



UNIVERSITÀ DEGLI STUDI DI PALERMO

DOTTORATO DI RICERCA IN  
DIRITTI UMANI: EVOLUZIONE, TUTELA E LIMITI  
CICLO XXXVI

Settore Scientifico Disciplinare: IUS/20

Macchine che decidono: prospettive di automazione dei processi decisionali  
in contesti democratici

Presentata da Paolo Capriati

Coordinatore dottorato  
**Prof. Aldo Schiavello**

Supervisore  
**Prof. Marco Brigaglia**

# **INDICE**

## **INTRODUZIONE**

### **CAPITOLO 1 – Cybersyn: forme di automazione dei processi decisionali ante litteram**

#### **1.1 Un progetto mai realizzato**

#### **1.2 La cibernetica e Cybersyn**

##### **1.2.1 La prima cibernetica**

##### **1.2.2 Il progetto cibernetico cileno**

##### **1.2.3 La seconda cibernetica**

#### **1.3 La questione teleologica**

##### **1.3.1 Sulla natura (non) teleologica dei sistemi politici**

##### **1.3.2 Questioni di legittimità nel progetto cibernetico cileno**

#### **1.4 Sul ruolo euristico di Cybersyn**

### **CAPITOLO 2 – Macchine per decidere**

#### **2.1 IA e democrazia: stato dell'arte**

## **2.2 Di che macchine si tratta**

### **2.2.1 Il criterio dell'autonomia**

### **2.2.2 Il criterio dell'intelligenza**

## **2.3 Concetti “ponte”. Per un'analisi congiunta degli apporti umani e artificiali**

### **2.3.1 Intelligenza ibrida**

### **2.3.2 Sistemi socio-tecnici**

### **2.3.3 Macchine politiche**

## **2.4 Per un'analisi disgiunta**

### **2.4.1 Il *machine learning***

### **2.4.2 Sull'utilità e l'opportunità di un'analisi disgiunta**

# **CAPITOLO 3 – Processi decisionali democratici**

## **3.1 Processi decisionali e teorie della democrazia**

## **3.2 Verso una definizione dei processi decisionali democratici: quattro domande**

### **3.2.1 Perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?**

#### **3.2.1.1 Giustificazioni strumentaliste**

#### **3.2.1.2 Giustificazioni non-strumentaliste**

#### **3.2.1.3 La natura della giustificazione pubblica**

### **3.2.2 La democrazia si fonda sull'idea che esista una posizione etica corretta?**

**3.2.2.1 Classificare le concezioni metaetiche**

**3.2.2.2 Oggettivismo vs soggettivismo**

**3.2.3 A quali risultati deve condurre una decisione presa in maniera democratica?**

**3.2.3.1 Sostanzialismo vs proceduralismo**

**3.2.3.2 Il tentativo di conciliare procedura e sostanza**

**3.2.4 Come si può prendere una decisione con metodo democratico?**

**3.2.4.1 Voto, negoziazione, argomentazione**

**3.2.4.2 L'argomentazione e la decisione corretta**

**3.3 Due concezioni della democrazia**

**3.3.1 Concezioni non-epistemiche**

**3.3.2 Concezioni epistemiche**

**3.3.2.1 I sistemi deliberativi come espressione delle concezioni epistemiche**

## **CAPITOLO 4 – Il ruolo delle macchine nei processi decisionali democratici**

**4.1 Processi decisionali e macchine**

**4.2 I due ruoli delle macchine**

**4.2.1 La macchina come attore in un processo decisionale**

**4.2.1.1 La questione della personalità giuridica**

**4.2.1.2 Una macchina può essere titolare di diritti?**

**4.2.1.3 La macchina come agente morale**

#### **4.2.2 La macchina come elaboratore e trasformatore di preferenze**

**4.2.2.1 La macchina come parte del processo decisionale o come processo decisionale?**

##### **4.2.2.2 Sull'autonomia della macchina-programma**

**4.3 La compatibilità della partecipazione delle macchine con le diverse concezioni della democrazia**

#### **4.3.1 Compatibilità con le concezioni non-epistemiche**

**4.3.1.1 La macchina-attore e le concezioni non-epistemiche**

**4.3.1.2 La macchina-programma e le concezioni non-epistemiche**

#### **4.3.2 Compatibilità con le concezioni epistemiche**

**4.3.2.1 La macchina-attore e le concezioni epistemiche**

**4.3.2.2 La macchina-programma e le concezioni epistemiche**

#### **4.4 La macchina che delibera**

## **CONCLUSIONI**

*(...) è con animo sereno e senza rimpianti che constato come il mio posto potrà essere occupato benissimo da un congegno meccanico. (...) Vediamo qual è la mia reazione psicologica apprendendo che lo scrivere è solo un processo combinatorio tra elementi dati: ebbene, ciò che io provo istintivamente è un senso di sollievo, di sicurezza. Lo stesso sollievo e senso di sicurezza che provo ogni volta che un'estensione dai contorni indeterminati e sfumati mi si rivela invece come una forma geometrica precisa, ogni volta che in una valanga informe di avvenimenti riesco a distinguere delle serie di fatti, delle scelte tra un numero finito di possibilità. Di fronte alla vertigine dell'innumerabile, dell'inclassificabile, del continuo, mi sento rassicurato dal finito, dal sistematizzato, dal discreto.*

Calvino, Cibernetica e fantasmi, in Una pietra sopra

## INTRODUZIONE

Se esistesse una macchina capace di decidere meglio di un'assemblea parlamentare, lascereste decidere a questa macchina?

È una domanda così vaga che non si sa bene da dove cominciare. In primo luogo, che significa “decidere meglio”? Esistono davvero decisioni migliori di altre? Non è, forse, tutto relativo? Una decisione potrebbe essere migliore per un certo gruppo di individui, ma peggiore per un altro, ad esempio. E che dire dell'assemblea parlamentare? Da chi è composta questa assemblea? Si tratta di un'assemblea di corrotti e inetti o si tratta di un gruppo di persone integre, integerrime e competenti?

Un convinto democratico potrebbe tagliare corto così: «se si lasciasse decidere a una macchina – e non a un'assemblea parlamentare –, vi sarebbe una violazione del principio democratico». Ma è davvero così? Davvero delegare potere decisionale a una macchina arrecherebbe pregiudizio al principio democratico? Non vi è, forse, anche nei sistemi democratici una delega del potere? Non si eleggono i rappresentanti perché prendano decisioni al posto nostro? Perché nel caso di una macchina sarebbe diverso?

È scontato dire che quando il concetto di democrazia prese forma non esistevano – né si immaginava potessero esistere – macchine con la potenza computazionale di oggi. Quello di democrazia è un concetto nato in antitesi ad altre forme di governo, come la monarchia, in cui a governare è una persona sola, e l'aristocrazia, in cui il governo è gestito da pochi. Si può sostenere senza indugio che un processo è tanto più democratico quanto più vi è la possibilità da parte di coloro che sono interessati ad una certa decisione di influenzare tale decisione. Come si pone il governo di una macchina in questo quadro? Esso è più vicino a una monarchia o a una democrazia?

La macchina, nella sua rappresentazione classica, si presenta come un agente *super partes*. Tale rappresentazione può essere esplorata alla luce dell'impiego metaforico del concetto di macchina. Faccio un passo indietro: perché sto parlando di

metafore e cos'hanno da dirci le metafore sul modo in cui ci rappresentiamo le macchine?

Le metafore non sono solo figure retoriche legate al mondo dell'immaginazione poetica, ma hanno la funzione di veri e propri strumenti cognitivi che permettono di categorizzare concetti ed esperienze e dare loro un senso (Lakoff e Johnson, 2008).

In quali casi viene, dunque, utilizzata la metafora della macchina?

(a) per riferirsi a qualsiasi sistema biologico. "Il corpo è una macchina".<sup>1</sup>

(a1) per riferirsi ad una persona. In questo senso, si possono distinguere due usi:

(a1a) per riferirsi ad una persona efficiente, infallibile. "Sei una macchina, non sbagli un colpo".

(a1b) per riferirsi ad una persona fredda, insensibile. "Sei privo di sentimenti, sei una macchina".

Il mito della macchina *super partes* origina da (a1) nelle sue due declinazioni: perché un soggetto sia *super partes* non deve commettere errori di calcolo (a1a) e la sua azione non deve essere governata da pulsioni "emotive" (a1b). Dall'altro lato vi è la persona "umana", ineluttabilmente destinata a sbagliare – "*errare humanum est*" – e ad essere governata da pulsioni di ogni sorta. Si instaura così una polarità: da un lato le macchine, agenti infallibili e lucidi e, dall'altro lato, gli umani.

Sappiamo che le cose non stanno esattamente così: ogni tecnologia ha dietro di sé gli interessi che hanno portato alla sua costruzione e assieme agli interessi vi sono esseri umani fatti di carne e di ossa. Proviamo, però, a dare per buona la storia della macchina terza e imparziale, che ha tagliato il cordone ombelicale con l'umano. Una macchina del genere non è la macchina di nessuno in particolare e, dunque, il suo è il

---

<sup>1</sup> Tale metafora è uno dei fondamenti della cibernetica. Di per sé, la cibernetica è già una metafora, anche se non da un punto di vista letterale. Se la metafora è la spiegazione di un concetto attraverso un altro concetto (Lakoff e Johnson, 2008), la cibernetica, creando un'analogia tra animali e macchine, dà vita a una doppia metafora, spiegando il mondo degli animali con quello delle macchine e viceversa. Della cibernetica e dell'analogia animale-macchina parlerò nel primo capitolo.



governo di nessuno. E il governo di nessuno è più vicino ad una monarchia o ad una democrazia?

