANTE 1999. In short, I believe that there is no reason not to speak of defeasibility from the perspective of the system of norms, as long as norms are understood in the sense in which they are usually understood in the language of jurists. When a rule is established, the legislator may have formulated a general mandate or permission (or the conditions of validity of an act or of a rule) and added to it some explicit exceptions (which, indeed, has nothing to do with defeasibility) and he may also (having taken them into consideration or not) have left others unexplicit. When that rule has to be applied to solve a controversial case, interpretative activity will, of course, have to be carried out. But defeasibility is also a phenomenon that is present in the practice of the establishment of rules. The legislator can (must) count on the existence of this phenomenon.

[24] Recall what was said above (fn. 5) regarding Alchourrón's opinion. Also for MacCormick all or almost all legal rules (or instead, the formulations of rules) are refutable (I believe that the expression "rebatible" and "rebatibilidad" used in the Spanish translation is correct), in the sense that «[the rules] should be considered as stating 'ordinarily necessary and presumptively sufficient conditions' for the normative consequences they attach to the operative facts they stipulate». The reason why this is so is that «the principles and the implicit values of such a system interact with the more specific provisions to be found in the texts of statutes or in the more narrowly defined *rationes* of binding precedents» (MACCORMICK 2005, 251, 241).

[25] The prohibition of torture is often given as an example of an indefeasible norm. Perhaps it could be said that examples of indefeasibility refer to institutional actions. But, in any case, for what I am trying to defend here, the thesis that many of the norms (and, therefore, of the rules) can indeed have implicit exceptions is enough. Juan Carlos Bayón is right when he says that the possibility of implicit exceptions to rules existing or not (for reasons of principle) is a contingent question. Indeed, a legal system (or the practice of rule application) could exclude that possibility, or limit it a lot (it could be Schauer's "entrenched model" of rule application). But it seems to me that this is not what happens in our constitutional law systems... Schauer, by the way, has a very nuanced opinion in this respect: he thinks that sometimes rules are treated (by the applicators) as not defeasible and that defeasibility is not always desirable (which seems to presuppose that, in general, it is), see SCHAUER 2012, 85, 87. He distinguishes (a distinction that seems useful to me) regarding whether defeasibility is an essential feature of law, between a descriptive, a prescriptive and a conceptual level. His conclusion: «Defeasibility may well be a desirable component of some parts of some legal systems at some times, but it is far from being an essential property of law itself» (2012, 88). In other words, I conclude myself, rules cannot be completely opaque regarding the underlying reasons, but neither can they be completely translucent. And another (I think equivalent) way of saying the same thing: in normal cases the applicator does not (should not) consider the possibility of whether implicit exceptions exist, but he also cannot completely exclude the possibility of extraordinary (or very extraordinary) circumstances happening. See BAYÓN 2001, 54.

"discovered" by the interpreter) as non-peremptory reasons, to serve as ingredients in a deliberation, and that is why they can be defeated; whereas, regarding rules, this (that they are defeated) can only occur very extraordinarily. Moreover, this distinction does not exactly correspond to the often drawn distinction between easy cases and hard cases. Easy cases are those that can be solved with rules, that is, when the interpretation of the text—including, of course, possible explicit exceptions to a general command or permission—does not raise doubts; principles play here no other role than that of certifying—it is not properly a question of deliberating—that the solution to the case can be obtained by simply applying a pre-existing rule. Hard cases, on the other hand, are those that require balancing and in which, therefore, principles play a relevant role: either because, in the absence of an applicable rule, one must resort to principles, or because the rule has to be corrected (to broaden or restrict its scope) and this can only be done by appealing to principles.

And all of the above leads us to the following. When trying to control people's behaviour by means of general rules, the legislator has to cope with the open, contingent character of the future, and has to do so by trying to harmonise (balance) two fundamental values: one is that of giving as much certainty as possible to the addressees of the rules, that is, they should be in a position to know in advance the (legal) consequences of their behaviour; and the other is to avoid that such application of pre-existing rules produces counterproductive effects, that is, effects that are contrary to the aims and values that inspired the legislation, to the reasons underlying the rules. Rules essentially fulfil the first function, that is, they are in a very special way mechanisms of certainty; and principles fulfil the second, they allow the openness of the system, they avoid what would otherwise be excessive rigidity. But they act together, that is, legal practice needs to have both rules that are established with relatively closed cases and which can only be defeated in very exceptional circumstances, and principles, with norms whose cases are open, so that their defeasibility, as I said before, is previously programmed. And if this is so, then it is pointless to conceive a legal system as consisting essentially of either rules or principles; both types of statements are necessary. However, depending on the subject matter and other circumstances, it is possible that sometimes regulation must be done fundamentally by means of rules (for example, when establishing criminal offences), while on other occasions it is necessary to leave more room for principles (for example, when regulating matters such as assisted human reproduction, which is highly dependent on technological changes that happen in a practically incessant pace, that cannot be anticipated and, therefore, that prevent a regulation in very specific terms).

4. *Defeasibility in the process of interpretation and application*

Let us turn now to the other instance, that of the application of the rules, of the, so to say, raw legal materials (which in reality are not only rules), for the resolution of hard, controversial cases. The "easy cases/hard cases" distinction does not correspond exactly (but only approximately), as we saw before, with the pair "cases solved exclusively using rules/cases that also require principles"; and it would be more accurate to say that the correspondence is between cases that do not require deliberation/cases that do. Because principles, as I said before, also play a role in determining that a case is easy. But for that, one only needs to take a simple glance and realise that the case is covered by some rule (or, better, by a group of statements including rules) that does not contradict any principle of the system; whereas, in hard cases, that is not enough: an in-depth look is needed, concerning rules and principles; deliberation is needed. This distinction coincides, by the way, with the one that psychologists are used to making today (see KAHNEMAN 2011) between quick thinking and reflective thinking. Thus, recognising a case as normal or easy and whose resolution requires a "simple look" would be a way of referring to system 1 of thinking

which, as we know, is intuitive thinking that includes both the use of heuristics and expert thinking; while there are problems (abnormal, hard cases) that cannot be solved in this way, but instead require an "in-depth look", which would be, in turn, the way of referring to *system 2* of thinking, to slow and reflective thinking, which Kahneman links precisely with deliberation. To put it more briefly: our *system 1* is the one that comes into operation when we have (when a judge has) to solve problems of rule application, while the solution of problems that involve principles (that involve deliberation) means activating *system 2*[26].

In legal theory, various typologies of hard cases have been constructed. A widely followed one is that of MacCormick, who, on the basis of the scheme of the judicial syllogism, differentiates between problems of proof and qualification (referring to the factual premise), and problems of interpretation and relevance (referring to the normative premise) (MACCORMICK 1978). It is, undoubtedly, of considerable interest, but it falls short, in my opinion (see ATIENZA 2013), because, in his scheme, MacCormick starts, as a major premise, from a type of norm, a rule of action, and does not consider other possibilities. In particular, he does not take into account a situation in which there is (let us say, at first) no rule, but the applicator simply has principles to solve the case. Such a situation is a particular instance of a hard case, which is what, strictly speaking, can be called a balancing problem. This is distinguished from a (more) simple question of interpretation, which would be solved by simply opting for one of the different possible meanings of an expression. But when it comes to balancing, there is something more, that is, the applicator, in the beginning, has only principles and, therefore, he needs to make a step from the principles to the rule. Otherwise, there would be nothing to oppose to speaking of "interpretation" in these situations, but it would be a special type of interpretation. And the classifications of hard cases must, of course, be understood in a flexible and instrumental way: nothing prevents that for the resolution of problems of the other indicated types some balancing must also be done; at least, in the broad sense of the term: when a decision or action has to be taken, and there are several possibilities, opt for the one in favour of which there are the heaviest reasons[27].

[26] The idea of connecting the distinction between cases that require deliberation and cases that do not with Kahneman's two systems of thought occurred to me when I was trying to face a critique from Bruno Celano. According to Celano, the proposal of distinguishing two tiers in legal reasoning (the level of rules and that of principles) fails because it is not possible to know, in a specific case, whether the applicable rule is a rule that must be reconsidered or not, without reconsidering it: «to establish whether a cursory glance is enough or whether it is necessary to look closely, it is necessary to look closely»; so, ultimately, the distinction between rules and principles would vanish, because it would always be necessary to deliberate, to ponder, and, in consequence, there would be no rules that fulfill the function of simplification of the decision-making process (that avoid deliberating). It seems to me, however, that Celano's proposal to distinguish between normal and non-normal cases (easy cases and hard cases) in a "psychological key" is actually not different from the above. According to him, «whether or not a certain rule, in a given case, is a reason to act [a justifying reason, this is the crucial point] depends on our psychological make-up»; «what reasons we have—what we should do, what judgment we should adopt—depends (when the alleged reasons in question are rules) on a background condition, a regime of normality. Given a properly justified rule, which applies to a given case, it is reasonable to follow it only under a regime of normality. Whether this condition is satisfied or not depends on mental facts»; «It is not up to *us* to determine, in each of the cases that fall under their antecedent, whether the rule is to be reconsidered: It is up to our *mind*» (see CELANO 2017, my translation; the same ideas are defended in Celano's chapter in this book). However, why would we have reasons to trust that our mind is able to tell the difference between non-normal and normal cases (those in which a rule must or must not be reconsidered), but, nevertheless, it is not able to tell the difference between those cases in which a simple glance is enough to consider that they are normal and those that we detect as non-normal and that, therefore, require a deeper look? In short, it is about establishing a distinction between those two situations or cases (which seems essential), and for this a theory such as Kahneman's is really useful (which, by the way, has a clear precedent in a 1924 article by John Dewey about the logical method and law, DEWEY 1924). (See ATIENZA 2017b, 428 f.)

[27] This would be the principle of practical rationality which Raz calls "principle P1" (RAZ 1975, 36) and which Bayón explains as follows: «siempre se debe hacer lo que se tiene una razón concluyente para hacer, esto es, lo que resulte en

The recourse to balancing is of particular importance (and visibility) when it is used to solve a conflict between rights, which in our legal systems happens with some frequency; precisely as a consequence of the phenomenon of the constitutionalisation of legal systems, and of the impossibility of fundamental rights being fixed in the Constitution only or almost exclusively by means of rules, without resorting to principles. In reality, it is about the problem, already raised by David Ross, of the transition from *prima facie* duties to real duties, but in law it is more complicated (than in morality) because of the importance that institutional elements have gained: what is "correct" legally speaking has a moral component, but not only, in the sense that the judgement of correctness also has to take into account the characteristic aims and values of legal practice. The whole recent discussion on (judicial) balancing could be summarised, in my opinion (ATIENZA 2017a, ch. 6), along these three questions: 1) what does balancing consist of?; 2) when should we resort to it?; and 3) is balancing a rational instrument or a simple excuse to act arbitrarily? And the answers, from my point of view, would be these.

Balancing is a type of reasoning structured in two phases. In the first one—balancing in the strict sense—we move from the level of principles to that of rules: therefore, creating a new rule that did not previously exist in the system in question. Then, in a second phase, the starting point is the created rule and the case to be solved is subsumed in it. What could be called the "internal justification" of this first step is a reasoning with two premises. The first premise simply states that, in relation to a given case, there are two applicable principles (or sets of principles), each of which would lead to solving the case in mutually incompatible ways: for example, the principle of freedom of expression, to consider this type of conduct permitted; and the principle of respect for privacy, to consider it forbidden. The second premise establishes that, given the particular circumstances of the case, one of the two principles (for example, the principle of freedom of expression) defeats the other, it has a greater weight. And the conclusion would be a general rule, expressed in terms of universality, linking the above circumstances with the legal consequence of the prevailing principle: for example, if circumstances X, Y and Z are present, then conduct C is permitted.

Naturally, the difficulty of that reasoning lies in the second premise, and this is precisely where we find Robert Alexy's famous "weight formula", which would be, therefore, the "external justification" of the second premise. This doctrine is well known, and I am not going to explain it here[28]. What I am interested in clarifying is that this approach, at least as it has been understood by many jurists (not so much by Alexy himself), constitutes a fairly clear example of what Vaz Ferreira called the fallacy of false precision (VAZ FERREIRA 1962; ATIENZA 2013, 162 ff.). For, as is well known, Alexy proposes to attribute a mathematical value to each of the variables in his formula and thus constructs an arithmetical rule that creates the false impression that the balancing problems can be solved by means of an algorithm, thereby concealing the fact that the key to the formula lies, as is quite obvious, in the attribution of those values: that is, in determining whether the effect on a principle is intense, moderate or slight, etc. However, if the Alexian construction were to be understood in a sensible way, we would have something like an argumentative scheme that includes diverse topics and which can be very useful when constructing the external justification of that second premise: what it would mean is that, when it comes to solving conflicts between goods or rights (or between the principles that express them:

cada ocasión del balance global de razones a favor y en contra sopesadas según su fuerza relativa». But given the existence of reasons not only of the first order, but also of the second order, there would be another principle "P2" which is stated as follows: «no se debe actuar según el balance de razones si las razones que lo deciden son excluidas por una razón excluyente no derrotada». And the principle that would gather the two situations (the true practical rationality) would be "P3": «siempre es el caso que uno debe, habida cuenta de todos los factores relevantes, actuar por una razón no derrotada» (BAYÓN 1991).

[28] Anyway, I have dealt with it on several occasions. See ATIENZA 2019.

X and Y) and we have to decide whether measure M is justified or not, we need to construct a type of argument that contains premises such as (it could also be presented as a group of "critical questions" to be asked): "measure M is ideal to achieve X"; "there is no other measure M' that allows satisfying X without harming Y"; "in the circumstances of the case (or in the abstract) X outweighs—is more important—than Y"; and so on (see ATIENZA 2019).

In relation to the question of when does a judicial body have to balance, the answer is that it has to do so when the rules of the system do not provide an adequate answer to a case (there is a gap at the level of the rules); that is, when it is faced with a hard case and the judge needs to resort (explicitly) to the principles. Here, in turn, it is important to distinguish between two types of gaps (I insist: gaps at the level of rules): normative gaps, when there is no rule, no specific guideline of conduct that regulates the case; and axiological gaps, when the rule exists but establishes an axiologically inadequate solution, so that in this second case, so to speak, it is the applicator or the interpreter (not the legislator) who generates the gap.

Well, if we understand that the law, the legal system, is not necessarily complete at the level of rules, that is, that it can have normative gaps, then there is no other option but to accept that the judge (who cannot refuse to solve a case) has to do so by resorting in these cases to principles, that is, by balancing. Whereas, in relation to axiological gaps, the judge could resolve without balancing, but would then run the risk of incurring in formalism, that is, he would not be able to comply, in those cases of evaluative imbalances, with the claim to do justice through the law. In other words, there are certain situations in which the recourse to balancing by judges is simply unavoidable (although not for all judges: there can be an established rule that, when a judge is faced with such a situation, he must defer the case to a higher body). Whereas in relation to the others (with the cases of axiological gaps) a distinction should, in my opinion, be made between three types of imbalances: a) between what is stated in (the wording of) the rule and the reasons underlying the rule itself: the purposes for which it was made; b) between the reasons underlying the rule and the reasons (values and principles) of the legal system as a whole or of a part of it; c) between the reasons underlying the rule (and eventually the legal system) and others coming from a moral system or some moral principle not incorporated in the legal system. Without going into detail, I think it could be said (that legal common sense tells us) that in the first case it is not difficult to justify balancing (without considering here whether any judge should do it or whether the operation should be reserved for judges of supreme or constitutional courts); that in the third it is never difficult, as it would mean to stop playing "the game of law"; and that in the second is where the most complex cases arise: sometimes balancing may be justified (sometimes not), but it will have to be done with special care and assuming that the burden of argumentation lies in the one who intends to establish an exception to the rule (the one who creates the gap).

The recourse to balancing presupposes, therefore, the phenomenon of the defeasibility of norms. And it is true, as GUASTINI (2008, 150) says, that both the identification of a normative gap and (if you like, the creation) of an axiological gap are operations that require interpretation. But in different ways. In relation to normative gaps, it must be determined that there is no rule of the system whose literal meaning refers to the case, and that naturally requires interpretation, but it could simply be a matter of what has been called (see above fn. 23) a noetic interpretation. And if it is so (if there is no applicable rule), then it will be necessary to resort to principles, that is, to intrinsically defeasible rules, to see which one is stronger, given the circumstances. Whereas in axiological gaps, the interpretation is much more complex (and controversial), since it deals with a deviation from the literal interpretation of the rule, on the grounds that there is some implicit exception[29]. And in order to justify the existence of this

[29] It would mean moving from a literal interpretation to a restrictive one. But, in reality, it could also happen that the transition was to a broadening interpretation: the formulation of the rule did not include something that it should have included. In other words, the problem consists of an imbalance between the wording of the rule and its

exception (that is, the transition to—the creation of—a new rule), one must turn to principles. In the article by Guastini to which I have referred several times (GUASTINI 2008), there are some examples which I think may serve to illustrate what I mean. One of them consists of a constitutional provision which establishes that "The President of the Republic may veto the promulgation of laws", and this provision is interpreted as referring only to ordinary laws, and not to laws of constitutional revision (which means creating the axiological gap and solving it in a certain way). For this, instead of "principles", Guastini prefers to speak of "legal theories" and "dogmatic theses", but this is obviously balancing: the reasons in favour of that restrictive interpretation are stronger than those in favour of sticking to the literal meaning.

Finally, arguing that balancing is a rational procedure, does not mean asserting that, in fact, it always is, that is, it seems obvious that it is possible to balance badly (to appeal to balancing to conceal arbitrary behaviour) or to balance when (or by whom) it should not be done. But on many occasions, when one examines the argumentation—the balancing argumentation—carried out, for example, by a court in a series of cases involving, let us suppose, a type of conflict between two certain principles, one can detect the existence of a type of rationality, which consists of the following (see ATIENZA 1996). On the one hand, in the construction of a taxonomy (based on the properties that are considered relevant) that makes it possible to establish increasingly specific categories of cases: for example, not only the conflict between principle P1 and P2, but also between principle P1 accompanied by circumstance X and principle P2 accompanied by circumstance Y, etc. On the other hand, in the elaboration of rules of priority: for example, when those two principles are confronted while these circumstances apply, the first principle prevails over the second. And finally in the respect, regarding the configuration of the taxonomy and the rules of priority, to the criteria of practical rationality: consistency, universality, coherence, adequacy of consequences, reasonableness... Properly understood, properly put into practice, balancing is not a purely casuistic, arbitrary mechanism. The person who ponders must have the pretension that the solutions that he is configuring will serve as a guideline for the future, as a mechanism of prediction, even though it is an imperfect mechanism, in the sense that new circumstances may always arise that had not been taken into account until then and which may force to introduce changes in the taxonomy and in the rules. In particular, the rules that are constructed by means of balancing inevitably have an open character, they are defeasible. But that, as we know, is a characteristic feature of practical rationality[30].

5. *Defeasibility, balancing and juridical common sense*

Sometimes there are many different ways of saying the same thing, or almost the same thing. And this is what happens, in my opinion, with many discussions that take place in the field of legal theory in general or of more specific legal theories: what we usually call—in the world of

underlying reasons, its justification. On this see ATIENZA & RUIZ MANERO 2000.

[30] Guastini, criticising Hart, states that the idea that «a rule that concludes with the expression 'unless...' is still a rule [...] seems to me to be totally absurd» (GUASTINI 2008, 154, fn. 34). And that would be because a "defeasible rule" « cannot be used as a premise in any normative reasoning» (my translation). The latter is true, in the sense that in the premise of a justificative judicial reasoning, what will appear will be that norm interpreted in a certain way (the "defeated" norm). But I think Guastini is forgetting that norms also fulfil other functions such as, for example, serving as a guide (and as justification criteria) for conduct (and not only for that of judges). And a defeasible rule does fulfill this function, even if the addressee knows that, *exceptionally*, things could be otherwise. For the rest, it seems to me that Juan Carlos Bayón is right when he states that Hart's affirmation is sustainable «siempre que quepa reemplazar los puntos suspensivos por criterios o pautas que de alguna forma sean internos al propio derecho» (BAYÓN 2001, 56). A defense of Hart's theses (basically in the same terms as Bayón) can be found in MACCORMICK 2005, 253 f.

continental law—legal dogmatics. This may be due to an excessive desire for originality, to the desire for imitating what happens in the "hard sciences", to the existence of different traditions or schools of thought which, in turn, may have their origin in different legal cultures (for example, those of continental law and those of common law, formalist or anti-formalist), to the growing climate of isolation in which theories of law are developed (and I believe that the tendency to "intellectual autism" is far from being exclusive to jurists), or to various other causes. It is possible, moreover, that "enlarging" a small difference is sometimes important: it allows a better understanding of some concept, some relevant aspect of the law and, as a consequence, it can serve to develop the knowledge (and improve the practice) of law. But I believe that other (many) times this is not the case, and in particular it is not usually the case for—let us say—ordinary jurists (not the theorists or legal philosophers) who, in my opinion, should be the privileged recipients of these theoretical elaborations: those who have to solve legal problems, of whatever kind, and who could supposedly find some help in legal theory to do so. It should also be taken into account that in law (there is a reason why it is also part of practical reason) happens something similar to what David Ross pointed out about ethics: theories of law cannot deviate much from what we might call the good common sense of jurists; the legal method must also consist of some version of what has come to be called "reflective equilibrium". In order to avoid, therefore, as far as possible, that this work might contribute to increasing the risk I am warning about, I will point out the conclusions that follow, in my opinion, from what has been written in the previous sections and which, it seems to me, can be perfectly integrated into this legal common sense.

1. Defeasibility and balancing are more or less new names for realities that are not. And they are not, because they obey the intrinsic needs of any legal system: to regulate human conduct by facing, as far as possible, the unpredictability of the future, avoiding excessive rigidity and contributing, in short, to making the—unavoidable—breach that will always exist between law and justice as narrow as possible.
2. Defeasibility means that general rules may in some cases have implicit exceptions and, thus, that reasoning with rules may be affected by this: there may be extraordinary circumstances that force us to modify a conclusion that would justifiably have been reached under—let us say—normal conditions.
3. In a broad sense, to balance means to deliberate, that is, when a decision or an action has to be taken, and there are several possibilities, to opt for the one in favour of which there are the heaviest reasons. This is what defines the activity of the legislator, whose deliberations—from the legal point of view—are carried out within very broad limits. However, the law-applicator can only resort to balancing in relatively exceptional situations, and has to carry out this operation within much stricter limits.
4. The above means that the justificatory legal reasoning is not exclusively of a classificatory (subsumptive) type. It cannot be so in the case of the legislator, for obvious reasons: legislating does not consist simply of including a law—a norm—under some constitutional precept (or one of a higher rank than that of the new norm). And, on occasions, it is not so in relation to the applicator of the law either: when there is no rule with sufficiently determined conditions of application to be able to say that the case is subsumed in the norm, or when the subsumption of the case in the conditions of application of some norm is not enough to justify the decision.
5. To clarify the above, it is necessary to resort to a distinction between rules and principles, even if legal sentences are not simply of these two types. But in legal systems there are both specific patterns of conduct that operate as peremptory or conclusive reasons—rules—and very open rules that operate only as non-peremptory or non-conclusive reasons—principles. The distinction need not be seen in rigid terms (the properties closed/open or

peremptory/non-peremptory are given as a continuum), but it allows to explain that when there is a rule that is strictly applicable to the case, the case is solved (or its solution is justified) by subsumption; whereas the latter does not happen if only principles are available.

6. In establishing a general *rule* of conduct (and this is also true for constitutive rules or definitions), the legislator usually lays down explicit exceptions: in relation to what is ordered, to the conditions that must be met for a rule or a valid act to be produced, for a definition to be satisfied... But one cannot entirely exclude the possibility of implicit exceptions, which can be attributed to diverse factors (careless drafting of the text, impossibility of predicting future contingencies, acceleration of social change, growth of legal requirements as a consequence of the culture of rights...). Recognising the existence of implicit exceptions means recognising that the rule in question (and the reasoning that incorporates it) is defeasible. As well as the necessity of having to carry out a balancing exercise in the process of its application.
7. In the case of *principles*, and given their nature as open norms, it does not make sense to speak of exceptions, but it does make sense to speak of balancing. Principles are not defeasible like rules (because they provide non-exclusive, non-peremptory reasons, they do not present the typical resistance of rules). But the balancing of principles in a certain case does lead to a rule, whose case contains the open conditions of application of the applicable principles as well as the specific (closed) conditions that justified giving priority to one of the conflicting principles, and whose legal consequence will be precisely the one stated in the prevailing principle. Such a rule is not only general, but also universalizable: what it establishes applies (or should apply) as long as the (generic) conditions laid down in its case are met.
8. The importance that is nowadays recognised, in the theory and practice of law, of the existence of rules with implicit exceptions (which can be defeated in extraordinary situations) and of principles whose application generally leads to a process of balancing, has to do with changes that affect the reality of our legal systems and is linked to what is usually called the phenomenon of constitutionalisation. In particular, if what justifies law—the supreme value of constitutionalism—is the guarantee of fundamental rights, this could not be achieved within the scope of a very formalist culture that denies—or tries to reduce to a minimum—these two phenomena, linked to each other: the acknowledgement of implicit exceptions (defeasibility) and the recourse to balancing.
9. But the fact that we must leave a considerable space for the use of these two instruments does not mean that we must not set limits to them, that anything goes and that the law is completely or fundamentally indeterminate. It is not, among other things because, if it were, we would in fact cease to have rights: if rules were easily defeated, and law-appliers could solve the cases they were presented with by resorting to a balancing exercise whenever they thought (even with good reasons) that they would thereby make fairer decisions, the idea of having a right would vanish.
10. Law must be seen as an authoritative enterprise with which certain ends and values are to be achieved. The jurist, in his practical and theoretical work, cannot forget either of these two components. The authoritative element (the materials established by the authorities recognised as having such power in a state under the rule of law) sets the limits within which this finalistic and axiological activity can be carried out. These materials are (have to be) interpreted (in the broadest sense of the latter expression), but interpreting is not the same as inventing, creating something *ex nihilo*. Interpreting law requires going back to some moral and political philosophy that accounts for the legal materials; or, rather, to the one that best accounts for those materials.
11. If we transfer the above premise to the problem of balancing, what follows is that this operation can only be carried out, in the application of the law, in extraordinary situations: a) when there

is no rule—specific guideline—applicable to the situation, in other words, we would be faced with what is usually called a normative gap; b) when such a guideline does exist, but what it establishes—according to the textual or literal meaning—entails a conflict of some importance with the principles (and values) of the legal system: or, said in other words, when there is an imbalance between what is established in the rule and the underlying reasons.

12. With regard to normative gaps, the use of balancing involves an easily recognisable logical scheme that is articulated in two phases: the first one concludes with the establishment of a (general and universalizable) rule; the second one consists of a simple subsumption. It is therefore a more complex procedure than simple subsumption (deduction), but it is nonetheless rational; the criteria of rationality that can be used for its control are, in addition to those of deductive logic, those characteristic of practical rationality, in which coherence must play a particularly important role.
13. Axiological gaps present a more complex situation. As the applicator always has at his disposal the possibility of solving the case by applying the rule "on his own terms", he will have to start by carrying out a balancing whose result is that the reasons for creating the gap are of greater weight than those existing for simply applying the rule. In short, he has to justify the existence of an exception in the norm—in the rule—which would be implicit. This cannot be done without resorting to principles and, therefore, to values; but those values cannot be other than those of the legal system of reference.
14. Defeasibility and balancing are mechanisms for the innovation of the law, but coherently, that is, in accordance with the authoritatively established purposes and values; and it should be remembered that, in constitutional states, this authority is of a democratic nature. Moreover, this process of innovation has an open character (as is generally the case with practical rationality), so that the new rules that are made (and the new interpretations of principles and values) will also continue to present the characteristic of defeasibility.

References

Aarnio A. 1990. *Taking Rules Seriously*, in «ARSP-Beiheft», 42, 180 ff.

Aguiló J. 2000. *Teoría general de las fuentes del Derecho (y el orden jurídico)*, Ariel.

Alchourrón C. 2000. *Sobre derecho y lógica*, in «Isonomía: Revista de Teoría y Filosofía del Derecho», 13, 11 ff.

Alexy R. 1989. *A Theory of Legal Argumentation: The Theory of Rational Discourse as Theory of Legal Justification*, Oxford Clarendon Press.

Alexy R. 2002. *A Theory of Constitutional Rights*, Oxford University Press.

Aristotle 1984a. *Nicomachean Ethics*, in Barnes J. (ed.), *The Complete Works of Aristotle*, Princeton University Press, 1729 ff.

Aristotle 1984b. *Rhetoric*, in Barnes J. (ed.), *The Complete Works of Aristotle*, Princeton University Press, 2152 ff.

Atienza M. 1996. *Juridificar la bioética. Bioética, derecho y razón práctica*, in «Claves de Razón Práctica», 61, 2 ff.

Atienza M. 2013. *Curso de argumentación jurídica*, Trotta.

Atienza M. 2017a. *Filosofía del Derecho y transformación social*, Trotta.

Atienza M. 2017b. *Epílogo (abierto)*, in Aguiló J., Grández P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 419 ff.

Atienza M. 2018. *Homenaje a Riccardo Guastini*, in Chiassoni P., Comanducci P., Ratti G.B. (eds.), *L'arte della distinzione: Scritti per Riccardo Guastini*, Marcial Pons, 15 ff.

Atienza M. 2019. *Argumentación legislativa*, Astrea.

Atienza M., Ruiz Manero J. 1996. *Las piezas del Derecho. Teoría de los enunciados jurídicos*, Ariel.

Atienza, M., Ruiz Manero J. 2000. *Ilícitos atípicos*, Trotta.

Bayón J. C. 1991. *Razones y reglas: sobre el concepto de "razón excluyente" de Joseph Raz*, in «Doxa. Cuadernos de Filosofía del Derecho», 10, 25 ff.

Bayón J. C. 2001. *¿Por qué es derrotable el razonamiento jurídico?*, in «Doxa. Cuadernos de Filosofía del Derecho», 24, 35 ff.

Campos R. 2023. *Metamorphoses of Global Law. On the interaction of law, time and technology*, Bloomsbury Publishing.

Celano B. 2017. *Particularismo, psicodeóntica. A propósito de la teoría de la justificación judicial de Manuel Atienza*, in Aguiló J., Grández P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 59 ff.

Chiassoni P. 2019. *Interpretation without Truth: A Realistic Enquiry*, Springer.

Dascal M. 1996. *La balanza de la razón*, in Nudler O. (ed.), *La racionalidad: Su poder y sus límites*, Paidós, 363 ff.

Dewey J. 1924. *Logical method and law*, in «Cornell Law Quarterly», 10, 17 ff.

García Salgado M.J. 2010. *La jurisprudencia de intereses de Philipp Heck*, Comares.

García Salgado M.J. 2019. *La ponderación de intereses como método*, in «Eunomía: Revista en Cultura de la Legalidad», 16, 283 ff.

Guastini R. 2008. *Variaciones sobre temas de Carlos Alchourrón y Eugenio Bulygin. Derrotabilidad, lagunas axiológicas, e interpretación*, in «Doxa. Cuadernos de Filosofía del Derecho», 31, 145 ff.

Hage J., Peczenik A. 2000. *Law, Morals and Defeasibility*, in «Ratio Juris», 13, 305 ff.

HART H.L.A. 1948. *The Adscription of Responsibility and Rights*, in «Proceedings of the Aristotelian Society», 49, 171 ff.

HART H.L.A. 1958. *Positivism and the Separation of Law and Morals*, in «Harvard Law Review», 71, 593 ff.

HOLMES O.W. 1897. *The Path of the Law*, in «Harvard Law Review», 10, 457 ff.

JHERING R. 1913. *Law as a Means to an End*, Boston Book Company.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

LACEY N. 2006. *A Life of H.L.A. Hart: The Nightmare and the Noble Dream*, Oxford University Press.

LIFANTE I. 1999. *La interpretación jurídica en la teoría del Derecho contemporánea*, Centro de Estudios Políticos y Constitucionales.

LUHMANN N. 1977. *Sociologia del diritto*, Laterza (Italian translation by A. Febbrajo).

LUHMANN N. 1990. *La differenziazione del diritto. Contributti alla Sociologia e alla Teoria del diritto*, il Mulino (Italian translation by R. De Giorgi).

MACCORMICK N. 1978. *Legal Reasoning and Legal Theory*, Oxford Clarendon Press.

MACCORMICK N. 1995. *Defeasibility in Law and Logic*, in BANKOWSKI Z., WHITE I., HAHN U. (eds.), *Informatics and the Foundation of Legal Reasoning*, Kluwer, 99 ff.

MACCORMICK N. 2005. *Rhetoric and the Rule of Law. A Theory of Legal Reasoning*, Oxford University Press.

PINTORE A. 1990. *Teoria analitica dei concetti giuridici*, Casa editrice Jovene.

PLATO 1997a. *Republic*, in COOPER J.M. (ed.), *Plato: Complete Works*, Hackett, 1000 ff.

PLATO 1997b. *Statesman*, in COOPER J.M. (ed.), *Plato: Complete Works*, Hackett, 323 ff.

POGGI F. 2021. *Defeasibility, Law, and Argumentation: A Critical View from an Interpretative Standpoint*, in «Argumentation», 35, 409 ff.

POSCHER R. 2022. *Aciertos, errores y falso autoconcepto de la teoría de los principios*, in GARCÍA AMADO J.A., DALLA-BARBA G.R. (eds), *Principios jurídicos. El debate metodológico entre Robert Alexy y Ralf Poscher*, Palestra Editores, 85 ff.

RAZ J. 1975. *Practical Reasons and Norms*, Oxford University Press.

RODRÍGUEZ J.L., SUCAR G. 1998. *Las trampas de la derrotabilidad: niveles de análisis de la indeterminación del derecho*, in «*Doxa.* Cuadernos de Filosofía del Derecho», 21, 403 ff.

ROSS D. 1930. *The Right and the Good*, Oxford Clarendon Press.

ROSS D. 1939. *The Foundations of Ethics*, Oxford Clarendon Press.

SCHAUER F. 2012. *Is Defeasibility an Essential Property of Law?*, in FERRER J., RATTI G. (eds.), *The Logic of Legal Requirements: Essays on Defeasibility*, Oxford University Press, 77 ff.

SIECKMANN J-R. 1990. *Regelmodelle und Prinzipmodelle des Rechtssystems*, Nomos.

TOULMIN S. 1958. *The Uses of Argument*, Cambridge University Press.

VAZ FERREIRA C. 1962. *Lógica viva*, Losada.

VON WRIGHT G.H. 1963. *Norm and Action,* Routledge & Kegan Paul.

Defeasibility and Practical Errors

RAFAEL BUZÓN

1. *Introduction* – 2. *Legal positivism* – 3. *Legal postpositivism* – 4. *Conclusions*

1. *Introduction*

Defeasibility is usually understood as the possibility that rules contain implicit exceptions. It is a traditional problem in philosophy of law, which was already present in Aristotle and his idea of equity, and which Hart baptized in 1948 with the name of defeasibility (HART 1948), generating a profuse discussion ever since. A discussion that has produced a very wide bibliography[1] mostly in the context of analytical legal positivism[2]. Nevertheless, it seems to me that the so-called problem of defeasibility is not strictly a problem of ambiguity of the term, nor of vagueness of the concept, but a theoretical problem. What I mean is that the concept of defeasibility is affected, as most concepts, by ambiguity and vagueness (both extensional and intensional), but that we are not dealing with semantic difficulties that can be solved by fixing a list of descriptive properties, but instead with difficulties that arise from the fact that the properties of the concept are centrally evaluative. This is so because the problem of defeasibility in law is inevitably linked to the discussion on the necessity of its foundation[3]. And in this discussion, the theories of law from which the problem is approached take on special relevance.

I will lay out the problem as a theoretical confrontation between positivism and post-positivism, as referents of the two most relevant theories of law in contemporary discussion. For that purpose, I will follow the traditional classification within legal positivism between exclusive legal positivism, inclusive legal positivism and ethical positivism. Later on, I will focus on two of the most refined positivist proposals: the deep conventionalism of Juan Carlos Bayón and the inclusive legal positivism of José Juan Moreso and his theorization of defeaters. Once these positions have been analyzed, I will try to show that the problem of defeasibility is essentially a practical problem that tries to avoid a practical error and that, for this reason, legal postpositivism is a much more adequate theory to explain and operate with the phenomenon.

2. *Legal positivism*

Roughly speaking, legal positivism is characterized by being a theory of law that conceives of the legal system as a set of rules understood as norms that correlate the closed description of a case with a normative solution. Among these rules there are logical relationships of deducibility[4], so that the model of legal reasoning is fundamentally subsumptive. For a legal positivist, being loyal to a rule is, basically, to be so to its expression and to its meaning. These rules are not identified by their content, but by their form, and for that reason they hold that the origin of rules is what determines their legality. All the law is based on conventions, on sources

* All translations made of literal quotes in Spanish are my own.

1 A good presentation of the topic and of the various contemporary currents can be found in CARPENTIER 2014. More recently in GARCÍA YZAGUIRRE 2022.

2 A sample of the most relevant positions, not only positivist, can be found in FERRER & RATTI 2012.

3 As happens with many of the legal concepts, for example, that of punishment. See TORRES ORTEGA 2020, 330.

4 The title (and content) of a recent writing of two contemporary legal positivists is very illustrative: *La derrotabilidad jurídica como relación sistemática compleja* (DOLCETTI & RATTI 2016).

of law, and everything beyond conventions is not legal. Hence a clear or easy case is one which has a single conventionally acceptable answer, and a controversial or hard case is one which has more than one conventionally acceptable answer. In this last case, as the legal system (the applicable convention) does not determine one single answer, the judge has the discretion to decide (AGUILÓ 2007).

However, this characterization of legal positivism is too coarse and does not account for the multiplicity of currents that it holds. Without going into detail, we could say that the three main currents are exclusive legal positivism, inclusive legal positivism and ethical positivism, all of them developed in the context of Hart's conventionalist turn in his *Postscript* (HART 1984; on this point GARCÍA FIGUEROA 2019). These three theories differ, in essence, with regard to their characterization of the rule of recognition and, in the end, to the criteria of legal validity. Exclusive legal positivism holds that the identification of law cannot depend on its adequacy to morality; inclusive legal positivism holds that the dependence on morality to identify the law is contingent; and, finally, the position of ethical positivism is that the identification of law must not depend on its adequacy to morality.

Within these theories, I am interested in highlighting two particularly refined approaches: the deep conventionalism of Juan Carlos Bayón and the inclusive legal positivism of José Juan Moreso and his theory of defeaters.

Bayón, following Michael Moore's terminology, wants to take legal conventionalism to its last consequences, calling his theory "deep conventionalism" (BAYÓN 2002b, 51 ff.). According to conventionalism, which Bayón understands as the minimum content of legal positivism, the limits of the law are the limits of our conventions and, therefore, the identification of the law is, in principle, a mere matter of social facts. However, for deep conventionalism, an agreement in all cases by all subjects would not be necessary to affirm the existence of a convention, but instead it would be enough to have simply «the agreement regarding certain paradigmatic cases that are recognized as correct applications of the rule». That is, what defines a rule as correct would not be the explicit agreement on its particular applications, but instead the background of shared criteria. For that reason, to identify the law it is not enough (as for classic conventionalism) to limit our analysis to the simple observation of uncontroversial social facts, but it is rather necessary to accept that the determination of the content of the law will take the form of a coherentist deliberation.

Bayón's aim, as GARCÍA FIGUEROA (2018) shows, is none other than isolating the theory of law from moral philosophy in order to identify the law in a premoral stage, oblivious to the need of correctness[5]. But this theoretical thesis has, at least, two consequences. The first one is that it radically contrasts conventionality with correctness, understanding the latter as what is not conventional, with the clear objection that there are partially conventional moral theories, such as ethical constructivism. And the second one is that the outline of the limits of the convention is two-dimensional. If we follow the distinction proposed by VEGA (2021) between "dintorno", "contorno" and "entorno" [6] to analyze the theories of law in their conceptions of the limits of what is juridical, we can see how Bayón, who tries to outline the sphere of autonomy of the law, has a circular and negative idea of limit, that is, a self-referential one. Thus, Bayón

[5] As Bayón says «however complex and controversial the reasoning aimed at establishing the content of our conventions may be, it should not be confused with genuine moral reasoning, which is precisely the one that operates as a critical instance from which to evaluate the content of any kind of conventional rule» (BAYÓN 2002b, 54, fn. 53; my translation).

[6] The "dintorno" refers to the unity of the legal category, the "contorno" to the demarcation of said category and the "entorno" to the environment of the "dintorno" and the "contorno". The image of a circle may be the one that best illustrates this distinction. The "dintorno" would be the inside part of the circle, the "contorno" would be the line with which you separate the "dintorno" from the "entorno" and the "entorno" is what lies beyond the drawn line.

draws the limits of conventions only regarding their closure in relation to its "dintorno". He is concerned only about the outline of its "contorno". However, the "contorno" must necessarily be defined by the "entorno" and, therefore, it must be set up from the outside too. The drawing of a limit, in this case the limit of a convention also has to be externally demarcated, because the "contorno" is formed both by the "entorno" and the "dintorno", which requires a more complex reasoning in order to show the diversity of the relationships "entorno-contorno-dintorno" than that of the simple demarcation between inside and outside. The separation between the system of rules of a conventional basis and the critical ethico-political ideals is based on the self-closing postulate of the conventional system, creating a "dintorno" held by itself, which, as Vega states, could be called a Münchhausen strategy.

And, as far as our problem is concerned, the pertinent question is: what is a practical error for a deep conventionalist? Answering this question already presupposes the adoption of the internal point of view and the internal point of view is precisely what distort deep conventionalism. Even if practical error is defined on the basis of the agreement on certain shared criteria based on certain paradigmatic cases of error, it is necessary to leave the convention, at the risk of falling in a consensus by convention and not by conviction (DWORKIN 1977). We have a problem, precisely, with the limits of convention, and convention itself cannot be what determines the result, just like the acceptance of a convention cannot come from the convention itself[7]. In short, modulating an ancient adage, we could say that the exception confirms the convention.

On the other hand, Moreso adheres to inclusive legal positivism. Let us remember that inclusive legal positivism is that theory of law that maintains the contingent connection between law and morality. More precisely, it maintains that for the identification of the law it is neither necessary nor impossible to appeal to moral criteria. This theory has been argued mainly by WALUCHOW (1994), COLEMAN (2001) and, among Spanish scholars, by MORESO (2001). For the issue at hand, Moreso has proposed an image of law that challenges the postpositivist approach[8].

Moreso argues, along Hart's lines, that the conception of law in two levels[9], those of rules and principles, collapses and must be replaced by an image with a single level composed of rules and defeaters (MORESO 2020, 87). Let us go step by step.

First, focusing on Atienza and Ruiz Manero, Moreso argues that the distinction in two levels of legal reasoning does not destroy the objection, announced by RAZ (1972, 823-854), that in the two-level approach not only the principles are *pro tanto* guidelines, but that rules would be so too. If one of the main functions of principles is to make exceptions to rules, then rules would not have closed conditions of application, and so they would only apply when they were not

[7] Even though we understood by convention the background of shared criteria. Curiously, Bayón recognized this fact in 1991: «the last operative reasons of a justificative legal reasoning (of one that takes into account the fact of the existence of legal rules) cannot be legal reasons, that is, the legal norms themselves, including the last norm which is the rule of recognition of the system: because either they are considered as practical judgements that are accepted for their content (and then their acceptance is indistinguishable from that of an ordinary moral judgment, i.e., of one that is not dependent on the fact of the existence of those rules), or else the fact of their existence is taken into account as an auxiliary reason, in which case the last operative reasons that give practical relevance to the existence of the last legal dispositions, and which it makes no sense to qualify in turn as legal, are presupposed as operative reasons. They also cannot be prudential reasons of the subject who develops the reasoning, since a prudential practical reasoning that takes into account the existence of legal norms is not apt to justify decisions that are imposed to others whatever their interests are: one can do what the law demands for prudential reasons, but one cannot appeal merely to their own interests to justify that another must do something. Thus, in the end, the operative reasons of a justificative legal reasoning must be moral reasons» (BAYÓN 1991, 737; my translation).

[8] Despite that, Moreso has recently indicated his coincidence with authors such as Atienza and Ruiz Manero in MORESO 2017, 205, and MORESO 2022, 564.

[9] Just as it was proposed by DWORKIN 1977, ch. 2 and then by ALEXY 1986, ch. 3 and ATIENZA & RUIZ MANERO 1996.

defeated by the force of a principle. This argument, according to Moreso, would cause the second floor of the building to collapse into the first one: principles end up usurping the place of rules, and there is only space left for a jurisprudence of reasons (MORESO 2018, 125 ff.).

Secondly, Moreso's proposal is built on the supposed failure of the previous one. He proposes an image of a single level where rules coexist with defeaters. For Moreso, defeaters are mechanisms available to the law to activate access to underlying reasons (MORESO 2021, 569), to authorize the return to the deep level of reasoning (MORESO 2016, 17) or, in short, to resort to moral reasoning (MORESO 2020, 87). Some examples of defeaters in law would be defences and excuses in criminal law, the conditions of invalidity in contract law, or indeterminate legal concepts.

In this respect, MORESO (2020b, 182) compiles a list of defeaters along the line of POLLOCK (1974; 1986), SINNOT-AMSTRONG (1988; 1999), MONTAGUE (1995) and BAGNOLI (2018):

1. *Cancelling defeaters*. For example, when my friend lends me a book and he says that I can keep it. In this case my duty of giving it back is cancelled.
2. *Excusing defeaters*. Moreso uses the case of coercion in criminal law: the action is still obligatory or forbidden but the author is not responsible.
3. *Overriding defeaters*. When there is a conflict of duties, but one of them prevails over the other[10].

Eventually, Moreso advocates for a model in which rules and the way of activating their exceptions can coexist in the same building without collapsing into a jurisprudence of reasons (MORESO 2018, 129). According to him, this is the way in which the law tries to preserve the ideal of the rule of law, combining legal certainty with formal justice and equity (MORESO 2018, 116).

Although interesting, Moreso's approach seems to have some problems. To begin with, the adoption of a one-level model with rules and defeaters does not take into account one of the fundamental reasons why postpositivist scholars use a two-level system precisely within the framework of the constitutional rule of law, that is, because rules are nothing but the result of the balancing of principles, so that, by definition, they cannot be at the same level. Principles give sense to rules and justify them. Also, as Ródenas has shown developing the posture of Atienza and Ruiz Manero, it is

> «perfectly conceivable that a subject, before applying a rule, deliberates on whether the result of applying it is compatible with the compromise between reasons that is expressed in the latter, without for that stopping to consider said rule as a peremptory reason for action. Whoever, before applying a rule, asks himself about its scope or exceptions (in the terms that I have pointed out here) does not question the compromise or judgment of prevalence between reasons that the rule expresses. Hence an agent can consider a rule as peremptory and not apply it to a case that he considers excluded or outside of its scope» (RÓDENAS 1998, 118 ff.).

Furthermore, it seems that the reconstruction of Moreso only accounts for explicit exceptions, that is, for the mechanisms that are already provided by the law to defeat rules. Nevertheless, the genuine problem of defeasibility appears, precisely, when there is no applicable legal convention. These problems generate, at least, the following questions: what is the place for principles in the model proposed by Moreso? In the case of the absence of an applicable legal convention, what happens in this zone of entrance of the defeaters: discretionality or right answer in material terms? And, finally, the most general and important question of all: what are

[10] Here Moreso makes two distinctions: (a) First distinction: (i) Strong: without residue; (ii) Weak: with residue (duty of compensation). (b) Second distinction: (i) Rebutting (from prohibited to obligatory, from obligatory to prohibited); (ii) Undercutting (from obligatory or prohibited to facultative).

really the defeaters? Regarding this last question, it may be reasonable to answer that the defeaters are mechanisms to avoid possible practical errors working as a conveyor belt between the level of rules and the level of principles. I will come back to this point later.

3. *Legal postpositivism*

By postpositivism I mean, in very general terms and for the point at hand, the theory of law (represented by Dworkin, Alexy, Nino, the last MacCormick, and Atienza) which holds that, besides rules, there are also in law other patterns of conduct such as principles, that are norms that establish what ought to be without specifying when their normative solutions are applicable. Consequently, the model of legal reasoning will be subsumption in the case of rules and balancing in the case of principles. Now the relationships between norms will not only be relationships of deducibility but also of justification: principles ground rules, they justify them, and thus rules are not seen as the product of purely authoritative acts of creation, but as the result of acts of developing, concretizing and balancing principles. Law is based on sources, on conventions, but not everything within the law is convention: beyond the criteria of formal validity there are criteria of material validity, which bring as a corollary the problem of implicit law. Taking this issue seriously implies that in law there are no unregulated relevant cases or, seen from another perspective, that the Dworkinian regulative ideal of "(only) one right answer" is assumed: cases are easy if there is an answer provided by the legal system which is logically consistent with other rules of the system *and*, at the same time, which has evaluative coherence with the principles of the system itself; and cases are hard when the system does not directly provide a predetermined solution and, therefore, it is necessary to unfold an intense argumentative activity to find it. From this point of view, discretion is not pure freedom of choice, but a responsibility of the adjudicator; indeterminacy is not confused with uncertainty[11].

If we follow postpositivist reasoning, defeasibility ceases to be seen as a problem of systematic relationships between rules and becomes a practical problem: the problem of implicit law. Implicit exceptions are not the result of the creative and discretionary activity of the judge, but rather justified reasons (arguments) brought out, precisely, from implicit law. And, consequently, the fundamental questions of postpositivism turn on the problem of whether there is or not a rational legal method to explicitly state implicit law. From the perspective of legal postpositivism, defeasibility is not seen as a problem of such dimensions (it is said that defeasibility can challenge the basis of exclusive positivism) but as a consequence of law's being a social practice: the fact that there is not an answer more or less directly predetermined by the system does not necessarily mean that there is no answer, but that it is more difficult to find it, and that it can usually be found because there are principles in the law. If defeasibility is seen as a problem of coherence, of adjusting the directive dimension of rules to its value dimension, the "one right answer" thesis still makes sense, at least as a "regulative idea". Moreover, in most legal systems there are mechanisms to adjust the value dimension of law to its directive dimension[12].

It seems that some positivist scholars stop at the Dworkinian "interpretative" stage, when they confirm that there is more than one conventionally acceptable answer to a case and that, consequently, the judge has to choose discretionarily one of those answers. They disregard the

[11] As Aguiló argues, following Dworkin, the thesis of judicial discretion leads us to the thesis of indeterminacy of law in controversial cases, while the one right answer thesis leads us to the thesis of uncertainty in hard cases (Aguiló 2021, 18).

[12] It is the case, for example, of atypical illicit acts (abuse of right, legal fraud, and deviation of power). See Atienza & Ruiz Manero 2000.

idea of a "postinterpretative" moment, the moment of judging which of these alternative answers is the best, the correct one[13].

But let us continue pulling on the thread of exceptions. The idea of implicit exception itself is directly related to implicit law, which only makes sense to talk about when we try to solve cases using legal rules. For legal positivism, implicit exceptions, if they are not the product of logical derivation from other explicit rules, since they do not refer to any convention (there is no behavioral guide), are inevitably discretionary acts of judicial creation.

In contrast, for legal postpositivism, implicit rules are those that are neither logical consequences of the explicit rules nor products of acts of will (AGUILÓ 2000, 183). If we look at the problem from the point of view of the sources of the law, which revolves around the connection between legal norms and the processes of which they are a product, talk about "implicit law" (product) and talk about "legal method" (process) cannot be separated. Having said that, from this point of view, the legal method takes into account, in addition to the directive dimension of rules, the value dimension of law. This means that, next to the criteria of normative logical *consistency* typical of legal positivism (nonconflicting duties), we need to add the criteria of evaluative coherence (nonconflicting values).

Considering duties independently of their orientation towards certain purposes has the consequence of accepting that for the law ritualist behavior is not a deviated form of behavior[14]. If we reject this image of law and we accept that every duty is linked to at least one purpose from which it gets its cause, law appears to us as an institution primarily aimed at the protection of certain goods.

In this light, defeasibility, therefore, is just another way of talking about the problem of the rationality of legal decisions. If we look at it from the perspective of building and justifying particular solutions using general rules, instead of that of finding in general rules the built-in solutions to particular cases, there appears the figure of the legal agency, the legal operator or adjudicator, which is completely overshadowed in the positivist vision described above. In these cases, the judge has to unfold an intense argumentative activity to provide reasons for the existence of a normative problem, a gap, from which comes the force of apparently self-contradictory assertions such as that "the exception confirms the rule". However, the justification of the problem goes hand in hand with the argumentative operations that are oriented to its solution. It is not merely a chronological procedure: there is a constant interaction between the various stages of the argumentative process (there is a recurrence both to the system, to the raw legal materials, and to the various normative propositions that are constructed). The judge has to construct the rule that is going to be the normative premise of his legal reasoning, he has to formulate and justify the rule that resolves the case.

Postpositivism endorses a strong view of practical reason and of practical error. However, the notion of practical error needs to be developed. For present purposes, the idea of practical error can be elucidated resorting to Atienza's three dimensions of argumentation (ATIENZA 2006): formal, material and pragmatic.

