Symposium on Assessing Agency, Moral and Otherwise: Beyond the Machine Question

In conjunction with the 2018 Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB 2018)

4th April 2018

BOREALIS AI
RBC Institute for Research

AISB Convention 2018
Detailed Programme

Wednesday 4th April

	Seminar Room 3	Seminar Room 4	Seminar Room 5
9:30 – 10:30			PLENARY KEYNOTE **Kerstin Dautenhahn** *Will the future be bright for Social Robots?*
10:30 – 11:00	*Coffee Break (Break-out area)*		
	AI, Robots and Public Engagement	Assessing Agency, Moral and Otherwise: Beyond the Machine Question	The 5th Computational Creativity Symposium
11:00 – 12:30	11:00 - Aurora Voiculescu 11:30 - Owen McAree	11:00 Invited Speaker: **Simon McGregor**: *Constructing Agency - The Blob and Ball Thought Experiment* 12:00 Bryony Pierce: *Are the notions of agency and responsibility relevant to questions about machine ethics?*	11:00 Invited Speaker: **Pablo Gervás**: *Automated Targeted Writing: Tailoring and Crafting Literary Pieces with a Given Message* 12:00 Pablo Delatorre: *Outcome Inference based on Threat Resources in Suspenseful Scenes*
12:30 – 13:30	*Lunch (Break-out area)*		

AISB Convention 2018
Detailed Programme

	Seminar Room 3	Seminar Room 4	Seminar Room 5
	AI, Robots and Public Engagement	Assessing Agency, Moral and Otherwise: Beyond the Machine Question	The 5th Computational Creativity Symposium
13:30 – 15:00	13:30 - Rebecca Davnall 14.00 - Joe Rolf	13:30 Antonio Chella: *Rilkean memories for a robot* 14:30 John Preston: *Machine agency, moral relevance, and moral agency*	13:30 Maximilian Droog-Hayes - *Automatic Detection of Narrative Structure for High-Level Story Representation* 14:00 Juan Alvarado - *Understanding MEXICA: an analysis of an Engagement-Reflection system* 14:30 Pablo Gervás - *Storyfying Observed Events: Could I Dress This Up as a Story?*
15:00 – 15:30	*Coffee Break (Break-out area)*		
15:30 – 17:00		15:30 Irina Zudina: *The ethics of developing AI: Why advanced AI has moral status* 16:00 Aurora Voiculescu: *Runaway concepts for robotics and AI: Law, technology and the posthuman* 16:30 Marta Ziosi: *The three worlds of AGI: Popper's theory of the three worlds applied to artificial general intelligence*	15:30 Philipp Wicke - *Storytelling by a Show of Hands: A framework for interactive embodied storytelling in robotic agents* 16:00 Eugenio Concepcion - *Afanasyev: A collaborative architectural model for automatic story generation* 16:30 Anna Jordanous - *Creativity vs quality: why the distinction matters when evaluating computational creativity systems*

AISB Convention 2018
Detailed Programme

Thursday 5th April

	Seminar Room 3	Seminar Room 4	Seminar Room 5
9:30 – 10:30			PLENARY KEYNOTE **Alan Winfield** *The Ethical Roboticist*
10:30 11:00	*Coffee Break (Break-out area)*		
	Digital interventions for cybersecurity	AI, Games and Virtual Reality	Emotion Modelling and Detection in Social Media and Online Interaction
11:00 12:30	11:15: Kholoud Althobaiti, Kami Vaniea and Serena Zheng: *Faheem: Explaining URLs to people using a Slack bot* 11:45: Keshav Kapoor, Karen Renaud and Jacqueline Archibald: *Preparing for GDPR: Helping EU SMEs to Manage Data Breaches* 12:15: Rosemary J. Thomas, Matthew Collinson and Judith Masthoff: *Caught by Phishing Emails? How can Argumentation Schemes be Used to Protect Users?*	11:00 Swen Gaudl - *POSH-SHARP: A Lightweight ToolKit for Creating Cognitive Agents* 11:30 Alexader McDiarmid & David C. Moffat - *Evolving Game Bots for Personality & Balance* 12:00 Discussion	11:00 George Langroudi, Anna Jordanous and Ling Li *Music Emotion Capture: sonifying emotions in EEG data* 11:30 Robert Harvey, Andrew Muncey and Neil Vaughan *Associating Colours with Emotion Detection in Social Media Tweets* 12:00 Francesca D'Errico and Isabella Poggi: *The lexicon of feeling offended*
12:30 13:30	*Lunch (Break-out area)*		