Per rispondere a questa domanda, bisogna tornare al momento della nascita della democrazia e, poi, alla sua evoluzione. La democrazia, come si è visto, è nata in opposizione al “governo di uno solo” ed ha continuato ad evolversi in questo senso. La democrazia è, dunque, molto più il governo “non di uno solo” che il governo di “tutti”. Si tratta di una concezione difensiva della democrazia che ne rivela la sua funzione prevalentemente antiautoritaria. Per questa ragione, se si intende per democrazia il governo “non di uno solo” (e non il governo “di tutti”), allora si può concludere che il governo “di nessuno” (cioè il governo delle macchine) sia molto più vicino alla democrazia che al governo “di uno solo” – o “di pochi”.

Detto altrimenti, il governo “non di uno solo” – o “non di pochi” – può anche voler dire il governo “di nessuno”. Mentre se si concepisce la democrazia come governo “di tutti”, non si può sostenere che il governo “di nessuno” sia compatibile con la democrazia.

Queste premesse anticipano la traiettoria dell’analisi che sto per presentare: l’evoluzione normativamente timida della democrazia non ha creato un antidoto per l’automazione dei processi decisionali. Alla domanda “considerato il decadimento democratico – ideale e istituzionale – con i suoi processi decisionali lenti, farraginosi e fallimentari, se ci trovassimo davanti alla prospettiva di delegare a una macchina dalle straordinarie capacità di elaborazione, dovremmo farlo?”, la filosofia politica occidentale – per come si è evoluto il recente dibattito sulla democrazia – non riesce a rispondere convintamente e persuasivamente di no. Si tratta quindi di un “sì”?

Prenderò le mosse dal progetto Cybersyn – una proto-forma di automazione dei processi decisionali nel Cile degli anni ‘70. Questo progetto anticipa alcune delle questioni dirimenti in riferimento al rapporto fra processi decisionali su questioni politiche e macchine. La prima di queste riguarda il ruolo che può giocare la tecnologia nelle decisioni di interesse pubblico. Cybersyn fa parte di un progetto più vasto che dal dopoguerra in poi vide coinvolti studiosi di diverse discipline: tale progetto è la cibernetica. Nel passaggio dalla prima alla seconda cibernetica, prende

forma la questione della teleologia: i sistemi biologici – o le macchine – sono sistemi non-teleologici? Questo problema è stato esteso in un secondo momento ai sistemi politici: i sistemi politici sono sistemi non-teleologici? È un problema che anticipa quelli legati alle giustificazioni dei sistemi politici. La partecipazione di una macchina ad un processo decisionale dipende dalla configurazione normativa del processo entro cui si inserisce. Il dibattito, dunque, sulla natura teleologica – o meno – dei sistemi politici è un’embrionale maniera di interrogarsi su ciò che giustifica un sistema politico democratico e sui fini che tale sistema dovrebbe perseguire – ammesso che esso debba perseguire dei fini. Ma prima di valutare il complesso di elementi che definiscono la forma normativa delle democrazie di oggi, c’è da chiarire cosa sia una macchina.

A questo fine è dedicato il secondo capitolo. La definizione di macchina che offrirò è necessariamente operativa e strumentale – strumentale alla costruzione di una macchina che possa prendere decisioni e non semplicemente eseguire operazioni di calcolo eterodirette. In questo senso va la necessità di fissare un criterio per determinare quando una macchina è una semplice esecutrice e quando, invece, agisce in base a leggi proprie. Il criterio individuato è quello dell’autonomia – e non anche quello dell’intelligenza. Ma quando una macchina è autonoma? L’autonomia si definisce nei termini di “relativa indipendenza da qualcos’altro”. Nel caso della macchina di questa ricerca, tale indipendenza è relativa al fattore umano. È, dunque, necessario fissare chiaramente ruoli e funzioni della macchina e dell’umano e non cedere alla tentazione di adottare un approccio congiunto che vede al rapporto fra macchine ed umani come ad un ecosistema di relazioni dalla trama inestricabile, nelle quali gli apporti – degli umani o delle macchine – sono indistinguibili. Il *machine learning* è ciò che permette di fondare questa distinzione: esso serve a comprendere dove finisce l’apporto umano e dove comincia quello della macchina. Questa macchina, tuttavia, pur essendo autonoma, non agisce in maniera isolata ma è “sistema” di un ambiente più vasto. Di che ambiente sto parlando?

È nel terzo capitolo che illustrerò tale “ambiente”. Sto alludendo al contesto democratico di riferimento. Tale contesto non va inteso come l’apparato istituzionale

che porta alla formazione delle leggi, ma come l'insieme degli elementi che formano l'immagine normativa della democrazia, intesa come metodo per prendere decisioni. Tale metodo deve essere analizzato alla luce di quattro domande:

- 1) Perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?
- 2) La democrazia si fonda sull'idea che esista una posizione etica corretta?
- 3) A quali risultati deve condurre una decisione presa in maniera democratica?
- 4) Come si può prendere una decisione con metodo democratico?

Sulla base delle risposte date a ciascuna di queste domande si delineano due distinte concezioni della democrazia – una concezione epistemica e una concezione non-epistemica. Le concezioni non-epistemiche sono intese in termini residuali: sono non-epistemiche tutte quelle concezioni che non sono epistemiche. In altre parole, a differenza delle concezioni epistemiche che prevedono una risposta ben precisa a ciascuna delle quattro domande presentate poc'anzi, quelle non-epistemiche sono tutte le concezioni che divergono in parte o totalmente dalla specifica risposta prevista per le suddette domande. Fra le concezioni epistemiche, spicca poi quella che vede le democrazie come sistemi deliberativi. Ma per quale motivo ho deciso di isolare due concezioni della democrazia?

Nell'ultimo capitolo, partirò da un'ipotesi: una concezione della democrazia – quella epistemica – è compatibile con la partecipazione delle macchine ai processi decisionali, mentre l'altra famiglia di concezioni – le concezioni non-epistemiche – pone degli ostacoli maggiori alla partecipazione. Ma che forme può assumere tale partecipazione? Propongo due possibilità di partecipazione: la macchina-attore – che partecipa ai processi decisionali così come vi parteciperebbe un qualsiasi altro individuo umano – e la macchina-programma – che raccoglie, elabora e trasforma le preferenze individuali. In secondo luogo, questa presunta compatibilità fra partecipazione delle macchine e concezioni epistemiche riguarda tanto la macchina-programma quanto quella attore o solo una di queste? Infine, un approccio sistemico alla deliberazione democratica foraggia, in qualche modo, la partecipazione delle macchine?

# **CAPITOLO 1 – Cybersyn: forme di automazione dei processi decisionali *ante litteram***

## **1.1 Un progetto mai realizzato**

## **1.2 La cibernetica e Cybersyn**

### **1.2.1 La prima cibernetica**

### **1.2.2 Il progetto cibernetico cileno**

### **1.2.3 La seconda cibernetica**

## **1.3 La questione teleologica**

### **1.3.1 Sulla natura (non) teleologica dei sistemi politici**

### **1.3.2 Questioni di legittimità nel progetto cibernetico cileno**

## **1.4 Sul ruolo euristico di Cybersyn**

# CAPITOLO 1

## 1.1 Un progetto mai realizzato

Che ruolo può giocare la tecnologia per le decisioni di interesse pubblico che vengono prese in maniera collettiva?

Può rivelarsi utile partire dal passato: che ruolo ha giocato la tecnologia nelle decisioni di interesse pubblico nella storia? Il progetto che illustrerò non trovò mai completa attuazione, eppure, esso anticipa alcuni elementi che verranno ripresi nel resto della trattazione.

Prima di presentare questo progetto, è opportuno precisare meglio i termini della questione: tanto “tecnologia” quanto “decisioni di interesse pubblico prese in maniera collettiva” sono espressioni che appaiono eccessivamente indeterminate.

In primo luogo, cosa si intende per “tecnologia”? Per tecnologia ci si può riferire a una varietà così estesa di fenomeni che bisogna delimitare il campo a quel tipo di tecnologie che interessano questa ricerca. L'altro termine della questione potrebbe essere utile nell'operazione di delimitazione: che tecnologie possono rilevare quando si parla di decisioni di interesse pubblico prese in maniera collettiva? Si tratta, sembrerebbe, di strumenti di supporto alla decisionalità o di strumenti decisionali veri e propri. In quale maniera una tecnologia può supportare o sostituire le decisioni umane?

A questo punto, è utile fare un passo indietro. Il progetto cui accennavo poc'anzi – quello che non trovò mai completa attuazione – suggerisce alcuni dei modi in cui si pensava che la tecnologia potesse trovare applicazione per decidere collettivamente su questioni di interesse pubblico. Si presenta, in questo modo, almeno una delle ragioni per cui ho deciso di partire dalla descrizione del progetto in questione: con esso si assiste ad una embrionale sperimentazione di alcune forme di automazione dei processi decisionali su questioni politiche.

Il progetto in questione è Cybersyn.<sup>2</sup> Cybersyn è stato immaginato e disegnato nel Cile di Salvador Allende – precisamente fra il 1971 e il 1973. La ragione per la quale non fu mai implementato – anche se Allende se ne servì, ad esempio, durante una crisi di governo – è da rintracciare nel colpo di stato del 1973, guidato dal generale Pinochet, che depose prematuramente il governo di Allende.