Very briefly, according to the formal dimension, legal reasoning is a set of non-interpreted sentences, that is, a reasoning in which we abstract from the truth or correctness of their content. Its aim is to determine whether from certain sentences with a certain form (premises) you can pass on to other sentences (conclusion). Thus, the criteria of validity are given by the

[13] AGUILÓ 2021, 17. For an extensive treatment of this problem and of the evaluation criteria of judicial argumentations, see ATIENZA 2013, ch. VII.

[14] For example, if we followed the thesis of Alchourrón and Bulygin, we would have to hold that the judges of the case *Riggs v. Palmer* acted incorrectly when they denied the inheritance to the grandson that had murdered his grandfather. That is the position, for example, of MARTÍN FARRELL 2014.

rules of inference. Under the material dimension, legal reasoning is conceived as a theory of premises understood as good reasons. Its aim is answering questions such as what action is due or what beliefs are valid as premises and conclusions. It focuses, then, on the semantic aspects of language, on content, for the assessment of which we have justification criteria, maxims of experience, scientific laws, etc. Finally, the pragmatic dimension focuses on the use of argumentation, that, from this perspective, is seen as an activity, as a practice and a social relationship, and it is divided into rhetoric and dialectics.

On the basis of these assumptions, it is possible to classify practical errors into formal, material and pragmatic (dialectical and rhetorical) errors. However, it seems to me that it is necessary to take into account the codetermination between the three dimensions, not only to avoid unjustified reductionism, but also to understand that in practical errors —even if, of course, they can be linked to the formal and pragmatic dimensions— what matters the most is the material dimension, because all practical errors presuppose the interplay between law, morals and politics.

Going back to the thread of the exposition, what is important for the problem of defeasibility is to realize how naturally postpositivists approaches incorporate it. There are, no doubt, various illustrations of this. Let us take, for example, the concept of rule in the sense of «the normative premise of a finished legal reasoning» formulated by AGUILÓ (2000, 185 ff.; my translation).

According to Aguiló, this notion of rule 1) is a construction of legal agents; 2) contains all relevant factors to the solution of the particular case; 3) has no specific normative hierarchy; 4) has the "force of law", of the whole legal system; and 5) its purpose is to justify the particular judgement taken by the judge.

That the rule is constructed by legal agents is a corollary of what said above. The judge has to decide to use the authoritative materials he has identified as legal. His process of building or constructing the rule is the step—or set of steps—that goes from the «normative statement to be interpreted» to the «interpreted normative statement» (i.e., the rule)—the link between them being an «interpretative statement» that solves a problem of interpretation. To state it more clearly: the construction of the rule is the step from «the rule that the judge has the duty to apply» to «the rule that the judge [finally] applies».

That this (final) general rule contains all of the relevant factors for resolving the case refers to the fact that the judge, in constructing the normative premise of her reasoning, has to articulate and resolve all the tensions of the systematic relationships that arise at the time of application: hierarchical, semantic, chronologic, genetic, etc., but also the tension dependent on the indeterminacy of both law and the facts[15].

The consequence of having resolved in the previous step all systematic relations, including hierarchical ones, is that the final rule has no hierarchy: it is the solution all-things-considered, once the whole legal order has been revised and once the one voice with which the law as a whole has to speak has been articulated. Hence, we affirm that this construction has the "force of law" and that its main purpose is to justify the judgement, i.e., how that particular case is decided.

This concept of rule is thus an attempt to account for the transition from the rules of objective law to the result of their interpretation after all things have been considered[16]. The fact that the law must speak with a single voice is independent of the fact that objective law is

[15] It must be noted that a good part of the theorists of defeasibility confine it exclusively to the scope of solving systematic relationships, trying to solve the problem in terms of rule validity and closing their reasoning with a last master rule, either presupposed by the theorist (Kelsen's *Grundnorm*) or empirically observable as a social practice (Hart's rule of recognition). But, as Nino observed, what is really interesting is not whether a rule belongs to a legal system in a descriptive sense, but if the rule must be applied to justify an action or decision (NINO 1992, 49).

[16] The idea behind this reasoning can be summarized in the expression «the construction of the case», which was extensively treated by the hermeneutic doctrine of authors such as Larenz or Hruschka (see TARUFFO 2011, 100 ff.).

unitary. It is a matter of translating the unity of law into action: all systematic and value relationships in tension have already been resolved and, therefore, the conclusion of this reasoning is a practical proposition, the representation of an action as justified, correct.

What I am trying to show is that defeasibility in law cannot be separated from the idea of problem. Defeating a rule always presupposes the detection of a problem resulting from a judgement of relevance made by the judge, and this problem always consists in a practical error. Legal reasoning begins by being systematic: if the system provides an answer to the case the judge will limit himself to constructing the normative premise by giving meaning to the relevant legal materials provided by the system itself[17]. If the system does not give an answer or gives an evaluatively unacceptable answer, we consider that there is a problem. And again, the fact that this problem does not have a predetermined solution does not mean that the law is indeterminate, but that it is uncertain.

4. *Conclusions*

After presenting the problem of defeasibility from the point of view of the theories of law, two proposals of contemporary legal positivism have been analyzed: Bayón's deep conventionalism and Moreso's inclusive legal positivism and his theory of defeaters. Facing the insufficiencies of both theories to account for the problem of defeasibility, the basic theses of legal postpositivism have been presented, going deeper in the idea of practical error and in the concept of rule as a premise of a finished legal reasoning, as proof that the problem of defeasibility is nothing other than that of the rationality of legal decisions, of making implicit law explicit, and that in this regard there are rational criteria that help us to solve hard cases and avoid making practical errors.

[17] I am referring, again, to the process that lies between identifying the normative materials and the decision to use those normative materials. That is, the step from the raw legal materials to normative premises. As Nino puts it, this is the step we make when we pass from the judgement "the constituent C has prescribed: 'no one can be detained without a written order of the competent authority'", that does not allow to justify any action or decision, to "no one can be detained without a written order of the competent authority", that does allow it. He explains that, in order to make this step, it is not only necessary to presuppose a principle such as "the constituent C is a legitimate authority and must be obeyed", but it is also necessary a rule of interpretation that allows to remove the quotation marks in the first sentence (NINO 1992, 82).

References

AGUILÓ J. 2000. *Teoría general de las fuentes del Derecho (y el orden jurídico)*, Ariel.

AGUILÓ J. 2007. *Positivismo y postpositivismo. Dos paradigmas jurídicos en pocas palabras,* in «Doxa. Cuadernos de Filosofía del Derecho», 30, 665 ff.

AGUILÓ J. 2021. *El tribunal se retira a deliberar. Un desafío teórico para juristas prácticos*, in «Revista jurídica de Les Illes Balears», 20, 11 ff.

ALEXY R. 2001. *Teoría de los Derechos Fundamentales*, Centro de Estudios Políticos y Constitucionales.

ATIENZA M. 2006. *El derecho como argumentación,* Ariel.

ATIENZA M. 2013. *Curso de argumentación jurídica*, Trotta.

ATIENZA M., RUIZ MANERO J. 1996. *Las piezas del Derecho. Teoría de los enunciados jurídicos,* Ariel.

ATIENZA M., RUIZ MANERO J. 2000. *Ilícitos atípicos*, Trotta.

BAGNOLI C. 2018. *Defeaters and Practical Knowledge,* in «Synthese», 195, 2855 ff.

BAYÓN J. C. 1991. *La normatividad del Derecho: deber jurídico y razones para la acción*, Centro de Estudios Políticos y Constitucionales.

BAYÓN J.C. 2002a. *Derecho, convencionalismo y controversia*, in NAVARRO P.E., REDONDO M.C. (eds.), *La relevancia del derecho. Ensayos de filosofía jurídica, moral y política*, Gedisa.

BAYÓN J.C 2002b. *El contenido mínimo del positivismo jurídico*, in ZAPATERO V. (ed.), *Horizontes de la filosofía del Derecho. Homenaje a Luis García San Miguel*, Volume 2, Servicio de Publicaciones de la Universidad de Alcalá.

CARPENTIER M. 2014. *Norme et exception. Essai sur la défaisabilité en droit*, Fondation Verenne.

DOLCETTI A., RATTI G.B. 2016. *La derrotabilidad jurídica como relación sistemática compleja,* in «Analisi e Diritto 2016», 35 ff.

DWORKIN R. 1977. *Taking Rights Seriously*, Harvard University Press.

FERRER J., RATTI G.B. 2012. *The Logic of Legal Requirements: Essays on Defeasibility*, Oxford University Press.

GARCÍA FIGUEROA A. 2018. *El convencionalismo jurídico o la irrelevancia del juspositivismo*, in «Persona y Derecho», 79, 71 ff.

GARCÍA YZAGUIRRE V. 2022. *Conflictos entre normas y derrotabilidad*, Editorial Colex.

HART H.L.A. 1948. *The Adscription of Responsibility and Rights,* in «Proceedings of the Aristotelian Society», 49, 171 ff.

HART H. L. A. 1994. *The Concept of Law*, 2nd Ed., Clarendon Press.

MARTIN FARRELL D. 2014. *Positivismo jurídico: Dejen que herede Palmer*, in «Lecciones y ensayos», 93, 63 ff.

MONTAGUE P. 1995. *Punishment as Societal Defense*, Rowman & Littlefield.

MORESO J.J. 2016. *Con la plomada de Lesbos. Celano sobre Rule of Law y particularismo*, in «Revista Iberoamericana de Argumentación», 13, 1 ff.

MORESO J.J. 2017. *Atienza: dos lecturas de la ponderación*, in AGUILÓ JOSEP, GRÁNDEZ PEDRO (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, pp.

MORESO J.J. 2018. *Imágenes del Derecho*, in «Persona y Derecho», 79, 115 ff.

MORESO J.J. 2020a. *Lo normativo: variedades y variaciones*, Centro de Estudios Políticos y Constitucionales.

MORESO J.J. 2020b. *Towards a Taxonomy of Normative Defeaters*, in BERTEA S. (ed.), *Contemporary Perspectives on Legal Obligation*, Routledge.

MORESO, J.J. 2022. *Nuevas variaciones para mis críticos*, in «Eunomía. Revista en Cultura de la Legalidad», 22, 558 ff.

NINO C.S. 1992. *Fundamentos de derecho constitucional*, Astrea.

POLLOCK J.L. 1974. *Knowledge and Justification*, Princeton University Press.

POLLOCK J.L. 1986. *A Theory of Moral Reasoning*, in «Ethics», 96, 506 ff.

RAZ J. 1972. *Legal Limits and the Principles of Law*, in «Yale Law Journal», 81, 823 ff.

RÓDENAS A. 1998. *Entre la transparencia y la opacidad. Análisis del papel de las reglas en el razonamiento judicial*, in «Doxa. Cuadernos de Filosofía del Derecho», 21, 1, 99 ff.

SINNOT-AMSTRONG W. 1988. *Moral Dilemmas*, Blackwell.

SINNOT-AMSTRONG W. 1999. *Some Varieties of Particularism*, in «Metaphilosophy», 30, 1 ff.

TARUFFO M. 2001. *La prueba de los hechos*, Trotta.

TORRES ORTEGA I. 2020. *Sobre la fundamentación del castigo. Las teorías de Alf Ross, H.L.A. Hart y Carlos Santiago Nino*, Centro de Estudios Políticos y Constitucionales.

VEGA J. 2021. *Dintorno, entorno y contorno del Derecho. Ensayo metateórico sobre los límites de la categoría jurídica*, in «Anales de la Cátedra Francisco Suárez», 55, 535 ff.

Intuitionism, Practical Reasoning and Defeasibility

DANIEL GONZÁLEZ LAGIER

1. *The relevance of cognitive sciences for practical reasoning: some examples*

The term "cognitive sciences" usually refers to a set of studies (stemming from psychology, neuroscience, artificial intelligence, linguistics, cognitive anthropology, and so on) that, from an experimental and scientific point of view, deal with the mind and the capacities related to the acquisition of knowledge and decision making. The cognitive sciences have recently undergone enormous development, largely supported by neuroscientific techniques for observing brain function. These disciplines can shed much light on important aspects of practical reasoning in general, and legal reasoning in particular, as their contributions show general characteristics of human thought processes that are clearly reflected in the latter. The following are examples of the importance they may have for, on one hand, establishing the possibility of human rationality (and the type of rationality) and, on the other, concerning discussions on the theory of law and legal argument:

a) Kahneman and Tversky have shown that human reasoning proceeds by means of heuristics and shortcuts that lead to numerous biases and errors (KAHNEMAN 2011, KAHNEMAN 2003). These biases lead us away from the answers that would be correct in accordance with a standard conception of rationality based on deductive, inductive and probabilistic rules of inference. This type of research therefore appears to assume a pessimistic conception of human rationality (which they call limited rationality). We simply cannot know whether a belief of ours is true or false, or whether it is based on correct reasons, as biases are usually unconscious. Legal reasoning would not escape these biases, which would affect all areas, from the legislative to the judicial. For example, studies have been made of how the way the facts in the prosecution's indictment are classified creates an *anchoring bias* in the judge (FARIÑA et al. 2002), so it would be necessary to consider whether and how these can be overcome (institutional modifications, argumentation strategies, etc.). The study of biases in legal reasoning could be seen as a complement, from a psychological perspective, to the traditional analysis of fallacies carried out by the theory of argumentation.
b) Overcoming the pessimism of bias-based studies, evolutionary psychology has argued that the human mind consists of a set of information processing and decision-making modules or systems that have been designed by natural selection. This ensures that human beings are rational, at least in the sense that their decisions and actions are (considered as a whole) instrumentally adequate to meet goals relevant to adaptation and survival, i.e. biological goals (GARCÍA CAMPOS 2011). A vision similar to this is behind the authors who have been constructing new disciplines, such as "neuroethics" and "neurolaw", which—with the support of neuroscientific research—argue that moral reasoning (and legal reasoning, to the extent that it is permeated by moral considerations) is determined by intuitions and emotions formed by a process of natural selection. In other words, they have an important

adaptive value for human beings (for example, we could say that we believe solidarity with the members of our community is an important moral value because supportive and cooperative behaviour favours the survival of the species) (González Lagier 2017). Marc Hauser, for example, has even postulated the existence of a moral module or organ that is decisive in our deliberation when we are faced with moral dilemmas (Hauser 2006).

c) However, the notion of rationality of evolutionary psychology and neuroscience is not the same as the "tradition of biases and heuristics". It could be said that evolutionists assume an instrumental or consequentialist notion of rationality based on efficiency for survival, while Kahneman, Tversky and many others assume a deontological conception based on rules or standards (of deductive and inductive logic). Is there a way to accommodate both types of rationality in human behaviour? It has been suggested that "dual systems" or "dual processing" theory attempts precisely this type of reconciliation (García Campos 2012). To this end, the theory postulates that human beings have two reasoning or decision-making systems: one of them is fast, intuitive, non-verbal and unconscious (we are aware only of the result of reasoning), evolutionarily older and shared with other animals. Following the convenient terminology used by Kahneman, we can call this System 1. The other is slow, intentional, methodical, conscious, verbal (or verbalisable), developed later and characteristic of human beings (System 2) (Kahneman 2011). The first system would correspond to a strategic rationality – the result of natural selection – while the second would correspond to rule-based rationality, more connected to the social and cultural dimension of human beings (García Campos 2009, 78). What is rational in one sense need not coincide with what is rational in the other, so there may be a conflict between the two systems. Obviously, this discussion is relevant to understanding how legal reasoning works and how the two types of decision processing come together: For example, Manuel Atienza has suggested that the articulation between System 1 (the intuitive) and System 2 (the slow and rational) could account for the interplay between rules (which would guide quick and intuitive decisions) and principles (whose application would require more complex reasoning) (Atienza 2017, 429).

d) Even when a dual conception of reasoning that leaves room for reason (in the sense of deductive, inductive procedures, etc.) is maintained, it will not necessarily play an important role in the correctness and justification of decisions. For example, Jonathan Haidt, in the view of practical reasoning that he calls "social intuitionism", accepts (as we shall see below) the thesis of the two systems of decision-making, but considers that System 2 (the rational one) does not really fulfil a control function for System 1, at least when it comes to making moral decisions. In most cases, in his view, System 2 performs only a function of a *posteriori* rationalisation of intuitively made decisions, without attempting to change them:

> «Once people find evidence to support them, even a single piece of bad evidence, they often stop the search, since they have a "make-sense epistemology" [...] in which the goal of thinking is not to reach the most accurate conclusion but to find the first conclusion that hangs together well and that fits with one's important prior beliefs» (Haidt 2001, 821).

If Haidt is right, the distinction between context of discovery and context of justification, with which an attempt has been made to refute the objections of sceptics about the justification or correctness of judicial decisions, is no longer important. It can no longer be said that what matters in fact is not how a conclusion is reached, but whether it is justified in the light of rational considerations because these would not be oriented towards correctness, but rather motivated by consistency with the results of System 1, whatever they may be.

In general, three levels or areas of possible relevance of cognitive sciences can be distinguished: descriptive, conceptual and normative. At the first level, cognitive sciences would provide a description of how, in fact, reasoning and decision-making processes take place in the

mind (e.g., which brain modules are involved in moral problem solving). We could say that, insofar as these are empirical disciplines, this is their most natural function. At a conceptual level, cognitive sciences can provide knowledge to be taken into account in the construction of concepts hitherto considered to be philosophical, such as the mind, rationality, reasoning, decision, volition, emotion, and so on. At a normative level, they may help to construct criteria for evaluating different reasoning strategies and suggest which ones are more appropriate in relation to certain purposes, i.e., in some sense they may help to choose which is the best strategy for reaching a rational justified decision (e.g., some psychologists suggest that probabilistic reasoning is improved if problems are presented in frequency format rather than in percentage format, from which it would follow that they *should be* presented in frequency format) (García Campos 2014, 26). Clearly, it is at the conceptual and normative level that there is the greatest interconnection between cognitive science and philosophy. And it is the possibility that the cognitive sciences have something to say (and the extent to which this is the case) at the normative level that generates the most controversy. A common thesis is that a distinction must be made between the explanation and the justification of a decision or a belief. While cognitive sciences can be very useful in providing explanations of our way of reasoning and deciding, by themselves they can contribute little to its justification, because this would imply moving from facts to rules, which would be forbidden by several arguments (such as George Moore's naturalistic fallacy argument and "Hume's guillotine").

In this paper I will try to discuss the plausibility, for legal reasoning, of the intuitionist model of moral judgement put forward by the psychologist Jonathan Haidt. Of course, it is not the only model proposed from the cognitive sciences, nor the only one that resorts to intuitions to explain decision-making, but it is one of the most influential and well-known in moral and legal philosophy (although Haidt is thinking primarily of moral reasoning, it is not difficult to project his ideas on to legal reasoning). In particular, I will take as a "test bed" the problem of the defeasibility of rules. To do so, I will proceed as follows: firstly, I will present the intuitionist model; secondly, I will briefly explain what the defeasible nature of legal reasoning consists of and how it could be explained by the intuitionist model. I will then make some critical remarks about this model. Finally, I will discuss whether a model of reasoning based on cognitive science can have a normative scope, in the strong sense. Whether, as well as having explanatory relevance, it can also provide guidelines for the justification or correctness of normative reasoning.

2. *The "intuitionist" model of normative reasoning*

In his paper *The emotional dog and its rational tail: a social intuitionist approach to moral judgment*, Jonathan Haidt proposes a model of explanation of practical reasoning that he calls "social intuitionism": intuitionism because of the role it gives to intuitions and emotions, and social intuitionism because of the role it gives to the influence of social and cultural aspects.

The main thesis of social intuitionism is, according to Haidt himself, «that moral judgment is caused by quick moral intuitions, and is followed (when needed) by slow, ex-post facto moral reasoning» (Haidt 2001, 817). The author illustrates his thesis with an experiment and justifies it with several indications of the causal ineffectiveness of reason for the formation of moral judgement. The experiment is as follows:

> «"Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each

> of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?"».
>
> «Most people who hear the above story immediately say that it was wrong for the siblings to make love, and they then begin searching for reasons [...]. They point out the dangers of inbreeding, only to remember that Julie and Mark used two forms of birth control. They argue that Julie and Mark will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them. Eventually, many people say something like "I don't know, I can't explain it, I just know it's wrong"» (HAIDT 2001, 814).

Haidt believes that this type of situation (this "bewilderment" at "feeling" that something is wrong but not being able to justify why) is characteristic of our moral attitudes and that a theory explaining moral judgement must account for these cases. In his opinion, intuitionist theories are better placed to explain them: «Intuitionism in philosophy» he argues «refers to the view that there are moral truths, and that when people grasp these truths they do so not by a process of ratiocination and reflection, but rather by a process more akin to perception, in which one "just sees without argument that they are and must be true"» (HAIDT 2001, 814). Moral judgements are, then, caused directly by these intuitions and reason plays no causal role here.

The indications Haidt presents for doubting that reason usually plays an important role in the formation of moral judgements are as follows:

1) Firstly, it appeals to the growing consensus among psychologists on the fact that, when subjects are trying to solve a problem (and moral problems would be no different from any others on this point), not only is the reasoning process activated, but, simultaneously, another much faster process is activated that brings a quicker solution. This is the "dual processing model". However, Haidt believes there is scientific evidence that, of the two processes, the intuitive and emotional one sets the tone in problem solving, basically through an automatic assessment of situations and people. For example, with respect to the formation of opinions about other people (including judgements about their character or about the moral correctness of them or their actions), the evidence shows that «People form first impressions at first sight [...], and the impressions that they form from observing a "thin slice" of behavior (as little as 5 s) are almost identical to the impressions they form from much longer and more leisurely observation and deliberation» (HAIDT 2001, 820). These first impressions subsequently determine the rest of our moral opinions about these people.
2) Secondly, some research provides evidence that both processes do not occur simultaneously, but that the role of reason is usually limited to justifying or giving meaning, *a posteriori*, to the intuitively issued opinion. And not only is it a *post hoc* reasoning, but also, in Haidt's opinion, it is not reasoning that is genuinely open to evidence and reasons, but one "motivated" by factors such as the desire to maintain good relations with others or to avoid the anxiety that can be produced by seeing our own worldview threatened. Thus, the reasoning process is «more like a lawyer defending a client than a judge or scientist seeking truth» (HAIDT 2001, 820), in the sense that it is not impartial reasoning. Evidence is sought exclusively in favour of our intuitive beliefs.
3) Also, and thirdly, there are also neuropsychological experiments (for example, those by Damasio) that can be taken as indications that the connection between moral reasoning and moral action is quite weak, and that emotions (fundamentally empathy) are much more effective as motivators of moral behaviour, thus proving Hume, Adam Smith and others right.

From his intuitionism and the residual role of reasoning, Haidt draws three important conclusions:

a) First, that the reasoning process only works objectively (and not merely as *a posteriori* rationalisation) under very specific circumstances: it requires the person to have time and capacity, motivation to be accurate (truth-oriented), to lack an *a priori* judgement to defend, and to have neither an interest in maintaining the relationship nor an interest in maintaining coherence (HAIDT 2001, 822). Haidt does not therefore deny that sometimes reasoning (in the classical sense) can be the cause of our moral judgements («particularly among philosophers» Haidt writes, «one of the few groups that has been found to reason well») (HAIDT 2001, 819). But these are rare circumstances. In «more realistic circumstances, moral reasoning is not left free to search for truth» (HAIDT 2001, 822). Often, we resort to reasoning only with the intention of arousing in others the same intuitions that we have ourselves.
b) Secondly, that «our moral life is plagued by two illusions»: the illusion that our moral judgement is due to our reasoning (the "*wag-the-dog illusion*") and the illusion that our arguments change the opinions of others (the "*wag-the-other-dog's-tail illusion*"): it «is like thinking that forcing a dog's tail to wag by moving it with your hand should make the dog happy») (HAIDT 2001, 823).
c) Finally, the lesson Haidt draws is that the roots of human intelligence lie not in our rational ability to search for and evaluate evidence, but «in what the mind does best: perception, intuition, and other mental operations that are quick, effortless, and generally quite accurate» (HAIDT 2001, 822).

But what does intuition really consist of? Drawing on the work of various authors, Haidt proposes the following traits to distinguish between the reflective process and the intuitive process (HAIDT 2001, 818):

The process of intuition	*The process of reflection*
Fast and effortless	Slow and with effort
Unintentional and works automatically	Intentional and controllable
Inaccessible: you are aware only of the result	Accessible to the consciousness and visible
No attention required	Requires attention, which is always limited
Parallel distribution	Sequential
Pattern matching: thinking is metaphorical, holistic	Symbolic manipulation: thought tries to preserve the truth; it is analytical
Common to all mammals	Typical of humans over two years of age and of some apes trained in a certain way
Depends on the context	Does not depend on the context
Depends on the bases (brain and body) that support it	Independent of its bases (can be implemented in any organism or machine acting according to certain rules)

Haidt is also concerned with the mechanisms by which intuitions are "fixed" and with their origin and development. Regarding the former, he draws on Damasio's hypothesis of somatic markers and Lakoff and Johnson's explanation of the role of metaphors in the way we interpret the world.

As is known, according to Antonio DAMASIO (1994) experiences of the world are associated with emotional sensations of pleasure or pain, so they are "marked" as positive or negative. This allows a quick (and automatic) assessment of analogous situations that may arise in the future. There comes a point where the mere thought of a particular action is sufficient to provoke an "as if" response in the brain, whereby the person weakly experiences the same bodily feelings that he or she would experience if performing the action. This "marking" of situations as positive or negative allows the brain to very quickly rule out possibilities for action associated with negative sensations and select those associated with positive sensations much more quickly than rational analysis can do.

Meanwhile, the linguist George Lakoff and the philosopher Mark Johnson have argued that human thought processes are largely metaphorical (LAKOFF & JOHNSON 2003). Metaphors construct a framework of concepts that largely determine our way of interpreting and living in the world. For example, "argument is war" ("Your claims are *indefensible*", "*He attacked every weak point* in my argument", "His criticisms were *right on target*", "*I demolished* his argument", "I've never *won* an argument with him", "If you use that *strategy*, *he'll wipe* you out") (LAKOFF & JOHNSON 2003, 12 f.); "time is money" ("You're *wasting* my time", "This gadget *will save* you hours", "I don't *have* the time to *give* you", "How do you *spend* your time these days?", "I've *invested* a lot of time in her", "I don't *have enough* time *to spare* for that", "He's living on *borrowed* time") (LAKOFF & JOHNSON 2003, 15 f.); "rational is up" while "emotional is down" ("We put our *feelings* aside and had a *high-level intellectual* discussion"; "He couldn't *rise above* his emotions") (LAKOFF & JOHNSON 2003, 24). Many of these metaphorical associations are built from our experiences as physical bodies: for example, we associate the purity and cleanliness we need in food with moral goodness, and corruption and filth with vice HAIDT 2001, 825). In short, what our author seems to want to suggest is that these associations (which Lakoff and Johnson seem to consider automatic and which, in turn, determine our judgements) are behind many of our moral intuitions:

> «Moral intuition, then, appears to be the automatic output of an underlying, largely unconscious set of interlinked moral concepts. These concepts may have some innate basis [...], which is then built up largely by metaphorical extensions from physical experience. Metaphors have entailments, and much of moral argument and persuasion involves trying to get the other person to apply the right metaphor. If Saddam Hussein is Hitler, it follows that he must be stopped. But if Iraq is Vietnam, it follows that the United States should not become involved [...]. Such arguments are indeed a form of reasoning, but they are reasons designed to trigger intuitions in the listener» (HAIDT 2001, 825).

As well as the mechanism for fixing intuitions, Haidt proposes a hypothesis about the evolutionary origin of many of our moral intuitions. His proposal tries to explain them, as Darwin did, as a development of the social instincts of animals. The process would go from mere regularities ("descriptive rules", he calls them) of many species to the "prescriptive rules" (rules for which compliance is reinforced by the reactions of others) that appear in primates and which, with the appearance of language, reach a higher stage (although «the cognitive and emotional machinery of norm creation and norm enforcement was available long before language existed», HAIDT 2001, 826). These are rules related to reciprocity, empathy, altruism, loyalty and so on, and, given the parallelism «between the social lives of humans and chimpanzees, the burden of proof must fall on those who want to argue for discontinuity—that is, that human morality arose ex nihilo when we developed the ability to speak and reason» (HAIDT 2001, 826). However, society carries out "pruning" work on these innate intuitions evolution has inscribed in human beings, modelling them with the influence of others,

immersion in customs and socialisation processes (this explains why children end up with a morality specific to their group or community). Surprisingly, Haidt states that «a culture that emphasized all of the moral intuitions that the human mind is prepared to experience would risk paralysis as every action triggered multiple conflicting intuitions» (HAIDT 2001, 827) (surprisingly, because then evolution does not seem to be wise or adequate enough to ensure survival by moral guidelines alone).

Haidt even proposes six innate intuitions (and their opposites) as the axes of an "intuitive ethics" (which would explain the coincidence of moral rules in different cultures): 1) care (harm); 2) fairness (cheating); 3) loyalty (betrayal); 4) authority (subversion); 5) sanctity (degradation) and 6) liberty (oppression) (HAIDT 2013, ch. 7). Different combinations in the importance assigned to each of these intuitions or emotions would give rise to the different moral systems (HAIDT & CRAIG 2004).

3. *The defeasibility of rules and the intuitionist model*

Legal reasoning is a case of general reasoning with certain particular features. One of these is the importance of what we can call its defeasible character. Defeasibility is not an exclusive aspect of legal reasoning (nor, probably, of normative reasoning), but it can perhaps be considered to have a special importance in the latter. Be that as it may, it is at the centre of a good part of the discussions of current legal theorists[1].

I will call defeasibility the fact that the application of legal rules is necessarily open to unforeseen exceptions, if it is to be rational (fair, reasonable, justified). The norms (rules) are drafted with a general scope; to mark the limits of their field of application, a series of generic properties are taken into account (such as, for example, being over 18 years of age, having acted negligently, having caused damage, having an income greater than x, etc.). These properties (the conditions of application of the rule, to use von Wright's terminology) make it possible to establish the cases in which it is *prima facie* justified to apply the rule and in which cases it is not. In other words, the scope of application of the rule, including its explicit exceptions. However, as it is impossible for legislators to foresee all the properties that will be relevant in order to satisfactorily regulate a case, it may be that the application of the rule to a specific case (to which it should initially be applied) is not reasonable, that is, it appears to be unjustified as a relevant property that had not been considered is present in that case. Obviously, in the case of law, the reasoning that shows the application of the rule to that particular case is not justified must, in turn, be based on other legal standards (values, principles, reasons underlying the rule, etc.). Thus, as well as the explicit exceptions provided for in the formulation of the rule, other exceptions not provided for appear after a more in-depth analysis of the case to be resolved. Following a terminology proposed by Celano, we can therefore distinguish between "normal" and "non-normal" cases of application of the rule (in the latter the rule would be displaced). But what criteria do we use to distinguish between the two, and do the cognitive sciences have anything to contribute to this problem?

In my opinion, projecting the intuitionist model to the problem of defeasibility is a good testing ground for assessing its role in the analysis of legal reasoning. In the philosophy of law, in several works Bruno Celano and Marco Brigaglia have suggested a "psychologized" theory of defeasibility, in some points close to intuitionism (although Haidt's intuitionism goes further).

> «The hypothesis is the following: the fact that certain cases are normal and others are not—on this, as has been said, depends the answer to the question whether it is reasonable, in such a case, to

[1] See, for example, the contributions in FERRER & RATTI 2012.

> reconsider the rule (this is, I emphasize again, a normative question)—depends on psychological facts: certain cases are normal or abnormal only because they appear to us as such. Which cases appear to us as normal and which do not is a clear question of psychology. So: whether or not a certain rule, in a given case, is a reason to act – a justifying reason; this is the crucial point – depends on our psychological make up» (CELANO 2017, 101, my translation; see also CELANO & BRIGAGLIA 2017).

For these authors, mental states can be conscious and unconscious. What allows us to detect that a case is not normal is a feeling of "surprise" or "bewilderment" (to use Haidt's expression), which is the conscious manifestation of certain unconscious mental processes.

To analyse this type of proposal, it may be useful, once again, to distinguish their descriptive, conceptual and normative scope:

a) *Descriptive level:* From the intuitionist model it is possible to account for the ability to distinguish between normal and non-normal cases of application of a rule based on the notion of intuition and emotion. It is intuition that alerts us that something is wrong with the application of the norm and we "discover" or "realise" that a case is or is not normal through a "feeling" of approval or disapproval of the possibility of the rule being applied to that case. According to Haidt's intuitionism (going beyond that suggested by Celano and Brigaglia), the intuitionist model could further postulate that such intuition would be part of (or might be traced back to) a set of moral principles that have an evolutionary origin, subsequently shaped by social and cultural factors (especially in the case of law, given its institutional character). What these intuitions and emotions would be telling us is that the application of the rule to that case would violate some relevant moral and/or legal principle which would ultimately be linked to the survival of the species. On the other hand, legal reasoning requires that the "abnormal" case (as well as the proposed solution) should be justified on the basis of legal standards (i.e., principles or values recognised in the legal system), and not on the basis of feelings. But, if Haidt is right, this justification will be an *a posteriori* rationalisation that does not change the direction of the initial intuition.
b) *Conceptual level*: Therefore, we can call "normal cases" those in which the application of the solution established in the rule generates a feeling of approval, and non-normal cases (cases in which there are implicit exceptions) those in which the application of the solution established in the rule generates a feeling of disapproval. The feeling (or intuition) could be taken not only as a warning (a symptom) that we are facing a case with an implicit exception, but also as a conceptual criterion for identifying such a case.
c) *Normative level:* At this level it would be argued that it is justified to displace (defeat) the rule when its application to a particular case generates a feeling of disapproval (on the other hand, it would be justified to apply the rule when it does not generate a feeling of disapproval). As can be seen, this step implies a reduction of the normative level—the justification level—to the psychological level.

Is this an adequate reconstruction of the defeasible nature of legal reasoning? I believe it inherits some general problems from the intuitionist model.

4. *Some doubts about the plausibility of the intuitionist model: descriptive claims*

As we have seen, the intuitionist model "à la Haidt" can be characterised using four theses:

1) *Intuitionism:* our moral judgment is dominated, or strongly influenced, or largely conditioned, by quick, unconscious, unintentional intuitions, a long way from reflection.

2) *Emotivism:* these moral intuitions depend fundamentally on our emotions or are the expression of these emotions.
3) *Innatism:* these are innate intuitions and emotions, transmitted genetically, although they can be culturally moulded.
4) *Darwinism:* moral intuitions are seen as mechanisms that evolution has selected because they ensure the survival of the species.

Haidt's explanation of how we reason in normative contexts is based on these theses. However, in my opinion, all of them raise certain problems, which could cast doubt on whether the intuitionist model is a good description or explanation of our reasoning processes in these contexts. Several of these objections arise from a philosophical analysis of the data used by the cognitive sciences to draw their conclusions (which shows, by the way, the need for an interaction between both viewpoints). Let us take a look at them:[2]

1) *Intuition and reason:* Haidt's first thesis seems to establish too radical a dichotomy between intuition and reason. In other words, he seems to be assuming that one excludes the other, as if they were two independent information processing systems, in contrast to the views of other authors, such as Kahneman and Damasio. Haidt's intuitionist model also adds that decisions are made in System 1, while System 2 provides only an *a posteriori* rationalisation of decisions. However, it also seems a plausible thesis that the two systems collaborate in decision making, either because System 2 can correct System 1 (Kahneman) or because emotions simplify and order information and alternative actions based on past experiences, avoiding complexity that would paralyse reason (Damasio). This hypothesis of collaboration between the two systems fits well into philosophical analyses of intuition such as that of Mario Bunge, who has shown that the idea of "intuition" encompasses many different phenomena (modes of perception; forms of imagination; sudden, rapid or incomplete inferences; capacity for synthesis; capacity to evaluate a situation and to choose the best alternatives, and so on) (BUNGE 2013, ch. III.1) and several of these forms of intuition cannot be seen as opposing or excluding reason: rapid or incomplete inference is embryonic or primitive reasoning and "synoptic apprehension", as Bunge points out, «is not a substitute for analysis, but a reward for careful analysis». But, in addition, the different types of intuition—even those that cannot be seen strictly speaking as reasoning, not even incomplete reasoning—are especially encouraged by the continuous exercise of reasoning, problem analysis, experience in an activity, or dedication to studying a discipline. In short, many intuitions would not be possible without collaboration between the two systems.
2) *Normative intuitions and emotions:* Some difficulties also arise over the relationship between normative judgements and emotions. The argument behind this linkage is as follows: On one hand, brain imaging shows that when we reason morally, areas of the brain related to emotion are strongly activated; on the other, psychological tests show that in most cases we solve moral dilemmas intuitively. The two things must be related: intuitions, therefore, arise from emotions—from the activity of the affective areas of the brain. But in my opinion this way of reconstructing moral decisions raises several problems:

a) First, the evidence we have for the role of emotions in decision-making is actually indirect. What neuroscientists can prove is that, when we are faced with a moral dilemma, the areas of the brain related to emotions are activated in a particularly pronounced way. But this does not allow us to infer that emotions generate moral judgment. It could be that it is the moral

[2] A development of these criticisms can be found in GONZÁLEZ LAGIER 2017.

judgment that generates the emotion: for example, realising, through analysis, the unfairness of a situation may generate indignation; or, in a moral dilemma, being aware that any solution will cause some sort of harm may provoke regret.

b) Secondly, the notion of emotion apparently assumed by the intuitionist model converts emotions into mere feelings (of pleasure or displeasure) that lack propositional content. This way of understanding emotions and their relation to morality departs from the conceptions of emotion most widespread today among philosophers and many psychologists. For these conceptions (which, to a large extent, support Aristotle's conception of emotions), at least three distinct dimensions must be distinguished: (i) A cognitive dimension (in a broad sense, ranging from a belief to a mere perception); (ii) an affective or purely phenomenological dimension (the sensation of pleasure or pain) and (iii) a motivational dimension (a tendency towards action). Cognitive theories of emotion focus on the first aspect, "somatic" and mechanistic theories of emotion focus on the second, and behaviourist theories on the third element. Finally, non-reductivist theories try to account for all aspects of emotions. If emotions are identified exclusively with the second aspect, many problems remain to be solved: (a) the possibility of unconscious or sensationless emotions is not accounted for; (b) the possibility that emotions can be part of rational (teleological) and not only causal explanations of behaviour is not accounted for; (c) the two-way nature of the relationship between emotions and beliefs is not accounted for; (d) the possibility of evaluating emotions as reasonable or unreasonable (depending on the underlying belief) is not accounted for; and, above all, (e) there is, once again, a marked distancing between emotions and reason (González Lagier 2009, ch. II).

c) *Innatism:* Nor is the thesis—closely linked to the previous ones—of the innatism of moral intuitions free from objections. The arguments in favour of innatism again rest on the automatic nature of the response and the inability to give reasons (Haidt), together with the coincidence of responses despite the diversity of the respondents (Hauser). However, to be sure that our moral beliefs are innate, we would have to entirely rule out the possibility that the automatic responses are due to the acceptance of deeply rooted but culturally transmitted principles. As Adela Cortina, commenting on Haidt's experiments, points out:

> «The fact that respondents answer intuitively—that is, immediately and automatically, without being aware of how they have come to formulate the judgment—and that on many occasions they can give no reason why an action seems good or bad to them, may well be explained by the fact that they have learned it socially and have not subjected it to review» (Cortina 2006, 86, my translation).

Many rules of social morality (such as incest, in Haidt's example) are inherited from the social environment, which, most of the time, inculcates them without giving reasons to justify them. On the other hand, the universality of rules or principles must also be taken with caution: many supposedly universal moral behaviours are not, in fact, moral ones, or their universality is independent of their moral nature. For example, Marc Hauser has postulated the universality of the offspring care principle (which obviously has a clear evolutionary explanation). But should we also draw the conclusion that foraging or fleeing from predatory animals also has moral value (Barta 2013, ch. III)? Or reproduction? What I mean is that the nature of many habits given as examples of universally accepted behaviours is not (or not exclusively) moral.

d) *Darwinism:* We have seen that another of the characteristic theses of this attempt to account for morality from a biological and neuroscientific point of view is, in fact, moral Darwinism. Moral intuitions are seen as mechanisms that evolution has selected because they ensure the survival of the species. This thesis, however, suffers from a degree of ambiguity. To clarify it, this is useful to distinguish, once again, between the (descriptive) claim to explain

> morality and the (normative) claim to justify it. We can call the former descriptive moral Darwinism and the latter prescriptive moral Darwinism. Descriptive moral Darwinism, in turn, can try to explain human beings' *capacity* to demonstrate ethical behaviour (to evaluate behaviours as right or wrong from a moral point of view) or (more ambitiously) to try to explain *the content* of morality; in other words, why we believe certain behaviours are right or why some principles or values are so widespread (AYALA 2013, 61). Descriptive moral Darwinism appeals to the idea that behaving morally or adapting one's behaviour to certain principles is a trait that has facilitated the evolution of the species and its survival. The prescriptive version adds that, as this is so, such principles are justified (we will return to the prescriptive issue in the next section).

Descriptive moral Darwinism is merely a hypothesis which has not been sufficiently confirmed. Take the claim that what explains the human ability to evaluate behaviours as good or bad and to adjust behaviour to certain principles is that this ability is an evolutionary advantage. To accept it conclusively, one would have to reject the (also plausible) alternative hypothesis put forward by Francisco Ayala according to which ethical behaviour is only indirectly a result of evolution in as far as it is a consequence of the development of human intelligence; in other words, what has adaptive value and has been favoured by evolution is human intelligence, not the ability to behave morally (which is a consequence of human intelligence) (AYALA 2013, 66). If, on the contrary, the claim is that moral codes are determined by evolution, the problem is that it is not possible to find a set of relevant principles that are universal, but not formulated in an excessively vague and empty way. Moreover, it is possible to find types of behaviour, such as aggressiveness and territoriality, that are evolutionarily important and cannot be accepted as examples of moral behaviour.

In short, I believe that intuitionism, at least this version of it, is more of a hypothesis still lacking confirmation and in need of refinement than a well-founded explanation of the functioning and basis of practical reasoning.

5. *The problem of justification and normativity*

Studies of practical reasoning carried out in the cognitive sciences can adopt three positions regarding the ideas of correctness or justification of reasoning—in other words questions of normativity: (1) limiting themselves to trying to give a description and explanation and giving up their normative claims; (2) trying to draw conclusions about justification from those descriptions or (3) denying that it makes sense to speak of ideas such as justification, correctness or normativity. Michel Ruse has explained how cognitive sciences can lead to this third option:

> «Is it not the case that sometimes, when one has given a causal explanation of certain beliefs, one can see that the beliefs, in themselves, neither have a foundation nor could ever have such a foundation? [...] Once we see that our moral beliefs are simply an adaptation put in place by natural selection, in order to further our reproductive ends, that is an end to it. Morality is no more than a collective illusion fobbed off on us by our genes for reproductive ends» (RUSE 1991, 506).

It seems to me that what lies behind these positions that challenge the idea of justification and normativity is the following: if practical reasoning is *necessarily* determined by intuitions, heuristics and unconscious and uncontrollable biases—that is, if it cannot be otherwise—then there is no point in asking when it is correct or under what conditions it is justified. It is possible to speak of regularities and divergences (which would have to be explained causally), but not of norms, rules or criteria for correctness. In short, this is a sceptical stance on

normativity. Here I will set aside this option, focusing on the problems with the second alternative: the claim that the cognitive sciences can have a strong normative scope.

5.1. *Normative claims based on the cognitive sciences*

Can the cognitive sciences say anything about when reasoning is correct and when it is not? Haidt seems to have only descriptive pretensions, but other authors have gone further and have tried to find in moral intuitions—or directly in the functioning of the brain—normative criteria with which to evaluate the correctness of our moral theories or principles about how we should live. Some examples include:

Neuroscientist William Casebeer, who has claimed that the Aristotelian moral theory of virtue is more plausible from a neurobiological point of view than the moral theories of Kant or John Stuart Mill. His argument is that each of these theories implicitly contains a specific moral psychology that demands different cognitive capacities. Thus, Kant's theory would seem to require «at least the ability to check universalized maxims for logical consistency in a manner that is separable from the taint of affect and emotion»—an ability that corresponds to the functions of the frontal region of the brain. Mill's utilitarian theory requires the ability to perform utilitarian calculations and cultivate emotions that move us to procure the happiness of others, which involves the pre-frontal, limbic and sensory regions of the brain. The Aristotelian ethics of virtue, finally, would be the most demanding, because it requires us to educate our character so that our appetites are in line with good reasons. This implies a "global psychology" that requires the coordinated intervention of the aforementioned regions of the brain. Our author believes there is scientific evidence to tentatively accept that "moral cognition" brings different brain systems and networks related to both cognition and emotions into play in a coordinated way (i.e., the pre-frontal, frontal, limbic and sensory regions: what could be called "the area of moral cognition"), showing that «there is clear consilience between contemporary neuroethics and Aristotelian moral psychology» (CASEBEER 2003, 845). This makes it possible to rule out the plausibility of the other theories.

For his part, Michel Gazzaniga writes: «I would like to support the idea that there could be a universal set of biological responses to moral dilemmas, a sort of ethics built into our brains. My hope is that we soon may be able to uncover those ethics, identify them, and begin to live more fully by them» (GAZZANIGA 2005, xix).

Meanwhile, Patricia Churchland suggests that, just as health «is a domain where science can teach us, and has already taught us, a great deal about what we ought to do», so too in «the domain of social behavior [...] we may learn a great deal from common observation and from science about conditions favoring social harmony and stability, and about individual quality of life» (CHURCHLAND 2011, 190).

5.2. *Is it possible to infer normative conclusions from purely descriptive statements? Arguments against Hume's Law*

I will now make some comments on the normative scope of the cognitive sciences.

As is well known, in a famous passage, Hume states that:

> «In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprized to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, it is necessary

> that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it» (HUME 1960 [1739-40], 469).

This passage has usually (although not unanimously) been interpreted as prohibiting the inference of statements "about what ought to be" from statements "about what is the case". Thus, some of the authors who defend the normative relevance of the cognitive sciences try to argue against the validity of Hume's Law. Let us examine in more detail these arguments, which are usually of two types:

1) The first group involves *arguments based on counterexamples*: A frequent way of showing that it is possible to ground rules or values in descriptions is to present examples of arguments in which this derivation is apparently made. This is the strategy followed, in a famous article, by John Searle (SEARLE 1964). Cognitive scientists have also resorted to this type of argument. Marc Hauser, proposes the following:

> «FACT: The only difference between a doctor giving a child anesthesia and not giving her anesthesia is that without it, the child be in agony during surgery. The anesthesia will have no ill effects on this child, but will cause her to temporarily lose consciousness and sensitivity to pain. She will then awaken from the surgery with no ill consequences, and in better health thanks to the doctor's work.
> EVALUATIVE JUDGMENT: Therefore, the doctor should give the child anesthesia» (HAUSER 2006, 3)

However, this type of argument appears to fall into one of the following errors: (1) it confuses what is good or owed from a technical point of view with what is good or owed from a moral or normative point of view; (2) it presents as a complete argument what is, in fact, an incomplete argument that includes a hidden premise: the very norm or value judgement from which the conclusion is derived.

To avoid the first error, it should be noted that every time a statement includes the term "must" it is not necessarily a genuinely normative statement. Sometimes "must" expresses a conjecture ("must be so" can mean "probably is so"). At other times, it can be replaced by "has to" and expresses a practical necessity. It is important to distinguish between *deontic* or genuine *duties* and *technical*, prudential *duties* or *practical necessities*. Many of the examples given as derivation of "must" from "is" do not conclude genuine deontic duties, but rather practical necessities. As von Wright points out, we can

> «find two main answers to the question of why a certain thing should or may or may not have to be done. One is that there is a rule ordering or permitting or prohibiting the doing of the thing. The other is to say that the ends and necessary connections make the doing or not doing of the thing a practical necessity (or not)» (VON WRIGHT 1963, 74).

Regarding the second error, it is easy to see that the arguments we formulate in everyday contexts often do not include all their premises. It is even feasible to think that in some cases it is impossible in practice to state all the premises necessary to reach the conclusion. But what follows from this is the *defeasibility* or revisability of the conclusion, not that its correctness is independent of the implicit premise. In order to be correct, Hauser's example presupposes a normative premise that unnecessary suffering should be avoided.

2) The second type of argument *restricts Hume's Law to deductive arguments*: it is claimed that what Hume's Law proscribes is deductively deriving a duty from existence, but there are other types of acceptable inferences, such as induction or inference of the best explanation,

> by means of which it is possible to move from descriptions of facts to rules (CASEBEER 2003, 842 f.; CHURCHLAND 2011, ch. 1). One way of arguing that Hume's Law refers exclusively to deductive inferences is to see it as a consequence of the logic conservation principle: in a deduction it is not possible to conclude something that was not already included in the premises. Deductions can make us aware of new facts, but they must already be in the premises. We cannot therefore deduce ought-to-be statements from descriptive propositions alone. This can happen with anything. As Pigden observes «You can't get 'hedgehog' conclusions from hedgehog-free premises (at least, not by logic alone)» (PIDGEN 1991, 423). However, unlike deductions, inductions and inferences to the best explanation do extend our knowledge, so the principle of conservation does not apply to them. Therefore, if Hume's Law is only a manifestation of the principle of conservation of deductive arguments, then those who argue that it is not applicable to non-deductive inferences are right.

But is that all Hume's Law is? Probably not. It can be argued that there are important differences between descriptive statements and normative statements—sometimes we speak of a "logical gulf" between them, or between facts, on the one hand, and rules and values, on the other. Thus, descriptive statements have a downward direction of fit[3] (i.e., words to world: words are intended to fit the world) while normative statements have an upward direction of fit (i.e., world to words: the world is intended to fit the words). Descriptive statements are true or false, while rules or values are not. Statements expressing duties presuppose an internal point of view, in which deontic terms (obligatory, forbidden) are used, whereas if a description refers to a rule or a duty, it does so from an external point of view in which deontic terms are merely mentioned. All these differences between descriptions and rules mean that the former cannot serve as reasons, or express reasons, to justify the latter. It is not only that deductive justification requires that what is to be deduced is included among the premises, it is that, even when it is admitted that induction or abduction can have a justificatory scope, no descriptive statement can, by itself, be a reason justifying a prescriptive statement. It can provide an explanatory reason why we accept certain rules or values. But if one wants to conclude the justification of a rule from descriptions and by means of non-deductive arguments, one has the burden of proof.

If the above considerations are correct, when it is proposed that we should follow patterns of behaviour that have adaptive value, either these are simply recommended as prudent measures to maintain the survival of the human species but without any genuinely normative character, or it is assumed that the survival of the species is a morally valuable end, in which case the normativity stems not from the facts, but from this value assumption.

[3] An explanation of the idea of "direction of fit" can be found in SEARLE 1975.

References

ATIENZA M. 2017. *Epílogo (abierto)*, in AGUILÓ REGLA J., PEDRO GRÁNDEZ P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 429 ff.

AYALA F. 2013. *Evolución, ética y religión*, Universidad de Deusto.

BARTA R. 2013. *Cerebro y libertad. Ensayo sobre la moral, el juego y el determinismo*, Fondo de Cultura Económica.

BUNGE M. 2013. *Intuición y razón*, Debolsillo.

CASEBEER W.D. 2003. *Moral Cognition and Its Neural Constituents*, in «Nature Reviews Neuroscience», 4, 840 ss.

CELANO B. 2017. *Particularismo, psicodeóntica. A propósito de la teoría de la justificación judicial de Manuel Atienza*, in AGUILÓ REGLA J., PEDRO GRÁNDEZ P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 59 ss.

CELANO B., BRIGAGLIA M. 2017. *Reasons, rules, exceptions: toward a psychological account*, in «Analisi e diritto 2017», 131 ff.

CHURCHLAND P. 2011. *Braintrust*, Princeton University Press.

CORTINA A. 2011. *Neuroética y neuropolítica. Sugerencias para la educación moral*, Tecnos, 2011.

DAMASIO A.R. 1994 *Descartes' Error: Emotion, Reason and the Human Brain*, Avon Books.

FARIÑA F., ARCE R., NOVO M. 2002. *Heurístico de anclaje en las decisiones judiciales*, in «Psicothema», 14, 1, 39 ff.

FERRER J., RATTI G. (eds.) 2012. *The Logic of Legal Requirement. Essays on Defeasibility*, Oxford University Press.

GARCÍA CAMPOS J. 2009. *Justificación y racionalidad desde la teoría dual del conocimiento,* «Ideas y valores», 139, 61 ff.

GARCÍA CAMPOS J. 2011. *Razonamiento y racionalidad desde la psicología evolucionista*, in «Metatheoria», 2, 1, 79 ff.

GARCÍA CAMPOS J. 2012. *Convergencias y divergencias en las teorías duales de sistemas*, in «Andamios», 9, 19, 283 ff.

GARCÍA CAMPOS J. 2014. *Normatividad y psicología cognitiva. La propuesta naturalizada de M. Bishop y J. D. Trout*, in «Revista de Filosofía de la Universidad de Costa Rica», 53, 135, 25 ff.

GAZZANIGA M.S. 2005. *The Ethical Brain*, Dana Press.

GONZÁLEZ LAGIER D. 2009. *Emociones, responsabilidad y Derecho*, Marcial Pons.

GONZÁLEZ LAGIER D. 2017. *A la sombra de Hume. Un balance crítico del intento de la neuroética de fundamentar la moral*, Marcial Pons.