AISB Convention 2018
Detailed Programme

	Seminar Room 3	Seminar Room 4	Seminar Room 5
	Digital interventions for cybersecurity	AI, Games and Virtual Reality	Emotion Modelling and Detection in Social Media and Online Interaction
13:30 15:00	13:30 **Invited Speaker**: Lynne Coventry, Director of PaCT Lab, Northumbria University 14:15: Lynsay Shepherd and Karen Renaud *How to Design Browser Security and Privacy Alerts* 14:45: Tiffany Skinner, Jacqui Taylor, John Dale and John McAlaney: *The Development of Intervention E-Learning Materials and Implementation Techniques for Cyber-Security Behaviour Change*	13:30 Edward J. Powely - *Towards Immersive 3D Visualisation of Game AI Algorithm* Symposium Final Discussion & Future	13:30 **Invited Speaker**: Diana Maynard: *Analysis of hate speech in the GATE social media toolkit* 14:40 Lubna Alharbi, Floriana Grasso and Phil Jimmieson: *An experiment with an off-the-shelf tool to identify emotions in students' self-reported accounts*
15:00 15:30	*Coffee Break (Break-out area)*		
15:30 17:00	15:30: John Paul Vargheese, Matthew Collinson and Judith Masthoff *How can persuasion reduce user cyber security vulnerabilities?* **15:45 - 17:00: Activity: Learning Cafe**	LEGO ROVERS (family workshop – by ticket only)	15:30 Francesca D'Errico, Marinella Paciello and Matteo Amadei: *Prosocial words in social media discussions on hosting immigrants. Insights for psychological and computational field* 15:50 Invited Speaker: **Pietro Cipresso** - *The role of a virtual body in modeling emotions for Social Media and Online Interaction: The BodyPass project*

AISB Convention 2018
Detailed Programme

Friday 6th April

	Seminar Room 1	Seminar Room 3	Seminar Room 4	Seminar Room 5
9:30 10:30				PLENARY KEYNOTE **Ruth Aylett** *Giving Embodied Social Agents a Culture*
10:30 11:00	*Coffee Break (Break-out area)*			
	Philosophy after AI: Mind, Language and Action	Social Interactions in Complex Intelligent Systems (SICIS)	Cybernetic Serendipity Reimagined	18th workshop on Computational Models of Natural Argument (CMNA)
11:00 12:30	11:00 Steve Battle, *Principles of robot autonomy* 11:30 Giusy Gallo, Claudia Stancati, *Acting robots or ethical machines?* 12:00 Christopher Burr, Geoff Keeling, *Building machines that learn and think about morality*	11:00 I. Khan et al., *Adaptation and the Social Salience Hypothesis of Oxytocin: Early Experiments in a Simulated Agent Environment* 11:30 A. Wolfenden and N. Vaughan, *How effective is Ant Colony Optimization at Robot Path Planning* **12:00 Demos**	11:00 Introduction 11:15 Invited Speaker **Alison Pease**: *Modelling Serendipity in a Computational Context* 12:05 Katie Mccallum, Majed Al-Jefri and Kate Monson: *Making Sense of the Incomprehensible: Naivety as a Tool for Escaping Anthropocentric Narratives*	11:00 Introduction 11:10 Invited Speaker: **Federico Cerutti**, *On Scientific Enquiry and Computational Argumentation*