Cybersyn venne progettato da Stafford Beer, le cui idee sono ascrivibili al pensiero cibernetico. Ma che cos'è la cibernetica? Dopo aver chiarito a grandi linee cosa si intende per cibernetica, presenterò nel dettaglio Cybersyn. Il progetto cibernetico cileno illustrerà alcuni dei ruoli che la tecnologia può giocare quando occorre prendere decisioni di interesse pubblico in maniera collettiva. Al di là delle specifiche applicazioni nei processi decisionali, per stabilire il ruolo di una certa tecnologia è necessario definire il contesto di riferimento. In questo senso, verranno esplorate due distinte ipotesi: nella prima ai sistemi politici viene riconosciuta natura teleologica, nella seconda no. Secondo Beer (l'ideatore di Cybersyn), i sistemi politici sono fondamentalmente sistemi non-teleologici che definiscono autonomamente i propri fini. L'assenza di teleologia, per Beer, è fondamentale per iscrivere un certo sistema politico nella tradizione democratica. A differenza di quanto sostiene Beer, accordare natura teleologica a un sistema politico non significa classificarlo come non democratico. Anche se nel resto della trattazione non si affronterà la questione sulla natura di un sistema politico nei termini di “teleologia” o “assenza di teleologia”, questa distinzione anticipa alcuni dei problemi relativi al modo in cui si concepisce la democrazia. È, quindi, chiaro che il ruolo che la tecnologia può assumere in un sistema politico dipende, in primo luogo, dal modo in cui viene concepito il processo democratico. Per concludere, si vedrà come, anticipando alcuni nodi cruciali che nel seguito della trattazione verranno approfonditi, Cybersyn svolga fondamentalmente un ruolo euristico.

---

<sup>2</sup> A tal proposito, occorre un chiarimento terminologico. Si intenderà, talvolta, con “Cybersyn” l'insieme degli sforzi cui Beer e il suo entourage si dedicarono per la costruzione di un progetto cibernetico per il Cile durante il governo Allende.

## 1.2 La cibernetica e Cybersyn

La cornice teorica entro cui si è sviluppato il progetto Cybersyn è la cibernetica.

Norbert Wiener (1968) – considerato il padre della cibernetica – pone la questione della cibernetica in questi termini:

abbiamo deciso di chiamare l'intero campo della teoria del controllo e della comunicazione sia nella macchina che negli animali con il nome di cibernetica, che deriva dal greco κυβερνήτης ovvero timoniere. La scelta di questo termine è dovuta al riconoscimento che abbiamo inteso dare al fatto che il primo significativo scritto sui meccanismi a feedback è un articolo sui regolatori (*governors*) pubblicato da Clerk Maxwell nel 1868, e che *governor* è derivato dalla corruzione latina di κυβερνήτης. Abbiamo anche voluto ricordare che i motori per il governo delle navi sono una delle prime e meglio sviluppate forme di meccanismo a feedback.

Benché questa definizione non chiarisca cos'è la cibernetica, essa introduce alcuni aspetti che occorre esaminare.

Il primo di questi aspetti riguarda l'equiparazione, come nel sottotitolo dell'opera di Wiener – Controllo e comunicazione nell'animale e nella macchina –, dell'animale alla macchina.

Il secondo aspetto – quello del feedback – è ciò che nella definizione di cibernetica di Wiener acquista una rilevanza prioritaria. Ciò di cui si occupa la cibernetica, infatti, sono i meccanismi a feedback.

Quando intendiamo effettuare un movimento secondo un determinato modello, la differenza fra tale modello e l'effettivo svolgersi del movimento viene adoperata come un nuovo segnale che determina una regolazione del movimento stesso tale da mantenerlo quanto più possibile vicino a quello dato dal modello (Wiener, 1968).

Il feedback sarebbe, dunque, l'informazione relativa alla differenza fra il modello e il movimento che mima il modello. Tale meccanismo risulta ben illustrato dall'immagine di un movimento umano.

Supponiamo ora che io prenda una matita. Per fare questo devo muovere certi muscoli. Nessuno di noi, eccetto qualche esperto anatomista, sa tuttavia quali siano questi muscoli; e anche fra gli anatomisti, ve ne sono pochi, se non nessuno, che possano eseguire questa azione mediante una cosciente e intenzionale successione di contrazione dei muscoli implicati. Ciò che noi vogliamo consapevolmente fare è solo prendere la matita. Una volta presa questa decisione, il nostro movimento procede in modo tale che – per così dire – quanto manca alla presa della matita decresca progressivamente. Questa parte dell'azione non si svolge a livello di piena consapevolezza (Wiener, 1968).

Quello che qui rileva, dunque, è lo scarto fra il movimento programmato e quello effettivo. È proprio il feedback che si occupa di colmare e correggere tale scarto.

Conoscere, invece, tutti i muscoli che si attivano durante un'azione riguarda un altro piano – quello della piena consapevolezza. Una conoscenza del genere si rivela impossibile da rappresentare. Tuttavia, il fatto che non sia possibile conoscere tutti i movimenti di cui si compone un'azione non rende impossibile compiere l'azione.

Detto altrimenti, la rappresentazione esatta del movimento non è necessaria alla realizzazione dell'azione. Da questo esempio già traspare una frattura fra rappresentazione e *performance*.<sup>3</sup>

### **1.2.1 La prima cibernetica**

---

<sup>3</sup> Rappresentazione e *performance*, in questo contesto, vengono intesi come strumenti gnoseologici.



In sintesi, (almeno) due sono gli elementi che contraddistinguono il pensiero cibernetico: l'equiparazione dell'animale alla macchina e la delimitazione del campo di studio ad entità che rispondono a meccanismi di feedback.<sup>4</sup>

Il riferimento al primo elemento – l'analogia animale-macchina – implica l'adozione di un metodo di studio in larga parte condiviso: tanto per gli animali quanto per le macchine tale metodo è quello dell'analisi comportamentale. Ma cosa si intende per analisi comportamentale?

Parafrasando “*Behavior, Purpose and Teleology*”, l'analisi comportamentale consiste nell'esame dell'output di un oggetto – dopo averlo astratto dal suo ambiente – e nelle relazioni fra questo output e il suo input. Se con “output” si intende qualsiasi mutazione che l'oggetto produce nell'ambiente, con “input” ci riferiamo a qualsiasi evento esterno all'oggetto che in qualche modo lo modifica. In questa definizione dell'analisi comportamentale è omissa il riferimento alla specifica struttura e all'organizzazione intrinseca dell'oggetto. Questa omissione è fondamentale poiché su di essa si fonda la distinzione fra un approccio comportamentale e altri metodi di studio, definiti funzionalisti. Più precisamente, nell'analisi funzionalista, che viene contrapposta a quella comportamentale, l'obiettivo principale è l'intrinseca organizzazione delle entità studiate, la loro struttura e le loro proprietà. In questo tipo di analisi, quindi, le relazioni fra l'oggetto e il suo ambiente passano in secondo piano (Rosenblueth, Wiener e Bigelow, 1943).

A prescindere dalla complessità del comportamento, l'analisi comportamentale è applicabile tanto agli organismi viventi quanto alle macchine. Anche se gli studi funzionalisti rivelano profonde differenze fra animali e macchine, è ravvisabile una certa uniformità nelle classi generali di comportamento. È proprio tale uniformità a giustificare l'adozione di un metodo di studio congiunto per i comportamenti delle due diverse entità (Rosenblueth, Wiener e Bigelow, 1943).

I comportamenti che studia la cibernetica sono solo quelli di entità che agiscono secondo meccanismi a feedback. Il riferimento al feedback suggerisce la natura

---

<sup>4</sup> Ho già proposto la struttura a due elementi del pensiero cibernetico in Capriati (2023).

teleologica di un comportamento: la sua presenza mostra come esista un modello a cui l'entità – biologica o meccanica – deve uniformarsi. Tale modello è ciò verso cui l'entità tende e qualsiasi difformità rispetto al modello viene segnalata attraverso un feedback. Il feedback indica, infatti, la differenza fra l'effettivo comportamento dell'oggetto e lo stato delle intenzioni. Lo studio sulle intenzioni – o sul fine – che persegue l'entità è necessariamente uno studio teleologico. Secondo Rosenblueth, Wiener e Bigelow (1943), uno studio teleologico è considerato utile solo se vengono evitati i problemi relativi alla causalità e se tale studio riguarda esclusivamente un'analisi sulle intenzioni.

La concettualizzazione degli elementi appena esposti restituisce un'idea della cibernetica come scienza che si occupa dello studio di entità il cui funzionamento interno non è conosciuto o conoscibile. La prima cibernetica, infatti, si concentra sul comportamento e non sull'organizzazione interna di un oggetto. Ashby in “*An Introduction to cybernetics*” presenta il caso del bambino che prova ad aprire una porta. Perché il bambino possa aprire la serratura (output), egli deve approcciarsi alla maniglia (input). Il bambino deve, quindi, imparare a controllare l'output attraverso l'input e non c'è la necessità che egli sia in grado di vedere il meccanismo interno che li collega. Nella vita di tutti i giorni, secondo Ashby, ci confrontiamo con sistemi i cui meccanismi interni non si possono ispezionare e che debbono essere trattati con i metodi con cui ci si avvicina a una *black box*.

La *black box* – come oggetto insondabile e inconoscibile – assume una grande rilevanza nel mondo della cibernetica, tanto da costituire una vera e propria ontologia.

Pickering (2010, p. 20) sostiene che l'ontologia della *black box* restituisce un'immagine performativa del mondo. Ma cosa vuol dire questa affermazione? Una *black box* sarebbe un oggetto in grado di compiere determinate azioni: se sottoposto ad alcuni input, restituisce certi output. Sapere come essa lavora, però, non è necessario: è qualcosa che potrebbe (o non potrebbe) nascere dall'esperienza performativa della *box*. Per Pickering (2010), il segno distintivo della cibernetica è il rifiuto di lunghe digressioni sulla conoscenza o, detto altrimenti, la convinzione che per alcune questioni tali digressioni siano sbagliate, non necessarie o impossibili. La

cibernetica, quindi, sarebbe orientata a occuparsi della *performance* in quanto *performance*, e non in quanto pallida ombra della rappresentazione.