HAIDT J. 2001 *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, in «Psychological Review», 108, 4, 814 ff.

HAIDT J., CRAIG J. 2004. *Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues*, in «Daedalus», 133, 4, 55 ff.

HAIDT J. 2013. *The Righteous Mind. Why Good People Are Divided by Politics and Religion*, Penguin.

HAUSER M. 2006. *Moral Minds: The Nature of Right and Wrong*, Harper Collins.

HUME D.1981. *A Treatise of Human Nature*, ed. by L.A. Selby-Bigge, Clarendon Press. (Originally published in 1739-40)

KAHNEMAN D. 2003. *Maps of Bouded Rationality: Psychology for Behavioral Economics*, in «The American Economic Review», 93, 5, 1449 ff.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Penguin. (*Pensar rápido, pensar despacio*, in

LAKOFF G., M. JOHNSON, 1980. *Metaphors We Live By*, The University of Chicago Press.

PIGDEN C.R. 1991. *Naturalism*, in SINGER P. (ed.), *A Companion to Ethics*, Blackwell, 421 ff.

RUSE M. 1991. *The Significance of Evolution*, in SINGER P. (ed.), *A Companion to Ethics*, Blackwell, 500 ff.

SEARLE J.R. 1964. *How to Derive an 'Ought' from an 'Is'*, in «Philosophical Review», 73, 1, 43 ff.

SEARLE 1975. *A Taxonomy of Illocutionary Acts*, in GUNDERSON K. (ed.), *Language, Mind and Knowledge*, University of Minnesota Press, 344 ff.

VON WRIGHT G. H. 1963. *The Varieties of Goodness*, Routledge & Kegan Paul.

Presumptions, Legal Argumentation, and Defeasibility

JOSEP AGUILÓ REGLA

1. *On the Nature of Presumptions: From "it is presumable" to "it must (shall) be presumed"* – 2. *"It is presumable". The hominis presumptions* – 3. *"It must (shall) be presumed". The norms of presumptions (legal presumptions)* – 4. *Presumption-rules and presumption-principles* – 5. *In the core of defeasible reasoning. The problem of the iuris et de iure presumptions* – 6. *What can cognitive sciences contribute to the argumentative use of presumptions?*

1. *On the Nature of Presumptions: From "it is presumable" to "it must (shall) be presumed"*

In general terms, the verb "to presume" means to assume or believe something because there are indications, signs or clues for doing so. According to this, the presumptions generally show the following three elements: a) one or some base facts (the indications, signs or clues), b) a presumed fact (what is suspected or believed) and c) a connection between these two kinds of facts.

It is clear that presumptions play an important role in our argumentations. Taking the aforementioned elements, it is easy to show how the argument works. Let us take the Toulmin's scheme of arguments (TOULMIN 1958). This scheme is a structure made up of *a claim* (a particular statement that one is seeking to defend), *grounds* or *data* (one or a number of particular statements which support the claim), *a warrant* (a general statement whose acceptance entitles passing from grounds to the claim) and *a backing* (general information related to the field in which one is seeking to argue). So, by using this scheme, the argumentative use of presumptions can be reconstructed in a completely natural way. The particular statement which expresses the particular *presumed fact* (what is suspected or believed) constitutes the claim. The statements which express the *base facts*, (the indications, signs or clues) constitute the grounds or data. The *general statement of presumption* constitutes the warrant whose acceptance justifies the acceptance of the presumed fact (the claim) supported by the acceptance of the base facts (the grounds or data). And, finally, the general information which supports the acceptance of the warrant constitutes the backing of the argument.

The argumentative role played by presumptions is unquestionable. Nobody doubts it. However, if we focus our attention on the generic statement of presumption (the *warrant*) it seems clear that we can find examples that appear to be propositions and others which appear to be norms. Indeed, while it is true that in some occasions the generic statements of presumption (the warrants) can be ambiguous, it is also true that there are typical (paradigmatic) ways to express these two alternatives. For example, while the use of the expression "it is presumable" generally serves to express the propositional nature of a statement of presumption, the expression "it must (shall) be presumed" is used to express its normative nature. In this way, while a statement of the type "if P (generic base fact) then Q (presumed generic fact) is presumable" appears to be a propositional statement; the form "if P (generic base fact) then Q (presumed generic fact) must (shall) be presumed" seems to be a normative statement (a norm).

In this paper, I propose to differentiate these two types of presumptions in law and in legal argumentation. On the one hand, the so-called *hominis* presumptions, that is, those made by people when they make factual inferences and, on the other, the presumptions established by legal norms (legal presumptions). In order to emphasize the differences between them, I will use these two expressions respectively: "it is presumable" and "it must (shall) be presumed" (AGUILÓ REGLA 2018)[1]. Then, once the notion of legal presumption has been properly clarified, I will try to show that

[1] «In the *presumptio juris* the law says the inference must be made -the *homo* has no discretion, he involuntarily

the distinction between rules and principles is applicable to presumption norms (to legal presumptions). Consequently, I will distinguish between norms of presumption that are rules (presumption rules) and norms of presumption that are principles (presumption principles). Finally, I will focus on the defeasibility of presumptive reasoning and how cognitive sciences can help detecting material fallacies.

2. *"It is presumable". The hominis presumptions*

2.1. Let us consider the following story as a starting point. A father hires the services of a private detective to find his son, because 15 years ago he left home leaving a goodbye note in which he said they would never see him again. After some time, the detective meets with the father to inform him about his investigations. The dialogue begins with this question:

- Did you find my son?
- No, but I regret to inform you that your son has died. He lived in Australia. Three months ago, he participated in a regatta and the ship sank in a place where the depths are abysmal. The rescue services did not find survivors or corpses. The crew members disappeared with the ship.

The detective concluded his report with these words: "Your son is dead and to continue investigating under these circumstances would be to scam you".

This story illustrates a presumption. The detective and the father presume the death of the son, they do not have any direct proof or evidence. It is easy to make explicit the three elements above mentioned.

a) A presumed fact: the death of the son.
b) One or more base facts: the sinking of the ship and the disappearance of the son.
c) A connection between both kinds of facts. Between the base fact and the presumed fact there is a statement of presumption; that is to say, a general statement whose acceptance authorizes the passage from some facts to other.

2.2. This type of presumptions belongs to theoretical reasoning (they are propositional in nature). They may, however, be part of practical reasoning. To be a fragment of practical reasoning does not change their theoretical nature at all. Presumptions share this trait with all factual inferences: they can take part of practical reasoning and do not leave therefore the propositional and/or theoretical realm of truth.

2.3. The truth judgments involved in *hominis* presumptions are always empirical and they have a probabilistic content. The general statement of presumption is accepted because it is considered to be well founded (that is, expressing a regularity, normality or high probability of truth). Then the primary function of this statement is to approximate us to the truth in a material sense. Therefore, sticking to what this statement establishes is the safest option.

2.4. The "security" of sticking to the general statement of presumption is a matter of probability. In this sense, presumptive reasoning shares the idea of defeasibility with all

makes the inference; while in the *presumptio hominis* the law says the inference may be made -the homo has a discretion, he voluntarily makes the inference» (FISK 1925, 22).

probabilistic reasoning. If new information appears, the conclusion can be rejected without rejecting any of the premises on which the presumption was based.

2.5. In the Law these presumptions are known as *hominis* presumptions (presumptions made by people). They share with all factual inferences the two properties just highlighted (their propositional nature and their defeasibility). So, does the phrase "it is presumable" (typical of *hominis* presumptions) contribute something different from the phrase "it is probable" (typical of factual inferences)?

2.6. "It is presumable" should be reserved for those cases in which the evidence (the basic facts, the evidentiary facts, etc.) is sufficient to consider a fact as proven (not only probable); and, as a consequence of this, "it is presumed" transfers the burden of proof (or of argumentation) to whoever intends to deny the conclusion, the "presumed fact". In any factual inference, the occurrence of some events is an indication of (a reason to believe in) the occurrence of other events. In *hominis* presumptions this is also the case, but there is an additional component: although they are materially defeasible, they are pragmatically conclusive. In other words: these presumptions not only serve the function of giving reasons to believe that certain events have occurred (true in the material sense); but also give reasons to consider certain facts as proven (true in a dialectical, pragmatic or procedural sense). For this reason, "it is presumed" fulfills the generic function of approaching the material truth and the specific function of establishing a pragmatic, dialectical or procedural truth, transferring the burden of proof to whoever intends to deny the occurrence of the presumed fact by proving an alternative version of facts. In this sense, the general statement of presumption is also a rule of presumption, a rule of the burden of proof and/or argumentation[2].

2.7. There are three ways to oppose the conclusion. a) Denying the empirical foundations of the general statement of presumption; that is, to challenge its warrant role (to *challenge* the presumption). b) Accepting the general statement of presumption (that is, to accept that it expresses a regularity with a high probability of truth), but denying the occurrence of the base facts; that is, to *block* the presumption. c) Accepting both the general statement of presumption and the occurrence of the basic facts, but rejecting the conclusion (i.e. "the son is alive"). Here we would speak either of *excepting* the warrant (the general statement of presumption), or of *defeating* the presumption.

Only in cases b) and c) (when the general statement of presumption is accepted as a sure way to get the truth), it is acceptable that this statement also operates as a rule of presumption (it operates in dialectical terms as a rule for the distribution of the burden of proof and/or argumentation). In case a), the rejection of the general statement of presumption in material terms also implies its rejection as a rule of presumption in procedural or dialectical terms (AGUILÓ REGLA 2018).

3. *"It must (shall) be presumed". The norms of presumptions (legal presumptions)*

3.1. Let us take another starting point. Let us consider any of the legal norms of our legal systems that establish the presumption of paternity with respect to children born during the marriage.

There is no doubt: they are legal norms, norms of presumption. Now we find the same elements that we saw in the previous type.

[2] The specific function of the speech act of "presuming" is none other than reversing the burden of proof and/or argumentation (WALTON 1993).

a) A presumed fact: "X is the father of Y".
b) Some basic facts: "Y is the son of X's wife and was born during the marriage".
c) A connection between them; that is, a presumption statement that is normatively imposed (what I will call a "presumption rule" from now on).

There is no doubt that we could equally apply Toulmin's scheme of arguments. But at this point the interesting question is another one: Does the normative imposition of the presumption change its "nature"?

3.2. To properly understand the nature of presumption rules, a practical role must be assumed. Here I am going to refer to the role of the judge and the role of the legislator.

3.3. For the judge, a rule of presumption is just another valid legal rule. What does it force him to do? It is clear that the one bound by the rule of presumption is someone called to act in some way, not someone called to believe in something. Thus, the primary function of the rules of presumptions is not to establish any material truth, but a procedural truth (in the sense of a truth in the process). Accepting and applying a rule of presumption does not require the belief in the occurrence of any fact, but to consider it proven under certain circumstances. In other words, presumption rules acquire their meaning in a context of institutional decision, within a decision-making process. In this sense, the rules of presumption (by establishing procedural truths) benefit (facilitate the claim of) one party and harm (make difficult the claim of) the other party. The rules of presumption, therefore, always incorporate an element of partiality or (formal) inequality between the parties: they break the egalitarian principle that constitutes the process. Two questions immediately arise: Is it justified to establish such an obvious inequality between procedural parties? What relationship do the rules of presumption maintain with truth in the material sense?

3.4. To answer these questions let us now consider the role of the legislator. Let us see a catalog of reasons with which the legislator could justify the establishment of a rule of presumption.

a) Reasons of procedural economy linked to the probability of the truth of the presumed fact under certain conditions. That is, what underlies an "it must be presumed" is the acceptance by the legislator of an "it is presumable".
b) Reasons of procedural fairness aimed at restoring the balance between the parties due to the extraordinary difficulty involved in proving some facts.
c) Reasons of procedural caution linked to the unequal gravity of the legal consequences for the parties. To minimize the risk of greater damage, a procedural truth is established and the proof is placed on the party that assumes the lesser risk[3].
d) Institutional reasons aimed at stabilizing expectations and legal situations. A rule of presumption may respond more to the claim of establishing an institutional regularity or normality in the future, than to try to account for an already existing regularity.

Although this catalog of reasons could be expanded, it is enough to show that it makes sense to dictate presumption rules even when there is no an "it is presumable" in its justification; and that the relationship between an "it must be presumed" (normative) and "it is presumable" (theoretical) is contingent.

[3] The partiality of a norm of presumption can be justified in various ways: by probabilistic considerations (Q is more/less frequent than −Q in case of P), for evaluative considerations (the consequences of assuming Q in P are more/less serious than those of presuming −Q) and procedural considerations (it is more/less easy to produce proves in favor of Q than of −Q, in case of P) (MENDONCA 1998, following ULLMAN-MARGALIT 1983).

3.5. The rules of presumption establish a procedural truth (a truth in the process)[4]. They constitute, therefore, points of departure and arrival in a decision-making process. Due to its normative nature, the presumption *cannot be challenged* by denying its empirical foundations, denying that it brings us closer to the material truth. The warrant of this presumption is a legal norm whose validity and applicability do not depend on it.

But the procedural truths established by the rules of presumption can always be defeated. There are two ways to oppose the presumed fact: one, to *block* the presumption by showing the falsity of the base fact(s); another, to *defeat* the presumption showing the falsity of the presumed fact or providing clues or reasons to believe in its falsity.

4. *Presumption-rules and presumption-principles*

4.1. The norms of presumption establish a procedural truth and its content to the judge is always the same: to take a fact as proven. In this sense, all presumption norms have a more institutional than substantive meaning: the truths they establish are materially defeasible, but pragmatically conclusive. The central notion is therefore procedural (institutional) truth, not material (substantive) truth. However, sometimes the duty to take a fact for granted is subject to a condition, while in others it is not. This allows us to distinguish within the presumption norms between presumption-rules and presumption-principles (AGUILÓ REGLA 2006 and 2018).

4.2. The presumption-rules are norms of presumption that respond to the typical conditional structure of the legal rules.

4.2.1. According to them, the duty of the judge to assume the procedural truth is subject to one condition: the proof of the basic fact established by the rule. If the base fact is not proven, then the duty to assume the presumed fact does not arise.

4.2.2. The judge's reasoning is strictly a subsumption: the particular facts proven by the party are subsumed in the generic case (base fact) provided by the presumption-rule. The falsity of the basic fact supposes the *blocking* of the application of the legal consequence foreseen by the presumption-rule (that is, the duty to assume the procedural truth).

4.2.3. The legal consequences generically established by the presumption-rules can be *defeated* in a particular case: the procedural truth is defeatable by the material truth.

4.3. The presumption-principles are norms of presumption that respond to the typical categorical structure of the legal principles[5].

[4] Laudan discusses what the presumption of innocence requires and what instructions jurors receive (and should receive) in criminal trials in the United States. He does not use the distinction between material truth and procedural truth that we have used here. Rather, he opposes "material" to "probatory" and projects them onto the innocent/guilty pair. The combination produces four pairs, and he insists on the asymmetry between innocence and guilt: (i) Material innocence: the defendant in a process is materially innocent only in the event that he did not commit the crime. (ii) Probatory innocence: the defendant is probatorily innocent if the accusation fails because does not satisfy the standard of criminal proof (not guilty). (iii) Material guilt: the defendant actually committed the crime. (iv) Probatory guilt: the accusation against the defendant satisfies or exceeds the standard of proof. The asymmetry consists in the fact that while evidentiary guilt supports the affirmation of material guilt (the law assumes that if guilt has been proven then one is really guilty), probatory innocence does not guarantee the inference about material innocence (LAUDAN 2005).

[5] In structural terms, a good way to characterize legal principles as opposed to legal rules is by attending to von Wright's notion of categorical norm. Considering the notion of condition of application of a norm («the condition

4.3.1. Here, the duty of the judge is not subject to the proof of any base fact. The same thing happens with legal principles in general: they operate whenever they are relevant. This means that a presumption-principle displays its effectiveness whenever it is relevant.

4.3.2. Since these presumptions are not subject to any conditions, they cannot be blocked. They can only be defeated.

4.3.3. The presumptions of innocence, good faith, constitutionality of statutes, legality of administrative acts, etc., are considered procedural principles: they define the process and translate into burdens of proof and/or argumentation for who alleges guilt, bad faith, unconstitutionality or illegality. In this sense, the presumptions-principle play a much more shaping institutional role of the process than the one played by the presumptions-rule.

5. *In the core of defeasible reasoning. The problem of the iuris et de iure presumptions*

5.1. In accordance with what has been said, all presumptions, no matter whether they are *hominis* or legal, are at the core of the idea of defeasible reasoning.

5.2. *Hominis* presumptions are defeasible both if viewed from the perspective of the genre to which they belong (evidence inferences), and by the specific difference that characterizes them (the reversal of the burden of proof).

5.2.1. *Hominis* presumptions share with evidentiary inferences the property of being inductive inferences and, therefore, of being defeasible. In particular, in the presumptions *hominis* it is perfectly possible to continue accepting the validity of the warrant of the presumption, the truth of the occurrence of the basic fact and, nevertheless, reject the truth of the presumed fact by showing its falsity or weakening its credibility.

5.2.2. If we look at what they contribute beyond the idea of evidentiary inference, then the question is also clear: reversing the burden of proof means admitting the possibility of defeating the presumption through new evidence.

5.3. The same thing happens with the norms of presumptions (legal presumptions), and it makes no difference whether they are rule-presumptions or principle-presumptions: in both cases, legal reasoning is defeasible. The norm of presumption remains valid even if the presumption is defeated in a particular case; that is, even if the presumed fact is shown to be false.

5.3.1. Presumptions-rules can be both *blocked* (showing the falsity of the base fact) and *defeated* (showing the falsity of the presumed fact).

which must be satisfied if there is to be an opportunity for doing the thing which is the content of a given norm, will be called a *condition of application* of the norm») von Wright considers that norms can be divided into *categorical* and *hypothetical*. «We shall call a norm categorical if its condition of application is the condition which must be satisfied if there is going to be an opportunity for doing the thing which is its content, *and no further condition*. We shall call a norm hypothetical if its condition of application is the condition which must be satisfied if there is going to be an opportunity for doing the thing which is its content, *and some further condition*. If a norm is categorical its condition of application is given with its content. From knowing its content, we know which its condition of application is. For this reason, special mention of the condition is not necessary in a formulation of the norm» (VON WRIGHT 1963, 73-75).

5.3.2. Presumptions-principle can be only defeated, not blocked.

5.4. Despite this, many jurists speak of indefeasible legal presumptions. This is the case of the so-called *iuris et de iure* presumptions or *absolute* presumptions. These jurists maintain that they are indefeasible presumptions because, they say, no evidence is admitted against the presumed fact. These presumptions, they say, can be blocked, but not defeated.

5.4.1. Let's consider an example. Let us imagine a rule that establishes that "sexual relations between an adult and a minor are always non-consensual and constitute sexual abuse". This norm can be blocked by denying some of the basic facts (there were no sexual relations, the accused person was not an adult when the sexual relations took place, or the supposed abused person was not a minor at the time of the sexual relations). But if it is not blocked, it is not possible to prove that the relations were consensual because the proof of that is prohibited. Many legal scholars would argue that this rule establishes an indefeasible presumption: the absence of consent is presumed and evidence to show that there was consent is prohibited.

5.4.2. In my opinion, this is a mistake. This type of rules does not force to presume any procedural truth. It only obliges to apply the legal consequences derived from the occurrence of certain facts (the base facts of the crime). Treating these rules as norms of presumption is a mistake that brings with it a lot of conceptual problems. The idea of an indefeasible presumption is clearly contradictory, paradoxical.

5.4.3. The purpose of this type of rule is, rather, to eliminate all traces of presumptive reasoning. Therefore, in conceptual terms, this type of norms must be opposed to (instead of being confused with) norms of presumption (SCHAUER 2009). In my opinion, the idea of a rule of presumption *iuris et de iure* (absolute presumption) is an error in theoretical terms.

6. *What can cognitive sciences contribute to the argumentative use of presumptions?*

6.1. Detecting material fallacies. A central theme of legal argumentation is the validity of the arguments. In Alicante, following a scheme proposed by M. Atienza, we usually distinguish among formal, material, and pragmatic validity (ATIENZA 2013, 110-117). Formal validity refers to the fact that the argument complies with the rules of inference, that is, with the requirements linked to internal justification of the argument. Material validity assumes that the argument satisfies certain methodological requirements for obtaining the premises, that is, linked to the so-called external justification of the argument. And pragmatic validity refers to compliance with certain rules of fair play in the dialectical interaction between the subjects who debate; if the rules are followed, the agreement or the victory reached will be valid.

In this sense, fallacies are always the violation of rules: of rules of inference in formal fallacies; of methodological rules in material fallacies; and rules of fair play in pragmatic fallacies.

As we have seen, presumptions always establish an argumentative connection between facts, between the base facts and the presumed fact. In presumptions *hominis* the connection between these facts is strictly cognitive and in the rules of presumption this is also the case in some of them. This connection between facts may or may not be justified. The justified use of presumptions can be seen as heuristics and the unjustified uses, as biases. In this sense, cognitive sciences can clearly contribute to the detection of material fallacies.

6.2. Justified presumptions as heuristics and unjustified presumptions as biases.

> «Heuristics are cognitive shortcuts, or rules of thumb, by which people generate judgments and make decisions without having to consider all the relevant information, relying instead on a limited set of cues that aid their decision making. Such heuristics arise due to the fact that we have limited cognitive and motivational recourses and that we need to use them efficiently to reach everyday decisions. Although such heuristics are generally adaptive and contribute to our daily life, the reliance on a limited part of the relevant information sometimes results in systemic and predictable biases that lead to sub-optimal decisions» (PEER & GAMLIEL 2013, 114, following KAHNEMAN et al. 1999 and KAHNEMAN 2011).

If in this paragraph we were to replace the word "heuristics" with the word presumptions, we would obtain a coherent paragraph that would provide us with an approximate explanation of the genesis and functionality of all those presumptions whose foundation is found in a theoretical "it is presumable". And, naturally, the risk of biases helps to explain why presumptive reasoning can always be defeated.

References

AGUILÓ REGLA J. 2006. *Presunciones, verdad y normas procesales*, in «Isegoría», 35, 9 ff.

AGUILÓ REGLA J. 2018. *Las presunciones en el Derecho*, in «Anuario de filosofía del Derecho», XXXIV, 201 ff.

ATIENZA M. 2013 *Curso de argumentación jurídica*, Trotta.

FISK O.H. 1925. *Presumptions*, in «Cornell Law Review», 11, 20 ff.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Penguin.

KAHNEMAN D., SLOVIC P., TVERSKY A. 1999. *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press.

LAUDAN L. 2005. *The Presumption of Innocence: Material or Probatory?*, in «Legal Theory», 11, 4, 333 ff.

MENDONCA D. 1998. *Presunciones*, in «Doxa. Cuadernos de Filosofía del Derecho», 21, I, 83 ff.

PEER E., GAMLIEL E. 2013. *Heuristics and Biases in Judicial Decisions*, in «Court Review: The Journal of the American Judges Association», 422, 49, 113 ff.

SCHAUER F. 2009. *Thinking Like a Lawyer: A New Introduction to Legal Reasoning*, Harvard University Press.

TOULMIN S.E. 1958. *The Uses of Argument*, Cambridge University Press.

ULLMAN-MARGALIT E. 1983. *On Presumption*, in «The Journal of Philosophy», 80, 3, 143 ff.

VON WRIGHT G.H. 1963. *Norm and Action. A logical Enquiry,* Routledge & Kegan Paul.

WALTON D. N. 1993. *The Speech Act of Presumption*, in «Pragmatics & Cognition», 1, 1, 125 ff.

PART VII.

Issues on Legal Evidence

Psychological Factors in the Evaluation of Legal Evidence

BARTŁOMIEJ KUCHARZYK

1. *Free evaluation of evidence*

In *Law's Empire* Ronald Dworkin casually hints that factual issues in legal cases are quite trivial: «If judges disagree over the actual, historical events in controversy, we know what they are disagreeing about and what kind of evidence would put the issue to rest if it were available» (DWORKIN 1986, 3). Meanwhile, fact-finding makes up the lion's share of everyday legal proceedings. Also many high-profile trials, with the case of O.J. Simpson at the forefront, consist mainly of disputes over evidence and facts (KADRI 2005). Herr Professor, the facts are different—evaluation of evidence is a genuine epistemological problem and often a tough practical task.

Law itself has been answering the question of how to try facts at least since, as is always the case, Roman times. The answers provided by specific legal systems could be presented on a continuum from free to legal evaluation of evidence. The latter was especially common in medieval Europe when legal acts contained detailed rules on what significance should be assigned to particular kinds of evidence, including ordeals and torture (LANGBEIN 2012). Some of those rules have survived to this day, fortunately merely in the form of lovely legal maxims such as *confessio est regina probationum* (confession is the queen of trials) or *testis unus, testis nullus* (one witness, no witness).

The idea of free evaluation of legal evidence appeared in the Digest of Justinian and as a legal principle was introduced within Napoleonic codes. Free evaluation of evidence means that the law does not prejudge if some evidence is stronger or weaker, or less or more reliable than others, let alone decisive. The basic criterion for deciding questions of fact is the inner conviction (judgment) of the fact trier. Contemporarily the principle of free evaluation of evidence is explicitly featured in the codes of civil and criminal procedure of many continental law countries[1]. Common law systems, although containing complex rules on the admissibility of evidence, also give judges and jurors freedom in the evaluation of admissible evidence (freedom of proof—TWINING 1997).

Freedom of evaluation is of course not unlimited. Firstly, the evaluation models adopted in individual legal systems are not at the edge of the above-mentioned continuum—they consist of the principle and more or less numerous exceptions, such as irrebuttable presumptions or evidentiary prohibitions (evaluation is free *in principle*). With regard to the latter (as well as to the common law admissibility rules), one could even say that the evaluation of evidence consists of two stages: the a priori assessment of admissibility which is rather formalized, and the a posteriori evaluation of credibility and significance which is free (again, in principle)[2].

[1] See e.g. art. 233 of *Kodeks postępowania cywilnego*, art. 7 of *Kodeks postępowania karnego* (Poland), art. 1358 of *Code civil*, art. 353, 427 and 428 of *Code de procédure pénale* (France), § 286 of *Zivilprozessordnung*, § 261 of *Strafprozeßordnung* (Germany), art. 116 of *Codice di procedura civile* and art. 192 of *Codice di procedura penale* (Italy).

[2] For this reason, in the presented paper I focus on the a posteriori evaluation, which does not mean, however,

Second, an arguably more fundamental legal principle exists: the principle of truth (substantive truth as lawyers sometimes say to the amusement of philosophers). The factual findings should be as true as possible, hence the evaluation can be free as long as it is rational. That is why courts of appeal may review the evaluation of evidence conducted at the first instance.

However, research in cognitive sciences, particularly psychology, shows that human thinking often falls short of rationality standards. In this paper, I present a working classification of psychological factors that may diminish the rationality of legal evidence evaluation and examples of the impact of such factors (psychological pitfalls in free evaluation of evidence) taken from research and theory in experimental psychology.

2. *A few remarks on the psychology of (legal) thinking*

According to contemporary cognitive psychologists, the human mind works in two general modes referred to as the first and the second cognitive system (STANOVICH & WEST 2000)[3]. System 1 ("intuitive cognition") includes processes that are relatively fast, effortless, unintentional, automatic, associative, metaphorical, holistic, pattern-based, context-dependent, embodied, and unconscious (one has no mental access to the process itself though may be aware of its result). This system is primal both evolutionarily (it occurs in some form in all mammals) and ontogenetically (it is innate although we develop it with experience), and almost constantly active, hence should be considered the default mode of the human mind. Within System 1 processes run in parallel (a lot of information is processed at the same time which increases the pace of work) and do not require our limited cognitive (attentional) resources. We intuitively determine, for example, the source of sudden noise, assess emotions in the voice and facial expressions of our interlocutors, or appraise newly met people. One of the fundamental mechanisms of System 1 is heuristics—simple but fallible implicit thinking strategies (see below, 3.2.2.1.1).

Cognitive processes encompassed by System 2 (reasoning, "rational system"—EPSTEIN 1994) are relatively slow (time-consuming), effortful, intentional, controllable, rule-based, logical, analytical, abstract (working on symbols), and consciously accessible (monitorable). Such processes are available only to intellectually fit humans after infancy and possibly to apes taught language. System 2 develops through cultural and formal learning, processes information serially (process by process), and requires attentional resources to function efficiently. It is primarily associated with complex cognitive activities—we use it for example, hopefully, to write papers or conduct non-elementary mathematical calculations. It should of course also be involved in endeavors such as evaluation of legal evidence or legal decision-making. However, due to high usage costs (cognitive resources, energy, time, etc.) System 2 is not constantly active. Moreover, its effectiveness strongly depends on an individual's cognitive abilities (intelligence, memory, especially working memory, cognitive control, etc.), acquired cognitive skills and knowledge as well as external factors influencing their mind's work (e.g. time pressure or cognitive load).

Both systems are involved in our thinking which is studied by psychologists as mental functions such as inference (reasoning), decision-making, evaluation (judgment), and problem-solving. People process information both automatically and intentionally, draw conclusions consciously, but also beyond awareness, make judgments and decisions based on analysis (that is, at least in the sense of mechanism, rationally) or intuition, solve problems sometimes by insight and sometimes step by step, etc. Importantly, it seems that the role of intuitive processes (System 1) is much greater than rationalists would like (HAIDT 2001).

that psychology is irrelevant in the decisions on admissibility, especially when they exercise judicial discretion (see e.g. rules 402 and 401 of the Federal Rules of Evidence).

[3] A less technical account may be found in Daniel KAHNEMAN's (2011) superb book *Thinking, Fast and Slow*.

The potential adverse impact of at least some of the System 1 processes (and errors in the operations of System 2) on the rationality of legal judgments regarding facts and the evaluation of legal evidence is evidenced by both the results of research conducted in legal contexts and experiments in general psychology. The former could be collectively referred to as the experimental psychology of law (part of empirical legal research). In general, the experimental psychology of law studies non-legal factors of legal decision-making. Participants of experiments in this field take part in simulations of court trials or are asked to make various types of judgments and decisions (in particular "verdicts") based on information they have read. Such research is of course not without methodological weaknesses (KRAMER & KERR 1989; BORNSTEIN 1999; BIENECK 2009; BORNSTEIN & KLEYNHANS 2018). It should also be noted that it is conducted primarily in the United States, where one of the key elements of the legal system is the jury. As a jury is usually composed of laypersons, experiments also involve mainly non-lawyers, and in the trial simulation experiments simulation is often limited to mock jury deliberation (DEVINE et al. 2001). The results of such studies should therefore be applied only with caution to systems where legal decisions are made by professional judges.

On the other hand, more and more research is currently being conducted (also in Europe) involving lawyers, including judges (RACHLINSKI & WISTRICH 2017). Furthermore, many psychological phenomena and regularities simply apply to the vast majority of people (and lawyers are people after all), often regardless of the context (professional, personal, etc.), education, and experience (expertise). For this reason, even where jurors play the role of potted plants (if they are involved in the trials at all), neither the results of research in the field of experimental psychology of law (even those with lay participants only) nor theories and experiments in general psychology should be disregarded. Evaluation of legal evidence (especially when free) is not only an element of legal proceedings but also a complex mental process.

3. *Psychological factors in the evaluation of legal evidence*

As the body of research that can be directly or indirectly related to the issue of legal fact-finding is extensive and grows fast, I propose a working classification of psychological factors that may influence the rationality of legal evidence evaluation. For the reasons explained in KUCHARZYK 2021, the classification is based first of all on foundherentism—a theory of epistemic justification developed by Susan HAACK (1993). According to Haack, an individual's belief (for instance, an evaluation) is influenced by evidential and non-evidential elements of its causal nexus. The former include their other beliefs, perceptual states, introspective states, and memory traces. The latter are all their other mental states, e.g. desires, fears, or being under the influence of psychoactive substances.

Within the psychological, functional conceptual grid, non-evidential elements can be associated with non-cognitive or to be more precise—as our cognition (especially System 1) is not independent of emotions, motivations, etc. (THAGARD 2008)—meta-cognitive factors. Respectively, evidential elements depend most of all on cognitive factors. Both kinds of factors can be divided into further categories (see below). Developing the classification I will also give examples of the problematic impact of such factors on rationality, that is psychological pitfalls in free evaluation of evidence. One should still remember that the proposed classification is tentative, has an academic (educational, guiding) character, and does not claim to be an exhaustive and exclusive division. The human mind works as a whole, so our mental processes (including both systems of cognition) interact in many complex ways. In addition, psychological categories are theoretical constructs ("explanations" of behavior)—therefore do not necessarily accurately reflect the mechanisms of the mind or brain, and, as the concepts invented by various researchers, are often not compatible with each other. It would also be a mistake to suppose that

meta-cognitive factors are the exclusive domain of System 1 and that cognitive ones matter only to System 2. Factors of both kinds may affect the work of both systems, in particular their cooperation and competition[4]. In general, the factors described in this paper enhance System 1 dominance and thus hinder the work of System 2 (the rational one).

Last but not least, one should see the psychological constructs and phenomena discussed below as *risk* factors. Their actual impact may depend, inter alia, on the types of evidence being evaluated or the type, model, and course of the proceedings. Examples of such relationships will be given in the descriptions of selected pitfalls.

3.1. *Meta-cognitive factors*

Meta-cognitive factors that may influence the evaluation of evidence include affective, motivational, individual (differential), social, and situational (external) phenomena, processes, and properties. In the following subsections, I present examples of factors of each type.

3.1.1. *Affective factors*

Arguably the most important affective factors of evaluation are moods, i.e. long-term feelings of low intensity and often unspecified cause, and emotions, i.e. correlated changes in physiological arousal, feelings, cognitive processes, and behavioral reactions (including facial expressions) of an individual appearing in response to a situation assessed by them, not necessarily consciously, as important (EKMAN & DAVIDSON 1994). Both moods and emotions can influence the evaluation of oneself (self-esteem), other people, objects, and events.

The influence of mood on thinking depends on its type (valence). Among other things, it is believed that a positive mood stimulates creativity, problem-solving, and divergent thinking, while a negative mood encourages more detailed information processing, cognitive effort, and convergent thinking (BAAS et al. 2008; FORGAS 2013). From a legal perspective particularly instructive in this regard is an experiment conducted by Joseph FORGAS and Rebecca EAST (2008).

Participants in this experiment were shown videos to put them in a good (happy), neutral, or bad (sad) mood. They then watched videos of four people denying that they had committed a theft (two of whom were lying). Finally, participants were asked to assess the truthfulness (innocence) of these four people. Respondents in a negative mood were more likely than others to believe that the people in the recordings were lying. More importantly, however, these participants were more accurate in evaluating "suspects' testimonies". While the average effectiveness of detecting lies (guilt) by participants in a neutral or positive mood corresponded to the random one (50%), the participants in a negative mood were significantly (statistically) more effective. (The efficiency of "truth detection" did not depend on mood though.)

However, the results of this and other experiments described in this paper should not be interpreted too literally. A cautious conclusion here would be to say that the evaluator's mood may affect the evaluation of the credibility of explanations, testimonies, statements, etc. The direction and strength of this influence may be determined by numerous moderating variables. Let us note for example that in the experiment by Forgas and East lying occurred in as many as half of the recordings. Perhaps if the proportion of suspects telling the truth had been higher, a negative mood would lead to erroneous (too harsh) assessments of credibility.

Even more complex is the question of how emotions affect evaluation (CLORE & HUNTSINGER 2007). Various emotions can modify the mode and content of the assessment in different ways.

[4] Some factors may be considered inputs, outputs, or elements of one of the systems (for example heuristics are one of the mechanisms of System 1).

For example, Simone Schnall, Gerald Clore, Alexander Jordan, and Jonathan Haidt (SCHNALL et al. 2008) investigated the impact of disgust on moral judgments (evaluations).

In their four experiments participants rated on seven- or nine-point scales—from "perfectly fine" to "extremely bad"—issues such as marriage between closest cousins or keeping a wallet found on the street. In each experiment, disgust was induced in half of the participants. For example, in the first experiment, there was an extremely foul odor in the lab. In each case, the judgments (mean scores) in the experimental group (that is among the participants who were "externally disgusted") were more severe than in the control group.

Furthermore, research in experimental psychology of law shows that legal evidence that evokes strong (usually negative) emotional reactions, such as gruesome autopsy photos or moving victims' testimonies, increases the severity of punishment, while information about the accused's tragic life history may—depending on what the evaluators think about the responsibility of the accused and what emotions it evokes in them—decrease or increase it (SALERNO & BOTTOMS 2009).

3.1.2. *Motivational and individual factors*

The evaluation process can also be influenced by the needs, desires, preferences, expectations, attitudes, and traits of the evaluator. Traits should be understood here as, relatively constant in time and between situations, predispositions to behavior consistent with certain patterns (MISCHEL & SHODA 1995). Interindividual variation (individual differences) concerns, among others, temperament, cognitive abilities (e.g. intelligence), and personality.

An important theoretical construct on the border of the psychology of motivation and the psychology of personality that significantly affects cognitive processes is the need for closure, i.e. the desire to obtain clear and certain answers and ambiguity aversion (WEBSTER & KRUGLANSKI 1994). The need for closure may motivate the search for information—people with a high need for closure want to remove ambiguity as soon as possible and for as long as possible (KRUGLANSKI & WEBSTER 1996). However, this can lead to resistance to change of inaccurate, quickly-made (based on partial data only) assessments, oversimplification of information processing, reduced hypothesis generation, and increased susceptibility to cognitive biases. The individual level of the need for closure is relatively constant over time and between situations but it remains under the influence of external variables—for example, it is intensified by time pressure (VAN HIEL & MERVIELDE 2003).

The influence of motivational mechanisms on evaluations, beliefs, conclusions, and decisions is also illustrated by three well-known psychological phenomena (illusions) aimed at maintaining positive self-esteem and well-being. The first is wishful thinking, i.e. the tendency to accept as true beliefs that are beneficial to us (KRIZAN & WINDSCHITL 2009). Wishful thinking is closely related to self-deception, which means taking actions (often unconsciously) to confirm or cause a desired belief (VON HIPPEL & TRIVERS 2011). Self-deception can involve distortions in perception, information processing, and judgment. The third important phenomenon is the reduction of cognitive dissonance (FESTINGER 1962). Having two contradictory or incompatible beliefs usually causes discomfort and the need to change one of them. Particularly strong may be postdecisional dissonance resulting from the choice of one of several comparably attractive options. In this case, a typical reaction is to overestimate the selected option and underestimate the rejected options (BREHM 1956; LEE & SCHWARZ 2010). Rationalization (ex-post justification) of a decision or evaluation may therefore be related to the cognitive omission of its true causes.

In a legal context, cognitive dissonance may be, for example, the result of the incompatibility of evidence with the initial hypothesis (theory or intuition) regarding the case (ASK et al. 2011). Potential ways to reduce this dissonance are belief change (rejection of the initial hypothesis) and its opposite: skepticism towards evidence. The latter, however, is asymmetric—the

reliability of "flexible", i.e. relatively subjective, evidence (e.g. witness statements), is rated lower than the reliability of evidence the interpretation of which is relatively objectified, e.g. DNA tests (Ask et al. 2008; Marksteiner et al. 2011). Therefore, the evaluation of flexible evidence inconsistent with the a priori hypothesis seems particularly prone to errors.

Style and efficiency of reasoning, assessing, problem-solving, etc. obviously depend on intelligence and cognitive abilities but are also (moderately) correlated with "non-intellectual" individual characteristics. The Big Five model (Costa & McCrae 1992)—arguably the most influential theory of personality—distinguishes five basic personality traits: neuroticism, extraversion/introversion, agreeableness, conscientiousness, and openness to experience, virtually each of them having cognitive aspects. For example, a high level of openness to experience is associated with intellectual curiosity, divergent thinking, creativity, and a preference for novelty and diversity (Komarraju et al. 2011), thus, in a way, is the opposite of the need for closure.

3.1.3. *Social and situational factors*

The aforementioned time pressure is not the only external factor that can influence thinking about evidence and facts. Keeping in mind the academic nature of the classifications of mental mechanisms and their moderators, it is worth distinguishing, as far as possible, the information context of evaluation and decision-making (see below, 3.2) from the "physical" (time, place, external stimuli, etc.) and social context.

The social context includes phenomena related to the actions, characteristics, or mere presence of other people. Some of them, such as groupthink syndrome (Janis 1982), concern collective evaluation and decision-making, but many affect individual assessments and decisions as well.

Conformism is a perfect example of a social mechanism that poses a serious threat to the freedom and rationality of the evaluation process. In Solomon Asch's (1956) classic experiments groups of people were asked to compare the length of lines shown to them, with rankings given individually. The actual participants in the experiment did not realize that all the other people in their group were instructed to give wrong answers at certain times (even when, and perhaps especially when, the right answer was obvious). Group pressure turned out to be so strong that real participants very often adapted their answers to the majority, and some even claimed that they saw what the group dictated. In legal contexts, conformism seems to be a problem primarily in the case of group decision-making, but even individual evaluations may be influenced by beliefs and expectations within the social environment (parties in the case, public opinion, superiors, politicians, etc.).

On the other hand, various types of characteristics of the assessed person may influence the evaluation of their statements (e.g. testimonies), in particular as to credibility. The key mechanism of this influence is the halo effect, i.e. extrapolating the assessment of one trait to other issues (Nisbett & Wilson 1977a). This effect is related to the general preference for unambiguous evaluations (the stronger, let us recall, the greater the need for closure). An example of the halo effect is ascribing high intelligence, desirable personality traits, and even happiness in life to physically attractive people (Eagly et al. 1991). People considered physically attractive are statistically more likely to find a job, earn more, or win elections (Hosoda et al. 2003; Verhulst et al. 2010).

Moreover, research in the field of experimental psychology of law suggests that attractive defendants get more favorable sentences (Stewart 1980) unless beauty has been a means to the crime (Sigall & Ostrove 1975). This relationship may be particularly strong if the accused is a woman (Efran 1974), but disappears in the case of the most serious crimes (Downs & Lyons

1991). Lower sentences can also be expected when the attractiveness of the perpetrator and the juror is similar (Darby & Jeffers 1988). See also Mazzella & Feingold 1994.

The halo effect is not limited to physical attractiveness. From the perspective of legal evidence evaluation, especially instructive are studies demonstrating the positive impact of a witness's self-confidence on the assessment of their credibility (Penrod & Cutler 1995; Bradfield & Wells 2000). One should note that the self-confidence of a witness often does not correlate or even correlates negatively with the accuracy of their testimony (Smith et al. 1989).

Confusing self-confidence with credibility is just one of many problematic tendencies in evaluating the accuracy of witness statements. Another example is ignoring the influence of external factors (e.g. suggestions) on the form and content of testimonies (Memon et al. 2003). The accuracy of testimonies is, as a rule, simply overestimated (Lindsay et al. 1981). It is worth a mention that contradictions in testimonies may lead to lower assessments of their credibility and accuracy without affecting legal decisions (Berman et al. 1995).

In the case of expert opinions (expert witness testimony), the influence of the status of the expert on the evaluation of their opinion may be significant (Pornpitakpan 2004). The evaluation is usually not independent of the assessments of previous opinions and the expert's reputation (the prestige of the institution to which they are affiliated may also matter). Empirical confirmation is found, for example, for common intuition about the benefits of having the label of a "good student" (Darley & Gross 1983). It is in particular for this reason that reviewers of scientific papers should not know the names of authors. However, anonymization is not always possible or even desirable. One should remember that the halo effect leads to errors of judgment when in fact there is no correlation between the issue being assessed and the issue affecting the evaluation (or there is a correlation opposite to the applied). Due to the specialist nature of the matter, a judge evaluating the credibility of an expert opinion often has few cognitive tools at their disposal. In such a situation, relying (among other things) on the expert's carefully assessed achievements seems reasonable. However, it is certainly advisable to limit the impact of personal relationships (including simple liking) between judges and experts.

The evaluation may also be influenced by variables related more to the situation of assessment than to its participants (situational factors). An obvious example here is distractors, i.e. stimuli that divert attention from the object of reflection. By engaging cognitive resources they hinder the use of System 2. Distractors can be external (e.g. noise) or internal (e.g. perseverative thoughts). Their influence is particularly strong in the case of cognitive overload (time pressure, fatigue, stress, excess of information and tasks, etc.—Lavie 2010).

Other situational factors may operate in more complex ways. As to trials, it has been demonstrated, for example, that the use of a computer presentation in court may increase the understanding of presented evidence, the evaluation of its credibility, and favor for the party presenting it (Hewson & Goodman-Delahunty 2008; Feigenson 2010). In the case of witness testimonies, their order may be important—the presentation of witnesses according to the chronology of events in the case generally gives better results (rather in terms of benefits for the particular party than the accuracy of factual findings) than according to the importance of their testimonies (Pennington & Hastie 1988 and 1990).

3.2. *Cognitive factors*

Affective, motivational, individual, social, and situational factors are in necessary, complex, and mutual relations with mechanisms traditionally defined as cognitive (thus the former are rather meta- than non-cognitive). Building on Haack's (1993) account of evidential elements of belief's causal nexus, I divide cognitive factors into experiential and belief-related, i.e. directly associated with the formation and content of beliefs.

3.2.1. *Experiential factors*

Experiential factors could be further divided into perceptual, introspective, and mnemonic. However, the vast majority of research, significant from the perspective of legal fact-finding, on perception and memory concerns not the process of evidence evaluation, but witnesses (in particular eyewitnesses) and their mistakes. Knowledge of witness psychology is certainly invaluable for the evaluator of testimony (lack of it may be considered another psychological pitfall), yet due to the limited volume of this paper, let me refer the Reader to the relevant literature (e.g. GUDJONSSON 1992; LOFTUS 1996, 2005, 2019; MEMON et al. 2003; WIXTED & WELLS 2017).

Introspection instead seems to be of little importance for legal fact-finding which is based primarily on external empirical evidence. Legal evidence is a document or testimony, but not a judge's thought, even if related to the case. On the other hand, introspection provides the subject of cognition with insight into their cognitive processes and beliefs. The quality (accuracy) of this insight, however, is debatable.

Research on introspection shows that people's sincere explanations for their actions may be untrue. Richard NISBETT and Timothy WILSON (1977b) gave numerous examples of this phenomenon in their classic work *Telling More Than We Can Know: Verbal Reports on Mental Processes*. We often lack access to what is happening in our mind, so we interpret our behavior using folk psychology—common (a priori) theories about the human psyche (GOLDMAN 1993; KOZUCH & NICHOLS 2011). In particular, some "reasonings" may be only post hoc justifications of intuitively made judgments or decisions (HAIDT 2001). From a legal perspective, this leads to questions about the reliability of the reasons given for legal decisions, including explanations of factual findings. As rightly pointed out by Richard POSNER (2010), judges' reports on their cognitive processes should be considered only a source of hypotheses.

3.2.2. *Belief-related factors*

The content of new beliefs (in particular evaluations) depends both on the way they are formed and the content of already-held beliefs. Therefore belief-related factors involve the principles of functioning and systematic errors (tendencies) of the human mind related to the formation of beliefs, as well as the beliefs themselves (knowledge, opinions, etc.).

3.2.2.1. *Belief-formation factors*

From the perspective of identifying psychological pitfalls in legal evidence evaluation arguably the most interesting belief-formation factors are heuristics and cognitive biases.

3.2.2.1.1. *Heuristics*

Heuristics, i.e. practical, intuitive, not necessarily conscious rules of assessment, decision-making, or problem-solving (GIGERENZER & GAISSMAIER 2011) are the basic mechanisms of shaping beliefs within System 1. Heuristics are simple, fast, and frugal (in terms of cognitive resources), but also fallible—not in all cases they bring the correct results.

In a classic article, Amos TVERSKY and Daniel KAHNEMAN (1974) described three commonly used heuristics of judgment under uncertainty: the representativeness heuristic, the availability heuristic, and the anchoring and adjustment heuristic (effect). Each of them can be presented as replacing the question to be answered with a simpler one (KAHNEMAN 2011). The fallibility of heuristics is related to the fact that even a correct answer to the latter question does not always solve the original problem.

The representativeness heuristic replaces the question "What is the probability that object X belongs to class Y?" with the question "To what extent is X representative of Y?". The person using it relies then on the similarity of X to their prototype or stereotype of Y.

One of the pitfalls related to the representativeness heuristic is the base rate fallacy (neglect). In a study by Kahneman and Tversky participants read the following description: «Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail» (Tversky & Kahneman 1974, 1124).

The participants were asked to rank a list of occupations that Steve could have, from most to least likely. Rare occupations with stereotypes fitting Steve's description (e.g. librarian) were rated as more likely than much more common professions "not matching" the description (e.g. farmer). Probability was therefore assessed more based on representativeness (similarity to the stereotype) than the base rates (ratios in the population, e.g. farmers to librarians).

The base rate neglect may be of fundamental importance for the rational evaluation of legal evidence. Participants in another Kahneman and Tversky study were presented with the following case:

> «A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (a) 85% of the cabs in the city are Green and 15% are Blue.
>
> (b) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time» (Tversky & Kahneman 1982, 156).

The participants were asked to assess the probability that the cab involved in the accident was blue rather than green. The most common answer was 80%. Most therefore based their estimates solely on the credibility of the witness, completely ignoring the a priori probability resulting from the ratio of both types of cabs in the city, i.e. the base rate. The correct solution—calculated with the use of Bayes' theorem—is only about 41%[5].

Contrary to the testimony of a fairly credible witness, it is then more likely that the accident was caused by a green cab. This experiment points out the potential dangers of disregarding statistical evidence (Wells 1992). Statistical information may prove valuable in the process of finding the facts of the case, and in particular help in evaluating the credibility and accuracy of other evidence. In a case such as the above, only by taking into account the base rate one can avoid a serious error in fact-finding and a violation of the rational evaluation of evidence principle. In practice, of course, it is not so much about specific numbers (which, in principle, are unavailable) but about thinking within System 2, and thus, among other things, checking the correctness of heuristic solutions, including taking into account statistical data and regularities.

Interestingly, when in another experiment the base rate was presented as the proportion of accidents caused by the cabs of both companies, the participants were much more likely to apply it. Humans are generally better at causal inferences (or inferences with a semblance of causality) than at probabilistic reasoning (Kunda et al. 1990).

The second of the heuristics described by Kahneman and Tversky in their 1974 paper—the availability heuristic—is based on replacing the question about the frequency of a particular

[5] Of course, this solution is precise only with a certain idealization, including in particular the assumption that all information from the description is true and accurate. Moreover, one must ignore the possibility that not all cab drivers in the city work for one of the two companies, that the cab came from another city, etc. However, even without idealization, taking into account the adequate base rate increases the probability of the conclusion being correct.

category or the probability of a particular event with the question about the ease of recalling or imagining respective examples. This ease is in turn influenced by individual experiences, emotions, media, and other factors. A well-known example of the availability heuristic in action concerns the fear of air travel. Statistics prove that the plane is the safest means of transport, but media coverage or movies on plane crashes strongly affect memory and imagination. In research by Paul Slovik's team, the majority of respondents believed, among other things, that whirlwinds were a more likely cause of death than asthma (LICHTENSTEIN et al. 1978). Meanwhile, in the United States deaths from asthma were at that time twenty times more common than deaths from whirlwinds. The frequency of unusual (rare but spectacular) events is overestimated, and the frequency of common (frequent) events is underestimated, which directly translates into probability assessments.

The availability heuristic can affect the evaluation of legal evidence. The authors of the Dahlem Workshop 2004 report on the role of heuristics in litigation indicate, for example, that attorneys try to present evidence in such a way that the facts and conclusions most favorable to their clients are the easiest to remember and the most emotionally charged (PIPERIDES et al. 2006). One should thus note that the actions of parties' lawyers generally hinder the effective use of heuristics by judges and jurors and they may be intentionally aimed at taking advantage of cognitive biases and fallacies. Emotional, picturesque evidence can have a much greater impact on the decision than abstract information of objectively greater probative value (GUTHRIE 2006).

Mental availability affects judgments on the likelihood of events, the relevance of ideas or information (KURAN & SUNSTEIN 1999), and perhaps even the truth of statements. An extreme example of this third relationship may be the illusory truth effect, i.e. the tendency to consider information repeated many times true or trustworthy (HASHER et al. 1977; POLAGE 2012). The evaluation of the truth of a statement is influenced, among other things, by the impression of its familiarity and comprehension. Repetition makes it easier to process information, increases its (subjective) comprehensibility, but above all makes it more familiar. The impression of truthfulness (credibility, accuracy) of a given piece of information may also be the result of recalling it from memory (OZUBKO & FUGELSANG 2011). Therefore, known (familiar), mentally accessible, and easy-to-process statements seem true to us. Moreover, such subjective and insignificant properties of information as the ease of pronouncing the name of its source may also have an impact on the evaluation of credibility (NEWMAN et al. 2014). One may note that even prior knowledge may not protect against the illusory truth effect—someone's (true) belief may change as a result of their repeated contact with its negation (FAZIO et al. 2015).

The third of the heuristics of judgment (assessment) under uncertainty described by Kahneman and Tversky is the anchoring and adjustment heuristic (effect). When estimating numerical values, people often adjust—usually not enough—their guesses starting from the first number that comes to their mind (the anchor). They associate more or less appropriate values with the task (for example as a result of suggestion or priming) and then adjust the estimate until it no longer seems to be too low or too high, but usually fail. People asked to quickly estimate the product 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8 give a much lower number than those given the inverse equation: 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 (in the first case the anchor is 1, in the second 8). In addition, both groups give estimates significantly lower than the actual product—40320 (TVERSKY & KAHNEMAN 1974). Note that the anchor may have nothing to do with the estimated value.