AISB Convention 2018
Detailed Programme

	Seminar Room 1	Seminar Room 3	Seminar Room 4	Seminar Room 5
12:30 13:30	*Lunch (Break-out area)*			
	Philosophy after AI: Mind, Language and Action	Social Interactions in Complex Intelligent Systems (SICIS)	Cybernetic Serendipity Reimagined	18th Workshop on Computational Models of Natural Argument (CMNA)
13:30 15:00	13:30 Dean Petters, Achim Jung, *From the Chinese Room argument to the Church-Turing thesis* 14:00 Davide Serpico, Marcello Frixione, *Can the g Factor Play a Role in AGI Research?* 14:30 David Mathers, *The Possibility of Indeterminate Cases of Consciousness and the Ethics of AI*	13:30 M. Mulvenna et al., *HILDA - A Health Interaction Log Data Analysis Workflow to Aid Understanding of Usage Patterns and Behaviours* 14:00 S. Monica and F. Bergenti, *Kinetic Models of Relevant Phenomena Related to Opinion Dynamics* 14:30 N. Vaughan, *Morphogenetic Engineering For Evolving Ant Colony Pheromone Communication*	13:30 Aleksandar Zivanovic and Edward Ihnatowicz: *Sound Activated Mobile (SAM) at Cybernetic Serendipity* 13:55 Roisin Loughran and Michael O'Neill: *Serendipity in Melodic Self-organising Fitness* 14:20 Joseph Corneli, Dave Murray-Rust and Benjamin Bach: *Towards Open-World Scenarios: Teaching the Social Side of Data Science* 14:40 Ian Helliwell: *Symbiosis - Electronic Music for an Analogue and a Digital Age*	13:30 Invited Speaker: **Rudi Palmieri**, *Argumentation patterns and strategies in financial communication genres* 14:30 Cliff O'Reilly, Yetian Wang, Katherine Tu, Sarah Bott, Paulo Pacheco, Tyler Black and Randy Harris - *Arguments in Gradatio, Incrementum and Climax; a Climax Ontology*

AISB Convention 2018
Detailed Programme

	Seminar Room 1	Seminar Room 3	Seminar Room 4	Seminar Room 5
15:00 15:30	*Coffee Break (Break-out area)*			
	Philosophy after AI: Mind, Language and Action	Social Interactions in Complex Intelligent Systems (SICIS)	Cybernetic Serendipity Reimagined	18th Workshop on Computational Models of Natural Argument (CMNA)
15:30 17:00	15:30 Jack R. Coopey, *Machinics: Philosophy of Computer Science and its relation to Programming, Towards a Contemporary Ontology of Programming as a new Form-of-Life and Technics* 15:50 Discussion panel		**Cybertalks** 15:30 Ricardo Melo and Miguel Carvalhais: *The Chance of Serendipity* 15:42 Dietmar Köring: *The Technobody - An animatronic artefact as manifestation of 2nd Order Cybernetics* 15:54 Paul Pangaro: *Colloquy of Mobiles 2018 Project* 16:06 Liss C. Werner: *Gordon Pask's 'Cybernetic Theatre': beyond tinkering with Architecture* 16:18 Discussion panel	15:30 Jean-Baptiste Corrégé, Emmanuel Hadoux and Ariel Rosenfeld - *From Psychological Persuasion To Abstract Argumentation: A Step Forward* 15:50 Sultan Alahmari, Tommy Yuan and Daniel Kudenko - *Policy generalisation in reinforcement learning for abstract argumentation* 16:10 Rosemary J. Thomas, Nir Oren and Judith Masthoff *A System for Automation of Message Generation using Argumentation Schemes* 16:30 Mathilde Janier and Patrick Saint-Dizier -

Rilkean Memories for a Robot

Antonio Chella

Abstract. The paper discusses the role of Rilkean memories, recently introduced by Rowlands, in the building of the autobiographic self of a robot.