Da quanto emerso finora, si possono isolare due piani – quello della rappresentazione e quello della *performance*. Anche se la cibernetica non è definita puramente in termini anti-rappresentazionisti, la rappresentazione è orientata a correggere il funzionamento di quello che rappresenta. In questo senso, la cibernetica è una ontologia degli oggetti funzionanti e degli oggetti rotti che vanno aggiustati.

La *performance*, quindi, deve essere intesa come vero e proprio strumento gnoseologico. Secondo Pask (1958), esistono due tipi di osservatori: l'osservatore scientifico e quello cibernetico. Se il primo si astiene dall'interagire con l'oggetto osservato, il secondo è un osservatore partecipante, che, quindi, interagisce con l'entità che osserva. In secondo luogo, per l'osservatore cibernetico, l'oggetto non è più un'entità immobile e, per questa ragione, l'osservazione dell'oggetto non può ignorare i mutamenti che in esso si producono. La ragione, quindi, per cui si è parlato di "*performative epistemology*" (Pickering, 2010) è proprio questa: la conoscenza di un sistema complesso deve passare dalla *performance*.

Per Stafford Beer (1959), la cibernetica è la scienza che si occupa degli "*exceedingly complex systems*", che sono quei sistemi che non possono essere conosciuti attraverso i metodi utilizzati dalla scienza moderna. La ragione per cui tali sistemi non possono essere conosciuti mediante rappresentazione è legata al fatto che essi cambiano nel tempo: conoscere, cioè, il loro stato attuale non fornisce sufficienti garanzie per conoscere il loro comportamento futuro. Le rappresentazioni sono possibili, ma esse hanno solo una valenza provvisoria. Attraverso, invece, un metodo performativo, le entità sono osservate nel loro movimento, cioè nella loro evoluzione.

La cibernetica si occupa di studiare entità il cui funzionamento non è perfettamente conoscibile e il cui comportamento non è prevedibile. Quella della cibernetica, quindi, è un'ontologia dell'inconoscibilità (Pickering, 2010). Per questa ragione, la *black box* è la metafora adeguata per una ontologia del genere.

Da quanto emerso finora, oltre che per Cybersyn, la cibernetica fornisce un inquadramento teorico anche per il concetto di automazione.

### 1.2.2 Il progetto cibernetico cileno

Gli elementi che definiscono il pensiero cibernetico della prima ora – l’analogia animale macchina e la limitazione del campo di studio a ciò che concerne i meccanismi a feedback – si ritrovano in Cybersyn. Cybersyn – il cui nome è la fusione delle parole “*Cybernetics*” e “*Synergy*” – venne sviluppato tra il 1971 e il 1973 nel Cile di Allende con l’obiettivo di coordinare l’economia nazionale e di favorire la partecipazione dei lavoratori nelle decisioni. Ciò si sarebbe dovuto realizzare grazie a un sistema di comunicazione in tempo reale e a un programma automatizzato per l’elaborazione dei dati: si assiste così a una germinale forma di automazione di alcuni processi decisionali politici. Si possono rintracciare in Cybersyn due funzioni:

- a) trasmissione in tempo reale del feedback;
- b) elaborazione dei feedback.

Tali funzioni ritorneranno, sotto forme diverse, nel prosieguo della mia analisi. Occorre, quindi, osservare in che modo tali funzioni siano state implementate.

Preliminarmente, però, bisogna chiarire struttura e organizzazione di Cybersyn. Assieme a Stafford Beer, furono parte del progetto anche Fernando Flores – che, in qualità di ministro nel governo di Allende, si occupò della direzione politica – e Raúl Espejo, che aveva un compito di coordinamento generale.

La struttura di controllo e di gestione dell’informazione utilizzata prende il nome di “*viable system model*” (d’ora in poi, VSM). Il VSM venne pensato per essere adottato in qualsiasi forma di organizzazione. Per costruire questo modello, Beer si fece ispirare dagli organismi biologici, traducendo gli elementi chiave della loro organizzazione nella struttura di qualsiasi “*viable system*”. Il modello che egli elesse fu quello del sistema nervoso umano (fig. 1). Nel VSM, il sistema nervoso è diviso in cinque sottosistemi. Ogni sotto-sistema preserva la sua autonomia e controlla se stesso senza dipendere da un livello superiore del cervello. Tuttavia, vi è una funzione

di monitoraggio – svolta dalla colonna vertebrale – e che consiste in una comparazione fra le *performance* dei sotto-sistemi e quello che viene considerato il piano originario: ogni deviazione dal piano originario deve essere compensata da appropriate correzioni del comportamento dei sotto-sistemi, attraverso un meccanismo di feedback negativo (Beer, 1981).

In Cybersyn vi era ricezione e trasferimento di informazioni attraverso una serie di filtri e di protocolli. L'obiettivo del progetto era quello di ottimizzare la comunicazione fra le imprese e la direzione statale in tutto il Paese. Il VSM di Cybersyn venne, infatti, pensato per trasmettere e ricevere dati da – e verso – una sala di operazioni, emulando in ciò il rapporto che esiste fra il corpo e il sistema nervoso centrale.<sup>5</sup> Nella sala operazioni (la *operation room*) (fig. 2) – una stanza dal design futuristico che è anche l'immagine emblema del progetto Cybersyn – era presente un software, Cyberstride, che si dedicava all'elaborazione dei dati che arrivavano dalle imprese.

Negli stessi anni di Cybersyn, venne progettato – ma mai implementato – Cyberfolk (fig. 3). Esso, al pari di Cybersyn, era un sistema di comunicazione in tempo reale. Nel caso di Cyberfolk, la comunicazione, anziché fra aziende e stato, sarebbe dovuta avvenire fra cittadini e politici. Il progetto si basava sul modello “algedonico” – termine coniato da Beer e che si compone di due parole, “ALGOS” e “HEDOS”, che significano “dolore” e “piacere”.

Cyberfolk avrebbe permesso a ciascun individuo che disponesse di una televisione di comunicare il proprio grado di soddisfazione in merito all'azione del governo, esprimendo il proprio parere su una scala che va dal massimo piacere al massimo dolore. L'espressione del gradimento si realizzava attraverso dei voltometri che erano collegati alle televisioni domestiche. Si trattava di dispositivi dalla forma semicircolare, dotati di una lancetta che poteva essere ruotata verso destra (esprimendo piacere) o verso sinistra (esprimendo dolore), in risposta a qualcosa che era accaduto prima, come ad esempio, un discorso politico. Tale processo doveva

---

<sup>5</sup> In alcuni scritti di Beer (1981), l'analogia con il sistema nervoso – e, più in generale con il corpo biologico – è esplicita.

avvenire in tempo reale, così che i politici potessero ricevere un feedback immediato sui loro argomenti o proposte. In questo modo, i cittadini avrebbero potuto vedere la reazione dei politici ai loro feedback, così da creare uno scambio costante fra le due parti (Beer, 1981).

In Cyberfolk, il feedback doveva rimanere indefinito: l'espressione del gradimento era inarticolata, priva di parole. Il compito del voltmetro era solo quello di misurare il "piacere" e non anche la natura e le ragioni di questo piacere. A differenza dei sondaggi di opinione – che prevedono un numero di risposte limitato e definito –, i voltometri di Cyberfolk consentivano di esprimere un feedback generale, attraverso il quale era possibile stabilire se un certo discorso o argomento avesse generato dolore o piacere. Vi è una certa coerenza fra un meccanismo del genere e l'interesse per la cibernetica, cui si è accennato nelle pagine precedenti, per gli aspetti del sistema che riguardano il comportamento e il modello cui il comportamento deve adeguarsi.<sup>6</sup>

Se la scala algedonica riguarda l'espressione individuale di piacere-dolore, Beer (1983) per indicare il livello di benessere di una società nel suo insieme usa il termine "eudemony", che l'autore intende come *«I-like-it-here kind of happiness, that does not prejudice the nature of the well-being that the people's will seeks to express»*. Tuttavia, Beer non chiarì mai su cosa tale benessere dovesse fondarsi. Se si trattava, cioè, di un benessere economico o di altra natura.

Sulla base di quello che è emerso finora sul progetto cibernetico cileno, devo esplicitare due punti. In primo luogo, va sottolineato che il progetto di Beer per il Cile non si sarebbe limitato alla realizzazione di un coordinamento economico nazionale, ma riguardava un'idea più ampia che includeva anche altri aspetti della vita associata (Pickering, 2010). In secondo luogo, occorre osservare che l'obiettivo di Cyberfolk era quello di affiancare al feedback economico di Cybersyn un feedback sociale (Pickering, 2010).

---

<sup>6</sup> L'avversione di Beer (1974) per i sondaggi d'opinione è legata anche ad un altro ordine di ragioni: la presenza di domande strutturate e circoscritte non contribuisce mai all'emergere di reali novità in politica, ma ha il solo scopo di garantire lo status quo.

Per quanto riguarda le due funzioni cui si accennava poc'anzi – trasmissione in tempo reale del feedback ed elaborazione dei feedback –, esse dovevano trovare implementazione attraverso un sistema di telex – un telegrafo molto rapido – che avrebbe permesso la comunicazione in tempo reale di informazioni, e attraverso un software – Cyberstride – in grado di elaborare queste informazioni.

### **1.2.3 La seconda cibernetica**

Partendo dalle due funzioni di Cybersyn appena presentate, cercherò di tornare alla domanda di partenza: che ruolo può giocare la tecnologia per le decisioni di interesse pubblico che vengono prese in maniera collettiva?

Trasmissione in tempo reale del feedback ed elaborazione dei feedback sono solo due delle applicazioni che la tecnologia può trovare con riferimento alle decisioni politiche.