Even professionals are not free from the influence of anchors. The anchoring effect was observed inter alia in studies involving real estate agents (NORTHCRAFT & NEALE 1987), physicians (BREWER et al. 2007), and—which is of importance in the context of legal fact-finding—judges. One of the experiments by Birte Englich, Thomas Mussweiler, and Fritz Strack (ENGLICH et al. 2006) involved experienced German judges and prosecutors (on average

over 13 years in the profession). They were asked to evaluate evidence (including the opinion of a forensic psychologist, explanations of the accused, and witness testimony) in a case of multiple theft and to estimate the appropriate penalty. At the same time, they were informed that the penalty proposed in the indictment was determined randomly (half of the participants were suggested 3 months of community service, the other half—9 months). Despite this information, the numbers from the indictment influenced the decisions. Participants who had been suggested 9 months proposed significantly harsher sentences (on average 6 months) than those who were suggested 3 months (on average 4 months). Professional lawyers (as a group) thus turned out to be—arguably within their expertise—susceptible to the anchoring effect despite receiving information on the randomness (irrelevance) of the anchor.

Evaluation (and misevaluation) of legal evidence can be also associated with several heuristics described by researchers other than Kahneman and Tversky, among others the affect heuristic (SLOVIC et al. 2007; RACHLINSKI 2006) and the take-the-best heuristic (GIGERENZER & GOLDSTEIN 1999; PIPERIDES et al. 2006).

To sum up, according to Kahneman and Tversky heuristics are rather effective in everyday life but, as they answer simplified questions, in more cognitively demanding contexts systematically lead to mistakes and, as a result, to the formation of false beliefs. They may therefore limit the rationality of fact-finding, evaluation of evidence, and decision-making. This view is to some extent questioned by, among others, Gerd Gigerenzer (GIGERENZER 2006; GOLDSTEIN & GIGERENZER 2002; TODD & GIGERENZER 2012).

3.2.2.1.2. *Cognitive biases*

Among belief-formation factors (and pitfalls) one can also include cognitive biases—general, mostly subconscious tendencies in human thinking leading (even in everyday life) to systematic, predictable errors (HILBERT 2012). In the context of the evaluation of legal evidence, the confirmation bias and the hindsight bias are especially worth discussing.

Confirmation bias consists in researching, memorizing, recalling, and interpreting information in such a manner as to confirm already-held beliefs or hypotheses (NICKERSON 1998). It is thus a tendency to check assumptions, test hypotheses, or evaluate evidence in a biased way. The possible explanations of this tendency are cognitive parsimony in relation to limited information processing resources (STANOVICH 2009) and wishful thinking, whereas the resulting issues include, for example, polarization of views and attitudes (aggravation of disputes even if the parties have the same data at their disposal), perseverance of beliefs (upholding beliefs despite the supporting evidence having been undermined), primacy bias (relying on prior more than on later information) and illusions of correlations[6] (perceiving non-existent relationships between phenomena, persons, properties, etc.—GOLDING & RORER 1972).

Confirmation bias can influence thinking in legal contexts (NICKERSON 1998, 193 f). Judges and juries may favor certain, especially intuitive, conclusions and disregard alternatives. Extensive and complex evidence is fertile ground for the polarization of views (MYERS & LAMM 1976). Confirmation bias is also strictly associated with the prosecutor's fallacy (THOMPSON & SCHUMANN 1987; LEO & DAVIS 2010). In general, it is an error in statistical inference consisting of the assumption that a low probability of event A occurring given event B [P(A|B)] implies a low probability of event B occurring given event A [P(B|A)]. As the name suggests, this error may easily appear in the arguments of the prosecution in criminal trials. A typical example is inferring that a low probability of finding a particular piece of evidence (e.g. a DNA match) when the accused is innocent indicates a low probability of the accused being innocent when

[6] They should not be confused with the correlation fallacy, i.e. the illusion of causality.

this piece of evidence appears (THOMPSON et al. 2003; FENTON et al. 2016). A real-life instance is the terrible Sally Clark case—see e.g. NOBLES & SCHIFF 2005.

The hindsight bias (also known as the saw/knew-it-all-along effect or creeping determinism) consists in evaluating past or present events (including the results of decisions one made) as more probable and predictable than they had been before occurring (ROESE & VOHS 2012). It is a manifestation of a more general tendency to overestimate one's own achievements, capabilities, and knowledge—overconfidence (MOORE & HEALY 2008). In some cases, it may be even associated with changes in the content of memories (STAHLBERG & MAASS 1997). The hindsight bias plays a role in historical, economic, or sociological analyses, but also in medical and legal practice. In the context of law, it primarily concerns the attribution of responsibility or the ability to predict certain events. The hindsight bias may result in the accused or the defendant being wrongly found responsible. On the other hand, the plaintiff (or even the victim) may be assessed as insufficiently prudent—events with negative consequences are retrospectively perceived as riskier (OEBERST & GOECKENJAN 2016). The hindsight bias is also one of the explanations of the inadmissible evidence effect, i.e. the impact of information disclosed in a trial and then excluded from it on the verdict (see e.g. KUCHARZYK 2017).

3.2.2.2. *Belief-content factors*

Evaluation of legal evidence may be also influenced by the very content of the beliefs held by the evaluator, from knowledge through opinions and views to stereotypes and prejudices (BODENHAUSEN 1988; YOURSTONE et al. 2008; SMALARZ et al. 2016). Fact-finding should be based on reliable knowledge and proper experience and take into account the information context (e.g. expert opinions) and normative context (in particular the procedural rules). Also with these "substantial" belief factors one can associate psychological phenomena that may hinder the rationality of evaluation.

Some pitfalls are related to the properties of expert knowledge (expertise), i.e. extensive and deep knowledge in a specific field (ERICSSON et al. 2018). In addition to the scope expert knowledge is distinguished by its structure (hierarchical with many levels of abstraction), high degree of proceduralization with preserved access to declarative knowledge (possibility of verbalization), and schemes (heuristics) for solving problems specific to the domain (experts quickly and aptly recognize patterns specific to such problems and usually use effective heuristics).

These properties generally give experts an advantage over laypersons and novices in the field. Experts recognize patterns (categorize problems) more accurately, analyze more information, and use the strategy of falsifying (rather than confirming) hypotheses more often. However, they sometimes lack flexibility when solving non-routine problems (LEWANDOWSKY & KIRSNER 2000). The reason for that may be automatization—as experience in the domain increases, relevant cognitive processes gradually become uncontrolled and unconscious (FRENSCH & STERNBERG 2014). When a problem is misrecognized as typical for the field, the triggered scheme may prove ineffective and at the same time hard to quit or modify. As legal fact-finding (evidence proceedings) is a kind of expert problem-solving, such rigidity may appear in authorities collecting evidence, expert witnesses as well as judges evaluating evidence (BLASI 1995; MENKEL-MEADOW 2000).

Moreover, experts, like virtually everyone, make mistakes in assessing their knowledge. They display both the hindsight bias (GUILBAULT et al. 2004) and the common overconfidence (ANGNER 2006).

In experimental studies, the accuracy of answers is, as a rule, significantly lower than the level of certainty declared by respondents (KORIAT et al. 1980). This effect depends neither on intelligence nor on expertise (LICHTENSTEIN & FISCHHOFF 1977). Of the answers given with 100% self-confidence, an average of 20-30% turns out to be false. Research involving experts (e.g.

physicians) shows that the subjective certainty of assessments increases with the acquisition of information. However, the accuracy often remains unchanged or even—when there is a lot of information (cognitive overload)—decreases. Hindsight bias and overconfidence occur of course also in lawyers (ANDERSON et al. 1997; GOODMAN-DELAHUNTY et al. 2010).

Finally, one should note that legal (in particular judicial) expertise seems to concern primarily the normative domain, i.e. it consists of the knowledge of legal norms and the ability to interpret and apply them. The expertise of judges (as a group) in evaluating evidence and fact-finding is, however, quite uncertain (SPELLMAN 2007; PORTER & TEN BRINKE 2009).

The question of whether lawyers reason about facts better when the context of the task involves legal norms or proceedings is therefore an interesting empirical research problem. A preliminary clue has been provided by Chris Guthrie, Jeffrey Rachlinski, and Andrew Wistrich, who in one of their experiments presented a group of federal magistrate judges with the following case (related to the Green/Blue cab problem):

> «The plaintiff was passing by a warehouse owned by the defendant when he was struck by a barrel, resulting in severe injuries. At the time, the barrel was in the final stages of being hoisted from the ground and loaded into the warehouse. The defendant's employees are not sure how the barrel broke loose and fell, but they agree that either the barrel was negligently secured or the rope was faulty. Government safety inspectors conducted an investigation of the warehouse and determined that in this warehouse: (1) when barrels are negligently secured, there is a 90% chance that they will break loose; (2) when barrels are safely secured, they break loose only 1% of the time; (3) workers negligently secure barrels only 1 in 1,000 times» (GUTHRIE et al. 2000, 808).

Participants of the experiment were asked how likely it was that the barrel that injured the plaintiff broke off as a result of the negligence of one of the workers. They answered by choosing one of four options (probability ranges): 0-25%, 26-50%, 51-75%, or 76-100%.

The correct (see above, nt.4) solution is (only) about 8.3%. However, accidents like this intuitively seem to be the result of someone's carelessness. Therefore, the 90% chance (the probability of an accident given negligence) may be easily misinterpreted as the probability of negligence given an accident (the prosecutor's fallacy), which may be seen as the probability that the accident was caused by negligence. In this experiment, the majority (about 60%) of the participants chose the wrong options and most of them went with 76-100%. On the other hand, the judges fared more than twice as well as physicians solving an analogous problem in a medical context (CASSCELLS et al. 1978)[7].

4. *Conscious evaluation of evidence*

The relative success of the participants in the above-described study may mean that some of them used System 2 instead of intuition (GUTHRIE et al. 2007). Intuitive thinking is the domain of System 1—the default and dominant (especially in cognitively difficult conditions) mode of operation of the human mind. Many errors in reasoning, evaluation, and decision-making may therefore be attributed to its automatic, fast, and unconscious processes. The rationality of free (and thus dependent on the individual mind) evaluation of evidence depends to a large extent on whether these errors can be avoided by the evaluator.

[7] One should however remember that the results of separate experiments can only be compared with extreme caution.

Therefore, it can be argued that conscious thinking is a necessary condition for the rationality of free evaluation of evidence. "Conscious" should be understood twofold here. Firstly, evaluation should be informed, i.e. conducted by a person who is aware of the existence of psychological pitfalls of evaluation (has basic knowledge of relevant theories and research) and does not underestimate them. Secondly, evaluation should be attentive (cognitively controlled), i.e. made with the use of System 2, which, if necessary, corrects its counterpart's mistakes.

The road to informed evaluation is simply through education. Unfortunately, even knowledge of eyewitness psychology is not common among judges (BJØRNDAL et al. 2021). Meanwhile, the psychology of thinking is a vast and dynamically developing field. The same applies to epistemology, which could be instrumental in the legal pursuit of rationality and truth.

Attentive assessment should therefore be facilitated all the more. One can try to use legal mechanisms for this purpose, but their effectiveness may be quite limited. It would seem, for example, that the introduction, where there is none, of the obligation to justify factual findings in writing, could encourage more careful reasoning. However, as mentioned above, such justifications may be as well derived from intuitive conclusions (LIU & LI 2019). On the other hand, the effects of experimental methods of attention, working memory, or cognitive control training on the overall functioning of the mind are not always straightforward (POSNER et al. 2015). Even *reasonable* external conditions (lack of distractors, pressures, etc.) may often be difficult to provide. Therefore, individual self-control, mindfulness, and epistemic ethos remain indispensable.

References

Anderson J.C., Jennings M.M., Lowe D.J., Reckers, P.M. 1997. *The Mitigation of Hindsight Bias in Judges' Evaluation of Auditor Decisions*, in «Auditing», 16, 20 ff.

Angner E. 2006. *Economists as Experts: Overconfidence in Theory and Practice*, in «Journal of Economic Methodology», 13, 1 ff.

Asch S.E. 1956. *Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority*, in «Psychological Monographs», 70, 1 ff.

Ask K., Rebelius A., Granhag P.A. 2008. *The 'Elasticity' of Criminal Evidence: A Moderator of Investigator Bias*, in «Applied Cognitive Psychology», 22, 1245 ff.

Ask K., Reinhard M.A., Marksteiner T., Granhag P.A. 2011. *Elasticity in Evaluations of Criminal Evidence: Exploring the Role of Cognitive Dissonance*, in «Legal and Criminological Psychology», 16, 289 ff.

Baas M., De Dreu C.K.W., Nijstad B.A. 2008. *A Meta-Analysis of 25 Years of Mood-Creativity Research: Hedonic Tone, Activation, or Regulatory Focus?*, in «Psychological Bulletin», 134, 779 ff.

Berman G.L., Narby D.J., Cutler B.L. 1995. *Effects of Inconsistent Eyewitness Statements on Mock-Jurors' Evaluations of the Eyewitness, Perceptions of Defendant Culpability and Verdicts*, in «Law and Human Behavior», 19, 79 ff.

Bieneck S. 2009. *How Adequate Is the Vignette Technique as a Research Tool for Psycho-Legal Research?*, in Oswald M.E., Bieneck S., Hupfeld-Heinemann J. (eds), *Social Psychology of Punishment of Crime*, Wiley-Blackwell, 255 ff.

Bjørndal L.D., McGill L., Magnussen S., Richardson S., Saraiva R., Stadel M., Brennen T. 2021. *Norwegian Judges' Knowledge of Factors Affecting Eyewitness Testimony: A 12-Year Follow-up*, in «Psychiatry, Psychology and Law», 28, 665 ff.

Blasi G.L. 1995. *What Lawyers Know: Lawyering Expertise, Cognitive Science, and the Functions of Theory*, in «Journal of Legal Education», 45, 313 ff.

Bodenhausen G.V. 1988. *Stereotypic Biases in Social Decision Making and Memory: Testing Process Models of Stereotype Use*, in «Journal of Personality and Social Psychology», 55, 726 ff.

Bornstein B.H. 1999. *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, in «Law and Human Behavior», 23, 75 ff.

Bornstein B.H., Kleynhans A.J. 2018. *The Evolution of Jury Research Methods: From Hugo Munsterberg to the Modern Age*, in «Denver Law Review», 96, 813 ff.

Bradfield A.L., Wells G.L. 2000. *The Perceived Validity of Eyewitness Identification Testimony: A Test of the Five Biggers Criteria*, in «Law and Human Behavior», 24, 581 ff.

Brehm J.W. 1956. *Postdecision Changes in the Desirability of Alternatives*, in «Journal of Abnormal and Social Psychology», 52, 384 ff.

Brewer N.T., Chapman G.B., Schwartz J.A., Bergus G.R. 2007. *The Influence of Irrelevant Anchors on the Judgments and Choices of Doctors and Patients*, in «Medical Decision Making», 27, 203 ff.

Casscells W., Schoenberger A., Graboys T.B. 1978. *Interpretation by Physicians of Clinical Laboratory Results*, in «New England Journal of Medicine», 299, 999 ff.

Clore G.L., Huntsinger J.R. 2007. *How Emotions Inform Judgment and Regulate Thought*, in «Trends in Cognitive Sciences», 11, 393 ff.

Costa P.T., McCrae R.R. 1992. *Four Ways Five Factors Are Basic*, in «Personality and Individual Differences», 13, 653 ff.

Darby B.W., Jeffers D. 1988. *The Effects of Defendant and Juror Attractiveness on Simulated Courtroom Trial Decisions*, in «Social Behavior and Personality», 16, 39 ff.

Darley J.M., Gross P.H. 1983. *A Hypothesis-Confirming Bias in Labeling Effects*, in «Journal of Personality and Social Psychology», 44, 20 ff.

Devine D.J., Clayton L.D., Dunford B.B., Seying R., Pryce J. 2001. *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, in «Psychology, Public Policy, and Law», 7, 622 ff.

Downs A., Lyons P. 1991. *Natural Observations of the Links Between Attractiveness and Initial Legal Judgments*, in «Personality and Social Psychology Bulletin», 17, 541 ff.

Dworkin R. 1986. *Law's Empire*, Harvard University Press.

Eagly A.H., Ashmore R.D., Makhijani M.G., Longo L.C. 1991. *What Is Beautiful Is Good, but...: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype*, in «Psychological Bulletin», 110, 109 ff.

Efran M.G. 1974. *The Effect of Physical Appearance on the Judgment of Guilt, Interpersonal Attraction, and Severity of Recommended Punishment in a Simulated Jury Task*, in «Journal of Research in Personality», 8, 45 ff.

Ekman P.E., Davidson R.J. (eds.) 1994. *The Nature of Emotion: Fundamental Questions*, Oxford University Press.

Englich B., Mussweiler T., Strack F. 2006. *Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*, in «Personality and Social Psychology Bulletin», 32, 188 ff.

Epstein S. 1994. *Integration of the Cognitive and the Psychodynamic Unconscious*, in «American Psychologist», 49, 709 ff.

Ericsson K.A., Hoffman R.R., Kozbelt A., Williams A.M. (eds.) 2018. *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press.

Fazio L.K., Brashier N.M., Payne B.K., Marsh E.J. 2015. *Knowledge Does Not Protect Against Illusory Truth*, in «Journal of Experimental Psychology: General», 144, 993 ff.

Feigenson N. 2010. *Visual Evidence*, in «Psychonomic Bulletin and Review», 17, 149 ff.

Fenton N., Neil M., Berger D. 2016. *Bayes and the Law*, in «Annual Review of Statistics and Its Application», 3, 51 ff.

Festinger L. 1962. *Cognitive Dissonance*, in «Scientific American», 207, 93 ff.

Forgas J.P. 2013. *Don't Worry, Be Sad! On the Cognitive, Motivational, and Interpersonal Benefits of Negative Mood*, in «Current Directions in Psychological Science», 22, 225 ff.

Forgas J.P., East R. 2008. *On Being Happy and Gullible: Mood Effects on Skepticism and the Detection of Deception*, in «Journal of Experimental Social Psychology», 44, 1362 ff.

Frensch P.A., Sternberg R.J. 2014. *Expertise and Intelligent Thinking: When Is It Worse to Know Better*, in Sternberg R.J. (ed.), *Advances in the Psychology of Human Intelligence. Volume 5*, Psychology Press, 157 ff.

Gigerenzer G. 2006. *Bounded and Rational*, in Stainton R.J. (ed), *Contemporary Debates in Cognitive Science*, Blackwell, 115 ff.

Gigerenzer G., Gaissmaier W. 2011. *Heuristic Decision Making*, in «Annual Review of Psychology», 62, 451 ff.

GIGERENZER G., GOLDSTEIN D.G. 1999. *Betting on One Good Reason: The Take the Best Heuristic*, in GIGERENZER G., TODD P.M., ABC RESEARCH GROUP, *Simple Heuristics That Make Us Smart*, Oxford University Press, 75 ff.

GOLDING S.L., RORER L.G. 1972. *Illusory Correlation and Subjective Judgement*, in «Journal of Abnormal Psychology», 80, 249 ff.

GOLDMAN A.I. 1993. *The Psychology of Folk Psychology*, in «Behavioral and Brain Sciences», 16, 15 ff.

GOLDSTEIN D.G., GIGERENZER G. 2002. *Models of Ecological Rationality: The Recognition Heuristic*, in «Psychological Review», 109, 75 ff.

GOODMAN-DELAHUNTY J., GRANHAG P.A., HARTWIG M., LOFTUS E.F. 2010. *Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes*, in «Psychology, Public Policy, and Law», 16, 133 ff.

GUDJONSSON G.H. 1992. *The Psychology of Interrogations, Confessions and Testimony*, John Wiley & Sons.

GUILBAULT R.L., BRYANT F.B., BROCKWAY J.H., POSAVAC E.J. 2004. *A Meta-Analysis of Research on Hindsight Bias*, in «Basic and Applied Social Psychology», 26, 103 ff.

GUTHRIE C. 2006. *Law, Information, and Choice: Capitalizing on Heuristic Habits of Thought*, in GIGERENZER G., ENGEL C. (eds), *Heuristics and the Law*, MIT Press, 425 ff.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A.J. 2000. *Inside the Judicial Mind*, in «Cornell Law Review», 86, 777 ff.

GUTHRIE C., RACHLINSKI J.J., WISTRICH A.J. 2007. *Blinking on the Bench: How Judges Decide Cases*, in «Cornell Law Review», 93, 1 ff.

HAACK S. 1993. *Evidence and Inquiry: Towards Reconstruction in Epistemology*, Blackwell.

HAIDT J. 2001. *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, in «Psychological Review», 108, 814 ff.

HASHER L., GOLDSTEIN D., TOPPINO T. 1977. *Frequency and the Conference of Referential Validity*, in «Journal of Verbal Learning and Verbal Behavior», 16, 112 ff.

HEWSON L., GOODMAN-DELAHUNTY J. 2008. *Using Multimedia to Support Jury Understanding of DNA Profiling Evidence*, in «Australian Journal of Forensic Sciences», 40, 55 ff.

HILBERT M. 2012. *Toward a Synthesis of Cognitive Biases: How Noisy Information Processing Can Bias Human Decision Making*, in «Psychological Bulletin», 138, 211 ff.

HOSODA M., STONE-ROMERO E.F., COATS G. 2003. *The Effects of Physical Attractiveness on Job-Related Outcomes: A Meta-Analysis of Experimental Studies*, in «Personnel Psychology», 56, 431 ff.

JANIS I. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, Wadsworth Cengage Learning.

KADRI S. 2005. *The Trial: A History from Socrates to O.J. Simpson*, Random House.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

KOMARRAJU M., KARAU S.J., SCHMECK R.R., AVDIC A. 2011. *The Big Five Personality Traits, Learning Styles, and Academic Achievement*, in «Personality and Individual Differences», 51, 472 ff.

KORIAT A., LICHTENSTEIN S., FISCHHOFF B. 1980. *Reasons for Confidence*, in «Journal of Experimental Psychology: Human Learning and Memory», 6, 107 ff.

KOZUCH B., NICHOLS S. 2011. *Awareness of Unawareness: Folk Psychology and Introspective Transparency*, in «Journal of Consciousness Studies», 18, 135 ff.

KRAMER G.P., KERR N.L. 1989. *Laboratory Simulation and Bias in the Study of Juror Behavior: A Methodological Note*, in «Law and Human Behavior», 13, 89 ff.

Krizan Z., Windschitl P.D. 2009. *Wishful Thinking About the Future: Does Desire Impact Optimism?*, in «Social and Personality Psychology Compass», 3, 227 ff.

Kruglanski A.W., Webster D.M. 1996. *Motivated Closing of the Mind: 'Seizing' and 'Freezing'*, in «Psychological Review», 103, 263 ff.

Kucharzyk B. 2017. *Inadmissible Evidence Effect in the Context of Polish Law*, in Stelmach J., Brożek B., Kurek Ł. (eds), *The Province of Jurisprudence Naturalized*, Wolters Kluwer, 195 ff.

Kucharzyk B. 2021. *Swobodna ocena dowodów. Analiza interdyscyplinarna*, Krakowski Instytut Prawa Karnego Fundacja.

Kunda Z., Miller D.T., Claire T. 1990. *Combining Social Concepts: The Role of Causal Reasoning*, in «Cognitive Science», 14, 551 ff.

Kuran T., Sunstein C.R. 1999. *Availability Cascades and Risk Regulation*, in «Stanford Law Review», 51, 683 ff.

Langbein J.H. 2012. *Torture and the Law of Proof: Europe and England in the Ancien Régime*, University of Chicago Press.

Lavie N. 2010. *Attention, Distraction, and Cognitive Control Under Load*, in «Current Directions in Psychological Science», 19, 143 ff.

Lee S.W., Schwarz N. 2010. *Washing Away Postdecisional Dissonance*, in «Science», 328, 709 ff.

Leo R.A., Davis D. 2010. *From False Confession to Wrongful Conviction: Seven Psychological Processes*, in «Journal of Psychiatry and Law», 38, 9 ff.

Lewandowsky S., Kirsner K. 2000. *Knowledge Partitioning: Context-Dependent Use of Expertise*, in «Memory and Cognition», 28, 295 ff.

Lichtenstein S., Fischhoff B. 1977. *Do Those Who Know More Also Know More about How Much They Know?*, in «Organizational Behavior and Human Performance», 20, 159 ff.

Lichtenstein S., Slovic P., Fischhoff B., Layman M., Combs B. 1978. *Judged Frequency of Lethal Events*, in «Journal of Experimental Psychology: Human Learning and Memory», 4, 551 ff.

Lindsay R.C.L., Wells G.L., Rumpel C. 1981. *Can People Detect Eyewitness Identification Accuracy Within and Between Situations?*, in «Journal of Applied Psychology», 66, 79 ff.

Liu J.Z., Li X. 2019. *Legal Techniques for Rationalizing Biased Judicial Decisions: Evidence From Experiments With Real Judges*, in «Journal of Empirical Legal Studies», 16, 630 ff.

Loftus E.F. 1996. *Eyewitness Testimony*, Harvard University Press.

Loftus E.F. 2005. *Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory*, in «Learning & Memory», 12, 361 ff.

Loftus E.F. 2019. *Eyewitness Testimony*, in «Applied Cognitive Psychology», 33, 498 ff.

Marksteiner T., Ask K., Reinhard M.A., Granhag P.A. 2011. *Asymmetrical Scepticism Towards Criminal Evidence: The Role of Goal- and Belief-Consistency*, in «Applied Cognitive Psychology», 25, 541 ff.

Mazzella R., Feingold A. 1994. *The Effects of Physical Attractiveness, Race, Socioeconomic Status, and Gender of Defendants and Victims on Judgments of Mock Jurors: A Meta-Analysis*, in «Journal of Applied Social Psychology», 24, 1315 ff.

Memon A.A., Vrij A., Bull R. 2003. *Psychology and Law: Truthfulness, Accuracy and Credibility*, John Wiley & Sons.

Menkel-Meadow C.J 2000. *When Winning Isn't Everything: The Lawyer as Problem Solver*, in «Hofstra Law Review», 28, 905 ff.

Mischel W., Shoda Y. 1995. *A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure*, in «Psychological Review», 102, 246 ff.

Moore D.A., Healy P.J. 2008. *The Trouble with Overconfidence*, in «Psychological Review», 115, 502 ff.

Myers D.G., Lamm H. 1976. *The Group Polarization Phenomenon*, in «Psychological Bulletin», 83, 602 ff.

Newman E.J., Sanson M., Miller E.K., Quigley-McBride A., Foster J.L., Bernstein D.M., Garry M. 2014. *People with Easier to Pronounce Names Promote Truthiness of Claims*, in «PLoS ONE», 9, e88671.

Nickerson R.S. 1998. *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, in «Review of General Psychology», 2, 175 ff.

Nisbett R.E., Wilson T.D. 1977a. *The Halo Effect: Evidence for Unconscious Alteration of Judgments*, in «Journal of Personality and Social Psychology», 35, 250 ff.

Nisbett R.E., Wilson T.D. 1977b. *Telling More Than We Can Know: Verbal Reports on Mental Processes*, in «Psychological Review», 84, 231 ff.

Nobles R., Schiff D. 2005. *Misleading Statistics Within Criminal Trials: The Sally Clark Case*, in «Significance», 2, 17 ff.

Northcraft G.B., Neale M.A. 1987. *Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions*, in «Organizational Behavior and Human Decision Processes», 39, 84 ff.

Oeberst A., Goeckenjan I. 2016. *When Being Wise After the Event Results in Injustice: Evidence for Hindsight Bias in Judges' Negligence Assessments*, in «Psychology, Public Policy, And Law», 22, 271 ff.

Ozubko J.D., Fugelsang J. 2011. *Remembering Makes Evidence Compelling: Retrieval from Memory Can Give Rise to the Illusion of Truth*, in «Journal of Experimental Psychology: Learning, Memory, and Cognition», 37, 270 ff.

Pennington N., Hastie R. 1988. *Explanation-Based Decision Making: Effects of Memory Structure on Judgment*, in «Journal of Experimental Psychology: Learning, Memory, and Cognition», 14, 521 ff.

Pennington N., Hastie R. 1990. *Practical Implications of Psychological Research on Juror and Jury Decision Making*, in «Personality and Social Psychology Bulletin», 16, 90 ff.

Penrod S., Cutler B. 1995. *Witness Confidence and Witness Accuracy: Assessing Their Forensic Relation*, in «Psychology, Public Policy, and Law», 1, 817 ff.

Piperides C. (Rapporteur), Allen R.J., Dhami M.K., Flessner A., Hastie R., Koehler J.J., Lempert R., Schulz J., Wagner G. 2006. *Group Report: What Is the Role of Heuristics in Litigation?*, in Gigerenzer G., Engel C. (eds), *Heuristics and the Law*, MIT Press, 343 ff.

Polage D.C. 2012. *Making up History: False Memories of Fake News Stories*, in «Europe's Journal of Psychology», 8, 245 ff.

Pornpitakpan C. 2004. *Psychological Factors in Evaluation of Legal Evidence. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence*, in «Journal of Applied Social Psychology», 34, 243 ff.

Porter S., ten Brinke L. 2009. *Dangerous Decisions: A Theoretical Framework for Understanding How Judges Assess Credibility in the Courtroom*, in «Legal and Criminological Psychology», 14, 119 ff.

Posner M.I., Rothbart M.K., Tang Y.Y. 2015. *Enhancing Attention Through Training*, in «Current Opinion in Behavioral Sciences», 4, 1 ff.

Posner R.A. 2010. *Some Realism About Judges: A Reply to Edwards and Livermore*, in «Duke Law Journal», 59, 1177 ff.

Rachlinski J.J. 2006. *Bottom-Up Versus Top-Down Lawmaking*, in Gigerenzer G., Engel C. (eds), *Heuristics and the Law*, MIT Press, 159 ff.

Rachlinski J.J., Wistrich A.J. 2017. *Judging the Judiciary by the Numbers: Empirical Research on Judges*, in «Annual Review of Law and Social Science», 13, 203 ff.

Roese N.J., Vohs K.D. 2012. *Hindsight Bias*, in «Perspectives on Psychological Science», 7, 411 ff.

Salerno J.M., Bottoms B.L. 2009. *Emotional Evidence and Jurors' Judgments: The Promise of Neuroscience for Informing Psychology and Law*, in «Behavioral Sciences and the Law», 27, 273 ff.

Schnall S., Haidt J., Clore G.L., Jordan A.H. 2008. *Disgust as Embodied Moral Judgment*, in «Personality and Social Psychology Bulletin», 34, 1096 ff.

Sigall H., Ostrove N. 1975. *Beautiful but Dangerous: Effects of Offender Attractiveness and Nature of the Crime on Juridical Judgement*, in «Journal of Personality and Social Psychology», 31, 410 ff.

Slovic P., Finucane M.L., Peters E., MacGregor D.G. 2007. *The Affect Heuristic*, in «European Journal of Operational Research», 177, 1333 ff.

Smalarz L., Madon S., Yang Y., Guyll M., Buck S. 2016. *The Perfect Match: Do Criminal Stereotypes Bias Forensic Evidence Analysis?*, in «Law and Human Behavior», 40, 420 ff.

Smith V.L., Kassin S.M., Ellsworth P.C. 1989. *Eyewitness Accuracy and Confidence: Within- Versus Between- Subjects Correlations*, in «Journal of Applied Psychology», 74, 356 ff.

Spellman B.A. 2007. *On the Supposed Expertise of Judges in Evaluating Evidence*, in «University of Pennsylvania Law Review», 156, 1 ff.

Stahlberg D., Maass A. 1997. *Hindsight Bias: Impaired Memory or Biased Reconstruction?*, in «European Review of Social Psychology», 8, 105 ff.

Stanovich K.E. 2009. *The Cognitive Miser: Ways to Avoid Thinking*, in Id., *What Intelligence Tests Miss: The Psychology of Rational Thought*, Yale University Press, 70 ff.

Stanovich K.E., West R.F. 2000. *Individual Differences in Reasoning: Implications for the Rationality Debate?*, in «Behavioral and Brain Sciences», 23, 645 ff.

Stewart J.E. 1980. *Defendant's Attractiveness as a Factor in the Outcome of Criminal Trials: An Observational Study*, in «Journal of Applied Social Psychology», 10, 1980, 348 ff.

Thagard P. 2008. *Hot Thought: Mechanisms and Applications of Emotional Cognition*, MIT Press.

Thompson W.C., Schumann E.L. 1987. *Interpretation of Statistical Evidence in Criminal Trials*, in «Law and Human Behavior», 11, 167 ff.

Thompson W.C., Taroni F., Aitken C.G.G. 2003. *How the Probability of a False Positive Affects the Value of DNA Evidence*, in «Journal of Forensic Science», 48, 1 ff.

Todd P.M., Gigerenzer G. 2012. *Ecological Rationality: Intelligence in the World*, Oxford University Press.

Tversky A., Kahneman D. 1974. *Judgment Under Uncertainty: Heuristics and Biases*, in «Science», 185, 1124 ff.

Tversky A., Kahneman D. 1982. *Evidential Impact of Base Rates*, in Kahneman D., Slovic P., Tversky A. (eds), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, 153 ff.

Twining W. 1997. *Freedom of Proof and the Reform of Criminal Evidence*, in «Israel Law Review», 31, 439 ff.

Van Hiel A., Mervielde I. 2003. *The Need for Closure and the Spontaneous Use of Complex and Simple Cognitive Structures*, in «Journal of Social Psychology», 143, 559 ff.

Verhulst B., Lodge M., Lavine H. 2010. *The Attractiveness Halo: Why Some Candidates Are Perceived More Favorably Than Others*, in «Journal of Nonverbal Behavior», 34, 111 ff.

Von Hippel W., Trivers R. 2011. *The Evolution and Psychology of Self-Deception*, in «Behavioral and Brain Sciences», 34, 1 ff.

Webster D.M., Kruglanski A.W. 1994. *Individual Differences in Need for Cognitive Closure*, in «Journal of Personality and Social Psychology», 67, 1049 ff.

Wells G.L. 1992. *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, in «Journal of Personality and Social Psychology», 62, 739 ff.

Wixted J.T., Wells G.L. 2017. *The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis*, in «Psychological Science in the Public Interest», 18, 10 ff.

Yourstone J., Lindholm T., Grann M., Svenson O. 2008. *Evidence of Gender Bias in Legal Insanity Evaluations: A Case Vignette Study of Clinicians, Judges and Students*, in «Nordic Journal of Psychiatry», 62, 273 ff.

Reasoning about Forensic Science Evidence

BARBARA A. SPELLMAN, ADELE QUIGLEY-MCBRIDE

Introduction

Forensic science evidence has been both a blessing and a curse to the legal system for the past century. In a "perfect world" for investigators, forensic evidence could be used to quickly and easily determine what happened at a scene (e.g., was there even a crime and, if so, what was its nature?). It would also be ideal if forensic evidence could help rapidly determine who was and was not to blame. This "perfect world" would feature judges who understand the reliability and value of different types of forensic evidence and, as a result, appropriately allow (or disallow) that evidence to be used to apprehend or convict someone. Lawyers would understand the weaknesses of forensic evidence and question expert witnesses in a way that made both the flaws and strengths of the evidence clear to factfinders during trials. Finally, factfinders—professional judges, lay judges, or random citizens—would learn during trial how to evaluate the forensic evidence properly and use it to render a verdict.

Of course, we do not live in the world described in the above story, nor are we perfect reasoners. Some of the faults of forensic evidence (e.g., it may contribute, in several ways, to the arrest and conviction of people who are factually innocent) lie in the science underlying the forensic disciplines. But it is mostly us—the people who collect, analyze, interpret, communicate, question, evaluate, and use such evidence—who are responsible for its misuse (Morgan 2023). Each of those tasks involves "the human element" (Dror 2015) and, therefore, is a fertile ground for cognitive science study.

This chapter is in three parts. The first explains some of what forensic analysts do, drawing on cognitive science to understand their reasoning processes and how they, or the systems in which they work, can produce faulty judgments. The second part lists some of what people (specifically non-scientists, e.g., judges, lawyers, random jurors) believe—accurately or mistakenly—about forensic science, and how those beliefs arise. The third part describes occasions in which forensic analysts and legal decision makers cross paths in the legal system,

such as when judges rule on the admissibility of forensic evidence, when lawyers question forensic experts, and when jurors interpret forensic testimony. Although the chapter is framed in terms of the US jury-based adversarial system, the analyses should be relevant across systems, even when different parties are responsible for the relevant tasks.

1. *The cognitive science underlying forensic science*

To understand how to approach forensic science evidence from a legal perspective, it is important to understand how forensic science evidence is constructed. We don't mean how the evidence is left by a relevant participant or observer involved in the case; rather, we mean how it is found, analyzed, and interpreted by forensic scientists in order to produce evidence that can be used by investigators or at trial. Because forensics analysts are engaged in tasks that involve reasoning, judgment, decision-making, memory, communication, and other cognitive skills, cognitive science can provide insights into how evidence selection and interpretation can go right—or go wrong.

Cognitive science has established that reasoning processes, most of the time, lead us to the right conclusions, but those same processes can also lead us astray. Part 1 of this chapter moves between cognitive science and forensic science, describing the limits of forensic science knowledge, some general characteristics of human reasoning, what forensic analysts do, and, importantly, how general characteristics of human reasoning are not necessarily optimal when reasoning in forensic science, or any science, contexts. This core information will help lawyers and judges to understand, elicit, present, question, and evaluate forensic science.

1.1. *Scientific foundations of the forensic sciences*

In the 1980's, forensic science was a booming business. Law enforcement believed that it provided reliable ways to identify people by analyzing fingerprints, voices, hair, footprints, handwriting, and bitemarks. They believed that they could use fibers and tire tracks to narrow down the list of potential suspects and could determine how and where a fire started by examining fire debris. About the same time, advances in DNA processing allowed defendants to request that their cases be revisited, eventually revealing that mistaken eyewitness testimony, false confessions, and bad forensic science analyses were all implicated in hundreds of wrongful convictions.

But is an error in forensics always the fault of forensic analysts or, sometimes, could it be in the science itself? Forensic scientists have claimed that a sample of DNA, a fingerprint, or a bitemark impression picked up at a crime scene can be compared to a known sample taken from a suspect and if the two "match" it is guaranteed that the samples came from the same person. Disregarding the possibility of human error (e.g., in labeling the samples)—could such a guarantee be made?

In 2009, the National Research Council published a report called *Strengthening Forensic Science in the United States: A Path Forward* (hereafter, "NRC Report"). It investigated the scientific foundations of many forensic techniques and found all lacking, with the exception of single-source DNA evidence. DNA analysis was invented by scientists, not for forensic use, and has fairly robust scientific underpinnings—it is very likely that every person (except for "identical" siblings) has distinct DNA and there are techniques that enable analysts to distinguish between DNA from different people. But what about fingerprints? There is little to guarantee their uniqueness. And even if every fingerprint were unique, it may be that not all similarities and differences would show up when fingerprint impressions are taken; every new impression from the same finger differs at least slightly from prior impressions. So, how can analysts be certain that slightly different fingerprints are similar enough to each other to say they are from the same person or different enough to say they are not?

Yet fingerprint analysis has more empirical support than most other forensic techniques. Consider, for instance, bitemarks—a technique that most people have heard of but that has no scientific foundation. To leave an impression, the bite must be made into something soft, like food or skin, and such things immediately begin to lose the impression and rebound towards their original shape (see SAUERWEIN et al. 2023). Overall, the NRC Report showed that the reliability of the forensic sciences was unknown and was likely substantially lower than anyone wanted to believe.

1.2. *Three important characteristics of human reasoning*[1]

Human reasoning processes are complicated. Sadly, we are so often disappointed in its failures that we forget how well our reasoning processes usually work. We are able to navigate, communicate, and interact within a complex world full of people, objects, and challenges, with few untoward results. The three general characteristics of reasoning described below are common and useful in our complex everyday life, but all can create trouble for forensic analysts.

1.2.1. *Perceptions are shaped by pre-existing knowledge and beliefs*

One ubiquitous characteristic of reasoning is that we believe that our perceptions of the world are direct and accurate representations of what is in the world ("naïve realism"). However, that assumption is incomplete: our perceptions and beliefs are shaped both by what is in the world and by what is in our heads, for example, previous knowledge, experiences, expectations, and desires. (See Figure 1.) Combining information from these two sources is done so automatically that we are not usually aware that our interpretation of what exists depends at all on what we already know or believe.

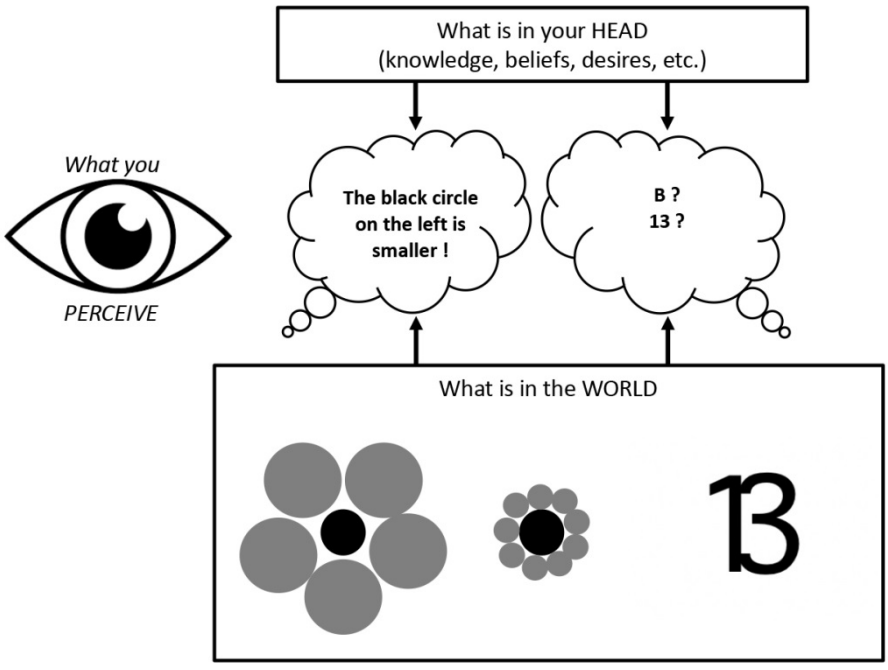


FIGURE 1.

[1] These topics appear in most Cognitive Psychology textbooks. More detail as related to forensic science can be found in SPELLMAN et al. 2022; DROR 2020; EDMONDS et al. 2017.

Consider the familiar visual illusion in the left of Figure 1. Although the two black circles are the same size, our visual system is automatically incorporating the surrounding context and comparing the nearby circles to them it such that the circle surrounded by the smaller ones looks larger than the circle surrounded by the bigger ones. The figure on the right illustrates how motivation or desire might affect perception. Some participants in a study were told to respond when they saw a number, others when they saw a letter; they would be rewarded for each one they got correct. Most of the stimuli were clearly either a number or a letter and participants generally answered correctly. However, participants in both groups responded when this ambiguous figure appeared (BALCETIS & DUNNING 2006).

1.2.2. *People create abstract knowledge structures including causal stories*

A second strong characteristic of human reasoning is that we create abstract knowledge structures: we automatically see patterns in our environment, and we group information into mental structures like categories and causal stories, which help us remember and make sense of the world. A simple example of causal story creation are inferences from one observation based on previous knowledge; for example, if you look outside and see someone carrying an umbrella, you infer that it is likely to rain, because you think that the only reason for the person to carry the umbrella is that they believe it will rain. On the other hand, if the sky were clear and the forecast were for sun, you might revise your hypothesis about whether the covered stick the person carried was actually an umbrella (maybe it's a parasol?), or perhaps consider that the person was traveling to somewhere that the forecast for rain.

Research shows that people tend to create stories with the most parsimonious (i.e., the simplest) explanations (READ & MARCUS-NEWHALL 1993) but also that once people have a story fixed in their mind, they will be likely to ignore or devalue disconfirming information (i.e., "confirmation bias"; see Part 1.4.2 below). Studies using stimuli based on real legal cases show that experimental participants who are pretending to act as jurors (typically from the US) will fit the facts they learn from testimony into a coherent story, and will discount some testimony (e.g., reports about the timing of events, guesses about the motivation of actors) to make their versions of the story more coherent (PENNINGTON & HASTIE 1991).

Creating causal stories turns out to be an important undertaking within the legal system: it is a task for investigators, who are trying to solve how a crime occurred; for lawyers, who are trying to convince others that their version of the story is the correct one; and for the judges or jurors who, after listening to testimony containing inconsistent facts, and arguments containing competing stories, must construct their own story of what did, or did not, happen, in order to come to a verdict.

Note that for amusement, people all over the world watch lots of examples of unrealistic causal story creation in television shows, where for the first 45 minutes the investigators (police or CSI or FBI) find wrong suspect after wrong suspect until, with a minute to spare, everything falls into place.

1.2.3. *Explaining (some) irrationality: "Two systems" of reasoning*

The third characteristic of human reasoning goes by many names. The gist is that people behave as if they have two different ways of learning, processing, and using information. Studies show that people often jump to conclusions that seem "obvious" in light of the available information, only to realize later that, if they had taken more time, they would have come to a different (and correct) answer (FREDERICK 2005).

Qualities that are often attributed to these two different types of processing are shown in Table 1. The Nobel Prize winner Danny Kahneman titled his popular book, *Thinking, Fast and Slow*, after this phenomenon. Part 2 of this chapter describes how people may unintentionally

learn information using the non-aware/conscious distinction and how people may be persuaded using the peripheral/central route distinction of PETTY and CACIOPPO (2012).

System 1 (Fast)	System 2 (Slow)
Automatic	Controlled
Non-aware	Conscious
Intuitive	Reflective
Heuristic	Logical
Peripheral Route	Central Route

TABLE 1. Characteristics of the fast and slow systems (adapted from KAHNEMAN 2011.)

Which type of processing people use at any given time depends on both the person and the situation. When people are motivated to think hard about something, they are more likely to use the slow system than if they are not motivated; and when people are under time pressure or stress, they are more likely to use the fast system than if they are not under such pressure.

1.3. *What forensic analysts do and where they do it*[2]

Forensic analysts portrayed in the media are usually collecting the evidence at the crime scene and performing a variety of different analyses in a forensic lab (e.g., deciding whether DNA or fingerprints come from a specific person, or whether a white powder is a certain kind of drug). In some shows, these people are also the arresting officers. In reality, forensic analysts are much more specialized and limited. Most forensic labs are affiliated with a local police department and do not cover all the forensic domains. Labs typically have many of the core "feature comparison" methods such as DNA, fingerprints, and firearms. The PCAST report (2016),[3] following up on the NRC Report (2009), addresses these methods and defines a feature comparison procedure as one in which an examiner «seeks to determine whether an evidentiary sample (e.g., from a crime scene) is or is not associated with a source sample (e.g., from a suspect) based on similar features» (46). Labs also often have analysts who test and identify whether trace evidence is a specific poison ("toxicology") or controlled substance ("seized drugs").

Forensic analyses also occur in medical examiners officers (e.g., an autopsy to determine time and cause of death). Some forensic analyses are even performed at a suspected crime scene itself because the relevant evidence cannot be fully or easily moved to a laboratory environment. For example, forensic analysts who are trying to figure out how a fire started in a building, or what actions created an unusual blood spatter pattern across a floor and wall, will return to a scene after other evidence is catalogued. And, of course, crime scene investigators are responsible for going to the scene of what might, or might not, be a crime, and deciding what is relevant evidence at the scene, so that it can be documented and preserved using photography, collection methods, or *in situ* analysis.

[2] Note that the term "forensic psychologist" refers to people with psychological training who evaluate individuals as to their mental abilities, state, and health. These may include competence to stand trial, potential danger to self or others, possibility of prior temporary insanity, experience of long-term mental illness, or other assessments. Forensic psychologists are not usually included as doing "forensic science" (e.g., forensic psychology is not addressed in the NRC Report), but many issues affecting the accuracy of their testimony are similar to issues affecting other forensic scientists and testifying experts generally.

[3] The PCAST report was written by the President's Council of Advisors on Science and Technology and entitled: *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (2016).

1.4. *Three dangers to forensic science reasoning*

The characteristics of human reasoning described above are useful in that they usually guide us towards correct or safe outcomes in everyday life. Understanding how these "normal reasoning" processes can also lead people to make bad judgments is important for understanding how forensic science, or forensic analysts, can go wrong.

1.4.1. *Exposure to task-irrelevant biasing influences*

As described in Part 1.2 above, humans take information from different sources and combine it automatically. But that is *not* what a forensic analyst is supposed to do with case information. For example, when deciding whether a fingerprint "matches" the suspect's print, the analyst should not know things like whether the suspect has prior convictions or whether he has confessed. Such information is not relevant to the process of comparing the features in those fingerprints. Yet dozens of studies show that exposure to "task irrelevant" information can improperly influence the decisions of real analysts (GARRETT 2022; KUKUCKA & DROR 2023). Forensic analysts are supposed to provide independent and accurate evaluations of the specific evidence they have been tasked with examining. The task of putting together and evaluating all the different pieces of case information to determine what happened at a crime scene and who was responsible is the job of the factfinder, not the forensic analyst. If an analyst has been biased by some other evidence or information (e.g., knowledge of a confession) (KASSIN et al. 2013), that information will become entwined with the forensic analyst's conclusion and, thus, incorporated into jurors' assessment of the case, even if the jurors have heard the confession and decided it is not believable, or have already considered that evidence (amounting to double counting).

Two techniques that are becoming more common for countering potential bias are *blinding*—keeping task-irrelevant information away from analysts—and *verification*—having a second analyst check the work of the first. Blinding can be done at the lab and case level. Labs can employ "case managers" who keep the paperwork and all other information about a case away from the analysts. Verification means that after an analyst reaches a conclusion, another analyst checks whether they believe that the conclusion is correct. Although these techniques seem simple, they are not free from potential biases, nor are they free from other problems. For example, in small forensic laboratories, it is difficult to keep tasks and information separated across people—for example, a verifier might know who the initial analyst was and be biased by that when verifying, and analysts who must also collect evidence will unintentionally gather task-irrelevant information during that process.

1.4.2. *Closing off hypotheses: Motivated reasoning and confirmation bias*

An important task for analysts, especially, but not only, those who are looking for causal processes (e.g., crime scene investigators, fire analysts; also jurors) is to keep an open mind regarding potential hypotheses for what may have happened. Jumping to conclusions is a consequence of the natural tendency to automatically combine information. People piece information together, try to find a way to explain everything, and once that has been achieved, are satisfied. But reaching a conclusion too quickly can have serious consequences in some settings. For example, in police investigations, sometimes early evidence suggests a particular suspect, then police choose to pursue only that suspect, and they end up ignoring evidence leading to other suspects. This conduct is known as "tunnel vision" and has led to wrongful arrests and wrongful convictions.

In reasoning, humans are motivated to be accurate and consistent but also, sometimes, to arrive at a particular conclusion; together these desires create what is commonly called "motivated reasoning" (KUNDA 1990). One consequence of motivated reasoning is "confirmation bias", the

cognitive scientists' name for "tunnel vision". Confirmation bias is when someone has a hypothesis that they believe is true, or that they want to be true, so they are motivated to keep it viable. They might seek out only information that confirms it and interpret any ambiguous evidence in a way that is consistent with their hypothesis. When confronted with disconfirming evidence, they may discount it or explain it away as somehow irrelevant, bad, or corrupted. Similarly, forensic analysts should not jump to conclusions. Rather, it is advisable to keep alternative hypotheses open for investigative purposes and for conveying the appropriate degree of certainty to other persons (e.g., factfinders at trial).

1.4.3. *Failures of metacognition: Overconfidence and bias blind spot*

Metacognition is thinking about the processes of other peoples', or ones' own, cognition. One common metacognitive bias, at least in Western samples, is overconfidence: people are often overconfident about their own abilities and tend to believe that their own arguments are stronger and more persuasive than others. More relevant here is the "bias blind spot," which refers to the finding that people believe that others are more biased than they are themselves (PRONIN 2007). When disagreeing with other people, we might believe that they have been influenced by stray information, or illogical reasoning, or confirmation bias, but people fail to see themselves falling prey to the same errors. However, people are poor introspective investigators—we are not conscious of everything we know or think, or how something might be affecting our decisions. For a forensic analyst, overconfidence in a conclusion, or a failure to recognize that one's judgment may have been improperly influenced, can have unfortunate consequences for justice.

1.5. *General forensic situations*

The accuracy of forensic output depends on the situation in which analysts are working as well as the quality of the forensic analyst's reasoning. Laboratories may be the source of biasing information; for example, analysts working in a "police lab" might be more motivated to support the suspicions of the police than analysts working in an independent lab. Analysts might see biasing information in case logs, on envelopes, or on bags containing evidence. These problems can be ameliorated by having "case managers" who provide a buffer between the analysts' work and task-irrelevant information. In addition to these cognitive bias problems, analysts might be under extra pressure in a high-profile case, generally working in a high-pressure laboratory or a lab with substantial backlog, or dealing with some temporary personal hardship. Such stressors can push people into rushing their decisions rather than carefully analyzing the evidence, which could easily result in more errors (BUSEY et al. 2022).