1 INTRODUCTION

It has been debated about the characteristics for an agent to be considered morally responsible for her actions. A generally recognized characteristic for moral agency is the capability for the agent to have a sense of self. According to this line of thinking, it has been debated whether a robot could ever be a morally responsible agent (see Gunkel [1], 46, for a discussion).
With the term "robot" we consider a humanoid robot, i.e., a mechanical entity with a human-like body shape, equipped with sensors like cameras, lasers, sonars, and with actuators like arms and legs, all controlled by a complex software system.
There is a long progression from a situation-action robot to a robot with a sense of self. In facts, the self of the robot is not something that is "uploaded," but, as in the humans, it develops in years after many interactions among the body of the robot, its control system, the users, the external environment, other robots and so on.
The current studies on the robot self essentially take into account the role of some model of episodic memory implemented in software by employing internal model methods or machine learning methods. However, the main role of the robot body in the building of the robot self is largely unexplored.
The paper claims that the development of the self of the robot, which is a complex issue involving several aspects, cannot ignore the aspects related with the body of the robot operating in the real world.

2 RILKEAN MEMORIES

Rilkean memories are a kind of autobiographical memories recently discussed by Rowlands [2,3], who took inspiration from the novel "The Notebooks of Malte Laurids Brigge" by Rainer Maria Rilke. Rilkean memories are:

> these memories that have become "blood," "glance and gesture," "nameless and no longer to be distinguished from ourselves" ([3], 54).

A Rilkean memory of an episode of the life of a person is related to the trace of that event left in the whole body of the individual, and not only in her brain.

[1] University of Palermo, Italy, email: antonio.chella@unipa.it

Rowlands discusses in detail how "this form of memory is typically, embodied and embedded; "it is a form of involuntary, autobiographical memory that is neither implicit nor explicit, neither declarative nor procedural, neither episodic nor semantic, and not Freudian." ([2], 141).
Rowlands points out that Rilkean memories of a person are responsible for the *style* of that person. An example is the motion style of the person: a person habit may be to walk on the left side of a path because of a traumatic episode during her life. The person may not be able to remember the traumatic episode explicitly, but the event entered into her blood, i.e., it becomes a part of her style.
Then, a Rilkean memory derives from episodic memory as a transformation of the act of remembering an episode in a behavioral and bodily disposition, even when the content of the episode is lost ([3], 73).
The whole self of a person is made up by different kind of memories, but, according to Rowlands, there is a strong relationship between the self of a person and the Rilkean memories. In facts, different people, with different histories and different episodes occurring in their life, acquire a distinct and recognizable personal style that survives even when the memories of the episodes occurring in the life of the person generating the memories, are lost. Then, the unity and persistence conditions that identify a person among the others is her style: "Rilkean memories … play a crucial role in holding the self together, in the face of certain well-documented facts." ([2], 154).

3 ROBOT RILKEAN MEMORIES

We claim that Rilkean memories may be considered even for robots, thus allowing it to build a robot self and to perform operations according to the acquired personal style.
In facts, the operation style of a robot is slowly acquired by the robot itself. The style does not depend uniquely on the software structures of its control system, as the outcomes of suitable neural networks storing the past interactions of the robot, but instead on the complex intermixing between the body of the robot and the software controlling the robot.
Consider, the NAO, a small humanoid robot built by SoftBank Robotics (previously Aldebaran) and commonly adopted in many research labs.
The RoboticsLab of the University of Palermo acquired two brand new NAOs in 2010. They were at the beginning mostly two instances of the same robot, and they were both capable of performing the same tasks in the very same ways. They had the same limitations due to their body constraints and due to their software characteristics.
The two brand new robots were employed for separate research projects with distinct purposes. As years went by, because of

their employment in different projects, the two NAO slightly changed their morphology in different ways. After performing and repeating several different actions, their motors started not to work correctly in idiosyncratic ways; their joints motions were not smooth enough and many bumps occurred in various parts of their bodies.