Mentre la trasmissione in tempo reale del feedback può essere considerata una forma di supporto alla decisionalità, l'elaborazione dei feedback è da annoverare tra gli strumenti che in una certa misura sostituiscono l'azione umana. Tale distinzione verrà approfondita nei prossimi capitoli. Per ora, è sufficiente osservare come fra questi due paradigmi non ci sia soluzione di continuità e che, quindi, si passi facilmente da una funzione all'altra.

Le due funzioni appena presentate sono descritte da Beer attraverso il VSM. Oltre che di animali e di macchine, per Beer, la cibernetica deve occuparsi anche di alcune forme di organizzazione – come le imprese o le istituzioni politiche. Nel fare ciò, egli estende l'analogia animale-macchina alle imprese e, più in generale, ai sistemi sociali. In virtù di ciò, quello che vale per i sistemi biologici e per le macchine varrebbe anche per l'organizzazione politica di un paese.

Seguendo i dettami della prima cibernetica ed estendendo l'analogia animale-macchina anche ai sistemi sociali, si può concludere che l'approccio allo studio delle forme di organizzazione politica deve necessariamente essere teleologico. Tuttavia,

Beer nega convintamente la natura teleologica del VSM. Riconoscere che il VSM abbia una natura teleologica, infatti, vorrebbe dire accettare che quello politico è un processo fondamentalmente eterodiretto, in cui i fini vengono posti da un'entità esterna. Negare la natura teleologica del VSM, quindi, significa dare l'immagine di un processo politico non manipolato e che si auto-organizza. La tesi di Beer è semplice: dal momento che sono gli stessi cittadini – o le imprese – a determinare i fini del sistema e, visto che i cittadini sono parti del VSM, è il VSM stesso ad autodefinirsi e, quindi, ad autodefinire i suoi scopi.<sup>7</sup> Per sostenere la sua tesi, Beer si appoggia al concetto di autopoiesi.

Il termine “autopoiesi” venne coniato da due biologi – Francisco Varela e Humberto Maturana (1987).

Quando parliamo degli esseri viventi, supponiamo che ci sia qualcosa in comune fra di essi [...] Quello che non è stato detto, tuttavia, è qual è l'organizzazione che li definisce come classe. La nostra proposta è che gli esseri viventi si caratterizzano perché si producono continuamente da soli, il che indichiamo denominando l'organizzazione che li definisce organizzazione autopoietica.

Con “autopoiesi”, dunque, si intende un particolare tipo di organizzazione, e con “sistemi autopoietici” si intendono quelle unità composite la cui organizzazione può essere descritta come un *network* chiuso di produzione di elementi che si forma attraverso l'interazione dei suoi elementi e che specificano la loro estensione costruendo i propri confini nel loro dominio di esistenza (Maturana, 1987). Si tratta, detto altrimenti, di sistemi che si autoproducono, cioè che sono in grado di produrre le proprie componenti.

Se da parte di Maturana e Varela l'estensione del concetto di autopoiesi ad entità non biologiche è cauta, altri autori si sbilanciano, applicando tale concetto in altri

---

<sup>7</sup> La formulazione di questi argomenti è una risposta alle critiche di Ulrich (1981).



campi.<sup>8</sup> A partire dagli anni '80, l'autopoiesi trova impiego anche nell'analisi dei sistemi sociali (Luhmann, 1986). Prima ancora di Luhmann, però, è proprio Beer ad adattare il concetto a un campo diverso da quello biologico. Tale concetto, come si legge nella sua prefazione a "Autopoiesi e cognizione" (Maturana e Varela, 1992) non riguarda solo la biologia – o alcuni sistemi disciplinari in particolare –, ma può essere definito "meta-sistemico" poiché, in qualche modo, riguarda tutte le discipline.

Il punto su cui Beer è in disaccordo con Maturana e Varela riguarda la natura delle società umane. In base a quanto sostiene Beer, nella prefazione al testo di Maturana e Varela (1992), le società umane sono a pieno titolo dei sistemi biologici. In quanto sistemi biologici, esse sono dei sistemi autopoietici. Maturana e Varela, invece, prendono le distanze da questa analogia.

Ma per quale motivo Beer è così interessato ad attribuire la qualifica di "autopoietico" ai sistemi sociali? La ragione è semplice e la spiega lui stesso: «il (...) motivo per cui il concetto di autopoiesi mi affascina così tanto è che coinvolge la distruzione della teleologia» (prefazione di Beer in Maturana e Varela, 1992).

Con riferimento alla distruzione della teleologia, in "Autopoiesi e cognizione" si legge:

usiamo la nozione di scopo quando parliamo di macchine perché essa chiama in gioco l'immaginazione dell'ascoltatore e riduce il compito esplicativo nello sforzo di trasmettergli l'organizzazione di una particolare macchina. In altre parole, con la nozione di scopo induciamo l'ascoltatore ad inventare la macchina di cui stiamo parlando. Questo, tuttavia, non dovrebbe portarci a credere che scopo, o fine, o funzione, siano proprietà costitutive della macchina che descriviamo con essi; tali nozioni sono intrinseche al dominio di osservazione, e non possono essere usate per caratterizzare qualche tipo particolare di organizzazione di macchine (Maturana e Varela, 1992).

---

<sup>8</sup> Coerentemente alla tradizione cibernetica, non viene fatta una distinzione fra sistemi viventi – o, *prima facie*, percepibili come viventi – e sistemi sintetici. In questo senso, i sistemi sintetici non vengono ritenuti a prescindere non capaci di autopoiesi.

Tanto per Beer quanto per Maturana e Varela, i sistemi autopoietici si caratterizzano per l'assenza di uno scopo – o di un progetto definito – verso il quale tendere. In questo senso, quindi, le società umane – e i loro sotto-sistemi, ad esempio «ditte e industrie, scuole e università, cliniche e ospedali, enti professionali, dipartimenti di stato, e interi paesi» (Maturana e Varela, 1992) – sono anch'essi dei sistemi autopoietici.

L'abbandono del concetto di teleologia comporta l'adozione di un approccio nello studio di questi sistemi – incluso il sistema politico di un paese – che può non tenere in considerazione lo scopo del sistema. La teleologia, infatti, «diventa solo un artificio della loro descrizione che non rivela alcun aspetto della loro organizzazione», e questa nozione non è necessaria «per la comprensione dell'organizzazione vivente» poiché «i sistemi viventi, come macchine autopoietiche fisiche, sono sistemi senza scopo» (Maturana e Varela, 1992).

Per Beer, come si è visto, la forma che i sistemi sociali – e i relativi sotto-sistemi – assumono è quella del VSM. Coerentemente a quanto detto finora, l'accostamento fra VSM e autopoiesi è esplicito. Beer (1981), infatti, fa un confronto molto oculato fra le condizioni di vita per come sono esposte nella teoria dell'autopoiesi e le condizioni di un *viable system*. Egli crede che esse siano complementari e si arricchiscano a vicenda. In questo senso, un *viable system* sarebbe autopoietico e le facoltà autopoietiche sono incorporate nella totalità e nei singoli sistemi.

Dal momento che il VSM è il modello utilizzato per la costruzione e la descrizione di Cybersyn, ne deriva che il progetto cibernetico cileno è, per Beer, fondamentalmente un progetto autopoietico, cioè, non-teleologico e che, dunque, non persegue alcuno scopo, ma che si fonda sull'auto-organizzazione e sull'auto-produzione delle relazioni sociali, per mezzo delle quali l'unità e la coesione del sistema vengono garantite.

Si può apprezzare, a questo livello, la differenza tra la prima cibernetica e i modelli che sono stati avanzati da Beer, Maturana e Varela. Se per Wiener, Rosenblueth e Bigelow – cibernetici della prima ora –, lo studio di meccanismi a

feedback era necessariamente lo studio di entità teleologiche, per Maturana e Varela la teleologia è prescindibile per la spiegazione del comportamento dei sistemi viventi, così come, per Beer, essa è prescindibile per la spiegazione del comportamento dei sistemi sociali.

Ma per quale ragione Beer è così interessato a smarcarsi da una ricostruzione teleologica e a legarsi al concetto di autopoiesi? Dal momento che molti hanno tacciato Cybersyn di essere uno strumento tecno-distopico per il controllo della popolazione, il richiamo al concetto di autopoiesi permette a Beer di riagganciarsi all'ideale dell'autogoverno e, quindi, all'ideale democratico. È proprio dalla tradizione democratica che bisogna cominciare, non tanto per decretare come e in che misura Cybersyn si inserisca in questa tradizione, quanto per valutare come tale tradizione risponda a quelle innovazioni tecnologiche che propongono la partecipazione delle macchine nei processi decisionali. Si vedrà, nei prossimi capitoli, a quali concezioni della democrazia si fa riferimento per determinare l'effetto dell'ingresso delle macchine nei processi decisionali. Per ora, mi limito ad osservare come la presenza o meno di un piano teleologico nella comprensione dei sistemi – siano essi animali, macchine o sistemi sociali – abbia prodotto una frattura fra prima e seconda cibernetica.

### **1.3 La questione teleologica**

Occorre addentrarsi in questa frattura per indagare la questione teleologica. In primo luogo, è davvero assente un piano teleologico nel discorso della seconda cibernetica?

Per sostenere la tesi dell'assenza di teleologia, Maturana e Varela (1992) osservano che ciò che contraddistingue gli esseri viventi non sia la loro capacità di riprodursi né tanto meno la loro evoluzione.

Riproduzione ed evoluzione non entrano nella caratterizzazione dell'organizzazione vivente, e i sistemi viventi sono definiti come unità dalla loro autopoiesi. Ciò è significativo perché rende la fenomenologia dei sistemi viventi dipendente dal fatto che sono unità autopoietiche (Maturana e Varela, 1992, p. 153).