2. *What factfinders already believe about forensic evidence*

Part 1 of this chapter describes some normal human reasoning processes, how they are involved in forensic science decision making, and what might go wrong when forensic evidence is analyzed and reported. Part 3 below describes whether and how forensic science evidence can be conveyed to factfinders so that the correct amount of weight is given to the information. Because people automatically use their pre-existing knowledge and beliefs to understand and interpret new information, we pause here in Part 2 to describe what ordinary people with no formal scientific or legal education already know and believe about forensic science and forensic analysts.

Through television, movies, the news, social media, and the occasional crime novel, forensic science is now very familiar to the general population—that is, to non-scientists and non-lawyers. But this way of learning exposes people to a plethora of exaggerated, misleading, and incorrect

information. Most people do not have the educational background necessary to critique complicated scientific results (KUHN 1989); thus, the general population is not equipped to effectively critique expert forensic testimony at the level required to identify the kinds of issues described in Part 1 above. As a result, the general population holds many misconceptions about forensic science, including believing that forensic evidence is objective, scientific, generally accurate and reliable, readily available at crime scenes, commonly presented in criminal cases, and able to shed light on the truth about what happened during a crime on its own (GARRETT 2022)[4].

2.1. *How people come to "know" about forensic evidence*

2.1.1. *Exposure to media*

As with most topics, people form their beliefs about forensic science using the information that is readily available to them—news media, television shows and movies, and social media platforms (HOUCK 2006). Importantly, legal professionals hold similar misconceptions about forensic science, even though their chosen career increases their likelihood of encountering forensic evidence and associated forensic analyses.

The sources of knowledge through which the public learns about forensic evidence can include accurate information. But, overwhelmingly, television shows, movies, and social media posts are not created to disseminate accurate information. Dramatic fictionalized accounts of how forensic evidence can be found, processed, and used, results in more viewers and more "likes" than descriptions of reality would receive. Yet although many of these productions are clearly fiction, they usually contain elements of truth (e.g., some aspects of legal procedure and the role of law enforcement are accurate, as are some pieces of information about forensics). Thus, it is not surprising that the average, non-expert consumer has trouble distinguishing what is true from what is not with regard to the forensic science in these shows (COLE 2015).

News outlets and documentaries can be misleading as well. News outlets typically do not employ people with advanced scientific backgrounds. Some news outlets research the underlying science so that they can accurately describe the methods and results in their articles and news segments, but this is not the norm. Most are less cautious with their language for various reasons (e.g., attracting viewers is a higher priority than ensuring their reporting accurately represents the science) or are simply unaware that they are portraying the science incorrectly. Furthermore, the stories that are "news-worthy" or entertaining enough to be reported or turned into a film or documentary do not reflect the everyday cases appearing in most trials. For most typical cases, there is no physical evidence at all, which is in stark contrast to what is observed in fictional shows like *CSI*, or some of the real trials that receive the most media exposure.

Thus, the information about forensic science that will typically come to mind when ordinary people call on their knowledge will be based on unreliable, exaggerated sources. Regardless of the exact source of any one person's knowledge about forensic evidence, the available sources are usually designed to prioritize entertainment rather than education.

2.1.2. *Creating "knowledge" through heuristic processes*

People can come to "know" something in two main ways—consciously and without awareness (see Table 1). Conscious learning occurs when a person is motivated to learn about a topic. When a person is motivated, they may purposely try to commit the information to memory and seek to

[4] This book describes the various sources of error that can occur in the forensic sciences, the reasons those errors persist in criminal investigations and trials, and the consequences of these failures for the accused, the criminal justice system, and beyond.

understand the underlying reasons that the information they are learning is correct. This is how people prepare for an exam or gather knowledge that they care about (e.g., researching pre-natal care when trying to get pregnant, reading cookbooks and watching cooking shows when attempting to make tastier food). Most people have not learned what they know about forensic science in this way—they have never been tested on the topic nor intrinsically motivated to ensure that they obtain and remember accurate information about forensic evidence and analysis.

Most learning is incidental. Without knowing it, people accumulate information and notice patterns in their experiences for use in future decisions and judgments. Consider that most people will have a view about global warming, the death penalty, or other controversial topics, but have never actively studied those topics. When people express views that have been formed outside of their awareness, they are drawing on information obtained through heuristic processes (KAHNEMAN 2011). In other words, their view is formed using information obtained from a variety of sources—information they have encountered often, information from people whom they like or respect, and information that is shocking or stood out. However, hearing something frequently, the likeability of the source, or the extreme nature of the information are not good proxies for accuracy. Gathering and using information in this way requires very few cognitive resources, but when inaccurate decisions or judgments have serious consequences, these are not the best sources of information for people to rely on.

2.1.3. *Changing knowledge through persuasion*

Is it possible to change someone's beliefs about forensic evidence? Persuading someone away from their original opinion can be achieved using the same processes that created that opinion in the first place. If people are motivated and paying attention, they can be persuaded by the quality of the information, how well the information is communicated, and the extent to which the reasoning is sound (CHAIKEN & MAHESWARAN 1994; PETTY & CACIOPPO 1981). People can be motivated when the topic is relevant to them, if they believe that they can learn and understand the topic, or if they are generally curious and like to seek out learning and enjoy cognitively engaging tasks.

When people are not motivated, they can be persuaded by more superficial features of the information and how it is communicated. Such features include peripheral or irrelevant information, the perceived level of expertise of the communicator, likeability or attractiveness of the communicator, how easily it feels to process the information, and how consistent the information is with their existing knowledge or beliefs (SPELLMAN & TENNEY 2010). When reading or watching the news, watching crime television shows, or enjoying a movie, most people are not trying to assess the quality of the information and are, instead, influenced by these more superficial aspects of the content that infiltrate our beliefs and opinions largely outside of our awareness.

2.2. *Common misconceptions about forensic evidence*

Given the unreliability of the sources from which most people learn about forensic science, it is no wonder that they have many misconceptions about forensic science processes and conclusions.

2.2.1. *Misconception 1: "Forensic analysis is objective and was created by scientists"*

Ordinary people tend to be impressed with science and scientists and believe that they act in the public interest. Moreover, to some degree, a deep appreciation of many sciences requires some scientific knowledge (FUNK et al. 2019). Because forensic analysis is described as a science, forensic techniques have become synonymous with other sciences to those interpreting the

forensic results and opinions. As a result, lay persons perceive forensic science as highly objective—neutral, impartial, and free from the influence of opinions, bias, and emotions.

However, except for DNA evidence, forensic techniques were not developed using scientific testing, methods, and validation; they were not even developed by scientists (NRC 2009). Instead, forensic disciplines were largely created by law enforcement and legal professionals for the purpose of finding suspects and providing evidence to incriminate those individuals. As a result, there are entire disciplines that were built on non-scientific foundations, and remain non-scientific in many ways (GARRETT 2022; see Part 1.1 above).

A related misconception is the belief that there is technology that analysts rely on to draw their conclusions about the evidence. This is true, to an extent. For fingerprints and DNA there are database-search algorithms that can produce candidates in cases where there is no suspect (e.g., Automatic Fingerprint Identification Systems (*AFIS*), the Combined DNA Index System (*CODIS*)). There is software that helps with DNA analysis—particularly when a DNA mixture is found at a crime scene. There are microscopes and image enhancement software that can improve analysts' ability to look at the details of evidence, and programs designed to digitize their analyses and observations. Ultimately, though, it is a person who makes a subjective judgment to reach a conclusion (DROR & MNOOKIN, 2010)[5].

2.2.2. *Misconception 2: "Forensic results are very rarely inaccurate"*

Many individuals believe that the error rate in forensic analysis is negligible or very low (MARTIRE et al. 2019). However, despite the call for research on error rates in the NRC report, for most forensic techniques there is no known error rate or empirical foundation for establishing accuracy and reliability (GARRETT 2022). Assessing error rates requires knowledge of ground truth, which is unknown in real cases. And analysts will almost never receive feedback about whether they made a correct judgment in real cases. In fact, forensic analysts can become more confident in their abilities because their work results in convictions. When the defendant they identified using their forensic test is convicted, they reason that their analysis must have been correct, not recognizing that their test results helped cause the conviction (i.e., using circular reasoning) (SPELLMAN et al. 2021).

There are a few ways to collect relevant information about accuracy and error rates. Forensic labs can test the accuracy of their output by introducing simulated evidence that mimics real cases, but for which ground truth is known, such as "black box" studies. These studies assess outcomes only, and aggregate error rates within a lab or larger group, but not the processes leading to outcomes (hence, "black box"). Blind proficiency tests are a measure of accuracy focused on the error associated with individual analysts. Such tests involve embedding fake but realistic cases, for which ground truth is known, into analysts' workflow and assessing the accuracy of their determinations.

When done well, "black box" and "blind proficiency" studies produce more precise error rates than classic proficiency tests, when analysts know they are being tested, which are notoriously easy. Understandably, laboratories want to avoid putting analysts in a situation where they must admit to having made errors in the past on the witness stand (ELDRIDGE et al. 2022; see Part 3.3.1 below for further discussion of the impact of providing error rates when giving testimony). However, there are some serious problems associated with "black box" studies and "blind proficiency" tests too.

Setting up such simulations—from evidence receipt through reporting conclusion—is logistically difficult (MEJIA et al. 2020). Analysts can often tell when they receive a simulated

[5] Note that technology can introduce other problems into the decision process. As recent research in many fields shows, big data can introduce biasing effects; see DROR et al. 2012 for a discussion of the biasing effects of AFIS.

case, which leads to more conservative responding that reduces the rate of false positive errors—the kinds of errors that can mistakenly incriminate someone. Plus, the samples are often more pristine than anything typically obtained from a crime scene and, thus, overestimate true performance in casework (GARDNER et al. 2020; KOERTNER & SWOFFORD 2018). There are also issues associated with the analyses and labelling of outcomes in error rate studies (BIEDERMANN & KOTSOGLOU 2021). As a result, when analysts are asked on the witness stand about error rates, they are either providing a subjective judgment about error in the absence of these studies, or they are referring to studies that are not representative of real casework.

2.2.3. *Misconception 3: "Forensic science reliability does not vary much across types of evidence"*

Not all forensic sciences are equal in terms of their reliability. Single source DNA profile comparisons tend to be very accurate, and the error rate for such analyses can be estimated using published empirical research. Other kinds of DNA analyses, such as Mitochondrial DNA analyses and DNA mixture analyses, are far less reliable, more difficult for the analyst, and require more subjective judgments from the analyst. Although forensic scientists and science educated people may understand the different between these two kinds of DNA evidence, lay persons typically do not and, thus, ascribe the high level of accuracy to those less reliable forms of evidence (CHIN & IBAVIOSA 2022; GARRETT 2022). The differences in reliability across forensic domains can be attributed to the foundational science in the field, the methods and procedures used, error rate studies, and studies examining the role of these forensic techniques in wrongful conviction cases (MORGAN 2023). Fingerprint analysis is believed to be fairly reliable and has a stronger research base and rigorous methodology than most other forensic sciences like bitemarks, shoeprints, voice analysis, or canine detection teams.

Regardless of the type of evidence, lay persons tend to overestimate the reliability of the discipline and underestimate the error rate (MARTIRE et al. 2019). Although lay persons can typically understand that some disciplines are relatively more accurate or reliable than others, they do not adequately adjust their perception of the evidence (LIEBERMAN et al. 2008). For instance, highly questionable evidence like bitemarks and hair analysis are considered much more reliable by lay persons than research suggests they are, yielding a larger discrepancy between their views and the actual credibility than seen for more credible evidence types (e.g., DNA, fingerprints).

2.2.4. *Misconception 4: "Forensic evidence is commonly available in criminal cases"*

As forensic evidence became more frequently used in criminal cases, and more frequently reported in the related news and entertainment media, people began to think that it was the norm for criminal cases to involve lots of forensic evidence (i.e., a result of the availability heuristic) (TVERSKY & KAHNEMAN 1974). Lawyers became nervous about a "CSI Effect" (named after the television show)—but they and the researchers who investigated the effect could not agree on the nature of the effect. Some researchers and lawyers (typically prosecutors) believed that jurors would expect every case to include forensic evidence, and if the current case did not, they would decide for the defendant. Other researchers and lawyers (usually defense attorneys) observed the increase in the use of forensic evidence in trials and believed that it would bias jurors to decide in favor of the side with the most forensic evidence (typically the prosecution) regardless of the other facts in the case (CHIN & IBAVIOSA 2022; COLE 2015).

A problem with the supposed CSI-effect is that ordinary people are not aware that evidence collected from a crime scene might not be of high quality to begin with or can become contaminated, thus complicating whether the evidence can be admitted in court and the conclusions that can be drawn from it (GARRETT 2022). Forensic testing is slow and expensive,

and not always helpful. Forensic testing is usually reserved for the most serious cases, and is certainly not commonplace.

2.2.5. *Misconception 5: "Forensic evidence can tell you whether a specific person committed a crime"*

Not only do people struggle to understand the forensic evidence and associated conclusions, but they also have trouble determining what can and cannot be "proven" by forensic evidence. For instance, physical evidence is still circumstantial evidence. The presence of someone's DNA or a fingerprint at a crime scene does not, in itself, prove that this person committed the crime. Forensic evidence can lead to an *inference* that a person had been at the scene at some point because they left behind something that is unlikely to be there unless they had been physically present. The condition of the evidence might reveal whether they had been there at about the time of the crime. Then, those facts can be used to further infer that that person was involved in the crime that took place.

Thus, forensic evidence alone can be fairly weak evidence that someone is the actual perpetrator without other corroborating evidence and reasons to believe they would do such a thing. For ordinary people to understand forensic evidence and what it can offer to a criminal investigation or case, they need to understand that it is just one piece of a puzzle and that, alone, forensic evidence is rarely enough to convict an individual. Even when considering DNA in a sexual assault case, generally seen as particularly incriminating, there are ways that the physical evidence can be wrong or misleading. Consider the example of Mr. Jama who was convicted of rape based on solely on DNA evidence taken from a rate kit at a hospital (*Jama* case, detailed in VINCENT 2010). It was later discovered that Mr. Jama could not have committed the crime and the positive test must have been an error. In fact, Mr. Jama's DNA had been tested at the same hospital two days prior. A formal inquiry was unable to conclusively determine how the relevant test sample had been contaminated by Mr. Jama's DNA but, ultimately, decided that must have been what happened in this case. The report suggests that DNA alone should never be sufficient for a conviction.

3. *Forensic science at trial: Using it appropriately and effectively*

The characteristics of human reasoning and how they contribute to errors in forensic science reasoning were described in Part 1, and common preconceptions and misconceptions among factfinders were described in Part 2. Given this factors, how and when should forensic science evidence be introduced and evaluated in trials? Trial procedures vary around the world, and we consider below if and how forensic science should be presented to factfinders—people who may be judges, lay judges, or citizen-jurors, but are neither scientists nor forensic experts. Judges, lawyers, and witnesses can all help ensure that forensic science information is communicated to factfinders in ways that result in an appropriate level of reliance on that information based on the underlying science. We use the language from the US system to label the function of legal actors in this chapter—a judge rules on the admissibility of evidence, lawyers ask questions of witnesses, and jurors determine the facts and the verdict in the case.

3.1. *Deciding whether a type of forensic science should be admitted at all*

In the US, it is the presiding judge's decision whether to admit proffered expert evidence. The US Supreme Court ruled in *Daubert v Merrell Dow Pharmaceuticals, Inc.* (1993)[6] that judges serve as "gatekeepers" and assess «whether the reasoning or methodology underlying the [expert

[6] *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 US 579 (1993).

scientific] testimony is scientifically valid». Rule 702 of the Federal Rules of Evidence also supports this approach. Judges are given guidance about what to consider when assessing the admissibility of expert scientific evidence, including whether the theory or technique can be, and has been, tested; whether the methods are standardized; whether the science has been subject to independent peer-review and published; whether there is a known error rate; and whether the findings are accepted in the relevant scientific community.

The *Daubert* opinion states «We are confident that federal judges possess the capacity to undertake this review»[7]. Despite Justice Blackmun's optimism, the lower Appellate Court re-tried the case using the new *Daubert* standard and noted that federal judges, who are "largely untrained in science", face a "daunting task" when ruling on the admissibility of expert scientific testimony[8]. Subsequent surveys of judges have revealed that many do not belive they have the training or knowledge needed to make such decisions about scientific evidence in general. A more recent survey inquired specifically about forensic science evidence, and judges indicated that they wanted more training and wished such training was more easily available (GARRETT et al. 2021).

A small number of judges have taken notice of the warnings in the NRC report (2009) and disallowed some types of forensic evidence. Yet, for the most part, US judges have continued to admit a wide range of forensic evidence, including those that have been heavily criticized by independent agencies (e.g., bitemarks, non-DNA hair comparisons). Why? One reason is that rejecting established precedent is difficult to justify in legal procedure/culture. If other courts have allowed this type of evidence in the past, why should things be changed for this new case? Judge Jed Rakoff, a US Federal Court Judge and longtime proponent of forensic science reform, suggests another reason—that trial judges, who are often past prosecutors, are «very hesitant to deprive the prosecution of evidence that may make the difference between conviction and exoneration» (RAKOFF 2023, 83). Thus, like all decisions made by people, judges' decisions about admissibility may be driven by their own biases.

3.2. *Describing numbers and conclusions*

Forensic science expert witnesses may be asked questions about the methods and procedures used, the reliability of their tests, and their own qualifications, etc. Typically, though, their featured testimony focuses on the results of forensic tests they performed (or written reports of such tests). Communicating the results of forensic tests to jurors so that they correctly understand them and can use them appropriately is complicated (BALI et al. 2020; ELDRIDGE 2019). One way to describe results is by reporting a number and describing the meaning of that number. For instance, when reporting the results of a DNA comparison, an analyst might say, "it is 5,500,000 times more likely that this blood came from the suspect rather than from a random other person" (reporting a likelihood ratio) or "there is only a 1 in a 1,000,000 chance that a random person would have these features" (reporting a random match probability).

There are problems with reporting numeric results. First, for any forensic discipline other than DNA, the science is unlikely to support the use of exact numbers. More importantly, though, people find it challenging to understand probabilities, especially ones that are very small or very large. Probabilities that seem near 0, or near 1 may be interpreted as being no different from 0 or 1, thus corrupting later conclusions. Finally, presenting very small random match probabilities leads some people to entirely misinterpret the very small numbers as the chance that the defendant is innocent.

The other common way to communicate forensic results is using categorical phrases designed to suggest a range of probabilities. For example, "almost completely certain" suggests that there is some

[7] *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 US 579 (1993).
[8] *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3D 1311 (9TH CIR. 1995).

probability of error, but it is very unlikely. "It's somewhat likely" suggests that whatever conclusion the expert is presenting may be unreliable. Other phrases imply that the results are definitive, such as "It's a match." These kinds of definitive terms are dangerous in the context of communicating with non-experts because what a non-forensic audience takes away from the testimony is something different from the expert's intended meaning (ELDRIDGE 2019). Research suggests there is low correspondence between the value experts intend to communicate with verbal categorical phrases and what jurors understand when hearing that language (e.g., MARTIRE et al. 2013; MARTIRE & WATKINS 2015).

Regardless of how the results are presented, it is a mistake not to clearly communicate the potential error rates to factfinders (KOEHLER 1996). As discussed in Part 1 and Part 2.2.2 above, no matter how good the science is, there is always a non-zero rate of error. In particular, in any discipline that relies on human decision-making, there will be error associated with the performance of an individual practitioner, as well as errors resulting from laboratory procedures. Whatever the nature of those potential errors, factfinders should be made aware of the potential for such errors in addition to whatever errors are possible in the science itself. When this information is disclosed, ordinary people appear to adjust their judgments about the credibility of the findings appropriately (GARRETT et al. 2020).

3.3. *Testimony*

As described in Part 2, factfinders come to court with pre-existing beliefs about the validity and reliability of forensic evidence. Studies of both jurors and judges indicate that when people rate the reliability of different types of forensic evidence, the relative rankings do align with actual reliability (i.e., current scientific belief about reliability of the evidence). However, although the ordering is correct, their beliefs are inflated and not grounded in a solid understanding of the techniques or the reasons they are or are not reliable (KOEHLER 2017; MARTIRE et al. 2019).

Given that people's attitudes and beliefs are very difficult to change, how can the overconfidence ordinary people have in forensic science be remedied? Experimental research has looked at a variety of techniques for recalibrating "mock jurors" (typically jury-eligible Americans who read vignettes or watch videos of simulated trials). These techniques can be implemented during cross-examination of witnesses or by adding additional witnesses. The success of such measures, however, depends on how "invested" a person is in their pre-existing views (see the descriptions of motivated reasoning and confirmation bias in Part 1.4.2 above). Jurors who acknowledge that their understanding of forensic science is limited should be more willing to change their beliefs than those who feel confident in their understanding. In addition, jurors who are strongly invested in the success of a particular outcome and believe that the testimony will threaten that outcome, will be less likely to change their beliefs.

3.3.1. *Cross-examination*

A multi-disciplinary group of research scientists, forensic scientists, and lawyers based in Australia authored: *How to cross-examine forensic scientists: A guide for lawyers* (EDMOND et al. 2014). Their suggestions include asking questions about the foundation of the forensic discipline, such as the validation of the technique, the likelihood of error, the forensic scientist's proficiency, exposure to potentially biasing information, the quality of the evidence in question, and whether the conclusions were verified by an independent analyst. Research with mock jurors shows that receiving this type of information can appropriately modify mock jurors' beliefs.

Attacking the source of information can be useful for debiasing (LEWANDOWSKY et al. 2012). For example, forensic examiners who admit to low performance on required proficiency exams are viewed as less credible and less persuasive than those who have shown high performance

(CROZIER et al. 2020). If the forensic examiner reveals that they were exposed to potentially biasing information (e.g., information about the defendant's prior criminal history or his confession), people justifiably find the expert less credible. However, if the analyst is asked whether they might have been biased in these situations, those who claimed that they were not improperly influenced seemed more credible than those who acknowledged the possibility that they had been influenced (KUKUCKA et al. 2020). This finding goes against existing research. Forensic examiners, like everyone else, tend to be blind to their own biases—they believe that others are biased but they themselves are not (KUKUCKA et al. 2017). In addition, bias occurs outside of a person's awareness and is difficult to prevent, so we know that forensic examiners cannot always "debias themselves". Yet, mock jurors believe experts can do this, perhaps due to the illusion that expertise immunizes people against such errors.

Of course, to do a good job of cross-examination, opposing counsel needs to have a sufficient degree of understanding of forensic science. And depending on how the legal system works, opposing counsel might need to find the time and financial resources to ask and pay for additional alternative testing.

3.3.2. *Experts for better or worse*

The existing research on calling opposing expert witnesses does not provide support for engaging in a game of "battling experts". The research suggests that ordinary people hearing from experts who completely contradict each other (e.g., "the time of death was 7 pm", "no, it was 11 pm") are not inspired to choose one view over the other—rather, it diminishes their view of both witnesses and their testimony and the science (MITCHELL & GARRETT 2021; SCOBIE et al. 2019).

However, rebuttal experts can teach jurors about aspects of forensic science that were not addressed by the initial expert. They can do that at the level of the forensic science generally—for example, explaining that fingerprint identifications can be wrong, especially when the latent prints are unclear—and then applying that information to the specific case (i.e., the print was of poor quality and not suitable for comparison). Given that information some jurors will decrease their confidence in the evidence (MITCHELL & GARRETT 2021). This technique follows the known debiasing technique of "providing an alternative narrative"—supplying information about how something might have happened differently (LEWANDOWSKY et al., 2012)—in this case why the two examiners came to different conclusions.

3.4. *Jury instructions*

Several researchers have suggested that jury instructions might be a good way of telling jurors about the limitations of forensic science, but the current set of forensic studies that use such instructions is limited and inconclusive. That said, jury instructions that target other kinds of evidence for which lay persons have similar, well-ingrained misconceptions, for example, eyewitness identification evidence, do not help jurors sort between good and bad, or reliable and unreliable, examples of such evidence (e.g., JONES et al. 2020; PAPAILIOU et al. 2015). Thus, it is unlikely that a short instruction from the judge, after all evidence has been presented, would be enough to overcome existing beliefs about forensic evidence and their already-formed opinions about the evidence in the case.

4. *Suggestions and conclusions*

No one is an expert in all forensic science disciplines. But judges may be called upon to rule on the admissibility of any forensic discipline—regarding the science itself and the testimony surrounding it. Lawyers (or judges) may need to question or cross-examine forensic practitioners or decide

whether to call them as witnesses. Ultimately, it is the factfinder—judges, lay judges, or jurors—who must be able to understand the appropriate probative value of the evidence and use it to evaluate the issue in a particular case. Although they should, not all legal professionals will be motivated to ensure that factfinders interpret and use this information correctly, but here we have provided background information and various techniques for those who do.

This chapter mainly focuses on the use of forensic science evidence in criminal cases; however, it is also relevant to some civil cases (especially for document and handwriting identification, and for experts in accident reconstruction). The cognitive science underlying the interpretation, use, and understanding of forensic science is broadly applicable to other kinds of decisions in legal contexts. For example, trials often involve non-forensic expert testimony, such as physicians who comment on the severity of injuries from a car crash, art dealers who opine on whether a painting is a forgery, real estate agents who judge the value of a piece of property, and even cognitive psychologists who present on the reliability of eyewitness memory. Understanding the general principles of reasoning, and the factors that could lead to forensic errors, is useful for evaluating any type of expert testimony. In the US, every law school has a course on Evidence Law where students learn about the admissibility of experts and expert evidence, but very few law schools offer courses specifically on scientific or forensic evidence (GARRETT et al. 2022). Judges have expressed a view that they would like more training about forensic science, especially early in their legal education, and access relevant continuing education classes or other aides to help them understanding this constantly and quickly evolving scientific field. We have no doubt that many lawyers do so as well.

We leave the reader with the following suggestions for more information about the intersection between cognitive science and forensic science.

Book for general background on how forensics are done and what can go wrong:

GARRETT B.L. 2022. *Autopsy of a crime lab: Exposing the flaws in forensics*, University of California Press.

Articles on forensics generally with a cognitive science approach:

DROR I.E. 2020. *Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias*, in «Analytical Chemistry», 92, 12, 7998 ff.

EDMOND G., TOWLER A., GROWNS B., RIBEIRO G., FOUND B., WHITE D., BALLANTYNE K., SEARSTON R.A., THOMPSON M.B., TANGEN J.M., KEMP R.I. 2017. *Thinking Forensics: Cognitive Science for Forensic Practitioners*, in «Science & Justice, 57, 2, 144 ff.

SPELLMAN B.A., ELDRIDGE H., BIEBER P. 2022. *Challenges to Reasoning in Forensic Science Decisions*, «Forensic Science International: Synergy», 4, 100200.

For general interest in forensic science and up-to-date news see:

What the US is doing:
https://www.nist.gov/forensic-science
https://forensiccoe.org

What the UK is doing:
https://www.gov.uk/government/organisations/forensic-science-regulator

References

Balcetis E., Dunning D. 2006. *See What You Want to See: Motivational Influences on Visual Perception*, in «Journal of Personality and Social Psychology», 91, 4, 612.

Bali A.S., Edmond G., Ballantyne K.N., Kemp R.I., Martire K.A. 2020. *Communicating Forensic Science Opinion: An Examination of Expert Reporting Practices*, in «Science & Justice», 60, 3, 216 ff.

Biedermann A., Kotsoglou K.N. 2021. *Forensic Science and the Principle of Excluded Middle: "Inconclusive" Decisions and the Structure of Error Rate Studies*, in «Forensic Science International: Synergy», 3, 100147. Available on: https://doi.org/10.1016/j.fsisyn.2021.100147.

Busey T., Sudkamp L., Taylor M.K., White A. 2022. *Stressors in Forensic Organizations: Risks and Solutions*, in «Forensic Science International: Synergy», 4, 100198.

Chaiken S., Maheswaran D. 1994. *Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgment*, in «Journal of Personality and Social Psychology», 66, 3, 460.

Chin J.M., Ibaviosa C M. 2022. *Beyond CSI: Calibrating Public Beliefs about the Reliability of Forensic Science through Openness and Transparency*, in «Science & Justice», 62, 272 ff.

Cole S.A. 2015. *A Surfeit of Science: The "CSI effect" and the Media Appropriation of the Public Understanding of Science*, in «Public Understanding of Science», 24, 2, 130 ff. Available on: https://doi.org/10.1177/0963662513481294.

Crozier W.E., Kukucka J., Garrett B.L. 2020. *Juror Appraisals of Forensic Evidence: Effects of Blind Proficiency and Cross-examination*, in «Forensic Science International», 315, 110433.

Dror I.E. 2015. *Cognitive Neuroscience in Forensic Science: Understanding and Utilizing the Human Element*, in «Philosophical Transactions of the Royal Society B: Biological Sciences», 370, 1674, 20140255.

Dror I.E. 2020. *Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias*, in «Analytical Chemistry», 92, 12, 7998 ff.

Dror I.E., Hampikian G. 2011. *Subjectivity and Bias in Forensic DNA Mixture Interpretation*, in «Science & Justice», 51, 4, 204 ff.

Dror I.E., Mnookin J.L. 2010. *The Use of Technology in Human Expert Domains: Challenges and Risks Arising from the Use of Automated Fingerprint Identification Systems in Forensic Science*, in «Law, Probability and Risk», 9, 1, 47 ff.

Dror I.E., Wertheim K., Fraser-Mackenzie P., Walajtys J. 2012. *The Impact of Human-technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts*, in «Journal of Forensic Sciences», 57, 2, 343 ff.

Edmond G., Martire K., Kemp R., Hamer D., Hibbert B., Ligertwood A., Porter G., San Roque M., Searston, R., Tangen J., Thompson M. 2014. *How to Cross-examine Forensic Scientists: A Guide for Lawyers*, in «Australian Bar Review», 39, 174 ff.

Edmond G., Towler A., Growns B., Ribeiro G., Found B., White D., Ballantyne K., Searston R.A., Thompson M.B., Tangen J.M., Kemp, R.I. 2017. *Thinking Forensics: Cognitive Science for Forensic Practitioners*, in «Science & Justice», 57, 2, 144 ff.

Eldridge H. 2019. *Juror Comprehension of Forensic Expert Testimony: A Literature Review and Gap Analysis*, in «Forensic Science International: Synergy», 1, 24 ff.

Eldridge H., Stimac J., Vanderkolk J. 2022. *The Benefits of Errors During Training*, in «Forensic Science International: Synergy», 4, 100207.

Frederick S. 2005. *Cognitive Reflection and Decision Making*, in «Journal of Economic Perspectives», 19, 25 ff.

Funk C., Hefferon M., Kennedy B., Johnson C. 2019. *Trust and Mistrust in Americans' Views of Scientific Experts*, Pew Research Center. Available on: https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/.

Gardner B.O., Kelley S., Pan K.D.H. 2020. *Latent Print Proficiency Testing: An Examination of Test Respondents, Test-taking Procedures, and Test Characteristics*, in «Journal of Forensic Science», 65, 2, 450 ff.

Garrett B.L. 2022. *Autopsy of a Crime Lab: Exposing the Flaws in Forensics*, University of California Press.

Garrett B.L., Cooper G.S., Beckham Q. 2022. *Forensic Science in Legal Education*, in «Journal of Law & Education», 51, 1, 1 ff.

Garrett B.L., Crozier W.E., Grady, R.H. 2020. *Error Rates, Likelihood Ratios, and Jury Evaluation of Forensic Evidence*, in «Journal of Forensic Sciences», 65, 4, 1199 ff. Available on: https://doi.org/10.1111/1556-4029.14323.

Garrett B.L., Gardner B.O., Murphy E., Grimes P. 2021. *Judges and Forensic Science Education: A National Survey*, in «Forensic Science International», 321, 110714.

Houck M.M. 2006. *CSI: Reality*, in «Scientific American», 295, 1, 84 ff.

Jones A.M., Bergold A.N., Penrod S. 2020. *Improving Juror Sensitivity to Specific Eyewitness Factors: Judicial Instructions Fail the Test*, in «Psychiatry, Psychology and Law», 27, 3, 366 ff. Available on: https://doi.org/10.1080/13218719.2020.1719379.

Kahneman D. 2011. *Thinking, Fast and Slow*, Macmillan.

Kassin S.M., Dror I.E., Kukucka J. 2013. *The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions*, in «Journal of Applied Research in Memory and Cognition», 2, 1, 42 ff.

Koehler J.J. 1996. *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, in «The University of Colorado Law Review», 67, 4, 859 ff.

Koehler J.J. 2013. *Proficiency Tests to Estimate Error Rates in the Forensic Sciences*, in «Law, Probability & Risk», 12, 1, 89 ff.

Koehler J.J. 2017. *Intuitive Error Rate Estimates for the Forensic Sciences*, in «Jurimetrics», 57, 2, 153 ff. Available on: https://www.jstor.org/stable/26322664.

Koertner A.J., Swofford H.J. 2018. *Comparison of Latent Print Proficiency Tests with Latent Prints Obtained in Routine Casework Using Automated and Objective Quality Metrics*, in «Journal of Forensic Identification», 68, 3, 379 ff.

Kuhn D. 1989. *Children and Adults as Intuitive Scientists*, in «Psychological Review», 96, 4, 674 ff.

Kukucka J., Dror I.E. 2023. *Human Factors in Forensic Science: Psychological Causes of Bias and Error*, in De Matteo D., Scherr K.C. (eds.), *The Oxford Handbook of Psychology and Law*, Oxford University Press. Available on: https://doi.org/10.1093/oxfordhb/ 9780197649138.013.36.

Kukucka J., Hiley A., Kassin S.M. 2020. *Forensic Confirmation Bias: Do Jurors Discount Examiners Who Were Exposed to Task-irrelevant Information?*, in «Journal of Forensic Sciences», 65, 6, 1978 ff.

Kukucka J., Kassin S.M., Zapf P.A., Dror I.E. 2017. *Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners*, in «Journal of Applied Research in Memory and Cognition», 6, 4, 452 ff.

Kunda Z. 1990. *The Case for Motivated Reasoning*, in «Psychological Bulletin», 108, 3, 480 ff.

LEWANDOWSKY S., ECKER U.K., SEIFERT C.M., SCHWARZ N., & COOK J. 2012. *Misinformation and Its Correction: Continued Influence and Successful Debiasing*, in «Psychological Science in the Public Interest», 13, 3, 106 ff.

LIEBERMAN J.D., CARRELL C.A., MIETHE T.D., KRAUSS D.A. 2008. *Gold versus Platinum: Do Jurors Recognize the Superiority and Limitations of DNA Evidence Compared to Other Types of Forensic Evidence*?, in «Psychology, Public Policy, and Law», 14, 1, 27 ff.

MARTIRE K.A., BALLANTYNE K.N., BALI A., EDMOND G., KEMP R.I., FOUND B. 2019. *Forensic Science Evidence: Naive Estimates of False Positive Error Rates and Reliability*, in «Forensic Science International», 302, 109877. Available on: https://doi.org/10.1016/ j.forsciint.2019.109877.

MARTIRE K.A., KEMP R.I., NEWELL B.R. 2013. *The Psychology of Interpreting Expert Evaluative Opinions*, in «Australian Journal of Forensic Sciences», 45, 3, 305 ff.

MARTIRE K.A., WATKINS I. 2015. *Perception Problems of the Verbal Scale: A Reanalysis and Application of a Membership Function Approach*, in «Science & Justice», 55, 4, 264 ff. Available on: https://doi.org/10.1016/j.scijus.2015.01.002.

MEJIA R., CUELLAR M., SALYARDS J. 2020. *Implementing Blind Proficiency Testing in Forensic Laboratories: Motivation, Obstacles, and Recommendations*, in «Forensic Science International. Synergy», 2, 293 ff. Available on: https://doi.org/10.1016/j.fsisyn.2020.09.002.

MITCHELL G., GARRETT B.L. 2021. *Battling to a Draw: Defense Expert Rebuttal Can Neutralize Prosecution Fingerprint Evidence*, in «Applied Cognitive Psychology», 35, 4, 976 ff.

MORGAN J. 2023. *Wrongful Convictions and Claims of False or Misleading Forensic Evidence*, in «Journal of Forensic Science», 68, 3, 908 ff. Available on: https://onlinelibrary.wiley.com/ doi/full/10.1111/ 1556-4029.15233.

PAPAILIOU A.P., YOKUM D.V., ROBERTSON C.T. 2015. *The Novel New Jersey Eyewitness Instruction Induces Skepticism but Not Sensitivity*, in «PloS one», 10, 12, e0142695. Available on: https://doi.org/10.1371/journal.pone.0142695.

PENNINGTON N., HASTIE R. 1991. *A Cognitive Theory of Juror Decision Making: The Story Model*, in «Cardozo Law Review», 13, 519 ff.

PETTY R.E., CACIOPPO J.T. 1981. *Issue Involvement as a Moderator of the Effects on Attitude of Advertising Content and Context*, in «Advances in Consumer Research. North American Advances», 8, 1, 20 ff.

PETTY R.E., CACIOPPO J.T. 2012. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, Springer Science & Business Media.

PRONIN E. 2007. *Perception and Misperception of Bias in Human Judgment*, in «Trends in Cognitive Sciences», 11, 1, 37 ff.

RAKOFF J.S. 2023. *Book Review: Is "Forensic Science" a Misnomer?*, in «Judicature Duke Edu », 106, 3, 80 ff.

READ S.J., MARCUS-NEWHALL A. 1993. *Explanatory Coherence in Social Explanations: A Parallel Distributed Processing Account*, in «Journal of Personality and Social Psychology», 65, 3, 429 ff.

SAUERWEIN K., BUTLER J.M., RECZEK K.K., REED C. 2023. *Bitemark Analysis: A NIST Scientific Foundation Review*, National Institute of Standards and Technology, Gaithersburg, MD, NIST Interagency Report (IR) NIST IR 8352. Available on: https://www.nist.gov/spo/forensic-science-program/ bitemark-analysis-nist-scientific-foundation-review.

SCOBIE C., SEMMLER C., PROEVE M. 2019. *Considering Forensic Science: Individual Differences, Opposing Expert Testimony and Juror Decision Making*, in «Psychology, Crime & Law», 25, 23 ff. Available on: https://doi.org/10.1080/1068316X.2018.1488976.

Spellman B.A., Eldridge H., Bieber P. 2022. *Challenges to Reasoning in Forensic Science Decisions*, in «Forensic Science International: Synergy», 4, 100200.

Spellman B.A., Tenney E.R. 2010. *Credibility in and out of Court*, in «Psychonomic Bulletin & Review», 17, 168 ff.

Thompson W.C., Scurich N. 2019. *How Cross-examination on Subjectivity and Bias Affects Jurors' Evaluations of Forensic Science Evidence*, in «Journal of Forensic Sciences», 64, 5, 1379 ff.

Tversky A., Kahneman D. 1974. *Judgment under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking under Uncertainty*, in «Science», 185, 4157, 1124 ff.

Vincent F.H.R. 2010. *Inquiry into the Circumstances Surrounding the Wrongful Conviction of Mr. Farah Abdulkadir Jama*, Victorian Government Printer. Available on: https://www.parliament.vic.gov.au/papers/govpub/VPARL2006-10No301.pdf.

The Division of Cognitive Labour in Law

MICHELE UBERTONE

> Our dependence upon the word of others can be shown to be extensive and deep. We exhibit such dependence, though seldom acknowledge it explicitly, in our confident knowledge claims and actions in everyday life as well as in our more theoretical pursuits. In everyday life, we automatically relay sporting scores and judicial verdicts, we accept new financial burdens on the basis of reported pay increases, and we plan holidays on the basis of geographical, transport and accommodation information from others. In the sciences, we talk of what is known and has been proved in hosts of instances where we have not done the proving or "done the knowing", and often this is in contexts where we wouldn't have the individual resources for the relevant investigations anyway (COADY 1994b, 225)

> There are two sorts of tools in the world: there are tools like a hammer or a screw-driver which can be used by one person; and there are tools like a steamship which require the cooperative activity of a number of persons to use. Words have been thought of too much on the model of the first sort of tool (PUTNAM 1975, 146)

0. *Introduction*

Any legal system—whether democratic or not—is generally believed to produce its effects in virtue of the fact that people know and understand its provisions. One could even say that law just *is* whatever is understood by officials and citizens to be law[1]. Officials must understand law to be able to enforce it. Citizens must understand law to be motivated by the threat of sanctions, to be able to identify and exercise their rights, to be able to identify and fulfil their obligations. The ability of citizens to understand the law is supposed to have an additional function in constitutional democracies: The political legitimacy of a constitutional democracy is believed to be grounded on the possibility of citizens to assess, accept, and, when necessary, criticise official and legislative action[2].

The standard story that lawyers seem to be constantly telling themselves about how citizens get to know and understand their normative environment, however, is scarcely believable. According to it, citizens *are* (or at least, if only they put enough effort into it, *can be*) perfectly informed and rational agents, able to comply to the desiderata of institutions by virtue of an autonomous and rational understanding of the legislative texts that the institutions produce as

[1] John Searle's social ontology, which we will briefly discuss later, can be interpreted in this sense. See SEARLE 1997, SEARLE 2010.

[2] On publicity as an essential aspect of the ideal of the rule of law see for example CELANO 2013.

well as of the factual context in which they are supposed to be applied. This theory often emerges as an implicit assumption in legal reasoning and legal doctrines. Consider, for example, the issue of the non-retroactivity of criminal law: if someone is judged according to a statute that came into force after the crime that person is accused of having committed, this is perceived as a major violation of that person's human rights. The European Court of Human Rights spends a lot of its time and energy to make sure that the non-retroactivity of criminal law is protected. Every citizen must have the right to know in advance, and in detail, the legal consequences of his or her actions. But for how many citizens is this right accompanied by a real ability to exercise it? Almost none[3].

This story has played an important role in our legal culture as a regulative ideal[4]. But to have a scientific comprehension of the functioning of institutions and to predict the practical effects that our legal decisions and policies of institutional design may have, we must recognize it as a fiction. It is not obvious why legal officials at all levels should conduct their legal reasoning and take their decisions *as if* citizens could have perfect individual knowledge of laws and legally relevant facts, as well as the ability to infer what their conjunction implies. Making factually false assumptions about people's cognitive abilities poses the risk of systematic distortions in practical reasoning and failures of instrumental rationality. To effectively protect the interests of real people, rather than hypothetical agents, judges' reasoning should maybe instead consider the real possibilities of understanding of the average individual and the real socio-linguistic dynamics through which he or she interprets legal concepts.

In this article, I would like to give a contribution to the discussion of this topic, suggesting that some implicit cognitive presuppositions of legal practice may be at odds with how science currently understands people to acquire and process knowledge. I would like to invite the reader to consider whether these implicit presuppositions have any essential significance in the field of legal practice, or whether a deeper understanding of the psychology by legal professionals could potentially eliminate them and maybe promote a more equitable and rational institutional design and administration of justice. The aim of the paper is to raise the question rather than provide an answer to it. I want to show how this question is important, taking a conceptual and philosophical perspective on it. The factual premises of the discussion will be sketched only in an impressionistic way and would deserve to be separately discussed on the basis of specific empirical studies. Showing the importance of the development of a line of empirical research in this direction is by the way an additional aspiration of this paper.

In the first section, I will try to briefly outline what I will call *cognitive perfectionism*. I will use this expression to refer to certain misconceptions, characteristic of folk psychology[5], regarding how

[3] In some cases, the prohibition of retroactive application of criminal law extends to technical details of criminal procedure that not only no average criminal knows before committing a crime, but often not even an average criminal lawyer can remember by heart. *Del Rio Prada v. Spain*, for example, concerned the retroactive application of an unfavourable jurisprudential change in the granting of a prison benefit. The appellant, Ines Del Rio Prada, had been convicted of serious crimes linked to Basque separatist terrorism and had applied for the benefit of the *redención de penas por trabajo*, which allowed a reduction in sentence for work done inside prison. However, the Spanish Supreme Court had adopted a new interpretation, known as the "Parot doctrine", according to which the reduction should be calculated differently. The Grand Chamber ruled that this sudden change of the case law that preceded Del Rio Prada's convictions constituted a violation of her human rights, as it was unpredictable by her at the time. In this regard, see VIGANÒ 2022.

[4] A regulative ideal can be defined as a normative horizon towards which we should strive as far as possible. Such a normative horizon is a state of affairs that we evaluate as desirable or correct. See MARTÍ 2005.

[5] Folk psychology is «in one sense, a putative network of principles constituting a commonsense theory that allegedly underlies everyday explanations of human behavior; the theory assigns a central role to mental states like belief, desire, and intention.[...] In another, related sense, folk psychology is a network of social practices that includes ascribing such mental states to ourselves and others, and proffering explanations of human behavior that advert to these states. The two senses need distinguishing because some philosophers who acknowledge the existence of folk psychology in the second sense hold that commonsense psychological explanations do not employ

individuals are supposed to develop their beliefs, construct concepts, and employ them in their actions. In short, cognitive perfectionism is the misconception that we have a significantly greater personal understanding of our environment than we actually do, and the misconception that our cognitive abilities can operate autonomously without relying on division of labour with other agents.

In the second section, I will discuss how cognitive perfectionism affects legal practice. To do so, I will discuss three examples of cognitive perfectionist doctrines, three myths so to speak of contemporary legal systems: the myth of legality, the myth of consent, and the myth of the judge as a "gatekeeper". I will also argue that law is traditionally characterised by two types of cognitive perfectionism, the idea that legal agents are perfectly rational (*rationality perfectionism*) and the idea that legal agents have perfect knowledge (*knowledge perfectionism*). Legal theory has mainly discussed the first aspect, while the second has traditionally been neglected. For this reason, the examples I chose are mainly intended to illustrate the second aspect rather than the first aspect.

In conclusion, I will raise the question of whether it would be desirable or even possible to eradicate cognitive perfectionism from legal practice or whether this is somehow essential to it. I won't give an answer to this question but I will try to briefly suggest some directions in which legal philosophical research about it could be developed.

1. *Cognitive perfectionism vs. Division of cognitive labour*

In any modern democracy, people "freely and consciously" select their representatives. Parliament transposes "the will" of the people into law. All laws get published so that everybody may "know" what they require. Judges "make decisions based on the law", thus implementing the will of the people in particular cases. Punishment is only addressed to criminals who "willingly" broke the law, "knowing" beforehand the legal consequences of their criminal actions. All of those are key ideas in contemporary legal systems, and all of them are deeply entangled with problematic assumptions about how the mind of various actors within the legal system works.

An important strand of recent legal philosophical literature has been concerned with identifying these folk-psychological assumptions in the law and comparing them with the scientific picture of the mind offered by neuroscience and contemporary cognitive psychology. Part of this literature consists in critiquing and deconstructing specific legal concepts. Some, for example, have criticised the concept of democracy, arguing that it is based on an inaccurate description of how people make decisions in the political context (they are much less selfish, rational and knowledgeable than we think) (BRENNAN 2016). Others have criticised the concept of criminal responsibility showing how it is rooted in moral intuitions that tend to disappear when we are exposed to a scientific explanation of the real psychological roots of crime (when we learn more about neuroscience we tend to be suspicious about the notion of free will and retribution and tend to find consequentialist theories of punishment more plausible) (SIFFERD 2004, SIFFERD 2006). Others still have criticised the concept of a human right, showing that talk of human rights can be "debunked" (talking of human rights is just a way to justify *ex post facto* intuitions that are independent from any relation with what we believe should be the sources of rights) (BUBLITZ 2021). A similar argument has been made with respect to the concepts of legal interpretation and

empirical generalizations, and hence that there is no such theory as folk psychology» (AUDI 2015). The first conception of folk psychology is sometimes referred to as the theory-theory of folk psychology and I will somehow subscribe to it when I will speak of cognitive perfectionism as a set of assumptions that we adopt to predict what both other people and ourselves can be expected to process information. See HUTTO & RAVENSCROFT 2021.This dominant view is sometimes referred to as the theory-theory of folk psychology and I will somehow subscribe to it when I will speak of cognitive perfectionism as a set of assumptions that we adopt to predict what both other people and ourselves can be expected to process information.

evidence (both of which have been described as ways of rationalising and making socially presentable insights causally independent of the premises that interpretive and evidential arguments should formally have) (HAIDT 2013). Other authors have attempted a more far-reaching operation by discussing in more general terms the relationship that exists or should exist between folk psychology and law (BRIGAGLIA 2015, HAGE 2021, TOBIA 2021, KUREK 2021).

In this article, I do not intend to discuss specific folk-psychological concepts adopted by law, nor the relationship between folk-psychology and law in general. Rather, I want to discuss a subset of characteristics that folk psychology ascribes to individuals and that I believe have profound implications for legal practice. Folk psychology gives us a distorted image of our ability to accumulate knowledge and process it rationally: it ignores the existence of a division of cognitive labour and thinks of individuals as cognitively perfect agents. I am calling this set of folk psychological assumptions "cognitive perfectionism". Agents are thought as "perfect" in the etymological sense of the word. "Per-fectus" in Latin means accomplished, complete, finished, and that is how we tend to think of our cognitive processes: as mechanisms accomplished in themselves, capable of functioning in relative isolation. We tend to think of the individual mind as capable of rationality and knowledge even if it is isolated from the physical and social context in which it evolved, whereas the cognitive processes that take place in our individual brains are better thought of as portions of broader, shared cognitive processes in which the ideals of rationality and knowledge can be realised. In particular, we think of individuals as capable of autonomous propositional knowledge, autonomous conceptual knowledge and perfect individual rationality. I will examine these three aspects separately, to make three points that can be summarised with three slogans: 1. *we know less than we think we know*, 2. *we master less concepts than we think we master*, 3. *we think less than we think we think*.

1.1. *Autonomous propositional knowledge*

I would like to start inviting the reader to make a little experiment. Please take a piece of paper and draw a bicycle. You don't need to be overly accurate. Just make sure to describe graphically, to the best of your understanding, how the bicycle works. Please don't look at pictures of bicycles on your phone. Just rely on your personal, individual knowledge. You can stop reading here: resume the reading once your drawing is complete...

So, do you know how a bicycle works? If you are like most people before this simple exercise, you *thought* you knew, but the attempt to draw one may have made you realise that your knowledge is far from accurate. In a 2006 experiment, a psychologist from the University of Liverpool, called Rebecca Lawson, asked subjects who claimed they understood well the functioning of bikes to complete a very simple drawing of one, and the results were quite comical. About half of the subjects, even some who used bikes every day, were unable to correctly represent how motion is transmitted from the pedals to the wheels via the chain (LAWSON 2006). An artist called Gianluca Gimini, independently from Lawson, had noticed this mismatch between how much we think we know about bikes and how much we actually know. He took some drawings of dysfunctional bikes, very similar to those produced by Lawson's subject and built them exactly as they were represented. Gilmini's work is a vivid representation of a key idea I would like to discuss here: *illusion of explanatory depth*. This is a metacognitive illusion: it is a proclivity to form false beliefs about the quality and nature of our own and other people's beliefs. We tend to greatly underestimate our reliance on other people's expertise. We fail to recognise how our commitment to the truth of even the most basic propositions is based on other people's commitment. We fail to recognise our inability to justify what we say and believe or even, as we will see in the next paragraph, our inability to identify the very content of the concepts we use. This is because information and evidence always feels "within reach" (ROZENBLIT & KEIL 2002). Our cognitive processes, as well as our behaviours, are so well functionally integrated with our social and technological environment, that we don't really need to draw a line between

what we individually know and what others know (CLARK & CHALMERS, 1998). Maybe I don't know my way home but the fact I can always rely on Google makes me feel as if I did (ELISEEV & MARSH 2023). Maybe I don't know exactly what the "carbon tax" is but the fact that a journalist I trust supports it makes me feel I also understand why it is a good thing.

This tendency to defer depends on the fact that our brain is in a sense a utilitarian. It does not perform a task if that task costs intellectual resources without producing equivalent intellectual profit. As Daniel Kahnemann puts it,

> «A general "law of least effort" applies to cognitive as well as physical exertion. The law asserts that if there are several ways of achieving the same goal, people will eventually gravitate to the least demanding course of action. In the economy of action, effort is a cost, and the acquisition of skill is driven by the balance of benefits and costs. Laziness is built deep into our nature. Generally, this means that when a cognitive task can be outsourced to our social or physical environment, our brain will outsource it. If I can form reliable beliefs about my way home deferring to my phone rather than to my own memory I will tend to do so» (KAHNEMAN 2011, 33).

Experiments along the same lines as Lawson's have also been run about the understanding of institutional policies, and the results were similar. Subjects holding strong views about policies regarding complex issues such as climate change were asked to explain them, and this had two effects. First, the request undermined the subject's belief that they understood the issue at hand: when attempting to provide an explanation, subjects realised they didn't really know that much; second, the request lead people to express more moderate views. While subsequent experiments failed to reproduce the second effect, the first one was confirmed. The request for an explanation reduces the people's belief that they understand.