To clarify this point, let us consider the event of uploading a software program on one NAO that, for some reasons, caused the NAO arm to bump against the wall and thus generating a permanent malfunction of a robot joint. Then, a Rilkean memory of this episode occurs: even when the software program that allowed the robot to bump against the wall has been modified or cancelled, the bumping episode has been transformed in the malfunction of the joint. Thus, even if the robot is unable to access the episode because its log file has been cancelled, the episode has been transformed in a Rilkean memory of the robot.

It also happened that, because of several traumatic events, a part of one robot broke and it has been substituted and renewed. The substitution of a NAO body part is a Rilkean memory of the sequence of events that allowed the previous body part to be broken.

Then, as the two robots have been employed in different experiments and for different goals, after so many years they now have slightly different bodies.

Let us consider for example the case of one of the two robots that shows a problem with the motor controlling the left arm, i.e., a Rilkean memory of a previous bumping event. The control system of this NAO must take into account this motion constraint to generate a suitable plan of operations that minimize the use of the impaired arm. The other robot does not present any problem with the left arm, but instead with the right leg, because of a different traumatic event. Similarly, the control system of the other robot generates a motion plan that minimizes the movements of the damaged limb.

Other traumatic events occurred during the operational life of a robot are transformed in the excessive warm-up of motors, the critical duration of batteries, the limitations in the motions of joints, the malfunctions of one camera, and so on.

Therefore, in all of these cases, the control systems of the robots have to take care of these incurring different Rilkean memories. For example, one of the two NAOs has a tendency of an excessive warm-up of a motor because of a previous episode of an excessive employment of this motor. This NAO must interrupt its actions quite often, and thus this robot performs its task in a characteristic rough and fragmented way. Another example is related with the cameras of the robot: the NAO has two cameras mounted on top and on the bottom of the head. One of the two NAOs has one camera broken because of a crash event, and then this robot has to move the head in a peculiar way to compensate the malfunction, which is a Rilkean memory of the head crash.

It should be noticed that the problems occurring with a real robot operating in real environment are difficult to simulate by employing a robot simulator. Therefore, the internal simulator method, which is typically adopted in the design of the software system control of a robot, is not able to deal with the Rilkean memories.

After years of operations, then two different performance styles emerge, that tightly depend on the intricate intermixing between the hardware, the software and the biographies of the two NAO.

When the two robots are reunited together and involved in the same task, the controller of each robot must take into account the different software capabilities and hardware constraints to generate two different plans for the two robots, for the very same task. Then, the two robots perform the same task in two different ways, according to their style acquired during their operational lives.

Therefore, the two robots earned different memories of past episodes of their lives and different functional styles. The bodies of the two robots summarize the autobiographic selves of the robots and their acquired different personalities.

4 CONCLUSIONS

Robot Rilkean memories are the traces of the occurring episodes left in the whole body of the robot even when the memories of the episodes is lost. In the current literature, the discussions about the self of a robot typically consider the software control system of the robot, while the role of the body has been of limited interest. Therefore, we maintain that a robot, after years of operations, acquires its functioning style, which is unique and different from the style of the other robots, even if from the same factory.

It should be noticed that the employment of robot simulators, which is a typical strategy in robot software design, does not take into account all the real-world issues related with Rilkean memories, as the degradation of the operations of the joints, of the camera, the motor because of occurring episodes. Then, the body of the robot is one of the main aspects of the autobiographic self of the robot.

ACKNOWLEDGEMENTS

The author wants to thank Mark Rowlands and Riccardo Manzotti for their comments on previous versions of this paper.

REFERENCES

[1] D.J. Gunkel. (2012). *The Machine Question*. MIT Press, Cambridge, MA.

[2] M. Rowlands. (2015). Rilkean Memory, in: *The Southern Journal of Philosophy*, **53**, 141–154.

[3] M. Rowlands. (2016). *Memory and the Self: Phenomenology, Science and Autobiography*. Oxford University Press, Oxford, UK.