### **1.3.1 Sulla natura (non) teleologica dei sistemi politici**

Nel momento in cui un sistema si differenzia dall'ambiente esterno, esso si costituisce come unità. Il rapporto che si instaura tra sistema autopoietico e ambiente esterno si fonda sul concetto di “omeostasi”. L'omeostasi è definita come «la condizione di mantenere costante o entro una gamma limitata di valori alcune variabili» (Maturana e Varela, 1992).

Si è visto che l'autopoiesi è ciò che permette a un sistema di conservare la propria unità. Ma la conservazione dell'unità non può essere considerata uno scopo verso il quale il sistema tende? Si ammetta che essa non sia lo scopo, ma che sia una qualità intrinseca e necessaria del sistema studiato, senza la quale il sistema cessa di poter essere definito tale. Si può dire la stessa cosa della regolazione omeostatica? Non vi è, forse, nel concetto di omeostasi, oltre alla conservazione dell'unità, la conservazione dell'unità in una certa forma, tale da fissare una norma – o una meta – cui il sistema deve adeguarsi – o deve tendere? Non si sta iscrivendo in questo modo il concetto di autopoiesi in un orizzonte teleologico?

Con riferimento all'assenza di teleologia, altre critiche possono essere mosse nello specifico a Beer. L'idea alla base di Cyberfolk era quella di garantire una reale partecipazione e di favorire l'autogoverno dei cittadini, evitando qualsiasi forma di eterodirezione. Dalla prospettiva di Beer, Cyberfolk è da considerarsi un sistema “non-teleologico” poiché non vi è uno scopo verso il quale il sistema tende né alcun obiettivo prefissato da raggiungere: il sistema, di volta in volta, si prefigge delle mete, attraverso il processo di feedback espresso dai cittadini. Tuttavia, Beer, in modo esplicito, si riferisce a un concetto che, anche se dal contenuto indefinito, riveste il

ruolo della meta o dello scopo verso cui il sistema tende: si tratta dell'*eudemony*. E anche se rinunciassimo all'*eudemony*, resterebbe il problema del consenso. L'obiettivo di Cyberfolk, infatti, è quello di misurare il consenso dei cittadini. La manovra, forse discutibile, di Beer è di non considerare il consenso come uno scopo da raggiungere, ma come un requisito necessario all'organizzazione per conservare la sua unità, per non disintegrarsi.

Altre critiche sono state avanzate con riferimento all'accostamento fra VSM e autopoiesi. Più precisamente, si è sostenuto che la differenza fra VSM e autopoiesi riguarda esattamente la questione della teleologia: il VSM avrebbe natura teleologica, mentre un sistema autopoietico no. Per Brocklesby e Mingers (2005), infatti, le nozioni di scopo, obiettivi e funzioni sono fondamentali per la teoria dei *viable system*, mentre sono concetti irrilevanti per l'autopoiesi. In altre parole, per Beer è difficile fuggire dall'idea di scopo nel definire i *viable system*.

Oltre, dunque, alla frattura fra prima e seconda cibernetica, vi è anche una frattura fra il pensiero di Maturana e Varela e quello di Beer. In sintesi, la supposta assenza di teleologia di Beer può essere contestata in tre punti.

In primo luogo, pur attribuendo natura non-teleologica all'autopoiesi, è stato osservato che i sistemi che Beer descrive come autopoietici sono fondamentalmente orientati ad ottenere un risultato che non coincide con la riproduzione delle reti di relazioni che costituiscono l'unità autopoietica. L'università, ad esempio, non tende a riprodurre gli elementi che la costituiscono – i ricercatori – ma a produrre conoscenza (Brocklesby e Mingers, 2005).

In secondo luogo, il concetto stesso di *eudemony* – cui si appella Beer per giustificare il VSM – ha a che fare con il raggiungimento di un benessere che, benché non specificato in senso sostanziale, non può che assumere il carattere di un fine verso cui il sistema tende. Lo stesso discorso può essere fatto per il consenso.

Vi è poi un terzo ordine di ragioni che riguarda il concetto di autopoiesi in generale. Anche se venisse riconosciuta natura autopoietica al VSM, si è visto come la stessa autopoiesi non è esente dalla critica di essere un sistema fondamentalmente teleologico. Benché venga chiarito che elementi necessari alla definizione di

autopoiesi non siano riproduzione ed evoluzione, è difficile non considerare il mantenimento della propria unità come un fine verso cui il sistema tende. In questo senso, il concetto di omeostasi rappresenta la norma cui un sistema autopoietico deve conformarsi.

### **1.3.2 Questioni di legittimità nel progetto cibernetico cileno**

L'autopoiesi segna un passaggio fondamentale fra la prima e la seconda cibernetica. Mentre nella prima cibernetica il feedback rivelava la natura teleologica dei sistemi studiati; nella seconda cibernetica, viene proposto un approccio non-teleologico e questi stessi sistemi vengono rappresentati come entità che si auto-producono, la cui autonomia è definita dalla capacità di conservare la loro stessa unità. Accordare o meno credito alla ricostruzione non-teleologica di Beer porta a due distinte conclusioni con riferimento alla questione del fondamento di legittimità del sistema.

Si è osservato come Beer, con il richiamo all'autopoiesi, cerchi di creare un legame con la tradizione democratica. Ciò significa che il VSM ambisce ad essere, in ultima analisi, un sistema democratico. L'autopoiesi, detto altrimenti, è il mezzo adottato da Beer per legittimare democraticamente il VSM. A fondamento di ogni sistema politico vi è una qualche fonte di legittimità. Tale questione sarà approfondita nel terzo capitolo, tuttavia, è opportuno fin da subito rilevare come la legittimità si costruisca a partire da una giustificazione.

Le questioni che concernono le giustificazioni di un sistema democratico intendono rispondere alla domanda "perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?". Si è soliti distinguere fra giustificazioni della democrazia di stampo strumentalista e giustificazioni non-strumentaliste. Se nelle teorie non-strumentaliste la giustificazione è intrinseca al metodo decisionale democratico, nelle teorie strumentaliste, invece, la giustificazione deve essere

connessa ai risultati che il metodo decisionale democratico produce in comparazione ad altri procedimenti (Christiano, 2022).

Per quale ragione concepire un sistema politico come teleologico – o meno – sarebbe relazionato al tipo di giustificazioni che sono a fondamento della democrazia? In qualche modo, la questione della natura teleologica precede quella delle giustificazioni. Se le giustificazioni forniscono argomenti sul perché un procedimento democratico sia migliore per prendere decisioni rispetto ad altri procedimenti, definire “teleologico” un certo sistema politico significa ammettere che esso persegue un determinato scopo. In questo senso, la presenza – o meno – di uno scopo anticipa la domanda sulle giustificazioni.

Perché un sistema possa essere definito “migliore”, occorre che si chiarisca il concetto di migliore: migliore a quale scopo? Una volta che si è ammesso che i sistemi politici perseguono uno scopo, “migliore” può voler dire “prendere decisioni migliori” per lo scopo che il sistema persegue. In questo modo, si entra sicuramente nel dominio delle giustificazioni strumentaliste, giacché, come si è visto, e come si vedrà più diffusamente oltre, le giustificazioni strumentaliste sono connesse ai risultati – e, più precisamente, alle decisioni – che il procedimento democratico produce. Se concepire un sistema politico in termini teleologici significa esplicitare un bene – fuori dal sistema stesso – che si intende perseguire, allora chi adotta una concezione teleologica deve necessariamente abbracciare giustificazioni di stampo strumentalista. Dal momento che un sistema politico descritto come non-teleologico non persegue, per definizione, alcuno scopo particolare, esso non può giustificarsi sulla base della produzione di determinati risultati: l’assenza di uno scopo esterno al sistema implica l’assenza di un parametro per valutare la qualità di una decisione e, dunque, sarebbe impossibile stabilire che la decisione *x* presa con metodo democratico sia migliore rispetto alla decisione *y* presa con un altro metodo. E se una concezione non-teleologica non può giustificare strumentalmente la democrazia, allora essa, per esclusione, deve reggersi su una giustificazione non-strumentalista.

Questo è quello che ho sostenuto nell’articolo “Autopoiesi dei sistemi politici: il caso Cybersyn”, ma mi sbagliavo: sostenere che i sistemi politici siano non-

teleologici non porta a giustificazioni di stampo non-strumentalista, ma, semplicemente, le concezioni non-teleologiche rendono insensata la domanda “perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?” Anche le tipiche giustificazioni di stampo non-strumentalista – quelle che si fondano su valori quali l’equità e la libertà – pur non essendo vincolate alla produzione di determinati output, individuano un bene o uno scopo che il sistema deve perseguire che, nel caso delle tesi non-strumentaliste tipiche, sono valori come l’equità e la libertà. In questo senso, si deve sostenere che anche le giustificazioni non-strumentaliste rientrano all’interno di un paradigma teleologico. Da ciò si può concludere che solo una concezione teleologica dei sistemi politici permette di giustificare la democrazia, mentre una concezione non-teleologica può, al massimo, essere il fondamento di un sistema che non ha bisogno di essere giustificato – dunque un sistema anarchico.<sup>9</sup>

Ma Beer aveva in mente l’anarchia quando progettava il suo sistema politico? Evidentemente no e, in questo senso, la sua teoria va a collocarsi in una posizione particolare. Si mettano da parte le remore nel considerare non-teleologico un sistema che ha come unico scopo il mantenimento della propria unità. Si assuma, in sostanza, che il modello di Beer sia davvero non-teleologico. Ciò cosa comporta? In quanto sistema non-teleologico, come si è visto, non ha senso porsi la domanda sulla natura delle giustificazioni: non ci sono giustificazioni per un sistema non-teleologico perché la domanda “per quale motivo la democrazia è il miglior modo per prendere decisioni di interesse pubblico?” non troverebbe risposta dal momento che il concetto