According to cognitive scientists Sloman and Fernbach, we are often prey to an illusion of "explanatory depth": We *think* we know how the world around us works, but we don't. They give the example of toilets:

> «Take a minute and try to explain what happens when you flush a toilet. Do you even know the general principle that governs its operation? It turns out that most people don't. [...] Its most important components are a tank, a bowl, and a trapway. The trapway is usually S- or U-shaped and curves up higher than the outlet of the bowl before descending into a drainpipe that eventually feeds the sewer. The tank is initially full of water.
>
> When the toilet is flushed, the water flows from the tank quickly into the bowl, raising the water level above the highest curve of the trapway. This purges the trapway of air, filling it with water. As soon as the trapway fills, the magic occurs: A siphon effect is created that sucks the water out of the bowl and sends it through the trapway down the drain. It is the same siphon action that you can use to steal gasoline out of a car by placing one end in the tank and sucking on the other end. The siphon action stops when the water level in the bowl is lower than the first bend of the trapway, allowing air to interrupt the process. Once the water in the bowl has been siphoned away, water is pumped back up into the tank to wait for next time. It is quite an elegant mechanical process, requiring only minimal effort by the user. Is it simple? Well, it is simple enough to describe in a paragraph but not so simple that everyone understands it. In fact, you are now one of the few people who do» (SLOMAN & FERNBACH, 2018, 14 ff).

Even the simplest objects require complex networks of people with different expertise to be produced and function. No one is an expert of all aspects of a given problem. In almost everything, we rely on other people's expertise.

> «Our point is not that people are ignorant. It's that people are more ignorant than they think they are. We all suffer, to a greater or lesser extent, from an illusion of understanding, an illusion that we understand

how things work when in fact our understanding is meager» (SLOMAN & FERNBACH 2018, 12).

Sloman, along with others, has further investigated this topic experimentally by developing what they call «the community of knowledge hypothesis». The community of knowledge hypothesis is the idea that we fail to distinguish our own knowledge from other people's knowledge. We think we understand a topic, but this confidence is derived from the fact that we can rely on others to make statements or perform actions related to that topic. Sloman and Rabb tested this hypothesis by conducting a series of experiments in which they asked participants to rate their own understanding of novel natural phenomena. The experiments showed that people's perception of understanding is increased when they were told that experts could fully explain the phenomenon and that explanation was public and accessible (SLOMAN & RABB 2016). Similar effects have been found in relation to understanding of policies (RABB et al. 2021). Rabb, Fernbach and Sloman argue that knowledge is collective in two senses: not only in the obvious sense that is it is acquired through testimony, but also because it is stored outside of our heads and processed through the help of others: «individuals retain detailed causal information for a few domains and coarse causal models embedding markers indicating that these details are available elsewhere (others' heads or the physical world) for most domains» (RABB et al. 2019, 821).

It is as if each of us was a computer connected to the internet. We keep most of our files in the cloud (in the community of knowledge) and we download them to our heads only when and only in the limited measure in which this is necessary to us. And, in this process, we don't have a clear perception of what files are in the cloud and what would be accessible also off-line. This is reflected in our understanding of concepts, as we will see in the next section. People confuse their own ability to distinguish subtle differences of word meanings with the knowledge of how to access those meanings through other people[6].

1.2. *Autonomous conceptual knowledge*

There is then a mismatch between our naive way of understanding how knowledge is produced and stored and the real way it is produced and stored. A great deal of philosophical work has been done to demonstrate this mismatch. However, although the issue of relying on others' testimony to form our beliefs has been a topic of discussion since the era of David Hume and Thomas Reid, contemporary philosophical literature since the 1970s has highlighted an additional, deeper dimension of our dependence on the knowledge of trusted individuals. There are two ways in which our knowledge can be said to depend on the knowledge of others: one has to do with the commitment to the truth of propositions[7], and the other has to do with the structure of the concepts we use to think[8]. We rely on others not only to determine which propositions to believe (epistemic deference), but also to determine the conceptual content of our own beliefs (semantic deference). Our illusion of explanatory depth therefore consists not only in deluding ourselves that we autonomously know the reasons behind our beliefs, but also that we have autonomous access to the conceptual content expressed by words we use every day (KOMINSKY & KEIL 2014). This distinction can be illustrated with an example[9].

[6] KEIL 2005, KEIL et al. 2008, KOMINSKY & KEIL 2014, KEIL & KOMINSKY 2015.

[7] On epistemic deference, that is on testimony as a source of knowledge, see for example: ADLER 2012, HARDWIG 1985, HARDWIG 1991, COADY 1994a, COADY 1994b, GOLDMAN 1999, GOLDMAN 2011, GALLAGHER 2013.

[8] On semantic deference see SPERBER 1985, 54 ff., MARCONI 1997, RECANATI 1997, RECANATI 2000, 261 ff., DE BRABANTER 2006, DE BRABANTER et al. 2007, SHEA 2018. Also see the related literature on semantic externalism and two dimensionalism: KRIPKE 1980, PUTNAM 1975, BURGE 1979, CHALMERS 2002, CHALMERS 2003.

[9] In this example I will continue to use the typical language of folk psychology and in particular I will refer to the folk-psychological notion of belief: however, I will modify this notion in such a way as to make it compatible with the existence of a division of cognitive labour in the processing of concepts.

Imagine you go to the doctor and you are diagnosed with arthritis. The diagnosis is likely to produce in you the belief that you probably have a condition called "arthritis". If you trust your doctor and you are as most people this belief will be deferential in both ways I mentioned. First: your commitment to the truth of the proposition that you have arthritis will depend on the doctor's commitment. You will not understand and scrutinise all of the medical reasons that lead the doctor to believe that you have arthritis. Your reason for believing that you have arthritis will be the fact that you doctor believes it. If you were to learn your belief about the doctor's belief is wrong, you will have a reason to amend also your belief about arthritis. In this sense, your belief would be *epistemically deferential*. Second: you don't really know what the word "arthritis" means. You couldn't tell if "arthritis" is just a type of "arthrosis" or "arthrosis" a type of "arthritis". You don't possess the necessary conceptual competence to identify what exactly the word refers to. This means that not only don't you know how to justify individually the statement that you have arthritis but don't even know what are the truth conditions that that statement is supposed to represent. In your idiolect, the word "arthtitis" essentially means something like "whatever disease the community of doctors calls 'arthritis'". Your very concept of "arthritis" is deferential. It has the same referent of the doctors' technical concept: it represents the same thing. But it has a different sense: the doctor's criteria to determine whether something is arthritis and your criteria are different[10]. The doctor relies on technical criteria, you rely on metalinguistic criteria. For the doctor the fact that makes a condition "arthritis" is that it is associated with certain symptoms, causes and consequences, for you it is the fact that doctors call it "arthritis" and identify it as such. In this sense, your belief that you have arthritis would be semantically deferential.

Except in the rare cases in which *we* are the experts, our beliefs are not transparent to us. They are deferential in these two senses. Both type of deference are justified by what we have called the law of least effort. Assessing the truth of propositions individually and using transparent (i.e. not deferential or opaque) concepts is often uselessly effortful. Deference allows us to obtain the same intellectual and practical profit, while minimising intellectual and practical costs. Given time and ability constraints, relying on the doctor's credentials and reputation is arguably the most efficient strategy to determine the truth value of the proposition that you have arthritis (as opposed to trying to acquire an expertise in rheumatology of your own). And even if you don't know the *technical concept* of arthritis, knowing the *word* "arthritis" is enough for you to go to the pharmacy and successfully ask for the medicines you need.

1.3. *Perfect individual rationality*

Not only do we need others to have access to propositional and conceptual information, we need others also to process that information and make decisions. In the last sixty years, cognitive psychology has more and more convincingly shown that, *pace* Aristotle, humans are not rational animals (KAHNEMAN 1994, STICH 1990). Individuals are quite bad at basic logic (EVANS 2002), probability (KAHNEMAN & TVERSKY 1972. TVERSKY & KAHNEMAN 1983), and rational decision-making (KAHNEMAN et al. 1982). We often take mental shortcuts that save us from wasting energy making autonomous decisions using the rules of logic. And these shortcuts often consist of relying on others: we rely on others to decide whether an inference is valid or not. Judges, who often consider themselves autonomous rational thinkers, are not exempt from this. Consider this quote from Judge Hutchinson, explaining how he reaches a solution in complex cases:

«[W]hen the case is difficult or involved, and turns upon a hairsbreadth of law or of fact [...] I, after

[10] Sense and reference are two possible meanings of "meaning". The "sense" of a word is its mode of presentation, the way it presents what it stands for, its reference.

> canvassing all the available material at my command, and duly cogitating upon it, give my imagination play, and brooding over the cause, wait for the feeling, the hunch—that intuitive flash of understanding which makes the jump-spark connection between question and decision, and at the point where the path is darkest for the judicial feet, sheds its light along the way»[11].

Where do these "judicial hunches" come from? Are they the result of unconscious legal syllogisms made by judges on the basis of their extensive knowledge of the law, which merely anticipate what will be made explicit in the judicial decision? Or are hunches selecting the solution to the case in some other non-logically justified way? Various strands of studies in cognitive science suggest that the latter may be the case. The way judges publicly justify their decisions often doesn't have much to do with the way they reached them in the first place.

An interesting study of bail decisions in the UK shows that judicial hunches may simply be the result of social influence: by default, judges decide based on how other legal officials they trust have previously decided or treated the same case. In bail decisions, judges are supposed to decide whether to grant the defendant unconditional bail or impose a sentence such as imprisonment on the basis of the defendant's character, trustworthiness to appear in court, not tampering with witnesses and not committing further offences. However, this study of two London courts (Court A and Court B) suggests that judges often base their decisions not on the law and the facts of the case, but on decisions previously made by the police, the Crown Prosecution Service and other judges. The study found that the average time spent by each judge on each case was less than 10 minutes. 95% of bail decisions in Court A were made using a fast and frugal heuristic based on 1) whether the prosecution had requested conditional bail, 2) whether a previous court had imposed conditions, 3) whether the police had imposed conditions or remanded the defendant in custody. Bail decisions in Court B followed a similar heuristic.

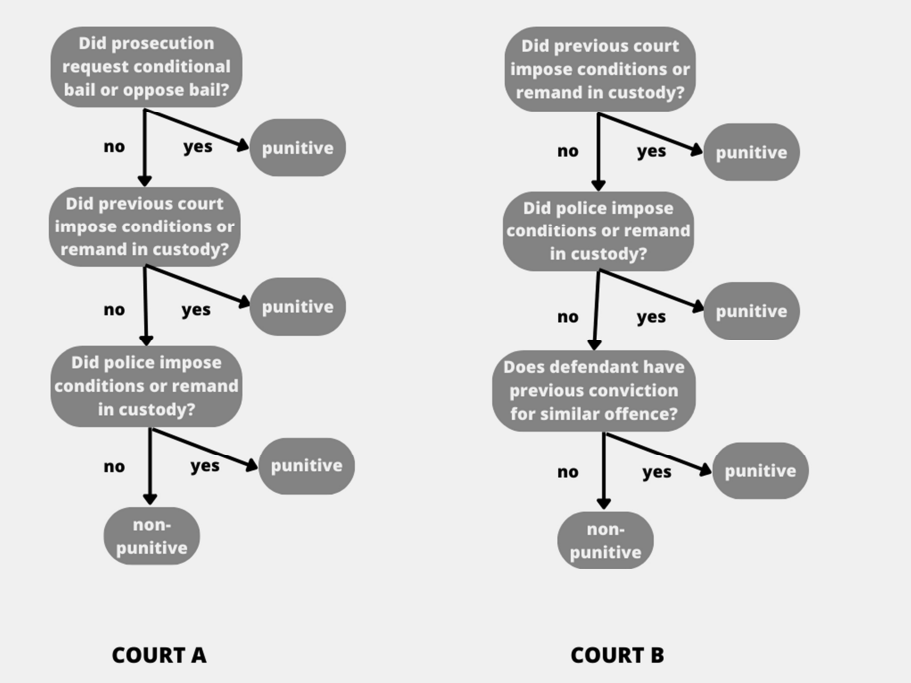


FIGURE 1. This diagram is taken from GIGERENZER 2008, which in turn is an adapted version of the diagram in the original published study, DHAMI 2003.

[11] Quoted in HAIDT 2013, 868.

According to the social psychologist Johnathan Haidt, the process of judicial decision-making reflects our general way of dealing with moral and practical issues, which is governed much more by intuitions of social origin than by logic. Haidt describes this process with the diagram reproduced in Figure 2, where A and B represent two different people, the circles represent different stages of their thinking, and the lines represent the causal links that may exist between these stages.

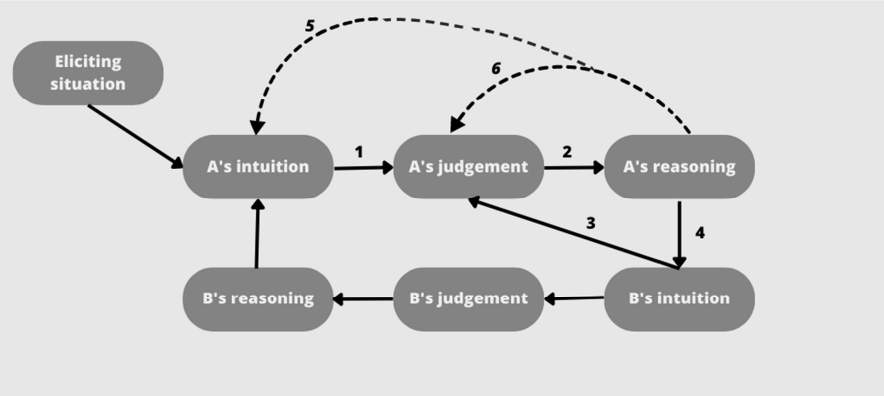


FIGURE 2. The social intuitionist model of moral judgment, taken from HAIDT 2001.

The non-dotted lines indicate the *most usual* causal links: *1. Our intuitions usually cause our judgement*; *2. Our judgement usually causes our (post hoc) reasoning; 3. Our judgement usually causes other people's intuitions to change; 4. Our (post hoc) reasoning also usually causes other people's intuitions to change*. The dotted lines indicate *rare* causal links: *5. Our private reasoning rarely causes a change in our own intuitions; 6. Our private reasoning rarely causes a change in our judgement* (HAIDT 2001, 815). So, according to this model, when a judge decides a case, his or her decision is *not* usually caused by a careful examination of the facts of the case and the applicable legal rules. What happens is that the judge has an unconscious intuition that leads him or her to adopt a certain hypothesis for a solution. Having reached a solution, the judge looks for facts and rules to justify it. But where does the initial unconscious intuition come from? Haidt's hypothesis is that the origin of intuition is at least partly social: we tend to align our judgement with the judgements of other people on the same issue.

This however is not always the case. When a piece of information coming from others conflicts with some of our pre-existing beliefs we often tend to disregard it. Confirmation bias is probably one of the best known and explained cognitive biases. It consists of the tendency of people to seek out, interpret and remember information in a way that confirms their pre-existing beliefs or hypotheses, while ignoring or discounting evidence that contradicts them. On the face of it, this mix of initial credulity and subsequent confirmation bias may have perverse consequences for the functioning of, for example, the criminal justice system. A prosecutor will tend to readily endorse an investigative hypothesis that does not conflict with his or her prior beliefs. In particular, he or she will tend to believe the testimony of the first witnesses interviewed. Once this hypothesis is established, however, he or she will do everything possible to confirm it by seeking additional evidence in its favour. Once the prosecutor has prosecuted, the judge will also tend to consider the case on the assumption that the investigative hypothesis is well-founded, contrary to the presumption of innocence (*in dubio pro reo*).

Although what I have said so far paints a rather discouraging picture of the human capacity to make good decisions, let alone to make a legal system work rationally, things may not be as bad as one might think. According to Hugo Mercier and Dan Sperber (MERCIER & SPERBER

2011, MERCIER & SPERBER 2017, MERCIER 2020), many behaviours that may seem dysfunctional when viewed at an individual level are actually quite functional when viewed from the perspective of the larger human community. According to the argumentative theory of reason, the theory developed by these two authors, our tendency to prefer evidence that confirms our pre-existing beliefs and ignore evidence that contradicts them, for example, stems from our evolutionarily developed ability to make the best possible case for a certain thesis, which in turn contributes to a greater efficiency.

According to Sperber and Mercier, our tendency to favour evidence that confirms our existing beliefs and ignore evidence that contradicts them suggests that our reasoning skills have evolved to persuade others and take sides in debates, rather than to prioritise our own accuracy and truth. Paradoxical as this may seem, confirmation bias is a more evolutionarily successful trait of the tendency to seek the truth in an unbiased and disinterested manner. Human communities that sought knowledge by dividing their cognitive labour into biased teams, favouring opposing theories and unwilling to abandon their thesis except with overwhelming arguments, proved evolutionarily more successful than communities formed by unbiased individuals genuinely interested in the search for truth. Individual irrationality produces collective rationality.

The division of labour in the exercise of rationality leads to resource savings, just as it does in various other aspects of life. The adversarial trial is a good example of this phenomenon. Each party in an adversarial trial has a bias towards their own position and a desire to win the case. This bias motivates them to carefully select and present evidence that supports their position and to challenge evidence presented by the opposing party. According to Sperber and Mercier's theory, this process of adversarial argument, in which both sides present their best arguments and evidence, can actually lead to more accurate and informed decisions by the judge. The process of adversarial argument allows for a full and thorough examination of the evidence, with each party scrutinising the other's evidence and arguments for weaknesses and flaws. In this way, the parties act as checks and balances on each other, ensuring that all relevant information is presented and evaluated by the judge. This process can help to mitigate the effects of individual bias by providing multiple perspectives on the evidence and arguments. In addition, the presence of a neutral third party (the judge) who is not invested in either side's position further contributes to the overall rationality of the process. In brief, even if confirmation bias can distort evaluations and attitudes and perpetuate false beliefs at the individual level, it can produce a division of epistemic roles that results in greater rationality at the collective level.

2. *Cognitive perfectionism in law: three legal myths*

The dramatic mismatch between our folk-psychological assumptions and the division of cognitive labour described in the previous section has a profound impact on legal practice. In this section, we will deal with cognitive perfectionism in law. The cognitive perfectionism of a norm or institution can be diagnosed by looking at its *ratio legis*, that is the official purpose that is believed to justify that norm or institution. We can say that a rule or institution is cognitively perfectionist if the result the legislator or the legal community intends it to achieve would only be achieved by it if the agents addressed were cognitively perfect, or at least endowed with rationality or knowledge far beyond the actual. For example, a norm is cognitively perfectionist if it assigns the right, duty, power, responsibility, etc. to φ, and φ is an action that a normal human being is incapable of performing due to his limited capacity to be rational or to accumulate and process knowledge. More generally the rule or institution is cognitively perfectionist if it assigns legal positions with respect to behaviours that can only be performed by unrealistically rational or knowledgeable agents.

There are two forms or two aspects of cognitive perfectionism. We can call them *rationality perfectionism* and *knowledge perfectionism*. Rationality perfectionism is the false assumption that agents

are capable of perfect individual rationality, whereas knowledge perfectionism is the false assumption that agents are individually capable of acquiring and processing unlimited or at least extremely large amounts of information. While rationality perfectionism is the aspect that has received the most attention in legal scholarship, knowledge perfectionism is a relatively neglected topic.

The idea that sanctions are the best way to prevent violations of the law is rationality perfectionist. It presupposes that violations (or non-violations) are ordinarily the result of a rational cost-benefit analysis. This idea has been criticised mainly by an extremely influential theory developed by Cass Sunstein and Richard Thaler on the basis of Herbert Simon's theory of bounded rationality (SIMON 1955), the work of Kahneman and Tversky on heuristics and biases (KAHNEMAN 2003, KAHNEMAN 2011), and the nascent behavioural economics (TEICHMAN & ZAMIR 2014). According to Sunstein and Thaler, traditional legal norms systematically fail to motivate their recipients because they erroneously assume their perfect rationality. According to them, legal norms traditionally set out to motivate recipients with rewards and sanctions, postulating that they in all aspects of life (from business to crime, from family life to interaction with the public administration) decide on the basis of a rational cost-benefit calculation. The fact that real agents tend to be irrational therefore renders law ineffective. Alternatively, Sunstein and Thaler propose techniques for influencing citizens that are based on their irrationality, which they call "nudges". Nudges focus on influencing individuals based on the emotional, automatic, and unconscious mechanisms of the human mind (often referred to as "System 1" thinking), rather than relying solely on conscious, rational deliberation, which requires effort and is less frequently activated (referred to as "System 2" thinking)[12]. An example of a nudge is the modification of organ donation legislation from an opt-in system, where explicit consent is required for donation, to an opt-out system, where explicit dissent is necessary to avoid donation. Traditional legislators may assume that this change would have no impact, as they assume the addressees of norms to act as *"homines oeconomici"*, and thus not to be influenced the status quo bias. However, a legislator well-versed in psychology knows that "choice architecture" plays a crucial role in shaping outcomes. Most agents will not activate their rationality (their System 2) to determine the fate of their organs. Guided by their intuitive autopilot (their System 1), they will simply not choose and accept the default option.

The literature has extensively explored the problem of legislators placing unwarranted trust in the rationality of individuals, but no comparable attention has been paid to the unwarranted trust placed in people's ability to individually process large amounts of knowledge. Knowledge perfectionism deserves to be examined separately because, as we shall see by discussing a few examples, it has effects independent of rationality perfectionism. In what follows I will provide three examples of the consequences of cognitive perfectionism in law, three legal myths, so to speak: The myth of legality, the myth of consent and the myth of the judge as gatekeeper. In discussing these examples, I will attempt to emphasise the aspects of knowledge perfectionism rather than those of rationality perfectionism, as I believe that the former, which the literature has neglected, may be of greater interest to the reader than the latter.

I use the word myths rather than legal fictions because the term "legal fiction" is usually employed to refer to descriptions of reality that legal practice explicitly treats as true, even though the agents involved in the practice know that they are not true (DEL MAR 2015). What I want to talk about are false representations of reality that are presupposed by practice and that agents often take to be true. An example of legal fiction is the idea that a corporation is a person. An example of a myth in the sense I mean is the mediaeval idea that God gave the emperor his legitimacy to rule. The ones I will discuss are myths and not fictions as they are not normally recognised as false by legal actors.

[12] THALER & SUNSTEIN 2003, THALER & SUNSTEIN 2008, SUNSTEIN 2011, SUNSTEIN 2013a, SUNSTEIN 2013b, THALER et al. 2013, SUNSTEIN 2014, SUNSTEIN 2015, SUNSTEIN 2017.

2.1. *The myth of legality* (ignorantia legis non excusat)

The first myth I would like to talk about to exemplify cognitive perfectionism is the very idea of legality. Our naive understanding of how legality works is associated with a specific folk psychological way of describing the relation between the contents of a legal system and the mental states of the community living under it. It is the view according to which the contents of a legal system correspond to mental representations which are shared by both citizens and legal officials.

This idea is well captured by John Searle's description of institutional reality[13]. According to Searle, human beings *constitute* institutional entities, including legal ones (such as valid contracts, laws and nation states) by collectively accepting them as existent. Institutions are identified by what Searle calls their *status functions*, which are functions they perform in virtue of being collectively mentally represented as having a certain status. In everyday life, we interact with all kinds of objects or people that we value in virtue of some kind of *function* they perform in society. Many times, the function is performed thanks to the physical structure or properties of the object or the intrinsic features or abilities of the person. Take the example of cars, hammers or pencils, on the one hand, or physicians, construction workers and janitors, on the other. But, according to Searle, there are other equally important entities that play equally important roles in our lives which perform their functions merely in virtue of being collectively recognised, through constitutive rules, as having a certain status. Law and all legal entities belong to this second category. Searle's favourite example is money. If we had a collective amnesia and we all forgot that the pieces of paper we have in our wallets are legally valid money, these would instantaneously cease to *function as* money. They wouldn't work anymore as means of exchange or as a reserve of economic value. They would cease to have all the interesting properties that normal pieces of paper which are not money lack. The same is true for contracts, nation states, university departments, as well as presidents, judges, and policemen. The key social functions of all these entities are performed only in virtue of the collective recognition of the status that society at large represents them as having.

Although Searle's theory describes our relationship with institutional reality in a sufficiently convincing manner, it totally overlooks the importance of the division of cognitive labour in the mental representation of legal concepts (ROVERSI et al. 2023). Institutional entities such as loan contracts, presidents, judges, etc. exist even though only an extremely small minority of individuals in the community have access to the constitutive rules of the corresponding status functions. What is important and problematic for our purposes is that, according to this view, the physiological functioning of a legal system presupposes that the members of the community governed by it can have shared mental representations of the *same* legal content. This idea, even if not necessarily with the analytical sophistications of Searle's terminology, tends to be presupposed by lawyers, when it comes to defining the ideal of the rule of law.

The idea of the rule of law is often traced back to Aristotle's Politics, where he discusses the problem of whether it is better to be ruled by men or by laws. His answer is that «it is proper for the laws good when rightly laid down to be sovereign, while the ruler or rulers in office should have supreme powers over matters as to which the laws are quite unable to pronounce with precision because of the difficulty of making a general rule to cover all cases» (ARISTOTLE 1932, 1282b; cf. WALDRON 2016). Since Aristotle, two main advantages of rule of law over rule of men have been traditionally pointed out. The first is that if a rule pre-exists the case to be settled, the decision over its regulation will be less likely to be arbitrary: it will more likely be

[13] SEARLE 1997, SEARLE 2010. Hart more realistically believes 1) that only a very small subset of the normative material that forms a legal system is the object of shared knowledge (the rule of recognition), 2) that the repositories of this shared knowledge are only a rather small subset of citizens (mainly legal officials) (HART 2012, 100 ff.). For a comparison between Searle and Hart on this point see ROVERSI et al. 2023.

based on reason and equality rather than extemporary whim or selfish desires or discriminatory preferences of the rulers. The second reason is that people will be able to conform to what the rule requires before the case regulated by it occurs. To put it differently, the rule of law can be seen as promoting two main complementary goods, or as having two main functions, one that we might call an *epistemic function* (communicating to citizens what is required of them) and one that we might call a *control function* (preventing officials from exercising arbitrary power in the service of their own interests rather than the general interest). The two functions can be conceptually separated (ENDICOTT 1999; FLETCHER 1998, 207). We could imagine a legal system that is perfectly capable of protecting citizens from the arbitrary decisions of officials, where decisions are made by tossing coins: in this case, the rule of law's control function would be protected, but not its epistemic function. On the other hand, one could imagine a legal system in which officials have private, selfish interests that they pursue in a constantly predictable way: in this system, their decisions would be perfectly predictable by citizens, but this would obviously not protect them from arbitrary government.

The stress on the epistemic function of the rule of law and the assumption that the two functions are somehow inseparable from each other and mutually supportive is arguably knowledge perfectionist. It assumes that control over the work of government should be carried out "from below" without considering the hypothesis that the control function could be better achieved through other forms of institutional design. Both advantages are predicated on the alleged informational consequences of regulating society with public, standing rules rather than case-by-case decrees. The rule of law creates an informational asymmetry, *blinding* the rulers and making us citizens able to *see* and determine how public action will affect our lives. It prevents the rulers and allows citizens to see the future. When I decide to violate a smoking ban, I can *know* the practical consequences that will follow, for instance that I will have a legal obligation to pay a 100 euro fine. The legislator (the creator of the smoking ban) on the other hand was *not* able to *know* that *I*, or any other specific person, would have to pay 100 euro. Standing public rules, unlike case-by-case decrees of the rulers, make the conditions of their application explicit beforehand. This prevents rulers from predicting who will profit or benefit from public action. At the same time, it allows us citizens to predict public action, criticise it based on the law, criticise the law itself, and finally act strategically based on our knowledge of the possible legal consequences of our actions. Under the rule of law, if I act unlawfully, I *know* (or I am at least able to know) that I do and what the possible consequences of this will be, and I *know* (or I am at least able to know) what it would take to avoid illegality and its consequences. All of this is accompanied by the idea that ignorance of the law does not excuse, that is, that everyone is burdened with the onus of knowing the law.

This account of the good that the rule of law is supposed to secure is characterised by knowledge perfectionism in two respects. First, because the epistemic function of the rule of law cannot realistically be realised in the way I have just described. None of us can know all the laws of a legal system in detail, not even if we activate our System 2, not even if we study all our lives to achieve this goal. Not only that, no one who has come to a certain legal conclusion with respect to a specific legal question can exclude that the conclusion he or she has come to is erroneous. This is because law is always defeasible and every legal conclusion based on a certain set of premises can change if a further premise is added to that set. And the possibility that a further premise may have to be added can never be ruled out since no one ever has complete knowledge of the sources of law of a certain legal system. Second, it is by no means certain that the pursuit of the epistemic function contributes to the realisation of the control function. The control function rule of law requires a multitude of highly specific laws, drafted in a technical manner, and subject to interpretation by expert jurists organised in communities that ensure mutual accountability (cf. AINIS 2002, 137; COLEMAN 1998). On the contrary, the epistemic function calls for a limited number of laws, characterised by simplicity and approximation, easily understandable by all.

2.2. *The myth of consent* (volenti non fit iniuria)

A second myth, relevant to all areas of law, is the myth of consent, i.e. the idea that a manifestation of our will is sufficient to adequately represent our interests and is proof that we have accepted the consequences to which we have declared our consent. Just consider for example the last time you expressed a legally valid consent on the internet. This was probably the last time you clicked "I accept" on some cookie policy on your phone. All websites owned in the EU, or targeted towards EU citizens, are now expected to comply with a law that requires them to ask for permission to use cookies. The idea behind the law is that the best way to protect citizens is to appeal to their rational faculties and inform them of what will happen to their data. The law seems to presuppose that if citizens accept the websites' cookie policy, they will have understood, rationally assessed what is in their best interest and accepted the terms and conditions. The law somehow seems to presuppose that the fact of clicking that button is sufficient evidence that the clicker has been informed of the legal consequences that will follow for the act of clicking. This is so even in cases where the amount of information to which one is "consenting" spans thousands of pages written in legalese that only a lawyer with a background in IT could reasonably understand. You are probably all familiar with this cookie law. If a website receives visitors from within the European Union, the *ePrivacy Directive* (EU cookie law) requires that it can only use cookies and trackers with their explicit consent. This is why every day, several times a day, you click 'I accept to allow cookies to be downloaded onto your device (phone, tablet or computer). However, it is likely that many of you do not know exactly what a cookie is or what it is used for.

The purpose (the *ratio legis*) of requiring consent is clearly protecting the consenting party. But in most cases consent (or dissent) cannot but be blind. Most of us would probably feel safer if the terms and conditions were approved by some kind of third-party expert body we can trust. The expectation that we are all-knowing beings thus leads the law to select a suboptimal strategy to protect our interests. This is a good example of how knowledge perfectionism in law would be capable of harming us even if we were *homines oeconomici*. Even when we are able to keep our System 2 active, we are often unable to match the model agent that law presupposes. For example, even if I were perfectly rational, this would not help me in understanding 200 pages of terms and conditions written in IT legalese. It would actually be irrational for me to acquire all the information necessary to understand it. The cost of understanding the cookie policy is higher than the expected utility derived from understanding it (STRAHILEVITZ & KUGLER 2016, MCDONALD & CRANOR 2008, REIDENBERG et al. 2014).

The myth of consent manifests itself in many aspects of legal practice, particularly in contract law, but it is precisely in the context of regulating our online lives that it risks producing the most damage in the near future[14]. This is because the digital services we use on a daily basis and blindly consent to without regard to the consequences[15] are increasingly the same tools through which we see and interpret the reality[16] we use to make decisions in our economic, political and personal lives. Although the Internet is fundamental to our lives, we do not always think about the kind of economic transaction we are involved in every time we consume it. While we are all aware of who pays for the service we receive when we go to a restaurant or a public library this is not so clear when we are on

[14] The reflections in the following pages on the power of influence that underlies the Internet economy are partly taken from an unpublished paper written with Marta Taroni: TARONI & UBERTONE 2021.

[15] Consent to the terms and conditions of online services is often extorted by providers using manipulative techniques based on biases called dark patterns, see KOCYIGIT et al. 2022.

[16] We increasingly perceive reality through the perspective of (often politically polarising) "filter bubbles" that are theoretically tailored to our "interests", but which are actually designed not to maximise our ability to understand, but to maximise our engagement with platforms and our consumption.

Google, Instagram, Twitter, TikTok, Facebook, Tinder, Grindr, Snapchat, LinkedIn, or YouTube. We pay for the restaurant, our taxes pay for the public library but who pays for services provided by the web giants we interact with every day? Although we use the internet far more often than we go to restaurants or libraries, only a minority of us can give a comprehensive answer to this simple question (although a clear answer to this question can be deduced from the terms of the contracts to which we are supposed to be parties). The answer in a nutshell is the following. Economic actors operating in the digital sector collect data about us: some of this data is used to improve products and services, while other is processed to obtain predictive algorithms, i.e. algorithms capable of predicting people's behaviour. By extracting behavioural data and profiling individuals, it is possible to target consumers with the right stimulus at the right time to induce behaviour that will benefit the platform and advertisers. A critical factor in the functioning of this industry is the use of behavioural psychology techniques to induce users to 1) devote as much time and attention as possible to the content proposed by the platforms (maximisation of time on screen); 2) provide (more) data about themselves; 3) purchase the goods or services proposed by the advertisers; 4) generally, behave online and offline, either by action or omission, in accordance with the behaviour desired by the advertisers or the platform. According to Shoshana Zuboff, those who claim that in the Internet economy the product "is us" are wrong. We, the users and consumers, are the means to obtain the predictive products and tools for large-scale manipulation (ZUBOFF 2019). These are then used to induce us to buy certain products and services or in some cases (think of the Cambridge Analytica scandal) to influence the exercise of our political rights, such as the right to vote.

2.3. *The myth of the judge as "gatekeeper"* (iudex peritus peritorum)

The final myth I would like to cite as an example of cognitive perfectionism in law relates to a more specific topic than the previous two: the use of science and expert testimony in the context of the courtroom[17]. According to evidence scholar Ronald Allen, there are two possible models of expert evidence: a deferential model and an educational model (ALLEN & MILLER 1993, ALLEN 2018). According to the deferential model, judges can limit themselves to verifying that the people called upon to testify as expert have sufficient credentials, have no conflict of interest and their theories are accepted by the scientific community. According to the deferential model, fact-finders will be allowed to accept the content of the expert testimony as true just on the basis of credentials, and even if they do not understand the scientific reasons that support it. According to the educational model, on the contrary, judges must be educated by experts in order to critically scrutinise the basis of the experts' opinions. The judge should be a "gatekeeper", invested with the mission to keep "junk science" out of the courtroom. When admitting evidence, the judge should always verify that the expert opinion supports fact-finding with substantial argument, not based on the mere authority of the expert and comprehensible to an ordinary person. According to this model, fact-finders, when assessing the evidence, must be educated by the expert, must understand the reasoning and theories underlying the expert's statements, and should disregard any appeal to authority. Throughout the twentieth century - but especially since the 1990s, following the famous US case of *Daubert v. Merrell Dow Pharmaceuticals* - the educational model has become dominant in both scholarship and jurisprudence (TARUFFO 2016, 337-339). Some have even argued that there is a need to invest in training judges in non-legal subjects so that they are able to scrutinise expert evidence without falling into a form of blind deference. However, there is a latent paradox in this kind of view. If fact-finders were to acquire sufficient scientific

[17] This is a subject I have already dealt with elsewhere, although in that context I had not treated it under the label of cognitive perfectionism. Here I will merely mention some of the arguments presented in two articles and a book, referring back to them for further discussion: UBERTONE 2019, UBERTONE 2022a, UBERTONE 2022b. On deference to experts in courts also see: CANALE 2021.

knowledge to check the actual basis of the expert's testimony, they would probably not need the expert's help in understanding the facts of the case.

As I have argued elsewhere, proponents of the educational model burden judges and fact-finders with a responsibility that could only be fulfilled by cognitively superhuman beings and fail to consider how the division of cognitive labour, and thus deference to experts, is an ineradicable feature of human cognition. This is a blatant example of cognitive perfectionism. The fact that this is a responsibility that cannot reasonably be assumed by an average judge is well understood by the judge who had to apply the educational model standard set by Daubert for the first time in a concrete case, i.e. the district judge who in the Daubert case had stayed the judgement and remanded the question of the admissibility standard to the Supreme Court. In a passage in the remand judgement, he expresses some scepticism about his own ability to personally examine the scientific basis of the evidence, and polemically ironises as follows

> «As we read the Supreme Court's teaching in Daubert, [...] though we are largely untrained in science and certainly no match for any of the witnesses whose testimony we are reviewing, it is our responsibility to determine whether those experts' proposed testimony amounts to "scientific knowledge", constitutes "good science", and was "derived by the scientific method" [...] Our responsibility, then, unless we badly misread the Supreme Court's opinion, is to resolve disputes among respected, well-credentialed scientists about matters squarely within their expertise, in areas where there is no scientific consensus as to what is and what is not "good science" and occasionally to reject such expert testimony because it was not "derived by the scientific method". Mindful of our position in the hierarchy of the federal judiciary, we take a deep breath and proceed with this heady task»[18].

The educational model, motivated by a concern to avoid attributing epistemic authority to experts, attributes wholly unwarranted epistemic authority to judges. Proponents of the educational model emphasise that by not deferring to experts, courts eliminate the possibility of committing fallacies of authority, but they fail to recognise that denying epistemic authority to experts logically implies the need to assign epistemic authority to the judge and jury. A version of this argument was put forward by the cognitive scientist George Lakoff in his polemical paper on Daubert. According to Lakoff (LAKOFF 2005), *Daubert* helped to reinforce, both in the public mind and in the courtroom, an authoritarian view of the judge and his role in the trial. According to this conception, the judge is a subject in whom the legal system must place unconditional trust, and who is better able than those recognised as scientists to verify the conformity of theories with the scientific method. Lakoff speaks of a mutation in the conceptual framework through which we interpret the roles of the various actors involved in the process and their interrelationships.

Any attempt to adhere to the educational model produces a regression to infinity. If, in order to believe what expert opinion tells us, we had to check the validity of the arguments on which it is based, then, for the sake of consistency, we would have to give the same treatment to the expert authorities on which it is based, and so on *ad infinitum*. Every discipline is made up of innumerable chains of deference, which obviously cannot be fully traced in the course of a trial (HARDWIG 1985, HARDWIG 1991). Requiring judges and jurors to educate themselves to the point of replicating the epistemic work of the expert will cause them to devote proportionately less time to epistemic tasks in the trial that can only be performed by them and not by experts, such as comparing the expert testimony with that of other witnesses or selecting the legally relevant elements of the testimony. Moreover, if judges and jurors were really to be required to have access to the reasons that the expert himself gives for his conviction, this would probably take an absolutely unbearable amount of time.

[18] *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311 (9th Cir. 1995).

We need experts only if we can use them as epistemic authorities. If we reject epistemic deference, we also ipso facto renounce the saving of resources that calling an expert to court should entail. This argument can be framed as a kind of transcendental argument. Experts are sources of knowledge. But a necessary condition for them to be sources of knowledge is that we can draw more from them than we must already possess in order to identify them as such. For this to happen, it is necessary that the test of expertise, whatever it may be, does not require independent possession of the information that the expert testimony is supposed to produce. If the test of expertise required such information, experts could never be sources of information.

3. *Conclusion: is cognitive perfectionism essential to law?*

"Ought" implies "can". If the responsibilities assigned to us by law are impossible to fulfil, there seems to be something deeply wrong with the rules that assign them to us. Similarly, there seems to be something wrong with a legal system that takes care to protect as rights prerogatives from which none of us can actually benefit because of our limitations (like, for example, "consenting" to the technical details of how our cookies are managed), and does not instead take care to protect us from the risks that arise from those very limitations (like the mental health risks of being systematically manipulated by the devices we use everyday). For all of these reasons, I can well imagine that the reader will have interpreted my account of cognitive perfectionism as a criticism of current legal practice. I believe, however, that underlying the unrealistic assumptions of legal practice concerning the functioning of the human mind that I have criticised there may be good reasons and that the possibility of abandoning those assumptions should be the subject of a serious philosophical discussion for which this article is by no means intended to be a substitute. In conclusion, therefore, I would like to present some sketchy considerations in favour of cognitive perfectionism, in order to suggest the kind of discussion that I think legal philosophy should develop in order to explore this issue. The issue, I think, is related to various classical problems of the subject: the nature of legal authority, the nomodynamic character of a legal system and the conflicting relationship between the rational and authoritative components of legal decisions emphasised respectively by natural law theory and legal positivism.

According to a highly influential legal positivist, Joseph Raz, the main service that legal rules provide to a society is the exercise of "authority", in the technical sense in which he uses that term (Raz 1985). Authority, for Raz, is the ability to provide a particular kind of reason for action, reasons that he calls "exclusionary," that is, reasons that exclude the possibility of other reasons being considered in deciding what to do. If I toss a coin to decide whether to go to the cinema to see film A or film B, my commitment to regard the result of the coin toss as authoritative implies my commitment, by virtue of the toss I have made, to exclude the consideration of other, first-order, otherwise relevant reasons. If I decide to give authority to the coin toss, I will choose the film to watch *solely* on the basis of whether the result was heads or tails. The fact that it came up heads or tails will not be an additional reason that needs to be compared and balanced with other reasons, such as whether film A is by a better director than film B, or whether film B is shown in a cinema closer to home. The result of the toss is a reason that supersedes the others.

Authority, defined in this way, is clearly not always reliable or useful. It all depends on who or what you choose to invest in as an authority. Very often, for example, it is not a rational choice to base one's actions or beliefs on the authority of a coin toss. Relying on authority is only rational if the authority is legitimate. According to Raz, an authority is legitimate if we are better able to achieve our ends by following its instructions than by not following them, or, more precisely, if we are better able to satisfy our "first-order" reasons by obeying the authority than by evaluating those reasons directly. For example, if I rely on the advice of a friend to decide whether to watch film A or film B, my friend is a legitimate authority only if he or she is

better able than I am to determine which film I will like best. This may be the case if the friend has similar tastes to mine and has seen both films, whereas I have not.

The interesting aspect of all this for our purposes is that to recognise an authority as legitimate is, on the one hand, to recognise that it is competent to rationally balance first-order reasons and, on the other hand, to accept that the rationality of that balancing is not open to challenge once the authority has issued its directives. We can then hypothesise that cognitive perfectionism may in some cases be a necessary consequence of the recognition of certain decision-makers as sources of authority in Raz's sense. Think of the authority of the consent expressed by the contracting parties: they are recognised as fully sovereign in the management of their rights, they are recognised as the highest authority in the representation of their own interests (except in cases of legal incapacity). The fact that the agreement between them is recognised as legally valid also implies that it is endowed with authority, and therefore that the individual parties to the agreement are endowed, within the limits set by law, with an incontestable authority to balance the reasons underlying the agreement.

Legal systems are characterised by a particular way of dealing with problems, sometimes described by philosophers as "nomodynamic" (KELSEN, 1949, 110 ff.). By this is meant that, unlike moral rules, legal rules systematically transform problems about what can or must be done into problems about who has the power to decide – who has the authority to decide – what can or must be done. This makes them, as Bruno Celano has pointed out (CELANO 2017, 230; CELANO 2018), systematically vulnerable to a paradox. The most striking example of this paradox relates to the concept of *res iudicata*, the unquestionable authority of final judgments. A legal rule or a decision adopted by the competent authority in accordance with the prescribed legal procedures, but which is wrong on the merits (contrary to the law or based on incorrect factual assumptions) may nevertheless produce its typical legal effects. The law treats judges of final instance as if they were infallible and, in this sense, cognitively perfect: in the name of stability of decisions, it accepts their judgments as necessarily correct. This always makes it possible for the law to accept as true propositions that are inconsistent with each other. For example, substantive law might provide that all persons who φ commit an unlawful act, and that Bill, who committed φ but was erroneously acquitted by a judge in a final judgement, did not commit an unlawful act. The recognition of authority always implies the possibility of such paradoxes. The same kind of unquestionability also manifests itself in other areas of law where authority is attributed to acts other than judicial decisions. For example, the law may provide in general terms that the content of contracts must of course be known to the parties (contracts are actually often thought of as the meeting of the minds converging on the representation on the same content), but at the same time it may provide for procedures which consider as the "content of a contract" the meaning of a text or set of texts which not only the parties may not know, but which may even be impossible for one or both contracting parties to know (think of the CEO of a large company who signs hundreds of contracts every day comprising tens of thousands of pages).

Recognizing legal authorities allows society to curb the discretion of political decision-makers and ensures social peace, but perhaps comes at the cost of accepting a certain degree of intrinsic irrationality in the legal system. The question arises as to what extent this irrationality is necessary, whether the service that legal authority essentially renders is compatible with a greater realism in the representation of agents and their faculties of understanding and action, whether it is possible for law to adopt a more realistic image of the human mind and at the same time ensure that inviolable areas of freedom and legal certainty are recognised. Would a law "tailored" to the real cognitive capacities of agents still be capable of exerting authority? Do we need law precisely to avoid having to discuss directly what behaviour is rational or irrational (in terms of first-order reasons)? Would eliminating cognitive perfectionism necessarily mean abandoning the rule of law and exposing ourselves to the dangers of the rule of men? Would it mean making law dangerously paternalistic or illiberal? As promised, the purpose of this paper was not to give answers but only to formulate questions, which is why I think I can stop here.

References

ADLER J. 2012. *Epistemological Problems of Testimony*, in ZALTA E. N. (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition). Available at: https://plato.stanford.edu/archives/fall2012/entries/testimony-episprob/.

AINIS M. 2002. *La legge oscura. Come e perché non funziona*, Laterza.

ALLEN R. 2018. *Fiddling While Rome Burns: The Story of the Federal Rules and Experts*, in «Fordham Law Review», 86, 4, 1551 ff.

ALLEN R., MILLER J. 1993. *The Common Law Theory of Experts: Deference or Education?*, in «Northwestern University Law Review», 87, 4, 1131 ff.

ARISTOTLE 1932. *Politics*, Harvard University Press (tr. by H. Rackham).

AUDI R. 2015. *Folk Psychology*, in *The Cambridge Dictionary of Philosophy*, Cambridge University Press, 365 ff.

BRENNAN J. 2016. *Against Democracy*, Princeton University Press.

BRIGAGLIA M. 2015. *Direzione normativa e teoria della mente*, in «Ragion pratica», 44, 103 ff.

BUBLITZ C. 2021. *Rights as Rationalizations? Psychological Debunking of Beliefs About Human Rights*, in «Legal Theory», 27, 97 ff.

BURGE T. 1979. *Individualism and the Mental*, in «Midwest Studies in Philosophy», 4, 73 ff.

CANALE D. 2021. *The Opacity of Law: On the Hidden Impact of Experts' Opinion on Legal Decision-Making*, in «Law and Philosophy», 40, 509 ff.

CELANO B. 2013. *Publicity and the Rule of Law*, in GREEN L., LEITER B. (eds.), *Oxford Studies in Philosophy of Law: Volume 2*, Oxford University Press, 121 ff.

CELANO B. 2017. *Due problemi aperti della teoria dell'interpretazione giuridica*, Mucchi.

CELANO B. 2018. *Lezioni di filosofia del diritto. Costituzionalismo, stato di diritto, codificazione, positivismo giuridico*, Giappichelli.

CHALMERS D. 2002. *The Components of Content*, in ID., *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, 608 ff.

CHALMERS D. 2003. *The Nature of Narrow Content*, in «Philosophical Issues», 13, 46 ff.

CLARK A., CHALMERS D., 1998, *The Extended Mind*, in «Analysis», LVIII, 1, 17.

COADY C. A. J. 1994a. *Testimony: A Philosophical Study*, Oxford University Press.

COADY C. A. J. 1994b. *Testimony, Observation and "Autonomous Knowledge"*, in MATILAL B.K., CHAKRABARTI A. (eds.), *Knowing from Words*, Springer, 225 ff.

COLEMAN B. 1998. *Are Clarity and Precision Compatible Aims in Legal Drafting?*, in «Singapore Journal of Legal Studies», December 1998, 376 ff.

DHAMI M. K. 2003. *Psychological Models of Professional Decision-Making*, in «Psychological Science», 14, 175 ff.

DE BRABANTER P. 2006. *Déference sémantique*, in PERRIN L. (ed.), *Le sens et ses voix: dialogisme et polyphonie en langue et en discours*, Université Paul Verlaine, 379 ff.

DE BRABANTER P., NICOLAS D., STOJANOVIC I., VILLANUEVA N. 2007. *Les usages déférentiels*, in BOUVIER A., CONEIN B. (eds.), *L'épistémologie sociale: Une théorie sociale de la connaissance*, Éditions de l'École des hautes études en sciences sociales, 139 ff.

DEL MAR M. 2015. *Introducing Fictions: Examples, Functions, Definitions and Evaluations*, in TWINING W., DEL MAR M. (eds.), *Legal Fictions in Theory and Practice*, Springer Verlag, ix ff.

ELISEEV D. E., MARSH E. M. 2023. *Understanding Why Searching the Internet Inflates Confidence in Explanatory Ability*, in «Applied Cognitive Psychology», 37,4, Special Issue, 711 ff.

ENDICOTT T. 1999. *The Impossibility of the Rule of Law*, in «Oxford Journal of Legal Studies», 19, 1 ff.

EVANS J. 2002. *Logic and Human Reasoning: An Assessment of the Deduction Paradigm*, in «Psychological Bulletin», 128, 978 ff.

FLETCHER G.P. 1998. *Basic Concepts of Criminal Law*, Oxford University Press.

GALLAGHER S. 2013. *The Socially Extended Mind*, in «Cognitive Systems Research», 25-26, 4 ff.

GIGERENZER G. 2008. *Moral Intuition = Fast and Frugal Heuristics?*, in SINNOTT-ARMSTRONG W. (ed.), *Moral Psychology, Vol 2: The Cognitive Science of Morality: Intuition and Diversity*, Boston Review, 1 ff.

GOLDMAN A.I. 1999. *Knowledge in a Social World*, Oxford University Press.

GOLDMAN A.I. 2011. *Experts: Which Ones Should You Trust?*, in GOLDMAN A. I., WHITCOMB D. (eds.), *Social Epistemology: Essential Readings*, Oxford University Press, 109 ff.

HAGE J. 2021. *Are the Cognitive Sciences Relevant for Law?*, in BROŻEK B., HAGE J., VINCENT N. (eds.), *Law and Mind*, Cambridge University Press, 17 ff.

HAIDT J. 2001. *The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment.*, in «Psychological Review», 108, 814 ff.

HAIDT J. 2013. *Moral Psychology and the Law: How Intuitions Drive Reasoning, Judgment, and the Search for Evidence*, in «Alabama Law Review», 64, 867 ff.

HARDWIG J. 1985. *Epistemic Dependence*, in «The Journal of Philosophy», 82, 335 ff.

HARDWIG J. 1991. *The Role of Trust in Knowledge*, in «The Journal of Philosophy», 88, 693 ff.

HART H. L. A. 2012. *The Concept of Law* (3rd Ed.), Oxford University Press.

HUTTO D., RAVENSCROFT I. 2021. *Folk Psychology as a Theory*, in ZALTA E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). Available at: https://plato.stanford.edu/entries/folkpsych-theory/.

KAHNEMAN D. 1994. *New Challenges to the Rationality Assumption*, in «Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft», 150, 18 ff.

KAHNEMAN D. 2003. *Maps of Bounded Rationality: Psychology for Behavioral Economics*, in «The American Economic Review», 93, 1449 ff.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Penguin.

KAHNEMAN D., SLOVIC P., TVERSKY A. (eds.) 1982. *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press.

KAHNEMAN D., TVERSKY A. 1972. *Subjective Probability: A Judgment of Representativeness*, in «Cognitive Psychology», 3, 430 ff.

KEIL F.C. 2005. *The Cradle of Categorization: Supporting Fragile Internal Knowledge Through Commerce with Culture and the World.*, in AHN W-K., GOLDSTONE R. L., LOVE B. C., MARKMAN A. B., WOLFF P. (eds.), *Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin*, American Psychological Association, 289 ff.

KEIL F., KOMINSKY J. 2015. *Grounding Concepts*, in MARGOLIS E., LAURENCE S. (eds.), *The Conceptual Mind: New Directions in the Study of Concepts*, The MIT Press, 677 ff.

KEIL F.C., STEIN C., WEBB L., BILLINGS V.D., ROZENBLIT L. 2008. *Discerning the Division of Cognitive Labor: An Emerging Understanding of How Knowledge Is Clustered in Other Minds*, in «Cognitive Science», 32, 259 ff.

KELSEN, H. 1949. *General Theory of Law and State*, Harvard University Press.

KOMINSKY J.F., KEIL F.C. 2014. *Overestimation of Knowledge About Word Meanings: The "Misplaced Meaning" Effect*, in «Cognitive Science», 38, 1604 ff.

KOCYIGIT, E., ROSSI, A., LENZINI, G. 2022. *Towards Assessing Features of Dark Patterns in Cookie Consent Processes*, in BIEKER F., MEYER J., PAPE S., SCHIERING I., WEICHE A. (eds.), *Privacy and Identity Management*, Springer, 165 ff.

KRIPKE, S., 1980, *Naming and Necessity*, Blackwell.

KUREK Ł. 2021. *Law, Folk Psychology and Cognitive Science*, in BROŻEK B., HAGE J., VINCENT N. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 55 ff.

LAKOFF G.P. 2005. *A Cognitive Scientist Looks at Daubert*, in «American Journal of Public Health», 95, 114 ff.

Lawson R. 2006. *The Science of Cycology: Failures to Understand How Everyday Objects Work*, in «Memory & Cognition», 34, 8, 1667 ff.

MARCONI D. 1997. *Lexical Competence*, MIT Press.

MARTÍ, J. L. 2005. *La nozione di ideale regolativo: note preliminari per una teoria degli ideali regolativi nel diritto*, in «Ragion Pratica», 2, 381 ff.

MCDONALD A.M., CRANOR L. F. 2008. *The Cost of Reading Privacy Policies*, in «A Journal of Law and Policy for the Information Society», 4, 543 ff.

MERCIER H. 2020. *Not Born Yesterday. The Science of Who We Trust and What We Believe*, Princeton University Press.

MERCIER H., SPERBER D. 2011. *Why Do Humans Reason? Arguments for an Argumentative Theory*, in «Behavioral and Brain Sciences», 34, 57 ff.

MERCIER H., SPERBER D. 2017. *The Enigma of Reason*, Harvard University Press.

PUTNAM H. 1975. *The Meaning of "Meaning"*, in ID., *Minnesota Studies in the Philosophy of Science*, University of Minnesota Press.

RABB N., FERNBACH P., SLOMAN S. 2019. *Individual Representation in a Community of Knowledge.*, in «Trends in cognitive sciences», 23, 891 ff.

RABB N., HAN J., SLOMAN S. 2021. *How Others Drive Our Sense of Understanding of Policies*, in «Behavioural Public Policy», 5, 454 ff.

RAZ J. 1985. *Authority, Law, and Morality*, in «The Monist», 3,1, 295 ff.

RECANATI F. 1997. *Can We Believe What We Do Not Understand?* in «Mind and Language», 12, 1, 84 ff.

RECANATI F. 2000. *Oratio Obliqua, Oratio Recta. An Essay on Metarepresentation*, MIT Press.

REIDENBERG J., BREAUX T., CRANOR L., FRENCH B., GRANNIS A., GRAVES J., LIU F., MCDONALD A., NORTON T., RAMANATH R., RUSSELL N., SADEH N., SCHAUB F. 2014. *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*, in «Berkeley Technology Law Journal», 30, 1 ff.

ROVERSI C., UBERTONE M., VILLANI C., D'ASCENZO S., LUGLI L. 2023. *Alice in Wonderland: Experimental Jurisprudence on the Internal Point of View*, in «Jurisprudence», 14, 143 ff.

ROZENBLIT L., KEIL F., 2002, *The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth*, in «Cognitive Science», 26, 5, 521 ff.

SEARLE J. R. 1997. *The Construction of Social Reality*, Free Pr.

SEARLE J. R. 2010. *Making the Social World: The Structure of Human Civilization*, Oxford University Press.

SHEA N. 2018. *Metacognition and Abstract Concepts*, in «Philosophical Transactions of the Royal Society B: Biological Sciences», 373, 1 ff.

SIFFERD K.L. 2004. *Psychology and the Criminal Law*, PhD Thesis, King's College London, University of London.

SIFFERD K.L. 2006. *In Defense of the Use of Commonsense Psychology in the Criminal Law*, in «Law and Philosophy», 25, 571 ff.

SIMON H.A. 1955. *A Behavioral Model of Rational Choice*, in «The Quarterly Journal of Economics», 69, 99 ff.

SLOMAN S, FERNBACH P. 2018. *The Knowledge Illusion: Why We Never Think Alone*, Penguin.

SLOMAN S., RABB N. 2016. *Your Understanding is My Understanding: Evidence for a Community of Knowledge*, in «Psychological Science», 27, 11, 1451 ff.

SPERBER D. 1985. *On Anthropological Knowledge*, Cambridge University Press.

STICH S.P. 1990. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, MIT Press.

STRAHILEVITZ L.J., KUGLER M.B. 2016. *Is Privacy Policy Language Irrelevant to Consumers?*, in «Journal of Legal Studies», S69, 45 ff.

SUNSTEIN C. 2011. *Empirically informed regulation*, in «University of Chicago Law Review», 78, 1349 ff.

SUNSTEIN C. 2013a. *Behavioral Economics and Paternalism*, in «Yale Law Journal», 122, 1826 ff.

SUNSTEIN C. 2013b. *Deciding by Default*, in «University of Pennsylvania Law Review», 162, 1 ff.

SUNSTEIN C. 2014. *Nudging: A Very Short Guide*, in «Journal of Consumer Policy», 37, 583 ff.

SUNSTEIN C. 2015. *The ethics of nudging*, in «Yale Journal on Regulation», 32, 413 ff.

SUNSTEIN C. 2017. *Is Cost-Benefit Analysis a Foreign Language?*, in «The Quarterly Journal of Experimental Psychology», 72, 1 ff.

TARONI M., UBERTONE, M. 2021. *Il diritto nella società della manipolazione*, unpublished manuscript.

TARUFFO, M. 2016. *La prova scientifica. Cenni generali*, in «Ragion Pratica», 2, 335 ff.

TEICHMAN D., ZAMIR E. (eds.) 2014. *The Oxford Handbook of Behavioral Economics and the Law*, Oxford University Press.

THALER R., SUNSTEIN C. 2003. *Libertarian Paternalism*, in «American Economic Review», 93, 175 ff.

THALER R., SUNSTEIN C. R 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness*, Yale University Press.

THALER R., SUNSTEIN C., BALZ J. 2013. *Choice Architecture.*, in «The Behavioral Foundations of Public Policy», 428 ff.

TOBIA K. 2021. *Law and the Cognitive Science of Ordinary Concepts*, in BROŻEK B., HAGE J., VINCENT N. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 86 ff.

TVERSKY A., KAHNEMAN D. 1983. *Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, in «Psychological Review», 90, 293 ff.

UBERTONE M. 2019. *La deferenza semantica nel processo*, in «Analisi e diritto»,1, 139 ff.

UBERTONE M. 2022a. *A Deference-Based Theory of Expert Evidence*, in «Archiv für Rechts- und Sozialphilosophie», 108, 241 ff.

UBERTONE M 2022b. *Il giudice e l'esperto: deferenza epistemica e deferenza semantica nel processo*, Giappichelli.

VIGANÒ F. 2022. *I principi di irretroattività e retroattività della legge penale nella più recente giurisprudenza costituzionale; la dimensione attuale della retroattività della legge penale più favorevole* in NATALINI A., DALLA LIBERA G., FUMO S. (eds.), *La successione delle leggi penali nel tempo. Report.*, Corte di Cassazione.

WALDRON J. 2016. *The Rule of Law*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition). Available at: https://plato.stanford.edu/archives/fall2016/entries/rule-of-law/.

ZUBOFF S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books.

PART VIII.

Law, Legal Reasoning and Artificial Intelligence

The Dual Challenge from AI and the Cognitive Sciences for Law and Legal (Reasoning) Practices

ANTONIA WALTERMANN

1. *Introduction*

What do a man claiming that a brain tumour caused his paedophilic behaviour, an AI system called DABUS being recognised as an inventor for the purposes of a patent application, and the modelling of legal reasoning in computational form have in common? They all *prima facie* challenge our legal (reasoning) practices (JONES 2013), albeit in seemingly very different ways. It may be useful, then, to be able to situate these and other challenges vis-à-vis each other and our legal (reasoning) practices to understand how they relate to each other, what it is they really challenge, and to critically reflect on them. That is what this chapter aims to do: the main claim of this chapter is that increasing insights into both human and artificial cognition challenge our understanding of legal reasoning as well as the core concepts we reason with. It proposes a framework that allows us to understand and situate these challenges, and briefly addresses how to respond to them.

To do this, the chapter proceeds as follows: following this introduction, section 2 offers the theoretical building blocks that are necessary for the analytical framework: it outlines the distinctions between human and artificial cognition, object- and meta-level or within and of a practice, and four ways in which extant legal (reasoning) practices can be challenged by insights from outside the law. It then combines these building blocks into an analytical framework within which different kinds of challenges from AI and the cognitive sciences can be situated—which section 3 then does. It considers several examples of challenges to our legal (reasoning) practices coming from the cognitive sciences and situates these within the analytical framework, thereby demonstrating a. the main claim of the paper and b., the way the framework works. Section 4, finally, considers different responses to these challenges, also touching on the question whether they really constitute challenges at all. The chapter concludes in section 5.

The value of this chapter does not lie in a systematic and substantive treatment of a particular field related to legal reasoning and the cognitive sciences. In fact, this chapter provides no such treatment at all. Instead, its value lies—ideally—in offering tools that allow readers to critically reflect on different challenges, to construct systematic, substantive treatments themselves, or to situate such treatments or challenges in a larger context. As such, the aim of this chapter is not primarily to make substantive claims (although it contains some), but to facilitate clear thinking.

* The author thanks Melina Poulin for her editorial and research assistance and the anonymous reviewer for the helpful comments. All remaining mistakes are, naturally, my own.

2. *Theoretical building blocks & analytical framework*

This chapter proposes a framework that can be used to identify, situate, and understand the different ways in which insights from the cognitive sciences broadly construed challenge law. To build this framework, we need a number of 'theoretical building blocks' without which our framework does not make sense. The building blocks that we need are:

- First, some terminological clarifications that set the stage for the analytical framework. This is predominantly the distinction between what I here call human and artificial cognition (section 2.1).
- Second, the distinctions between justifications (or concepts) within particular practices and outside (or of) the same practices (section 2.2).
- Third, four questions or challenges developed by Anne Ruth Mackor (section 2.3).

In section 2.4, we combine these building blocks into our analytical framework.

2.1. *Human and artificial cognition*

This chapter talks about a dual challenge from insights about human and artificial cognition: but what does this mean?

Cognitive science in general can be defined quite broadly as «the interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology» (THAGARD 2020). In this chapter, I distinguish between cognitive science that focuses on human cognition (e.g. the human mind, human intelligence, human reasoning, etc.) and artificial cognition (e.g. implementation of reasoning practices into artificial intelligence). I do this not because I think there are necessary differences between human and artificial cognition, but because the approaches of study and—often—the expertise of those in the field tends to differ (THAGARD 2020, 2). This does not take away that this distinction can be subsumed under the cognitive sciences more generally, or can be further refined.

2.2. *Of and within a practice*

Different scholars in different contexts have distinguished between different levels of abstraction with regard to different objects. In 1933, for example, the logician Alfred Tarski distinguished between object- and metalanguage for the purposes of establishing criteria for a truth definition: in Tarski's model, object language is the language under discussion, while metalanguage is the language used to discuss the object language (HODGES 2018). HAMPTON (1997) writes about this distinction between levels of inquiry and analysis, which she holds can be used for other areas of inquiry as well:

> «Consider the famous "liar's paradox," illustrated with the sentence, "This sentence is false." The sentence cannot be true when it tells us it is false; but if it is false, then given the assertion it is making, it would seem to be true. Alfred Tarski resolved this paradox by distinguishing two kinds of language, which he called the "object language" and the "metalanguage." The metalanguage is used to talk about the object language but is not itself part of that language. By understanding the predicates "is true" and "is false" to belong only to the metalanguage, we avoid the paradox» (HAMPTON 1997)[1].

[1] Note, however, that some authors argue—rightly, in my view—that an understanding of legal reasoning practices (our legal logic) can be better construed as not distinguishing between object- and meta-language (cf. HAGE 2005). Nonetheless, I think it is helpful for present purposes to make the distinction, without thereby wanting to make any

In 1955, the philosopher John Rawls distinguished between justifications within and justifications of a practice (RAWLS 1955, 3-32): with regard to the practice of punishment, he distinguished between applications of the practice, that is, the punishment of a particular individual, and the practice as such. In each case, one can ask about the justification: one might, first, ask about the justification within a practice, or, second, about the justification of a practice. What justifies the punishment of an individual person? Likely, one would here refer to the rules of the practice. What justifies the institution of punishment as such? Here, one cannot refer to the rules of the practice, but needs to look outside of the practice for justification.[2] In this connection, one can talk of justifications either internal and external (to the practice). This is also relevant in the next section.

For present purposes, we are not interested in establishing conditions for truth or in justifying punishment. Nonetheless, the *type* of distinction that both Tarski and Rawls make is useful for us. More precisely, it is useful for us to distinguish between two levels with regard to legal reasoning. These are:

1) the level of arguments we make within legal reasoning, including the conceptual building blocks of these arguments, such as LIABILITY, RESPONSIBILITY, GUILT, BLAMEWORTHINESS, TORT, CRIME, and much more.
2) the level of arguments we make about legal reasoning, including the building blocks we use to conceptualise these arguments, such as SYLLOGISM, VALIDITY, ANALOGY, or even LEGAL REASONING itself.

If we combine this distinction between the two levels with the distinction between human and artificial cognition of the previous section, we get the following:

	Within reasoning	Of/about reasoning
Human cognition		
Artificial cognition		

TABLE 1

The starting point of this paper is that insights from the cognitive sciences—or insights about human and artificial cognition, to use the distinction from 2.1—challenge both the concepts within our legal reasoning practices and of or about our legal reasoning practices.

2.3. *Mackor's four questions*

In 2013, Anne Ruth Mackor published a chapter asking what neuroscience can say about legal responsibility. In this chapter, she distinguishes between four questions, two internal to the practice of responsibility and two external to the practice. From least to most radical, these questions are:

1) whether the neurosciences have something to say about the proper application of conceptions of responsibility ('application');
2) whether findings of the neurosciences can change conceptions of responsibility (that is, the criteria for application of one or more concepts of responsibility) ('conception');

claims about the underlying (formal) logical structures of our reasoning.

[2] The picture is more complicated than this: different modes of justification are possible, such as conventionalist or coherentist. However, this is not the focus of this chapter, so we will not delve into it here.

3) whether neuroscientific research can show that responsibility practices are pointless to the extent that they cannot fulfil the goal they are intended to fulfil ('implication');
4) whether neuroscientific research can show that responsibility practices are fundamentally untenable because they rest on false presuppositions ('presuppositions') (MACKOR 2013).

The first of these question ('application') does not attack the practice or its concepts and their criteria of application in any way. It only asks whether our understanding of when the criteria for application are fulfilled is correct. In the context of neuroscience and responsibility, for example, this would be the question whether

> «neurosciences can help to show that specific (categories of) persons whom we used to think that fit the criteria of application, in fact do not fit the criteria, or, conversely, that specific (categories of) persons whom we thought not to fit the criteria, turn out to fit the criteria after all» (MACKOR 2013).

The first question thus leaves the criteria for application intact. This is different with the second question ('conception'), which calls the criteria for application of a concept into question. Do the neurosciences, to remain with that example, change the criteria for application of one or more concepts of responsibility (MACKOR 2013)?

The third question ('implication') calls into question the practice as such, by asking whether the practice does not have implications that show it to be pointless, i.e., because it fails to contribute to the realisation of the goals or purposes of the practice. In the case of neuroscience and responsibility, the question is «whether neurosciences can offer evidence to support the claim that holding people responsible is pointless» (MACKOR 2013). The fourth question, similarly calls the practice as such into question by asking whether it rests on false presuppositions that make it fundamentally flawed or untenable (MACKOR 2013).

The first two of these questions are internal ones because they are situated within the responsibility practices in question; the latter two are external ones because they assess the practice from an external perspective. Abstracting from the context of neuroscience and responsibility respectively, we can draw up the following framework:

Internal	External
Application Challenges to the application of concepts within the practice.	**Implications** Challenges our practice(s) by demonstrating that they cannot achieve their purpose
Conceptions Challenges the conceptions (applicability criteria) of concepts within the practice.	**Presuppositions** Challenges our practice(s) by demonstrating that they rest on wrong presuppositions and starting points

TABLE 2

The practice to which these questions or challenges apply can in principle be any practice. For present purposes, we are looking at our legal reasoning practice and, following section 2.2, challenges from both human and artificial cognition to both the concepts within the practice of legal reasoning as well as of the practice of legal reasoning. In section 3, we will apply these different questions to cases, which will make them more concrete and serve to illustrate how our analytic framework functions. Before doing so, however, the different building blocks need to be put together.

2.4. *The building blocks combined*

If we combine Table 1 and Table 2, we get the following picture:

	Within reasoning		Of/about reasoning	
Human cognition	Internal	External	Internal	External
	Application	**Implications**	**Application**	**Implications**
	Conceptions	**Presuppositions**	**Conceptions**	**Presuppositions**
Artificial cognition	Internal	External	Internal	External
	Application	**Implications**	**Application**	**Implications**
	Conceptions	**Presuppositions**	**Conceptions**	**Presuppositions**

TABLE 3: COMBINED

Table 3 shows the combination of the dual challenge from human and artificial cognition to and within our legal reasoning practices and Mackor's four questions: insights from either human or artificial cognition might challenge the application, conceptions, presumed implications or presuppositions of our legal reasoning practices or the concepts with which we reason. This gives us sixteen possible combinations and thus, sixteen possible positions along which to classify different (kinds of) challenges. We have four quadrants: the first (blue in the table above) is where insights about human cognition challenge the concepts we use within our reasoning practices; the second (red) indicates challenges from human cognition to the concepts of and about our reasoning practices; the third (yellow) brings us to insights from artificial cognition challenging concepts within our reasoning practices; the fourth (green) indicates challenges from artificial cognition to our reasoning practices as such. Within each quadrant, we can distinguish between internal and external challenges: challenges to the application or conception of the concepts in question are internal; if, however, the challenge is that the implications or presuppositions of the concept are fundamentally flawed, it is external.

This gives us our analytical framework in full. But how does it look in action and how does it help? In the next section, we will apply this framework by considering different examples and situating them on the 'map' above. We will see that attempting to do so will make us engage with the examples critically.

3. *Application*

In this section, we will consider a number of examples to see where they are situated in the analytical framework developed in the previous section. In doing so, we will see that insights from human and artificial cognition challenge both the concepts and practices that play a role within our legal reasoning and the concepts and practice of legal reasoning as such (the main claim of the paper). We will also gain a better understanding of how the analytical framework functions.

3.1. *Human cognition*

In this section, we will consider examples of challenges to law that arise from increasing insights about human cognition, and situate these examples within our analytical framework. We will consider three examples: the challenge raised by brain tumours causing paedophilia for criminal liability and sentencing, the charge that law undermines human flourishing, and the oft-discussed study suggesting that the time since the last (food) break is more decisive than factors such as risk of recidivism when it comes to whether inmates would get parole.

3.1.1. *Brain tumours and paedophilia*

There have been, at the time of writing, two cases (one in the US and one in Italy) of men who developed uncontrollable paedophilia, which coincided with the growth of a brain tumour. Once the tumour was removed, the paedophiliac urges and behaviour also ceased. In the US case, when the tumour regrew, the paedophilia also returned—and disappeared again after the tumour was removed once more.

What, if any, role should the insight that acquired paedophilia can be a consequence of cancer in the orbitofrontal region of the brain (BURNS, SWERDLOW 2003), play in sentencing of individuals? If we ask ourselves, for example, whether they should receive a reduced sentence, we are asking whether the criteria for application of a reduced sentence are fulfilled by a—in this case very specific—group of people who were not previously considered as fulfilling those criteria. That is an example of insights about human cognition challenging the application of concepts we use within our legal reasoning practices, i.e., it can be situated within the first cell ('application') of the first quadrant ('human cognition & within reasoning') of Table 3.

Similarly, consider the following argument in a news item published in response to the US case mentioned above:

> «The U.S. Supreme Court has ruled that executing mentally retarded murderers is unconstitutionally cruel because of their diminished ability to reason and control their urges.
> Chris Adams, a death penalty specialist for the National Association of Criminal Defense Lawyers, thinks the next logical step would be to include people who have brain tumors.
> "Some people simply don't have the frontal lobe capacity to stop what they're doing," he said» (NBC 2003).

This, too, can be situated in the first cell of the first quadrant of Table 3, i.e., it is an example of insights about human cognition challenging the application of the concepts with which we reason within the law. In essence, the argument is that (some) people with brain tumours satisfy the criteria that would make it unconstitutional to execute them, namely diminished ability to reason and control their urges.

Situating these questions and arguments in this way allows us to understand what precisely is—and is not—challenged.

3.1.2. *Does law undermine human flourishing?*

Alces and Sapolski, in a 2022 article, use the US case described above (calling the man in question by the pseudonym 'Mr Oft') and embed it in a larger and more encompassing argument. For them, it is one example of many neuroscientific insights that demonstrates that we are essentially—if complexly—mechanical entities:

> «Our normative systems conceive of law and morality as [...] a product of sufficient choice to attach blame, fault, and concepts of desert. But on what basis do we draw the distinctions between physical and normative malady: Are not both just (generally) distinguishable manifestations of mechanical causes? If human agents are essentially mechanical entities, on what basis could we find a normative difference between, say, tuberculosis and selfishness or insufficient ability to feel compassion for others?» (ALCES & SAPOLSKY 2021-22).

They hold that «extant legal doctrine and practices (civil as well as criminal) actually undermine human thriving: they are not merely a distraction; they are an impediment» (ALCES & SAPOLSKY 2021-22, 1081). This is because our current legal doctrines and practices are «relying on a misconception of what it means to be human» (ALCES & SAPOLSKY 2021-22, 1084).

According to them, it

> «is easy to see how Oft's tumor challenges a legal system equating intent with responsibility or volition. But it is not the sort of case that can be generalized easily enough to revolutionize legal thinking. This is because of the uniqueness of its clarity, where massively abnormal behaviour is caused by the singular and massive abnormality of a brain tumor, literally demonstrable at the scene of the crime. What contemporary science shows is that the intent behind our best and worst behaviors, and all those ambiguously in between, is as much the end product of factors outside our control as was Oft's intentional criminality. However, it is far harder to appreciate this than the case of Oft for at least three reasons: (1) unlike the singularity of his tumor, our behaviour mostly arises from a multitude of biological factors that subtly interact, (2) no single factor has remotely the overt sledgehammer causality of a tumor, and (3) many of the factors were set into action long before the behaviour occurred (with some even long before the individual in question was born» (ALCES & SAPOLSKY 2021-22, 1102).

According to Alces and Sapolsky, insights from contemporary science about human cognition and the causes of human behaviour conflict with the model of who we are as human beings that the law presupposes (ALCES & SAPOLSKY 2021-22, 1112). Applying our analytical framework to their argument, we can situate them in the fourth cell ('presupposition') of the first quadrant ('human cognition & within reasoning'): their challenge to the law is that it rests on wrong assumptions.

That is not the only challenge they raise, however: when they hold that the law, in making these wrong assumptions, undermines and actively impedes human thriving, this is another challenge, if one with an implicit premise. Implicitly, their argument suggests—and I think many would agree—that law does or should have as its aim to enable, rather than hinder, human thriving. If so, they raise a second challenge: that by relying on a misconceived view of humankind, law undermines the very purpose for which we have it. This thrust of their challenge can be situated in the second cell ('implication') of the first quadrant of our analytical framework.

3.1.3. *Hungry judges*

The previous two examples have both been situated in the first quadrant of our analytical framework, meaning that they have been examples of challenges from insights about human cognition to the concepts and practices that play a role within our legal reasoning. However, insights about human cognition also challenge our legal reasoning practices and the concepts we use to describe legal reasoning. This section is about one example of this.

A 2011 study by Danziger, Levav and Avnaim-Pesso has received much attention. The authors summarise their study as follows:

> «We test the common caricature of realism that justice is "what the judge ate for breakfast" in sequential parole decisions made by experienced judges. We record the judges' two daily food breaks, which result in

> segmenting the deliberations of the day into three distinct "decision sessions." We find that the percentage of favorable rulings drops gradually from≈65% to nearly zero within each decision session and returns abruptly to≈65% after a break. Our findings suggest that judicial rulings can be swayed by extraneous variables that should have no bearing on legal decisions» (DANZIGER et al. 2011, 6889-6892).

Bublitz, in assessing this study, writes that:

> «One intuitively shares their sentiment. Rulings have gone astray, they were influenced by something that ought not be there, something that affects or taints or contaminates decisions and renders them incorrect. Capricious rulings resulting from such extralegal factors appear as a paradigmatic instances of what it means to be 'at the whim of a judge'. Or, in the words of the former US Supreme Court Justice William Douglas: there will be no "justice under law if a negligence rule is applied in the morning but not in the afternoon." In other words, if justice is indeed what judges ate for breakfast, there is no justice» (BUBLITZ 2020, 6).

If this is correct, it provides a challenge to the core of our legal reasoning practice by demonstrating that core presuppositions of our legal reasoning practice—that judgments are rendered on the basis of legally relevant factors—are not borne out in reality. This can be understood as a challenge from human cognition to our legal reasoning practice, i.e., it can be situated in the fourth cell ('presupposition') of the second quadrant ('human cognition & about reasoning') of our table.

It bears mentioning, however, that Bublitz ultimately draws different conclusions in his chapter and that alternative explanations for the study have been offered (BUBLITZ 2020, 4),— which demonstrates the importance of critically assessing the challenges raised as well as the importance of replication. Whether and in how far the effect of the study and the conclusions drawn from it will bear out in future studies will remain to be seen, but at the same time, the challenge raised by this study should also not be underestimated.

3.2. *Artificial cognition*

In this section, we will consider examples of challenges to law that arise from increasing insights about artificial cognition, and situate these examples within our analytical framework. We will again consider three examples: the case of DABUS, an artificially intelligent system listed as sole inventor on a patent application; the question whether AI can engage in legal reasoning, particularly where legal reasoning is analogical; and the question whether AI can and should be considered an agent responsible for its acts in the eyes of the law.

3.2.1. *DABUS, the AI inventor*

In 2021, the Federal Court of Australia decided that an artificial intelligence system called DABUS could be considered an inventor for the purposes of an international patent application. Courts in other countries, such as the US, UK, and Germany, had rejected the patent application, as had Australian patent offices, but the Federal Court found that it was compatible with the Australian Patent Act to consider DABUS the inventor—but not owner—of the patent (MATULIONYTE 2021).

The Federal Court held that:

> «for the following reasons, in my view an artificial intelligence system can be an inventor for the purposes of the Act. First, an inventor is an agent noun; an agent can be a person or thing that invents. Second, so to hold reflects the reality in terms of many otherwise patentable inventions

> where it cannot sensibly be said that a human is the inventor. Third, nothing in the Act dictates the contrary conclusion»[3].

The Australian Federal Court here applied the concept of inventor in a way it had not previously been applied, while courts of other countries did not do so: we could say that the existence of DABUS and the patent application have challenged how the concept of INVENTOR is applied, with different courts responding differently to this challenge. Where should this be situated on the analytical framework? Is this a case of application—where the criteria for application are not challenged—or one of conception—where the criteria for application are in question?

The criteria for application in the eyes of the court, it seems to me, was whether DABUS could be said to have invented something «that satisfies all of the requirements of patentability in terms of novelty, inventiveness and utility»[4]. What was decisive for the Federal Court was whether the act of inventing was performed, not whether it was performed by a human or non-human being. In this connection, the argument was that

> «As an agent noun (like "computer", "dishwasher" or "lawnmower") the agent can be a person or a thing. In this context the primary judge noted that, whereas once the word "inventor", like "computer", might originally have been apt to describe persons when only humans could make inventions (or perform computations), now the term may be used to describe machines which can carry out the same function»[5].

If the sole criterion is carrying out the function and technological progress makes it possible for an AI system to carry out the function, this could make the case of DABUS an example fitting under the first cell ('application') of the third quadrant ('artificial cognition & within reasoning').

However, much of the debate of the DABUS case is precisely about whether this is the correct view. Let us consider why the patent application was initially rejected: the Deputy Commissioner of Patents who rejected the application before the applicant (Dr Thaler), sought judicial review of the rejection and did so for the reason that the ordinary meaning of "inventor" was taken to be inherently human (CURREY & OWEN 2021). On this view, then, DABUS does not challenge 'merely' the application of INVENTOR but also the criteria for application: is it a criterion for the correct application of INVENTOR that the entity in question is a human being?

Table 4 gives the two different views in tabular form:

Option 1	Option 2
Criterion for correct application of INVENTOR	Criteria for correct application of INVENTOR
(1) Something was invented	(1) Something was invented
	(2) The invention was brought about by human action

TABLE 4

From this perspective, the case of DABUS challenges the criteria for application of the concept INVENTOR, not just the application. This view was upheld in appeal by the Full Federal Court

[3] *Thaler v Commissioner of Patents*, 2021, Federal Court of Australia, sec 10.
[4] *Thaler v Commissioner of Patents*, 2021, Federal Court of Australia, sec 7.
[5] *Commissioner of Patents v Thaler*, 2022, Federal Court of Australia, sec 46.

which overturned the initial decision by the Federal Court (single judge)[6]. This demonstrates that what was at stake was not the application, but the conception of INVENTOR. That puts the DABUS case in the third cell ('conception') of the third quadrant ('artificial cognition & within reasoning').

The fact that on different construals, the DABUS case can be situated at different places in the framework demonstrates, in my view, the usefulness of the analytical framework: it demands critical reflection on what, precisely, is being challenged.

3.2.2. *Can AI engage in legal reasoning?*

In 2001, Cass Sunstein held that artificial intelligence cannot engage in legal reasoning. His argument rests, in part, on the state of technology at the time: the state of technology in 2001 was not (yet) far enough advanced that artificial intelligence could be said to be engaged in legal reasoning. However, he also held that claims, at the time, that AI could engage in legal reasoning rested on an «inadequate picture of what legal reasoning actually is» (SUNSTEIN 2001, 31). This second part of the argument is what I want to focus on here. Before I do so, two short notes are in order: first, Sunstein made this argument in 2001, so any claims about the state of the art of technology are outdated—but claims about legal reasoning are not, which is what makes this argument interesting for our purposes. Second, Sunstein's argument is US-centric and more obviously applicable to the common than to the civil law.

Sunstein argues that legal reasoning is often analogical. Reasoning by analogy, he posits, requires the following:

> «[A]nalogizers in law have to ask which case has relevant similarities to the case at hand. It is more accurate still to say that whether a case has relevant similarities to the case at hand depends on the principle for which the initial case is said, on reflection, to stand. It follows that the crucial step in analogical reasoning consists, not in a finding of "more" similarities, not in establishing "many" distinctions, and not even showing "relevant" similarities and differences, but instead in the identification of a principle that justifies a claim of similarity or difference. Because the identification of that principle is a matter of evaluation, and not of finding or counting something, artificial intelligence is able to engage in analogical reasoning only to the extent that it is capable of making good evaluative judgments» (SUNSTEIN 2001, 5).

This means that «the analogizer attempts to make best constructive sense out of a past decision by generating a principle that best justifies it, and by bringing that principle to bear on the case at hand» (SUNSTEIN 2001, 7) and Sunstein saw no reason to think that the AI systems of the time had the capacities to do so. He explicitly leaves open the possibility for this to change, but at the time of writing, mechanisms of abstraction and analogy are still out of reach (PAVLUS 2021).

For our present purposes, it is interesting to see where Sunstein's argument can be situated in our framework. In my view, the argument can be framed as an argument against considering the concept of legal reasoning (in sense of including analogical reasoning) applicable to artificial intelligence, i.e., as responding to a challenge from AI to the application of the concept. Sunstein's argument is not that the criteria for application should or have changed, which would situate the argument under *Conception* but it is an argument that application to AI is not (yet) appropriate. One could, however, construe it also as an argument for a different conception of legal reasoning, certainly if compared to the conception of legal reasoning Sunstein ascribes to those who claim(ed at the time) that AI can engage in legal reasoning. This would make the argument an example of *Conception* rather than *Application*.

[6] *Commissoner of Patents v Thaler*, 2022, Federal Court of Australia.

Bart Verheij made precisely that kind of argument in his 2020 presidential address to the seventeenth international conference on artificial intelligence and law when he held that some of the hurdles for the development of implementing legal reasoning in AI support the view that a particular model of legal reasoning—which he calls the subsumption model—is false:

> «[a]ccording to the subsumption model of law there is a set of laws, thought of as rules, there are some facts,—and you arrive at the legal answers, the legal consequences by applying the rules to the facts [...]. The case facts are subsumed under the rules, providing the legal solution to the case. It is often associated with Montesquieu's phrase of the judge as a 'bouche de la loi', the mouth of the law, according to which a judge is just the one who makes the law speak. All hurdles just mentioned show that this perspective cannot be true» (VERHEIJ 2020, 188).

Which are the hurdles that show this? Verheij lists the following: that legal reasoning is rule-guided, rather than rule governed; that legal terms have an open texture; that legal questions can have more than one answer, but still demand timely and reasonable responses; and that the answers to legal questions can change over time (VERHEIJ 2020, 187 f.).

Two years later, Francesconi hints at a similar challenge to our conception of legal reasoning when he writes that (far advanced) AI

> «actually opens up the possibility that a machine, on the basis of deductive rules, facts and categories, can reach the levels of complexity of human legal reasoning, until replacing it. But this perspective is not without question marks. For example, does the human judge argue only by deductive categories? Moreover, which role have the emotions [sic] in taking decisions? Will a digital judge, emotionally neutral, be fairer than a human judge?» (FRANCESCONI, 2022, 157).

While challenges to models (that is, conceptions) of legal reasoning as subsumption or of legal formalism are not new (LEITER 2005), attempting to model legal reasoning in computational form offer support of such challenges or new challenges of their own. ASHLEY (2002) offers a plausible explanation of how and why this is the case when he writes that:

> «the virtue of applying AI to research in legal or practical ethical reasoning "is that the nature of the subject forces additional explicitness and clarification, because ultimately its products must be encoded and run on a computer. Thus, hidden and unclear assumptions can often be exposed in such a context"» (ASHLEY 2002, 165, quoting SCHAFFNER 1990).

None of the above challenge the practice of legal reasoning as such, but raise questions about the application and conceptions of concepts of/within the practice.

3.2.3. *Legal responsibility for AI?*

With developments in the field of artificial intelligence, the question whether AI can and should be held morally and legally responsible has gained traction in academic and public debates[7]. I will here use only one argument within that debate as an example, and then only briefly: in 2017, Jaap Hage

[7] To give only a number of examples: Open Letter to the European Commission Artificial Intelligence and Robotics (http://www.robotics-openletter.eu/); Report of COMEST on Robotics Ethics, 2017 (https://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20COMEST%20ON%20ROBOTICS%20ETHICS%2014.09.17.pdf); ANDERSON & ANDERSON 2011; BROŻEK & JAKUBIEC 2017; BRYSON et al. 2017; CHOPRA & WHITE 2011; COECKELBERGH 2019; DAHIYAT 2021; FLORIDI et al. 2018; FLORIDI & SANDERS 2004; GUNKEL 2012; HIMMA 2009; SOLUM 1992.

has argued that it is possible and may be sensible to hold autonomous agents responsible, that is, legally liable. He summarises (part of) the argument as follows:

> «it is argued [...] that agency, responsibility and liability are not found in a mind-independent reality, but rather are attributed to elements of a social practice that will be called the 'practice of agency'. This practice may be based on the way human beings experience themselves and their fellow humans, but does not necessarily have a firm foundation in the 'real', mind-independent world. This practice might have been different from what it actually is and might attribute agency, responsibility and liability to autonomous systems just as easily as it actually attributes these characteristics to human beings. It is argued that a major reason to treat humans and autonomous systems differently in this respect—that humans act intentionally and on the basis of a free will—has lost much of its credibility in the light of modern science» (HAGE 2017, 256).

This argument is interesting for us to consider in light of the analytical framework of this chapter because Hage combines increased insights into artificial cognition and technological developments with insights into human cognition to build the argument. This shows the limitations of the framework of this chapter: not every challenge can be separated into neat analytical categories—rather, some challenges combine different categories. Nonetheless, we can ask what Hage ultimately challenges: his argument holds that our legal responsibility practice ('within reasoning' in our framework) rests on flawed presuppositions, namely «that human beings act intentionally and on the basis of a free will» (HAGE 2017). This makes Hage's challenge comparable to that of Alces and Sapolsky (discussed in section 3.1.2), although Alces and Sapolsky are not concerned with artificial intelligence in their own argument. Nonetheless, they are sharing a view of humankind, and what it means for our responsibility practice.

3.3. *First conclusions and some caveats*

In the sections above, we have considered six examples of how insights from the cognitive sciences—about human or artificial cognition—challenge the law, both when it comes to our concepts and practices of legal reasoning and the concepts and practices within our legal reasoning. Before we consider how to respond to these challenges in the next section, it is worth taking a moment to draw some first conclusions and make some additional remarks. I think the six examples above demonstrate a. that the cognitive sciences challenge the law in different ways (the main substantive claim of this paper) and b. that these challenges can be fruitfully situated in the analytical framework outlined in this chapter.

It bears mentioning, however, that insights from the cognitive sciences (regarding both human and artificial cognition) are not alone or the first to challenge law in these ways. Over the centuries, if not millenia, many challenges that fit within the above theoretical framework have been put forth already[8]. What is new are not the challenges, but the (epistemic) support for these challenges and the means for teasing/implementing legal reasoning techniques in artificial cognition. Equally, the analytical framework of this chapter is not the only possible way in which one can conceptualise and understand these challenges (cf. HAGE 2021).

It also bears mentioning that the insights from the cognitive sciences are, by and large, descriptive or, to put it in a different way, belonging to the realm of theoretical reason. We can distinguish between theoretical and practical reason as follows:

[8] MORSE (2015) holds, for example, that «[i]n principle, [...] neuroscience adds nothing new, even if neuroscience is a better, more persuasive science than some of its predecessors». And arguments about, for example, free will and determinism have existed long before the rise of neuroscience, cf. DILMAN 1999.

> «The aim of theoretical reason is the description and explanation and sometimes also the prediction of events. The aim of practical reason is not to describe and to explain and thereby to understand states of affairs or events in the world, but to answer the question, "*What should I do?*"» (MACKOR 2013)[9].

From the mere fact that something is the case, e.g. that a tumour causes paedophilic behaviour, we cannot derive any information about what should be done. This is also called *Hume's guillotine* (HUME 1978): it is impossible to derive what ought to be done *only* from what is. In order to derive ought-judgments, we need a premise containing an ought[10]. Imagine, for example, a father telling his child that she ought to eat vegetables. Asked "why", the father responds, "Because vegetables are healthy". But this leaves silent the premise that the child ought to eat healthy things. Why ought she do so? Again, an argument is required—and again, this argument will require an ought-premise.

Hume's guillotine has led some to conclude that challenges such as the ones outlined above are not a real threat to our legal practices, as these practices fall in the realm of practical, not theoretical reason (MACKOR 2013). This brings us to the last section of this chapter, which addresses possible responses to the challenges identified above.

4. *How do we respond to these challenges?*

So far, we have considered a number of examples of challenges from the cognitive sciences to our legal (reasoning) practices. We have also briefly touched on the idea that maybe, these challenges are not *really* challenges to our practices.

In this section, I want to address three different possible approaches to the question whether insights about human or artificial cognition really challenge our legal (reasoning) practices and if so, how we can respond. These approaches hold, respectively, that there is no challenge or inconsistency (section 4.1), that there is inconsistency, but no challenge (section 4.2), or that there is a challenge (section 4.3), in which case it needs to be addressed, which can be done on an individual (4.3.1) or a systematic basis (4.3.2).

4.1. *No inconsistency, no challenge*

One approach is to insist that there can be no inconsistency between our legal (reasoning) practices and insights about human or artificial cognition and therefore also no challenge, because our legal (reasoning) practices and insights about human or artificial cognition talk about different things. This might be because one belongs to the realm of practical and the other to the realm of theoretical reason, as we have seen above. It might also be because law can define its own concepts and as such, the cognitive sciences talk about different things than the law. For example, law determines what kinds of entities can count as INVENTOR in the eyes of the law, irrespective of whether we would regard these entities as capable of inventing things in non-legal terms. If artificial cognition develops to such a degree that we generally think that AI systems can invent things, this does not say anything about whether they can invent things in the eyes of the law. Similarly, from the 'mere' fact that artificial cognition

[9] Note that variations of «What should I do?» such as what others should do and what one should have done fall under practical reason as well.

[10] For a more nuanced explanation of Hume's guillotine and what taking Hume's guillotine for granted says about our ontological and epistemological beliefs (that is, beliefs about what exists and how we can have knowledge of it respectively), I refer readers to Jaap Hage's chapter in this volume on the nature of law and constructivist facts.

has developed to such a degree or that we generally think AI systems can invent things one cannot derive that the law ought to recognise AI systems as inventors.

This view can be illustrated using the metaphor of maps[11]. A map of Paris and a map of Rome cannot be inconsistent with one another and one cannot challenge the other, because these maps are not showing the same area at all. Analogically, if we take the cognitive sciences and our legal (reasoning) practices to be about different concepts or to belong to different realms altogether, we can say that there can be no inconsistencies between them and that the cognitive sciences cannot challenge our legal (reasoning) practices.

It is not obvious to me, however, that the cognitive sciences and our legal (reasoning) practices and the concepts used in each are maps showing different areas altogether[12]. Moreover, even if we maintain a clear distinction between practical and theoretical reason, insights belonging to the realm of theoretical reason can be embedded into practical reason arguments.

4.2. *Inconsistency, but no challenge*

A different approach would be to say that the two are maps that show the same area, but in different ways: one map shows rivers and bodies of water, the other is a roadmap. If we could put one map on top of the other or combine the two into one integrated map without problems, there is no challenge. They might portray the same landscape in different ways, but the map showing a river in one spot is not challenged by another map showing a road running parallel to the river. The picture changes when one map puts a road in the middle of a large lake. This is problematic, at least if the map does not also show that there is a bridge across the lake. In this case, then, there seems to be a challenge—unless we do not assume that the two maps of the same area should be consistent with one another. Say that one map is historical and the other is current. In this case, we can notice the inconsistency between the maps, but the historical map does not challenge the accuracy of the current map if the lake has simply dried out.

What does it look like to translate this analogy back to insights about cognition and our legal (reasoning) practices? It may be the case that our legal (reasoning) practices rely on flawed presuppositions about human cognition, or that in doing so, they sometimes fail to realise their own goals. If we can nonetheless argue that in most cases, our extant legal (reasoning) practices are better than alternatives that are more consistent with insights about human cognition, we have an argument why our legal (reasoning) practices ought not be consistent with insights about human cognition. If we could argue, for example, that it is more just to maintain present practices despite flawed presuppositions, rather than to adapt them to more accurate presuppositions, or that maintaining present practices maximises happiness, we have such an argument. This is the kind of argument that Strawson (1962) makes when he holds that determinism cannot displace the reactive moral attitudes that are deeply ingrained in us and form the basis of our responsibility practices (STRAWSON 1962). In a similar vein, proponents of retributivism as a basis for punishment may hold that to reject retributivism in favour of consequentialist justifications of punishment is undesirable because it presupposes a view of human beings that does not treat them as agents in their own right or respect their dignity (e.g., MURPHY 2017).

[11] Jaap Hage and I have used this metaphor already elsewhere: HAGE & WALTERMANN 2021. It is initially taken from HAACK 2008.

[12] For a much more extended version of argument as applied to legal responsibility practices, see HAGE & WALTERMANN 2021.

4.3. *Inconsistency & challenge*

If it cannot be demonstrated that the cognitive sciences and our legal (reasoning) practices are a. about entirely different things or b. ought not be compatible with one another, we are left with the option to recognise that there is an inconsistency that needs to be addressed, that is, that there is a challenge from the cognitive sciences to law. To return to the map metaphor, in this scenario we are looking at two concurrent maps of the same area that show an inconsistency. We now need to ask ourselves how we can address the inconsistency. A likely way to do so is to change one of the maps to make the two maps consistent with one another so that they could, in theory, be combined into one integrated map without any issue.

What could this look like?

4.3.1. *Individual*

We can address challenges from the cognitive sciences on an individual basis. If the cognitive sciences show that under certain circumstances, brain tumours in the orbitofrontal region of the brain cause paedophiliac behaviour, we can construct arguments about what this means for our practices of criminal law and criminal sentencing. If a study demonstrates that extra-legal factors such as time since the last meal break play a role in judicial decision-making, we can construct arguments about whether, why, and how to address this insight. For each challenge, we can construct arguments about what ought to change to address the challenge, that is, in order to bring our practice in line with the relevant insight about human or artificial cognition.

By definition, addressing a challenge on an individual basis will serve only to address that particular challenge. Particularly where the challenges from AI and the cognitive sciences challenge implications or presuppositions of law, however, a more integrated approach seems required.

4.3.2. *Systematic*

A more encompassing approach would be to develop a coherent theory of what our legal (reasoning) practices are, what they ought to be, and how insights from the cognitive sciences fit into this picture, and to use this theory as the backdrop of arguments about how to respond to new insights about human or artificial cognition that challenge our legal (reasoning) practices. This would be very demanding: such a theory would include not only state of the art knowledge about human and artificial cognition, but also encompass the aims of our legal (reasoning) practices, the values we pursue with them, an understanding of the nature of law and how it relates to the cognitive sciences, and more. Both the theory and our extant practices would then be revised on a continuous basis against the standard of coherence[13].

5. *Conclusion*

In this chapter, we have considered a number of examples of ways in which insights about human or artificial cognition (seem to) challenge our legal (reasoning) practices. We have developed an analytical framework within which to situate these challenges in order to facilitate clear and critical thinking about them. This framework is summarily represented in the table below:

[13] A lot more can and has been said about coherence. By way of example, HAGE 2013; LEHRER 1992; ZIPURSKY 1997.

	Within reasoning		Of/about reasoning	
Human cognition	Internal	External	Internal	External
	Application	**Implications**	**Application**	**Implications**
	Conceptions	**Presuppositions**	**Conceptions**	**Presuppositions**
Artificial cognition	Internal	External	Internal	External
	Application	**Implications**	**Application**	**Implications**
	Conceptions	**Presuppositions**	**Conceptions**	**Presuppositions**

We have also considered whether these seeming challenges really do challenge our legal (reasoning) practices. In this connection, we have considered a number of different possibilities, visualised in Figure 1:

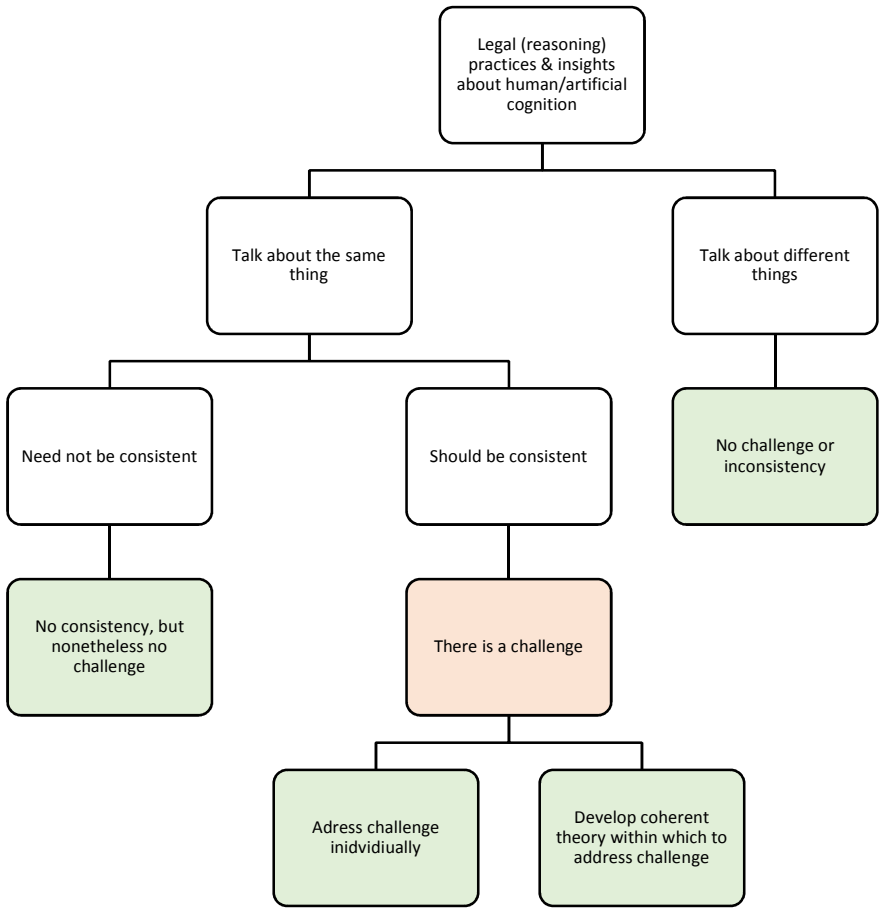


FIGURE 1

Increasing insights from the cognitive sciences about cognition do and will challenge our legal (reasoning) practices, both when it comes to the concepts and practices we use within our reasoning practices and when it comes to the concepts and the practice of legal reasoning itself. While this chapter has addressed this in terms of challenges, these increasing insights also offer chances: in particular, they offer chances for a more encompassing, more nuanced, better—in the sense of more grounded in and coherent with best scientific theories about cognition—understanding of our legal (reasoning) practices. It is my hope that the analytical framework of this chapter will contribute to realising this better understanding.

References

Alces P.A., Sapolsky R.M. 2021-22. *Nohwere*, in «William & Mary Law Review», 63, 4, 1079 ff.

Anderson S.L., Anderson M. 2011. *Machine Ethics*, Cambridge University Press.

Ashley K.D. 2002. *An AI Model of Case-based Legal Argument from a Jurisprudential Viewpoint*, in «Artificial Intelligence and Law», 10, 163 ff.

Beck S. 2015. *The Problem of Ascribing Legal Responsibility in the Case of Robotics*, in «AI & Society», 31, 473 ff.

Brożek B., Jakubiec M. 2017. *On the Legal Responsibility of Autonomous Machines*, in «Artificial Intelligence and Law», 25, 293 ff.

Bryson J.J., Diamantis M.E., Grant T.D. 2017. *Of, for, and by the People: The Legal Lacuna of Synthetic Persons*, in «Artificial Intelligence and Law», 25, 273 ff.

Burns J.M., Swerdlow R.H. 2003. *Right Orbitofrontal Tumour with Pedophilia Symptom and Constructional Apraxia Sign*, in «Archives of Neurology», 60, 3, 437 ff. Available on: https://pubmed.ncbi.nlm.nih.gov/12633158/.

Chopra S., White L. 2011. *A Legal Theory for Autonomous Artificial Agents,* University of Michigan Press.

Coeckelbergh M. 2019. *Artificial Intelligence: Some Ethical Issues and Regulatory Challenges*, in «Technology and Regulation», 2019. Available on: https://techreg.org/article/view/10999.

Currey R., Owen J. 2021. *In the Courts: Australian Court Finds AI Systems Can Be "Inventors."*, in «Wipo Magazine», 3. Available on: https://www.wipo.int/wipo_magazine/en/2021/03/article_0006.html.

Dahiyat E.A.R. 2021. *Law and Software Agents: Are They "Agents" by the Way?*, in «Artificial Intelligence and Law», 29, 59 ff.

Danziger S., Levav J., Avnaim-Pesso L. 2011. *Extraneous Factors in Judicial Decisions*, in «Proceedings of the National Academy of Sciences», 108, 6889 ff.

Dilman I. 1999. *Free Will: An Historical and Philosophical Introduction*, Routledge.

Floridi L., Cowls J., Beltrametti M., Chatila R., Chazerand P., Dignum V., Luetge C., Madelin R., Pagallo U., Rossi F., Schafer B., Valcke P., Vayena E. 2018. *AI 4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in «Minds and Machines», 28, 689 ff.

Floridi L., Sanders J. W. 2004. *On the Morality of Artificial Agents*, in «Minds and Machines», 14, 349 ff.

Francesconi E. 2022. *The Winter, the Summer and the Summer Dream of Artificial Intelligence in Law: Presidential Address to the 18th International Conference on Artificial Intelligence and Law*, in «Artificial Intelligence and Law», 30, 147 ff.

Gunkel D.J. 2012. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*, The MIT Press.

Haack S. 2008. *Putting Philosophy to Work: Inquiry and Its Place In Culture—Essays on Science, Religion, Law, Literature, and Life*, Prometheus Books.

Hage J. 2005. *Studies in Legal Logic*, Springer.

Hage J. 2013. *Three Kinds of Coherentism*, in Araszkiewicz M., Savelka J. (eds.), *Coherence: Insights from Philosophy, Jurisprudence and Artificial Intelligence*, Springer.

HAGE J. 2017. *Theoretical Foundations for the Responsibility of Autonomous Agents*, in «Artificial Intelligence and Law», 25, 255 ff.

HAGE J. 2021. *Are the Cognitive Sciences Relevant for Law?* In HAGE J., BROZEK B., VINCENT N. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 17 ff.

HAGE J., WALTERMANN A. 2021. *Responsibility, Liability, and Retribution*, in HAGE J., BROZEK B., VINCENT N. (eds.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge University Press, 255 ff.

HAMPTON J. 1997. *Political Philosophy*, Westview Press.

HIMMA K.E. 2009. *Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to Be a Moral Agent?*, in «Ethics and Information Technology», 11, 19 ff.

HODGES W. *Tarski's Truth Definitions*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), https://plato.stanford.edu/archives/fall2018/entries/tarski-truth/.

HUME D. 1978. *A Treatise of Human Nature*, Selby-Bigge edition, Oxford University Press. (Originally published in 1739-40)

JONES O.D. 2013. *Seven Ways Neuroscience Aids Law*, in «Vanderbilt Public Law Research Paper», 13-28. Available on: https://ssrn.com/abstract=2280500.

LEHRER K. 1992. *Coherentism*, in DANCY J., SOSA E. (eds.), *A Companion to Epistemology*, Blackwell.

LEITER B. 2005. *American Legal Realism*, in GOLDING M.P., EDMUNDSON W.A. (eds.), *The Blackwell Guide to the Philosophy of Law and Legal Theory*, Blackwell, 50 ff.