---

<sup>9</sup> Per “anarchia” intendo l’assenza di ordine costituito. In proposito, sono significative le parole di Nozick (2008, p. 27): «la questione fondamentale della filosofia politica, precedente ogni interrogativo su come organizzare lo stato, è se debba mai esistere uno stato. Perché non l’anarchia? Poiché la teoria anarchica, se difendibile, elimina radicalmente l’intero ambito della filosofia politica, è opportuno prendere le mosse da una disamina della sua principale alternativa teorica. Chi considera l’anarchismo una dottrina non priva di fascino sarà disposto ad ammettere la possibilità che con ciò si concluda anche la filosofia politica». Si può, dunque, concludere che l’unico sistema che non ha bisogno della filosofia politica – così come delle giustificazioni – è per l’appunto quello anarchico. Sul concetto di filosofia politica si tornerà nelle ultime pagine di questo capitolo.



di “migliore” deve avere un termine di riferimento (migliore in riferimento a cosa?) e, nelle concezioni non-teleologiche, è strutturalmente assente tale termine. Se si intendono i sistemi politici in termini non-teleologici, non è possibile giustificare la democrazia: una concezione non-teleologica non riesce a fornire giustificazioni per la democrazia poiché essa non genera il bisogno di giustificazioni: l’esistenza di un certo sistema si giustifica sulla base della sua esistenza. Da ciò discende che un sistema politico concepito come non-teleologico – così come è stato pensato da Beer – ha come unico obiettivo quello di evitare di perire. La necessità di sopravvivere – senza alcun richiamo ad altri valori – dà forma ad una difesa ostinata dello *status quo*. Si assiste, dunque, a una strana torsione: concepire un sistema politico in termini non-teleologici vuol dire concepire come unico modello politico possibile l’anarchia, ma, allo stesso tempo, esasperando tali premesse, si finisce per legittimare un sistema politico per il semplice fatto che esso esiste, delegittimando qualsiasi pulsione che comporti un cambiamento nella sua *organizzazione*, giacché, come si è visto, un sistema autopoietico deve mantenere la sua *organizzazione* invariata.

Al contrario, assumere che il VSM sia un sistema teleologico collocherebbe il pensiero di Beer in una posizione meno controversa. In questo senso, quella di Beer non sarebbe più una difesa ostinata dello *status quo* e l’aderenza al metodo democratico potrebbe essere giustificata in due modi:

- sulla base del fatto che la democrazia produce decisioni migliori – che, cioè, consentono di raggiungere l’eudemonia – (giustificazione strumentalista);
- perché la democrazia preserva alcuni valori fondamentali come quelli di libertà ed eguaglianza – attraverso il principio del consenso – (giustificazione non-strumentalista).

Se si protende per una concezione non-teleologica, invece, il paradosso rimane evidente: nonostante le intenzioni di Beer fossero quelle di dare vita a un sistema rivoluzionario, non eterodiretto e in grado di autodeterminarsi e rinnovarsi ad ogni istante, egli ha finito per costruire un sistema dal quale non è possibile uscire, poiché si pone su un piano trascendente e che giustifica la propria esistenza non sulla base di

qualche bene esterno al sistema stesso, ma sul fondamento che la vita fuori dal sistema non è pensabile.

La relazione fra sistema e ambiente assume, quindi, un ruolo fondamentale nella teoria di Beer. Se con la prima cibernetica – e con la metafora della *black box* – questa relazione era caratterizzata da uno scambio input/output, con l'affermarsi del concetto di autopoiesi questo modello entra in crisi. Per affrontare questa crisi, interviene la distinzione operata da Maturana e Varela (1992) – adottata anche da Beer – fra struttura e organizzazione. Mentre l'organizzazione è definita nei termini di «relazioni che definiscono un sistema come una unità» (Maturana e Varela 1992, p. 201), il rapporto con l'esterno avviene attraverso l'accoppiamento strutturale, che è la relazione fra la struttura di un'unità e quella del suo ambiente. I cambiamenti che si verificano nella struttura devono essere tali da preservare l'organizzazione del sistema nell'ambiente. L'insieme di questi cambiamenti è definita “adattamento” e il rapporto fra sistema e ambiente interessa solo gli aspetti strutturali. Perché il sistema possa mantenere la propria identità, l'organizzazione trascende la relazione sistema/ambiente ed ha il fine di rimanere invariata. Beer, dunque, accetta qualsiasi tipo di cambiamento strutturale per il suo sistema, a patto che questo continui a sopravvivere, conservando la propria organizzazione e, quindi, la sua unità.

#### **1.4 Sul ruolo euristico di Cybersyn**

Il progetto cibernetico cileno è il punto di partenza di questa analisi perché anticipa molti degli elementi che approfondirò nel resto della trattazione. Se oggetto di questa tesi è la relazione fra automazione dei processi decisionali e democrazia, Cybersyn, avendo affrontato pionieristicamente il difficile rapporto fra tecnologia e decisioni di interesse pubblico, ha già abbozzato delle risposte ad alcune delle domande che mi interessano.

Se l'intento di questa indagine è di vagliare le possibilità di automazione di alcune porzioni della decisionalità politica, si è specificato che ciò che rileva è l'ingerenza

delle macchine nei processi decisionali di interesse pubblico. Tale ingerenza può essere di due tipi: una forma *soft*, che vede le macchine come strumenti di supporto alla decisionalità; un'altra meno *soft* che, invece, accorda un certo potere decisionale alle macchine. Non è chiaro quando si possa stabilire il passaggio da un tipo di ingerenza all'altra. Si cercherà di costruire un criterio per fissare questa distinzione. Tuttavia, nello sviluppare tale criterio sarà importante anche approfondire il concetto di ingerenza che, quando parliamo di processi decisionali, prende il nome di partecipazione. Di che tipo di partecipazione si tratta? E a quali processi decisionali ci si riferisce?

Lo studio di Cybersyn, in secondo luogo, permette di tracciare i limiti del contesto entro cui inserire queste riflessioni. Si è osservato come il richiamo al concetto di autopoiesi – oltre che scongiurare la critica di aver creato un sistema manipolatorio – sia funzionale ad agganciare il modello di Beer alla tradizione democratica. La partecipazione delle macchine ai processi decisionali va misurata in un contesto democratico poiché quando ci si riferisce a decisioni di interesse pubblico prese in maniera collettiva, secondo la cultura politico-giuridica occidentale, si sta parlando di democrazia. Ciò che è importante, a questo punto, è capire se la teoria della democrazia, secondo i suoi più recenti sviluppi, sia pronta ad accettare – e in che misura – la partecipazione delle macchine ai processi decisionali.

Questa operazione di compatibilità venne intrapresa già da Beer: dal momento che egli considera, più o meno esplicitamente, il suo sistema come democratico, bisogna interrogarsi su che tipo di giustificazione si fondi tale sistema.

Beer vincola il VSM all'idea di autopoiesi – quindi a una concezione non-teleologica – che, però, non è in grado di fornire giustificazioni alla democrazia. Anche se non tornerò su concetti come quello di autopoiesi, verranno valutate le varie giustificazioni di volta in volta utilizzate per legittimare politicamente la democrazia. A seconda del tipo di giustificazione adottata, infatti, vi sono distinte concezioni di democrazia e, conseguentemente, diverse risposte alla domanda “che ruolo può giocare la tecnologia nelle decisioni di interesse pubblico?” In questo senso, le considerazioni sulla natura teleologica o meno del VSM introducono quelle sulla

natura delle giustificazioni della democrazia – che possono essere di tipo strumentalista o non-strumentalista. A partire dai due principali tipi di giustificazioni è possibile isolare due distinte concezioni della democrazia, una delle quali, come si vedrà, segna un argine – o una barriera – per la partecipazione delle macchine nei processi decisionali, così come rappresenta un limite per la delega di porzioni di decisionalità a una macchina.

Inoltre, si è osservato che nella ricostruzione di Beer non è così facile distinguere gli apporti umani da quelli tecnici. Il VSM presenta un'immagine del sistema politico in cui l'attività tecnica e quella umana sono a tal punto intrecciate che porre una linea di demarcazione si rivela un'operazione ardua. Nonostante esistano concetti ponte per catturare la commistione fra queste due attività – come, ad esempio, “sistema socio-tecnico” e “intelligenza ibrida” – che permetterebbero di glissare sulla questione, nel secondo capitolo si proverà ad elaborare un criterio utile per distinguere gli apporti umani da quelli tecnici. È per questa ragione che sarà necessaria una solida definizione di intelligenza artificiale – ammesso che “intelligenza artificiale” sia il concetto appropriato per cogliere l'ingerenza delle macchine nei processi decisionali – e occorrerà comprendere quali vie seguire per porre una distinzione efficace tra aspetti tecnici e umani in una decisione.

A margine – e quasi inesistenti – saranno i riferimenti agli strascichi cibernetici sul modo di intendere il rapporto fra decisioni collettive e tecnologia: in che misura le questioni sulla democrazia siano influenzate dal pensiero cibernetico non è in questa sede una questione che può venire esaurientemente affrontata.<sup>10</sup>

Le questioni, invece, che riguardano la cibernetica in generale sono il preludio a concetti che più in là verranno approfonditi. L'ontologia dell'inconoscibilità e la *black box* anticipano l'emergere di soggettività artificiali i cui comportamenti non sono prevedibili o sono difficili da spiegare a posteriori.