MACKOR A.R. 2013. *What Can Neuroscience Say about Responsibility? Taking the Distinction between Theoretical and Practical Reason Seriously*, in VINCENT N. (ed.) *Neuroscience and Legal Responsibility*, Oxford University Press, 53 ff.

MATULIONYTE R. 2021. *Australian Court Says that AI Can Be an Inventor: What Does It Mean for Authors?*, in «Kluwer Copyright Blog», 29 September 2021. Available on: http://copyrightblog.kluweriplaw.com/2021/09/29/australian-court-says-that-ai-can-be-an-inventor-what-does-it-mean-for-authors/.

MURPHY J. 2017. *Retribution*, in LUNA E. (ed.), *Reforming Criminal Justice: Punishment, Incarceration, and Release. Volume 4*, Arizona State University.

NBC NEWS 2003. *Pedophile Lost Urge after Surgery*, 1 November 2003 (Source: The Associated Press). Available on: https://www.nbcnews.com/health/health-news/pedophile-lost-urge-after-surgery-flna1c9478663.

PAVLUS J. 2021. *The Computer Scientist Training AI to Think with Analogies*, in «Artificial Intelligence Quanta Magazine», 6 August 2021. Available on: https://www.scientificamerican.com/article/the-computer-scientist-training-ai-to-think-with-analogies/.

RAWLS J. 1955. *Two Concepts of Rules*, in «Philosophical Review», 64, 1, 3 ff.

SCHAFFNER, K. F. 1990. *Case-Based Reasoning in Law and Ethics,* Presentation at the Foundations of Bioethics Conference, Hastings Center.

SOLUM L. B. 1992. *Legal Personhood for Artificial Intelligences*, in «North Carolina Law Review», 70, 4, 1231 ff.

STRAWSON P.F. 1962. *Freedom and Resentment*, in «Proceedings of the British Academy», 48, 187 ff.

SUNSTEIN C.R. 2001. *Of Artificial Intelligence and Legal Reasoning*, in «University of Chicago Law School», 8, 29 ff.

THAGARD P. *Cognitive Science*, in ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), https://plato.stanford.edu/archives/win2020/entries/cognitive-science/.

VERHEIJ B. 2020. *Artificial Intelligence as Law: Presidential Address to the Seventeenth International Conference on Artificial Intelligence and Law*, in «Artificial Intelligence and Law», 28, 2, 181 ff.

ZIPURSKY B.C. 1997. *Legal Coherentism*, in «Southern Methodist University Law Review», 50, 5, 1679 ff.

LegalTech in the Light of the Upcoming Artificial Intelligence Act

SUSANA NAVAS

1. *Artificial Intelligence and law. Computational models*

1.1. *Preliminary remarks*

Originally, the Artificial intelligence (AI) that was applied to Law, i.e. "*Artificial Legal Intelligence*" (ALI), was mainly focused on the study of the automation of legal reasoning and solving legal problems. Then, these studies were followed by working on computational models for legal argumentation.

GRAY (1997, 3) defined ALI as «the computer simulation of any of the theoretical practical forms of legal reasoning, or the computer simulation of legal services involving the communication of the legal intelligence». The ALI finds its origin in jurimetrics, i.e. the computerization of law. It was suggested in the late 1940s and early 1950s by the American School of Jurimetrics (BOURCIER & CASANOVAS 2003, 64-67).

Two computational models that are applied to legal reasoning to date are as follows:

i) *expert systems based on formal logic* by providing a set of rules
ii) *expert systems for conceptual information retrieval and cognitive computing* (*cognitive AI*):

- Expert information retrieval systems are automated systems that extract relevant legal information based on the association of concepts found in a text or document with other concepts, to solve the legal issue at hand.
- In *cognitive computing*, the algorithm not only selects, sorts, and summarizes the information for the end-users in a convenient fashion, but also explores and interacts with the data in unforeseen ways, providing creative solutions to legal problems. It, therefore, extends the capabilities and features of the existing expert systems by adding various techniques and approaches that fall under the domain of AI. Among those *machine learning* and its sub-area *deep learning* occupy a prominent place.

These two computational models differ based on the information source from which the data is extracted. In expert systems based on formal logic, the rules are fed into the system through binary coding instructions designed by the human engineers, and these rules are introduced into the machine by humans. Whereas in the expert systems dealing with information retrieval and cognitive AI, the system extracts knowledge from legal texts, documents, and jurisprudence on its own and provides a legal solution that explains and argues one or several other relevant cases and, in addition, it also makes predictions (ASHLEY 2017, 12 f., 34).

1.2. *Classical computational logic. Expert systems*

To the extent that legal reasoning and legal argumentation are based on Logic, computational logic can constitute a model that is capable of representing legal rules, inferring rules from case laws, and legal principles from the existing legal texts.

Former studies on the subject were carried out by the pioneers ALLEN (1957, 833-879) and NEWELL, SHAW, and SIMON (1959, 256-264), among others. However, their findings were not implemented in practice. In fact, the first attempt at automation was made in the 1980s concerning the logical representation of legal provisions.

At that time, it was thought that AI would have a major impact on the legal field, but the opposite was observed mainly because *logic-based knowledge and reasoning* were unable to adequately represent legal rules, insofar as they are interpreted differently. They are ambiguous, written as general clauses, use indeterminate legal concepts, or may contain contradictory legal propositions. Logic-based reasoning, on the other hand, focuses on true or false statements with little or no nuisance.

There are also other limitations, such as the existence of different legal systems depending on the jurisdiction, or the fact that inferences that are drawn always possess a certain degree of probability of being true, thus a certain degree of uncertainty always exists. Legal reasoning does not have to be true, rather it must comply with a certain level of probability. Moreover, in Law, we also work with "presumptions", which are difficult to be represented by classical computational logic.

In this light of thoughts, LEENES and LUCIVERO (2014, 193-220, 225) state that «Automating this flexibility in rule compliance is difficult, just as it is difficult to automate social activities and non-written rules that are embedded in drivers' practices». Difficult, yes, but not impossible, as LIEBWALD (2015, 301-314) points out.

Work has been done within the field of classical computational logic on models based on case-based knowledge and reasoning, prediction of legal solutions, and models of legal argumentation.

On the other hand, the information in classical computational logic is entered manually, i.e. the computer engineer or scientist has to add it into the system which is time-consuming and costly. For an expert system based on logical reasoning to be efficient, the focus should be put on a specific and limited area of law (e. g. product liability).

Classical computational logic can hardly select arguments for or against a certain legal proposition to deliver the best solution to a legal problem. Also, it must be kept in mind that the inference system changes when information is added, modified, or becomes invalid.

To provide the best possible solution, "something more" is needed. This "something more" is what scientists have been working on in recent decades and intensively in recent years. Efforts have been devoted to the study of computational models for legal argumentation. However, extensive further research still ought to be done in this field (OSKAMP & LAURITSEN 2002, 227-236; ASHLEY 2017, 129).

The development of a "fuzzy logic" may contribute to overcoming the barrier that formal logic presents. In this case, hermeneutics, or the science that studies the interpretation of legal rules should be taken into account to develop an AI system that is capable of formalizing and selecting the most appropriate interpretation criteria so that the application of a rule could generate an efficient and verifiable result. Nevertheless, systems with fuzzy logic remain unfinished and unreliable so far. Perhaps, analyzing how concepts are formed and ideas are associated in the human brain will allow progress to be made towards its computerization[1].

Additionally, a rule must be presented in an easily comprehensible fashion to the common person and the automation of legal reasoning using formal logic led to outcomes that were hard to understand for the average citizen and only accessible by the lawyer or another expert in the legal

[1] See in this regard the paper drafted by SCHORLEMMER et al., 2016.

field. This was a major obstacle to "automation". Moreover, authorities have been very reluctant to automate legal reasoning and the automated drafting of legal texts (BRANTING 2017, 5-27).

Although the classic computational model is superseded, it is still being applied in certain contexts. For example, the *Center for Computer-Assisted Legal Instruction and IIT Chicago-Kent College of Law's Center for Access to Justice & Technology* has a web-based system that helps litigants who don't have a legal representative or counselor draft and file a document before the court to sue or repel a legal action (ASHLEY 2017, 351-354).

1.3. *Artificial intelligence applications*

In a system of knowledge based on AI[2], the aim is to draw inferences from the analysis of collections of legal documents, whether they are written legal rules, judgments, contracts, public documents, or any other legal data. It also serves to find the best criterion for interpreting a rule or a legal term based on the analysis of multiple cases. The development of this new approach coincides with the development of statistical analysis techniques and the retrieval of correlations or patterns from a huge volume of data (*Big Data*), comprising legal data (BRANTING 2017, 5-27).

These computational models are working with i) presentation of legal concepts through ontologies and taxonomies; ii) retrieval of information; iii) learning from legal texts; iv) extraction or summary of information; v) extraction, a summary of legal arguments and predictions.

The three legal domains in which those models have been applied are:

i) Case law, an area that is particularly relevant for litigation. AI applications can provide important assistance to judges and, in general, to courts insofar as they can identify factual assumptions and extract principles from court decisions. In the Common Law countries, this application of AI is relevant because of the importance of the precedents, although it may offer advantages to judges and magistrates in Civil Law countries as well. Likewise, for argumentation, the analysis of the case laws enables the AI system to detect and extract certain arguments that are repeated by courts over time in cases of the same kind in a much faster and more reliable way than the "handwork" performed by legal experts[3].

For lawyers, the principles extracted from case analysis could be used for providing more accurate advice to clients and avoiding filing needless lawsuits. In this regard, the AI system could be considered more of an «auxiliary» to the legal practitioner than a "co-worker". Recently, the use of so-called "robo advisors" and "robot lawyers" indicates the possibility of legal advisors being replaced by AI in the future for legal counseling in low complexity cases[4]. Whether this automation is desirable at a corporate level or there are spurious interests involved, that is often the case, is another matter.

Concerning this area, perhaps the ongoing studies on *personalized automated assessments* could serve as a basis for automating part of the intellectual task of judges, for instance, in corruption or mass tort cases (GUTIÉRREZ et al. 2016).

ii) Document analysis. This field concerns the retrieval of information from a large volume of documents (e.g. identification of certain entities, lawsuits, quoted legal texts, and so on), the automated filling of case summaries, court decisions or legal documents, the constant updating of legal information, automated completion of contract forms or even typing court decisions (SOLAR CAYÓN 2019, 83 ff.). Nowadays, much of the interest in AI is focused on designing and

[2] I should forward to the AI definition presented in section 3.

[3] SHULAYEVA et al. 2017, 107-126. These authors discuss developments in automation with respect to case law, extracting legal arguments by distinguishing the *ratio decidendi* from the *obiter dictum*.

[4] I will refer to the LegalTech tools specifically aimed at consumers later on.

implementing systems that could quickly and effectively analyze large amounts of data.

iii)Analysis, drafting and auditing of legal texts, such as codes, acts, regulations, or ordinances. The rules of the legal system are systematically intertwined. In many cases, the meaning of a rule can only be fully understood when it is interpreted in the light of other rules that are part of the general or sectoral body of law in question. AI can be applied to the analysis of this legal system by relating rules to others, providing information about the best wording of a legal rule, and highlighting the influence or importance of a certain rule in the decision of certain cases by the courts. It also allows considering if the regulation of a specific case can be applied to another (e.g. the application of the rule of civil liability for the use of motor vehicles to the use of autonomous vehicles), or to identify rules that should be amended to maintain the coherence of the system, and so on.

In the particular case of legal drafting, the AI has been helping efficiently in carrying out various tasks such as searching for legal texts that may be relevant to a new rule that is to be drafted, generating working documents, linking of legal provisions or topics, numbering rules, giving written wording to the rule for edition, searching for legal terms or making up complex lists of legal information, etc.

Some of these tasks can take place in a phase before the drafting of the legal regulation since by employing AI it can be analyzed whether or not there is a need for this regulation, its potential applicability and economic impact. Once the regulation has been drafted and entered into force, the AI system may supervise its current application, any requirements for improvement, and analyze where it is necessary to calculate the economic cost of the application of the regulation in question. In short, AI can be used for auditing and quality control of regulations (see also BOURCIER & CASANOVAS 2003, 104-110).

Several AI systems already exist in the field of law. For example, IBM's Watson technology is used by the system called *Ross Intelligence* and it was created by a group of students at the University of Toronto. Also, its first cousin, *Debater*, who is also from IBM, serves the primary function of extracting legal arguments from a large database. Other systems used in the legal area are *Lex Machina* and *Ravel*. The former was acquired by LexisNexis and it makes predictions based on cases in patent and intellectual property law. It is based on the analysis of litigants' behavior. The second system that is mentioned was created by *Stanford Law School* students—joined by the *Harvard Law School* library—to scan a large portion of American case laws so that *Ravel* could visually relate a case with a legal concept.

These systems also include machine learning capabilities to predict outcomes. In this regard, they are constantly updating information while processing, learning from the environment, and steadily adjusting their results accordingly.

Moreover, these systems are based on an open-source architecture, which means that other intelligent tools applicable to the legal field could more easily be developed in the future. In the next section, the paper will focus on these "LegalTech" tools.

2. *Legaltech Ecosystem for Consumers*

2.1. *LegalTech and access to justice*

The term "LegalTech"—and what it stands for—is frequently used in our modern times among legal practitioners, particularly, lawyers[5]. It is described as applying new technologies by

[5] See, for instance, the Report *Future Lawyering 2020: emerging business areas: identifying opportunities* of the General

lawyers to provide their legal services for the tasks that are specifically performed by them even though the application of these tools in the field of justice is already taking place[6]. For some time now, legal practitioners have been using computer programs, sophisticated databases, and communications applications via *smartphones* (e.g., *WhatsApp* or email) that have eased their workload. They are, of course, aware of such tools.

Currently, the tech industry is going a step further by incorporating high-level technology into lawyers' daily work. The different techniques and systems that fall under the term "AI" will not only facilitate the work of the lawyers as a human-machine collaboration develops but will also carry out autonomously some of the activities which require human intervention.[7] Feasible scenarios include case analysis (e.g. e-discovery, big data analytics), drafting of legal documents such as lawsuits, automated drafting of contracts, or "legal" robots that provide legal information to clients.

As it is highlighted above, further research is being done on computational models about legal reasoning that allows the retrieval of legal arguments directly from legal materials (rules, judgments, journal articles,...). It is done so that predictions can be made about the outcome of, judicial decisions, complex legal questions are answered, or make decisions with legal relevance[8].

Based on the classification suggested by SOLAR CAYÓN (2019), the automation of legal services through AI-based tools can be classified into the following groups: i) legal research related to the tasks which require research, selection, and analysis of legal information; ii) compliance; iii) legal due diligence; iv) predictive analytics, breaking down the behavior of judges and courts to better gauge the success or failure rate of a lawsuit v) e-discovery or selection of evidentiary material; vi) automated production of personalized legal documents by applications; vi) online dispute resolution (ODR).

ROSS, inspired by IBM's Watson supercomputer, is capable of analyzing a huge amount of legal material within seconds and it can alert and update lawyers about any new significant information on their cases. Initially focused on insolvency, ROSS has been applied to new areas. ROSS uses a machine-learning AI-based technique.

Although, in the beginning, these collective set of technological tools, which can be directly used by consumers, were using the umbrella term "LegalTech"[9], the truth is that the expression "LawTech" has also been employed to draw clear differentiation between those tools that are designed for consumers and law firms separately (BUES & MATTHAEI 2017, 89-109). Between these two expressions, "LegalTech" and "LawTech", the former is winning the battle to refer to AI systems applied by lawyers (SALMERÓN-MANZANO 2021, 24), whether in B2B or B2C relationships, by justice (ENGELMANN et al., 2021, 317 ff.) and used by consumers. With particular regard to LegalTech tools in B2C relationships, it should be stressed that the consumer is not just a "client"[10] but a potential user of automated legal services who has access to legal information for a modest fee or in some cases is free of charge.

For each case, these automated legal services employ digital services and content. Consumers usually prefer to use online legal services because they are cost-effective. The digital element plays a very significant role in convincing consumers "to buy" such services so much so that if

Council of Spanish Lawyers published by Wolters Kluwer. Available at: https://www.abogacia.es/2019/05/10/ informe-abogacia-futura-2020-areas-de-negocio-emergente-identificar-oportunidades/.

6 NIEVA FENOLL 2018; BEN-ARI et al. 2017, 35. More recently, see the detailed study by BARONA VILAR 2021, 344 ff.

7 WAGNER 2018, 2-4; BARRIO ANDRÉS 2019, 37-66.

8 In this respect, reference should be made to the excellent monograph written by ASHLEY 2017, 10 ff. On these issues, see: NAVAS NAVARRO 2017, 24 ff.

9 Indeed, this vague term seems to include any technology applied in the field of Law, BECK 2019, 648.

10 The "client" would be the person (natural or legal) who requests a specific legal service from an expert, i.e., a "personalized" legal service (e.g., advice, handling of a specific case, and so on).

no such services are offered online they do not seek legal advice[11].

These consumers represent a "latent" (SUSSKIND R. & SUSSKIND D. 2016, 127) or even "non-existent" market (SOLAR CAYÓN 2019, 99) for the legal services sector. They are citizens who, because of their low income, cannot afford lawyers' fees. They are unable to get legal aid or make applications for small claims that they think are not affordable to pursue. Furthermore, consumers lack the basic skills or expertise required for legal drafting of a claim and filing it before the competent authority for cases of smaller claims as it is mentioned before[12].

Legal services provided online by tech companies (start-ups) are a particularly attractive alternative to traditional law firms for consumers because these tools are cost-effective for them[13]. Thus, those companies provide automated legal services or access to cloud services where consumers can fill customizable applications with or without the assistance of a chatbot. Such tasks are currently performed by legal practitioners.

2.2. *Technological tools available to consumers*

This section will discuss some of the legal tools that already exist in the "legal ecosystem" (GONZÁLEZ-ESPEJO GARCÍA 2019, 345 ff.), and they are accessible to the consumers. The saying in the finance sector "We need banking services, but not always banks" is rapidly becoming true for the automation of legal services as well, that is, "We need legal services, but not always lawyers".

Any research that has analysed smart technologies used in the legal field mentions a myriad of applications and tools that are directly accessible to consumers[14]. Using their PC, tablet, or smartphone (BRESCIA et al. 2015, 578-579), consumers can access online applications that can review contracts, small money claims from an airline, legal advice from bots, interact with virtual assistants, use freelance websites to hire legal experts, etc. Five scenarios will be further exposed in detail below. Although in theory, this paper presents them as independent technological tools, they produce "legal products" for the end-users when combined i.e. legal drafting, legal advice, intermediation, etc. Thus, the automated process of claiming some amount of money may be preceded by a chat with a virtual assistant who can assist the consumers with information about their rights.

High-profile AI systems, such as *Ross Intelligence,* that can process a large amount of data and are capable of making predictions about case decisions and judgments are only used by limited law firms (WAGNER 2018, 31 ff.). They are not available to consumers yet, but they will be in the future (SUSSKIND R. & SUSSKIND D. 2016, 41-43).

2.2.1. *Legal services intermediation platforms*

Intermediation platforms in the collaborative economy have also reached the area of legal practice. Nowadays, lawyers and their firms are no longer offering their services via law websites but they are also available on online platforms similar to *Airbnb*, *Uber,* or *Peopleperhour*.

[11] In the study developed in the United Kingdom between 2011 and 2013, in which there was a panel of consumers and a panel of experts (practitioners and academics) on the legal market and, specifically, on the legal education market, it is highlighted, based on the different interviews and surveys carried out, that the connection between accessibility, technology and cost-benefit determines that consumers clearly opt for the provision of online services for low-cost claims (LETR 2013, 99).

[12] As known, these claims can reach a maximum amount of 5000 Euros, if one wants to start the European procedure established by Regulation (EC) Nr. 861/2007 of the European Parliament and of the Council of 11 July 2007 establishing a European Small Claims Procedure (OJEU L 199, 31.7.2007).

[13] In this regard, it is worth to quote the Report commissioned by the American Bar Association (ABA), which reviews the state of the legal profession and the legal services market, highlighting, in the first part of the study, the difficulties that certain groups of the population have in accessing justice in the USA (ABA 2016, 10-19).

[14] Again, I should quote the study drafted by BENNETT et al. 2018, 22 ff.

Such platforms not only facilitate the relationship between lawyer and client through contract but also rank them based on the quality of their services. These platforms feature a range of law firms and have tools that can filter out lawyers with relevant expertise to match a client's needs. They also facilitate the clients in making the hiring decision by showing availability, response time, and rate of legal experts. Similarly, the consumer has access to the reviews left by other clients before hiring a certain professional (COUNCIL OF BARS & LAW SOCIETIES OF EUROPE 2018). The *Digital Services Act*[15] and *Digital Markets Act*[16] set out a range of obligations for online intermediation platforms that will also apply to the ones this paper is discussing.

The terms and conditions of use of these intermediation platforms set forth a disclaimer informing that they provide just legal information instead of legal advice. No contractual relationship between the lawyer and the client is generated by the fact that the former will answer online, where appropriate, some of the questions asked by the last.

The company running the platform is usually a technological *start-up*, instead of a legal services firm. Some global platforms for lawyers are *Rocket Lawyer, Anwalt.de, FlatLaw, Legalzoom, Avvo, Got.Law,* etc. The structure of these platforms is triangular ("two-sided market")[17] as long as they remain merely intermediaries. In some cases, consumers remunerate the performance of their services[18]. When platforms set conditions about the lawyer-client relationship, such as fees or working hours, they become contractual parties vis-à-vis the end-user and an employer vis-à-vis the lawyer, whose services the end-user has purchased. Hence, the doctrine emanating from the *leading case* in Europe regarding online intermediary platforms can be applied here[19]. As known, this is the case of Uber against the professional association Elite Taxis, which gave rise to the decision of the Court of Justice of the European Union on 20 December 2017[20].

The other issue can be the potential conflict between the use of these platforms by lawyers and the observance of legal ethics (COUNCIL OF BARS & LAW SOCIETIES OF EUROPE 2018, 7 ff.). About this, it should be noted that legal ethics are applied to lawyers only. In the new LegalTech scenario, it makes sense to refer to regulatory objectives in general in which a whole series of principles are taken into account, such as the protection of consumers and end-users.

2.2.2. *"Do-it-yourself" tools: Drafting of legal documents*

Several legal service providers allow consumers to download certain tools directly from their websites or a cloud space to draft their documents i.e. will deeds, sale deeds, contracts, agreements, lawsuits, etc. with or without the help of a virtual assistant as it is mentioned before (LÓPEZ-LAPUENTE GUTIÉRREZ & LAMELA DOMÍNGUEZ 2019, 234 ff.).

Those applications employ an AI system based on a decision tree that is designed of a series of questions i.e. filters that can narrow down the search by use of special keywords entered by the end-user. The programs designed for end-users are often user-friendly (BRESCIA et al. 2015, 572-573).

These systems employ natural language processing and machine learning and, although they are not superseding human legal advice yet, as far as they become more and more sophisticated, such replacement will be certain.

[15] COM(2020) 825 final.
[16] COM(2020) 842 final.
[17] ROCHET & TIROLE 2003, 1029; ARMSTRONG 2006, 668-691.
[18] This would be the case of "referral websites for lawyers" (COUNCIL OF BARS & LAW SOCIETIES OF EUROPE 2018, 6).
[19] Expanding this doctrine to similar cases is suggested by HACKER 2018, 80-96.
[20] C-434/15.

2.2.3. *Mass processing of identical cases. Small claims*

Another array of tools directly accessed on the website of a tech company by the end-user deals with small money claims related to flight delays, cancellations, lost luggage, fines in public car parks, delays in trains, etc. (BENNETT et al. 2018, Annex A).

On these websites, the end-user answers a series of questions, via a chatbot, and delivers information about their case similar to hundreds of others. Through an automated process, a document is drafted and signed by the user, authorizing the company to claim, on their behalf, to release the funds that are due. The claim can later be defended by a lawyer for the consumer in court, if necessary when no agreement is reached between parties. It is a remedy of last resort. Therefore, the websites not only offer legal services but also provide the end-users with digital legal content i.e. downloadable customized legal documents to serve their needs.

2.2.4. *From virtual assistants and chatbots to roboadvisors*

Both virtual assistants and chatbots are AI systems that permit humans to have "smart" conversations with robots. They can listen, understand, reason (*cognitive chatbot*) and answer questions (e.g. ChatGPT). They process natural language and usually operate with an AI system based on machine learning and decision trees (SOLANO GADEA 2019, 153 ff.). In the legal field, they can perform a variety of functions ranging from customer call centers to managing the company's operations and providing legal advice.

The end-user can interact with these assistants via smartphone applications, or through a chatbot. In this context, the voice is becoming increasingly important as some of these AI systems process natural language, giving rise to discriminatory biases in particular cases.

On the other hand, personal assistants such as Alexa are not unthinkable but those which can solve legal questions posed by consumers will be embedded intangible goods, which can be purchased both in physical and online shops[21].

2.2.5. *Legal Design*

Several technological tools are devoted to what is known as Legal Design, that is, the use of design techniques in indoor homes, architecture, or fashion but it is for drafting more intelligible contracts than the traditional "complicated" contracts.

Legal Design uses images in place of words to exploit the element of visualization. The "colorfulness" and "showiness" of these contracts are considered "more transparent" than the written expression (!). The tools for *Legal Design,* as well as the "product" that is generated with them, may just easily be accessed online.

The legal design of contracts raises many questions. One question deals with its interpretation as the hermeneutic criteria established in the legal system can hardly help to find the true meaning or purpose intended in the "visual" contract. To achieve the most righteous interpretation of such contracts, resources should be borrowed from other disciplines (e.g., psychology, pedagogy, education)[22], or new rules should be drafted that could better catch the parties' wills expressed by images or audios. The other issue relates to the criteria for determining what constitutes unfair clauses in such contracts.

It might be possible to interpret the images that represent the content of a contract or to determine that a clause is unfair by utilizing AI systems. However, this will end the supposed

[21] This is the case of the *RATIS chatbot* designed by German scientists (TIMMERMANN 2020, 150 ff), which is not for sale.

[22] BRUSCHWING 2021, 219-220.

transparency and easy understanding that is offered with Legal design causing non-transparency i.e. the opacity of the AI system applied. Paradoxically, in search of transparency, we will find opacity in its place.

Legal design, which might be extended to the design of legal norms, could drive quite worrying social engineering work (MAU 2019, 99 ff.) because of the extremely simplistic vision of the reality that this technique affords. The metric society in which we live and the personalization resulting from the use of social networks are also key contributors to such a new version of our real life.

Despite that, in the legal system, there are already some incipient examples of this school of thought, such as the icons used to highlight the level of risk for financial products. A potential implementation of the design in question in the legal system can be seen in Art. 12 para.1 and para. 7 of the GDPR[23], which refers to transparency, communication, and information of the data subject's rights regarding the processing of personal data. Art. 12 para. 1 states, among other aspects, that: «Information shall be provided in writing, or by other means, including, where appropriate, by electronic means». And Art. 12.7 highlights that: «the information to be provided to data subjects pursuant to Articles 13 and 14 may be provided in combination with standardized icons in order to give in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing. Where the icons are presented electronically that shall be machine-readable». These rules are potentially in favor of applying design thinking and, specifically, visual design to the information provided to individuals (BRUSCHWING 2020, 142-160).

As mentioned above, the (new) design of contracts may raise important questions. But it is especially relevant for a person with a particular disability. A certain design of the document in which visual tools are used to support disabled people to keep them informed and help them understand the given information for gaining their express consent about the matters that can affect them more as adat subject should be considered.

Moreover, such legal design could be considered "universal" in terms of the Convention on the Rights of Persons with Disabilities signed in New York on 13 December 2006[24], which defines it, in Art. 2, as: «the design of products, environments, programmes and services to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design». To this list, which is not exhaustive, should be added the design of documents in which legally relevant information is presented.

3. *The Application of the European Proposal for a Regulation on Artificial Intelligence (AIA) to the Legaltech Ecosystem*

How could the European Proposal for a Regulation on Artificial Intelligence (AIA), which was published on 21 April 2021[25] impact the automation of legal services and, particularly, LegalTech tools in B2C relationships?

Primary, it should be noted that the definition of AI as amended by the compromise text has a narrower scope compared to the original text of the AIA[26]. It means more traditional

[23] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation), OJEU L 119/1, 4.5.2016. Quoted as «GDPR».

[24] Available on: https://www.un.org/development/desa/disabilities/.

[25] The review of the AIA has led to a compromise text made public by the end of the same year (29 November 2021). Thus, the provisions I am going to quote in this section correspond to this compromise text [Presidency compromise text. Interinstitutional File: 2021/0106(COD)].

[26] Art. 3 (1): «'artificial intelligence system' (AI system) means a system that:

(i) receives machine and/or human-based data and inputs,

software systems and programming are excluded[27]. Yet, according to the list of techniques and approaches stated in Annex I of the AIA, which have embraced the term "AI", a broad spectrum of LegalTech applications will fall under it. Nevertheless, both the Committee on Legal Affairs and the Committee on Industry, Research and Energy' draft opinions suggest limiting the scope of application of the AiA to those AI systems that use the technique of machine learning and deep learning[28]. In this vein, both draft opinions embrace the AI definition expressed by the OECD under which «an AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy»[29]. According to the first Draft opinion mentioned Annex I of the AIA will refer just to: «Machine learning *and optimization* approaches, including *but not limited to evolutionary computing as well as* supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning»[30].

3.1. *Prohibited practices*

Art. 5 para 1 of the AIA gives a list of prohibited practices. For the technological tools at stake, the following are particularly relevant:

i) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness with the objective to or the effect of materially distorting a person's behavior in a manner that causes or is reasonably likely to cause that person or another person physical or psychological harm (lit. a)

ii) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, disability, or social or economic situation, with the objective to or the effect of materially distorting the behavior of a person pertaining to that group in a manner that causes or is reasonably likely to cause that person or another person physical or psychological harm (lit. b).

(ii) infers how to achieve a given set of human-defined objectives using learning, reasoning or modelling implemented with the techniques and approaches listed in Annex I, and
(iii) generates outputs in the form of content (generative AI systems), predictions, recommendations or decisions, which influence the environments it interacts with».

[27] Recital nr. 6 of the Compromise text of AIA made public on 29 November 2021.

[28] Draft opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD). Rapporteur: Axel Voss, 2.3.2022; Draft opinion of the Committee on Industry, Research and Energy for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 –2021/0106(COD). Rapporteur: Eva Maydell, 3.3.2022.

[29] OECD Legal instruments, *Recommendations of the Council of Artificial Intelligence*, adopted on 22.05.2019, C(2019)34 C/MIN(2019)3/FINAL, Available on: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[30] This amendment justification lies on «the justification for a lex specialis on AI by the Commission was based on the specific characteristics, such as autonomy and opacity, of (rather new) machine-learning and data-driven AI applications. It was argued that they are so far not adequately covered by existing laws. Their existence would therefore demand new laws. Symbolic AI (dominant from the 1950s-90s) is however already covered by numerous EU and national laws. Point (b) and (c) fall exactly in this category. It is therefore not justified to address them - again - within the AI Act. Their inclusion would be contradictory to the impact assessment as well as better regulation principles» (Amendment nr. 285, Draft opinion of the Committee on Legal Affairs).

iii) the placing on the market, putting into service or use of AI systems for the evaluation or classification of natural persons over a certain period of time based on their social behavior or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) Detrimental or unfavorable treatment of certain natural persons or groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected
(ii) Detrimental or unfavorable treatment of certain natural persons or groups thereof that is unjustified or disproportionate to their social behavior or its gravity (lit. c).

In this respect, the compromise text of the AIA is expanding the scope of Art. 5 by adding private individuals to its original scope which was previously applied exclusively to the public authorities.

Out of the variety of smart legal tools designed for the end-users, only those which involve profiling, ranking, or rating people by attributing a score should be regarded as prohibited practices if they breach fundamental rights or have the tendency to manipulate vulnerabilities (ENGELMANN et al. 2021, 321). For instance, this could be the case of online platforms on which lawyers offer services[31], websites that are offering users downloadable "do-it-yourself tools" (LÓPEZ-LAPUENTE GUTIÉRREZ & LAMELA DOMÍNGUEZ 2019, 234 ff.) or in the case of automated small claims services (BENNETT et al. 2018, Annex A).

In any case, other legal bodies should be applied, such as the GDPR and the Unfair Commercial Practices Directive[32]. On the other hand, the upcoming Digital Markets Act will establish limitations and prohibitions concerning intermediaries' platforms that will curb the recombination of data from different sources, which indirectly will lead to the decreasing, or even elimination, of profiling, scoring, and online behavioral advertising.

3.2. *High-risk LegalTech tools*

Can some of the technological tools - designed for end-users be considered "high risk" (Art. 6)? On one hand, the AIA contemplates AI systems that are safety components of other goods. On the other hand, it observes the AI systems themselves; also known as "*stand-alone AI systems*", that could be contemplated as products or systems as stated by the Draft Opinion of the Committee on the Industry, Research and Energy[33]. The LegalTech tools this contribution is dealing with is the latter, that is, the stand-alone AI system.

For a system to be considered a "high risk" AI system, the conditions to be met are different depending if we are dealing with an AI system, which is a safety component of a product or system or a stand-alone-AI system. In the first case, the system must be covered by the legislation that is harmonized with the AIA («*New Legislative Framework*», NLF), which is listed in Annex II. The so-called «New Legislative Framework» is composed of the following legal texts: Regulation (EC) Nr. 765/2008/EC of the European Parliament and of the Council of

[31] COUNCIL OF BARS & LAW SOCIETIES OF EUROPE 2018. Some global platforms for lawyers are Rocket Lawyer, Anwalt.de, FlatLaw, Legalzoom, Avvo and Got.Law.

[32] Consolidated text: Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (Unfair Commercial Practices Directive) (Text with EEA relevance). Available on: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02005L0029-20220528&from=EN.

[33] Amendment nr. 15.

9 July 2008, setting out the requirements for accreditation and market surveillance relating to the marketing of products[34]; Decision Nr. 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products[35] and Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and product conformity and amending Directive 2004/42/EC and Regulations (EC) Nr. 765/2008 and (EU) Nr. 305/2011[36]. Based on this regulatory framework, a set of rules (Directives and Regulations) have been adapted and thus, have become part of the NLF. While others are either in the revision process or are likely to begin soon, i.e. those quoted by Annex II. Moreover, the system is required to conform to the legislation by a third party before being placed on the market.

In addition, there is a whole range of harmonized[37] standards that are published in the OJEU[38].

In case of stand-alone-AI systems, Art. 6 para. 3 AIA[39] warns that these high-risk systems must be applied to specific areas that are expressly mentioned in Annex III. It should keep in mind that in addition to the technological tools applied in legal practice, other tech tools also exist, as I have mentioned before, which are intended to be used throughout the judicial process and that can be grouped under the term "e-justice" or "smart justice" (SUSSKIND R. 2019, 253 ff; BARONA VILAR 2021, 610 ff.).

The last domain mentioned in Annex III, number 8, concerns «the administration of justice and the democratic process» and includes «AI systems intended to be used by a judicial authority or on their behalf for interpreting facts or the law for applying the law to a concrete set of facts». Therefore, intelligent tools that are currently applied by legal practitioners would be left out. It does not seem that the area of «the administration of justice and democratic processes» could be interpreted so broadly as to encompass the automated "legal products" that are being discussed here[40]. Thus, they can not be considered "high-risk" systems under the scope of the discussion topic of this AIA.

However, as long as these systems involve machine learning and natural language processing, there is always a risk of biases and lack of transparency[41]. Of course, if the system is not fed with quality data, it could, for example, display racial, ethnic, or cultural biases by denying access to justice to a certain social group or make "unfair" decisions. The impact of such types of LegalTech tools on the fundamental rights of due process of law, right to defense, right to trial, among others, asserts that such tools should be classified as "high-risk" or be categorized as prohibited practices (ENGELMANN et al. 2021, 318 ff; NIEVA FENOLL 2018, 127 ff.). Based on this observation, it is recommended that Annex III should include not only the tech tools intended for court use but also the tools that are used in the legal practice by lawyers and law firms or those made by the tech companies for end-users (e.g. large generative AI models).

To partially avoid situations where it is uncertain whether an AI system would affect a specific area and thereby make it a high risk, the Draft Opinion of the Committee on Industry, Research and Energy suggested the introduction of a new rule; «In case there is uncertainty

[34] OJEU L 2018/30, 13.08.2008.
[35] OJEU L 218/82, 13.08.2008.
[36] OJEU L 169, 25.06.2019.
[37] More on this matter is available on: https://ec.europa.eu/growth/single-market/european-standards/vademecum_en.
[38] BLUE GUIDE PUBLICATION, (nt 75) 4.1.2.2.
[39] Axel Voss 2.3.2022.
[40] Nonetheless, it should be expanded in order to embrace ODR.
[41] The absolute absence of errors is no possible. Therefore, Art. 10 para. 3 AIA has been amended by the compromise text, made public on 13. January 2022, in order to make clear this concern. The proposed text considers that «Training, validation and testing data sets shall be relevant, representative, and to the best extent possible, free of errors and complete» [Interinstitutional File: 2021/0106(COD)].

over the AI system's classification, the provider shall deem the AI system high-risk if its use or application poses a risk of harm to the health and safety or a risk of adverse impact on fundamental rights of users, as outlined in Article 7(2)»[42]. Accordingly, not all of the LegalTech tools could be "high risk" despite their absence in Annex III.

The requirements that a high-risk AI system must fulfill in accordance with its specific puropose are the following: a risk management system (Art. 9); if the system uses machine learning, data sets must meet a range of quality requirements (Art. 10) to be considered as high quality. This can significantly reduce the number of errors and discriminatory biases; technical specifications must be documented (Art. 11); a mechanism to record the system motions must be implemented (Art. 12); transparent information for the users (Art. 13), who are not consumers but those who own and/or control the system (Art. 3 para. 4); human supervision (Art. 14); and accuracy, robustness, and cybersecurity (Art. 15).

3.3. *Low-risk LegalTech tools*

With regard to low-risk AI systems, which embraced the "limited" and "residual" risk AI systems regulated by the AIA, we must take into account, on the one hand, Art. 52, which states a duty of transparency about "certain" AI systems when they are interacting with end-users. In such cases, the end-users should be informed by the service provider that they are interacting with an AI system and not a person unless it is an obvious circumstance i.e. virtual assistant or a roboadvisor.

On the other hand, Art. 69 AIA deals with the AI systems that present "minimal or residual risk" and seeks to promote the development of codes of conduct with the clear intention that providers voluntarily comply with the requirements that are set out in Title III, Chapter 2 AIA and they have been referred earlier in this work.

Obviously, some of the intelligent legal tools that are based on decision trees (e. g. answers and questions), pose a low risk for the consumers. Nonetheless, it could be thinkable that some of these LegalTech tools embrace functionalities, some of them lead to qualify the system as "high risk" while others as "low risk" (AIDA 2022). Such hybrid AI systems are not brought under the umbrella of AIA. Therefore, there is uncertainty about the set of rules applicable to such AI systems. This gap should be bridged through the parliamentary drafting process.

In addition, the purpose of an AI system must also be taken into account. Indeed, there is a difference between a "general purpose" AI system and a "specific purpose" (intended purpose) AI system. The first comprises AI systems that are capable of executing general functions established in Recital nr. 70a added by the compromise text of 29 November 2021, such as voice or image recognition, pattern detection, video generation, translations, questions, answers, etc. On the contrary, the second refers to the AI systems that have an intended use. It is specified by the provider or by whoever introduces or puts it into use or service in the market and determines its terms and circumstances of use and creates instructions. This distinction is relevant to the extent that if the AI system is a general-purpose system, it should not meet the requirements required by the AIA. Whereas they will be mandatory for the intended purpose of AI systems (new Art. 52a). From the examples given by the compromise text, it can be deduced that only AI systems with minimal and residual risk are taken into consideration. It seems unlikely that a high-risk general-purpose AI system would not be complying with the requirements of the AIA just because it is considered a "general" purpose AI system. In short, the relationship between the classification of the type of an AI system based on the degree of risk it poses and its purpose is not as clear as it is desired to be as evidenced by the last AIA's version of the Council concerning large generative AI models.

[42] Amendment nr. 33.

4. *Conclusions*

The use of LegalTech tools to provide automated "legal information" to consumers strengthens consumer protection insofar as large segments of the population will have access to cost-effective legal services. It enables access to justice or other alternative dispute resolution mechanisms which, without the existence of "smart" tools, is expensive or out of reach for most people. The legal sector transformation is determining, as it is stated at the beginning of this paper, that while "legal services" are obviously a necessity, lawyers are not always needed, to the extent that those services could be provided by other market actors with the same or higher quality and efficiency (SANDEFUR 2019, 49-55.). However, it must be kept in mind that legal tech tools of the high-risk AI systems would be slow to implement.

This new legal practice scenario must be encouraged and initiated as soon as possible by liberalizing legal markets and allowing alternative business structures i.e. in the UK or Germany to participate. This, in turn, will provoke the review of the general statutes of the legal profession and the codes of ethics. It is important to add the following ethical aspects in the statute or the code of ethics:

(i) The lawyer's duty of long learning education, in particular, regarding legal technology;
(ii) The duty to record that the lawyer is assisted by an AI system or a "non-human assistant";
(iii) That the lawyer has to supervise the result generated by the system, at least at an early stage of the technology employed, in the same way as he or she supervises human assistants, e.g., paralegals and, finally; and
(iv) The duty to inform the customer that he or she is assisted by an AI system.

Another outcome of this transformation will be the emergence of new jobs for lawyers who will need training in technology or technologists trained in Law (SUSSKIND R. 2017, 133 ff.). The legal engineers who have the skill of representing legal rules through binary code, the legal technologist (bridging the gap between legal practice and the administration of justice and technology), the legal process analyst (splitting up cases and disputes into tasks that can be automated by deciding which provider can offer the best service in relation to each of them), the legal project manager (controls the process of legal decomposition and outsourcing that has been executed by the legal process analyst), the ODR specialist, the legal designer or the legal risk analyst would be more in demand than traditional lawyers.

Employers, on the other hand, would no longer be law firms but technology companies, legal know-how providers, legal process outsourcers, global finance companies, and many others.

A prospective regulation, both on professional services and professional bodies, should take full account of the digitalization of services that openly involve legal services now.

As a preliminary step, however, the review of Law degrees and LLM curricula to implement new skills (technology, business management, and innovation) is, in my opinion, an essential first step.

The application of LegalTech tools in the field of justice will be slower than in the legal practice but it is unavoidable. In any case, the principles highlighted in the *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*, adopted in Strasbourg in December 2018 by the European Commission for the Efficiency of Justice (CEPEJ 2018, 14) should be taken into account when developing and applying those LegalTech tools.

These principles are:

(i) The *principle of respect for fundamental rights*: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights.

(ii) The *principle of non-discrimination*: specifically prevents the development or intensification of any discrimination between individuals or groups of individuals
(iii) The *principle of quality and security*: about the processing of judicial decisions and data, use of certified sources and intangible data with models elaborated in a multidisciplinary manner, in a secure technological environment
(iv) The *principle of transparency, fairness, and justice*: make data processing methods accessible and understandable, authorize external audits, and
(v) The *principle under «users» control*: preclude a prescriptive approach and ensure that users are informed actors in control of the choices made.

References

ABA 2016. *Report on the Future of Legal Services in the United States*, by the American Bar Association (ABA) Commission on the Future of Legal Services, https://www.americanbar.org/groups/centers_commissions/center-for-innovation/past-work/commission-on-the-future-of-legal-services/.

AIDA 2022. *Identification and Assessment of Existing and Draft EU Legislation in the Digital Field*, Study requested by the AIDA special committee, January 2022. Available on: https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2022)703345.

ALLEN L.E. 1957. *Symbolic Logic: A Razor-Edged Tool for Drafting and Interpreting Legal Documents*, in «Yale Law Journal», 66, 833 ff.

ARMSTRONG M. 2006. *Competition in Two-Sided Markets*, in «RAND Journal of Economics», 37, 3, 668 ff.

ASHLEY K.D. 2017. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press.

BARONA VILAR S. 2021. *Algoritmización del Derecho y de la Justicia. De la inteligencia artificial a la Smart Justice*, Tirant Lo Blanch.

BARRIO ANDRÉS M. 2019. *Legal Tech y la transformación del sector legal*, in BARRIO ANDRÉS M. (ed.), *Legal Tech. La transformación digital de la abogacía*, Wolters Kluwer, La Ley.

BECK W. 2019. *Legal Tech und Künstliche Intelligenz*, in «DÖV», 16, 639 ff.

BEN-ARI D., FRISH Y., LAZOVSKI A., ELDAN U., GREENBAUM D. 2017. *Danger, Will Robinson? Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment*, in «Richmond Journal of Law & Technology», 23, 2, 13 ff.

BENNETT J., MILLER T., WEBB J., BOSUA R., LODDERS A., CHAMBERLAIN S. 2018. *Current State of Automated Legal Advice Tools*, in «Analysis and Policy Observatory», 2 May 2018.

BOURCIER D., CASANOVAS P. (eds.) 2003. *Inteligencia artificial y derecho*, Editorial UOC.

BRANTING L.K. 2017. *Data-Centric and Logic-Based Models for Automated Legal Problem Solving*, in «Artificial Intelligence and Law», 25, 5 ff.

BRESCIA R.H., MCCARTHY W., MCDONALD A., POTTS K., RIVAIS C. 2015. *Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice*, in «Albany Law Review», 78, 2, 553 ff.

BRUSCHWING C.R. 2021. *Visual Law and Legal Design. Questions and Tentative Answers*, in SCHWEIGHOFER E., KUMMER F., SARENPÄÄ A., EDAR S., HANKE P. (eds.), *Cybergovernance*, Weblaw, 219 ff.

BRUSCHWING C.R. 2020. *Humanoid Robots for Contract Visualisation*, in «UNIO - EU Law Journal», 6, 1, 142 ff.

BUES M.M., MATTHAEI E. 2017. *Legaltech on the Rise: Technology Changes Legal Work Behaviours, but Does Not Replace Its Profession*, in JACOB K., SCHINDLER D., STRATHAUSEN R. (eds.), *Liquid Legal: Transforming Legal into Business Savy, Information Enabled and Performance Driven Industry*, Springer.

COUNCIL OF BARS & LAW SOCIETIES OF EUROPE. 2018. *CCBE Guide on Lawyer's Use of Online Legal Platforms*. Available on: https://www.ccbe.eu/fileadmin/speciality_distribution/public/documents/DEONTOLOGY/DEON_Guides_recommendations/EN_DEON_20180629_CCBE-Guide-on-lawyers-use-of-online-legal-platforms.pdf.

CEPEJ 2018. *European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment*, 3-4 December 2018. Available on: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

ENGELMANN Ch., BRUNOTTE N., LÜTKENS H. 2021. *Regulierung von Legal Tech durch die KI-Verordnung*, in «Recht Digital», 7, 317 ff.

GONZÁLEZ-ESPEJO GARCÍA Mª.J. 2019. *El ecosistema Legal Tech en España* in BARRIO ANDRÉS M. (ed.), *Legal Tech. La transformación digital de la abogacía*, Wolters Kluwer, La Ley, 345 ff.

GRAY P.N. 1997. *Artificial Legal Intelligence*, Dartmouth Publishing Company.

GUTIÉRREZ P., OSMAN R., ROIG C., SIERRA C. 2016. *Personalised Automated Assessments*, in THANGARAJAH J., TUYLS K., JONKER C., MARSELLA S. (eds.), *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*, IFAMAAS.

HACKER Ph. 2018. *UberPop, UberBlack, and the Regulation of Digital Platforms after the Professional Association Elite Taxi Judgment of the CJEU*, in «European Review of Contract Law», 14, 1, 80 ff.

HONGDAO Q., BIBI S., KAHN A., ARDITO L., KHASKHELI M.B. 2019. *Legal Technologies in Action: The Future of the Legal Market in Light of Disruptive Innovations*, in «Sustainability», 11, 4, 1015. Available on: https://www.mdpi.com/2071-1050/11/4/1015.

LEENES R., LUCIVERO F. 2014. *Laws on Robots, Laws by Robots*, in «Law, Innovation and Technology», 6, 2.

LETR 2013. *Setting Standards: The Future of Legal Services Education and Training Regulation in England and Wales*, Report of the Legal Education and Training Review.

LIEBWALD D. 2015. *On Transparent Law, Good Legislation and Accessibility to Legal Information: Towards an Integrated Legal Information System*, in «Artificial Intelligence and Law», 23, 301 ff.

LÓPEZ-LAPUENTE GUTIÉRREZ L, LAMELA DOMÍNGUEZ A. 2019. *La automatización de contratos* in BARRIO ANDRÉS M. (ed.), *Legal Tech. La transformación digital de la abogacía*, Wolters Kluwer, La Ley, 234 ff.

MAU S. 2019. *The Metric Society*, Polity Press.

NAVAS NAVARRO S. (ed.). 2017. *Inteligencia artificial, tecnología, derecho*, Tirant Lo Blanch.

NEWELL A., SHAW J.C., SIMON H.A. 1959. *Report on a General Problem-Solving Program*, in *Proceedings of the International Conference on Information Processing*, 1 ff. Available on: http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General _Problem-Solving_Program_Feb59.pdf.

NIEVA FENOLL J. 2018. *Inteligencia artificial y proceso judicial*, Marcial Pons.

OSKAMP A., LAURITSEN M. 2002. *AI in Law practice? So Far, Not Much*, in «Artificial Intelligence and Law», 10, 227 ff.

ROCHET J.CH., TIROLE J. 2003. *Platform Competition in Two-Sided Markets*, in «Journal of the European Economic Association», 1, 990 ff.

SALMERÓN-MANZANO E. 2021. *Legaltech and Lawtech: Global Perspectives, Challenges, and Opportunities*, in «Laws», 10, 2, 24. Available on: https://www.mdpi.com/2075-471X/10/2/24.

SANDEFUR R.K. 2019. *Access to What?,* in «Dædalus, the Journal of the American Academy of Arts & Sciences», 148, 1, 49 ff.

SCHORLEMMER M., CONFALONIERI R., PLAZA E. 2016. *The Yoneda Path to Buddhist Monk Blend.* Available on: http://ceur-ws.org/Vol-1660/caos-paper4.pdf.

SOLAR CAYÓN J. I. 2019. *La inteligencia artificial jurídica. El impacto de la innovación tecnológica en la práctica del Derecho y el mercado de los servicios jurídicos*, Thomson Reuter, Aranzadi.

SHULAYEVA O., SIDDHARTHAN A., WYNER A. 2017. *Recognizing Cited Facts and Principles in Legal Judgements*, in «Artificial Intelligence and Law», 25, 107 ff.

SOLANO GADEA M. 2019. *Chatbots*, in BARRIO ANDRÉS M. (ed.). *Legal Tech. La transformación digital de la abogacía*, Wolters Kluwer, La Ley, 153 ff.

SUSSKIND R., SUSSKIND D. 2016. *El futuro de las profesiones. Cómo la tecnología transformará el trabajo de los expertos humanos*, Teell.

SUSSKIND R. 2017. *Tomorrow's Lawyers: An Introduction to Your Future* (2nd Ed.), Oxford University Press.

SUSSKIND R. 2019. *Online Courts and the Future of Justice* (1st Ed.), Oxford University Press.

TIMMERMANN D. 2020. *Legal Tech-Anwendungen*, Nomos Verlag.

WAGNER J. 2018. *Legal Tech und Legal Robots. Der Wandel im Rechtsmarkt durch neue Technologien und künstliche Intelligenz*, Springer.

contributors

Nikita Agarwal
Emory University

Josep Aguiló-Regla
University of Alicante

Manuel Atienza
University of Alicante

Anna M. Borghi
Sapienza University of Rome; Institute of Cognitive Sciences and Technologies, Italian National Research Council (CNR)

Marco Brigaglia
University of Palermo

Bartosz Brożek
Jagiellonian University in Krakow

Monica Bucciarelli
Turin University

Rafael Buzón
University of Alicante

Piotr Bystranowski
Jagiellonian University in Krakow

Bruno Celano
University of Palermo

Kristina Čufar
University of Ljubljana

Sofia de Jong
Leiden University

Francesco Ferraro
University of Milan

Daniel González Lagier
University of Alicante

Noam Gur
Queen Mary University of London

Miha Hafner
University of Ljubljana

Jaap Hage
Maastricht University

Marek Jakubiec
Jagiellonian University in Krakow

Bartosz Janik
University of Silesia in Katowice

Bartłomiej Kucharzyk
Jagiellonian University in Krakow

Łukasz Kurek
Jagiellonian University in Krakow

P.N. Johnson-Laird
Princeton University

Luisa Lugli
University of Bologna

Enide Maegherman
Maastricht University

Susana Navas
Autonomous University of Barcelona

Przemysław Pałka
Jagiellonian University in Krakow

Francesca Poggi
University of Milan

Maciej Próchnicki
Jagiellonian University in Krakow

Philippe Rochat
Emory University

Corrado Roversi
University of Bologna

Adele Quigley-McBride
Simon Fraser University

Barbara A. Spellman
University of Virginia School of Law

Niek Strohmaier
Leiden University

Kevin Tobia
Georgetown University

Michele Ubertone
Maastricht University

Caterina Villani
University of Bologna

Antonia Waltermann
Maastricht University

composed by rospeinfrantumi
published by Diritto & Questioni pubbliche
August 2023