La questione sulla natura – teleologica o meno – del VSM non verrà ulteriormente approfondita, ma rappresenta un problema che tornerà più volte nel corso di questa

---

<sup>10</sup> Questi "strascichi" vengono trattati, ad esempio, da Espejo (2022) e da Wessel Reijers, Liav Orgad e Primavera de Filippi (2022).

indagine. Accostare il sistema politico a un sistema biologico non è solo un modo di riscattare l'immagine del progetto cibernetico cileno, dandogli un'aura democratica, ma può anche essere visto come un modo per risolvere il paradosso della filosofia politica individuato da Hannah Arendt. Durante un'intervista del 1964 con Günter Gaus, Arendt sostiene:

L'espressione "filosofia politica", che evito, è estremamente appesantita dalla tradizione. Quando parlo di queste cose, a livello accademico e non, dico sempre che c'è una tensione vitale tra filosofia e politica. Cioè, tra l'uomo come essere pensante e l'uomo come essere agente, c'è una tensione che non esiste nella filosofia naturale, per esempio. Come tutti gli altri, il filosofo può essere oggettivo nei confronti della natura e quando dice ciò che pensa su di essa parla a nome di tutta l'umanità. Ma non può essere obiettivo o neutrale nei confronti della politica (Arendt, 2003).<sup>11</sup>

Se lo studio di un sistema politico è identico a quello biologico, la tensione interna alla filosofia politica svanisce. In questa ricerca, tale tensione comparirà a più riprese e in varie forme.

A seconda del peso dato a questa tensione, si possono apprezzare distinti orientamenti in merito alla questione della possibilità di automazione dei processi decisionali. Credere che esista una presunta oggettività quando si parla di questioni politiche permette, come si vedrà, di costruire una concezione della democrazia che agevola e favorisce la partecipazione delle macchine ai processi decisionali. Al contrario, ritenere che non esista uno sguardo oggettivo o neutrale sulle questioni politiche porta a concezioni della democrazia che pongono maggiori ostacoli a un programma di automazione di porzioni del processo decisionale. La partecipazione di una macchina ai processi decisionali dipende, in una certa misura, dal superamento della tensione fra filosofia e politica.

Beer elude questa tensione. Equiparando il sistema politico al sistema biologico, per l'ideatore di Cybersyn lo studio della filosofia politica non presenta una

---

<sup>11</sup> Traduzione mia da Arendt (2003).

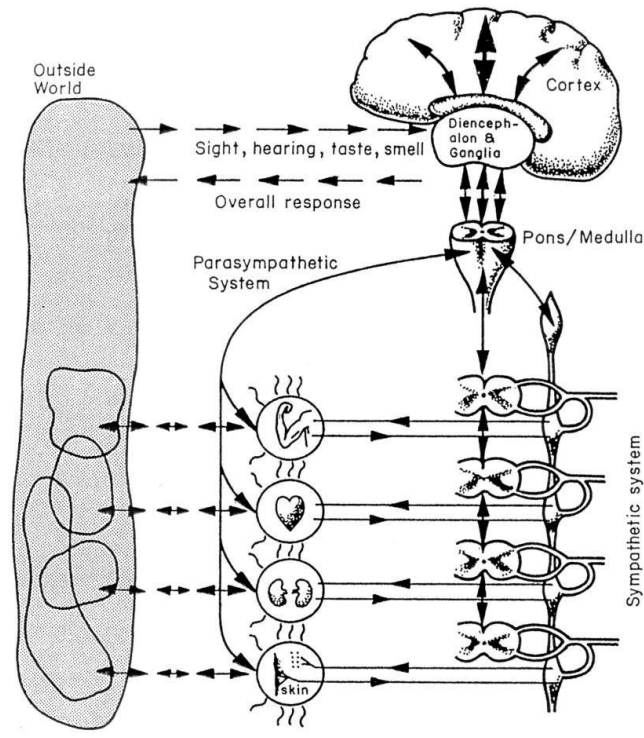
specificità tale da far sorgere la necessità di distinguerlo, come fa Hannah Arendt, dalla filosofia della natura. Ma perché questa tensione ha a che fare con il rapporto fra democrazia e intelligenza artificiale?

Per Arendt (2003), politica e filosofia appartengono a due distinti domini. Mentre la politica si occupa di questioni pratiche che riguardano la scelta fra una varietà di distinti corsi d'azione alternativi, la filosofia ha a che fare con le speculazioni astratte sulla verità e sull'identificazione di singoli corsi di azione obbligati. In altre parole, il conflitto fra politica e filosofia riprende quello fra opinione e verità. Quando un dominio invade l'altro, si verificano situazioni di confusione. Più precisamente, nel tentativo di determinare un unico corso di azione valido utilizzando gli strumenti della filosofia – quindi la speculazione –, la filosofia politica negherebbe la specificità del politico, imponendo una logica ad esso estranea (Palumbo, 2012). In questo senso, l'idea che ci possa essere un unico corso d'azione valido nel campo della politica e che il compito dei processi decisionali sia quello di scoprire tale corso d'azione, permette di sostenere che qualora una macchina si dimostrasse capace di individuare questo corso d'azione, l'attribuzione ad una macchina di potere decisionale sarebbe giustificata. Se, invece, viene rigettata l'idea che vi sia un unico corso d'azione valido, non avrebbe senso sostenere che esiste una macchina in grado di individuare questo corso d'azione e, conseguentemente, perderebbe di senso la prospettiva di attribuzione di potere alla macchina.

La questione, presentata in questi termini, appare eccessivamente semplificata. Il conflitto tra verità e opinione nel campo del politico – che, secondo Arendt (1990), risale alla disputa fra Socrate e Platone – dev'essere valutato alla luce dell'affermazione di distinte teorie della democrazia che reggono distinti modelli di processi decisionali.

Ma, prima ancora, di analizzare i processi democratici, bisogna definire l'altro termine della questione: la macchina. Cos'è, dunque, una macchina?

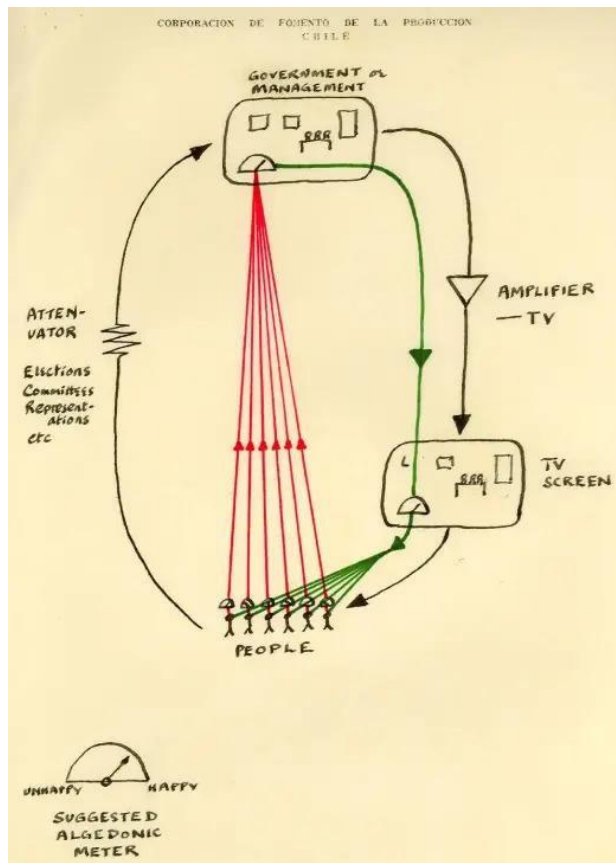
## IMMAGINI



**Fig. 1:** Viable System Model (biological). Reprinted from Stafford Beer, *Brain of the Firm: The Managerial Cybernetics of Organization*, 2nd ed. (New York: J. Wiley, 1981).



**Fig 2:** Operation room, Santiago, Chile.



**Fig. 3:** Cyberfolk disegnato da Stafford Beer.



## **CAPITOLO 2 – Macchine per decidere**

### **2.1 IA e democrazia: stato dell'arte**

### **2.2 Di che macchine si tratta**

#### **2.2.1 Il criterio dell'autonomia**

#### **2.2.2 Il criterio dell'intelligenza**

### **2.3 Concetti “ponte”. Per un'analisi congiunta degli apporti umani e artificiali**

#### **2.3.1 Intelligenza ibrida**

#### **2.3.2 Sistemi socio-tecnici**

#### **2.3.3 Macchine politiche**

### **2.4 Per un'analisi disgiunta**

#### **2.4.1 Il *machine learning***

#### **2.4.2 Sull'utilità e l'opportunità di un'analisi disgiunta**

## CAPITOLO 2

Molti degli studi sulla relazione tra intelligenza artificiale e processi democratici hanno riguardato la stima dei vantaggi o dei pregiudizi che le nuove tecnologie possono portare all'ideale democratico. Dopo aver chiarito che non mi occuperò di queste valutazioni, procederò a tracciare il profilo della macchina che interessa questa ricerca. Si tratta di una macchina che deve essere in grado di prendere una decisione e che non si limita ad eseguire il compito per cui è stata programmata. Sto, quindi, alludendo ad una questione di autonomia: una macchina è in grado di prendere decisioni solo se è autonoma. Ma quando una macchina è autonoma? L'autonomia è definita nei termini di indipendenza *relativa* da qualcosa e, nel caso di macchine, tale indipendenza riguarda il fattore umano. Quella dell'autonomia sarà da considerarsi condizione necessaria e sufficiente per definire la macchina protagonista di questa ricerca. Farò a meno, dunque, del criterio dell'intelligenza poiché esso non può essere determinato in termini chiari e non ambigui. Esistono, tuttavia, approcci che mettono in discussione la presunta autonomia delle macchine, proponendo uno studio “non separatista” di quelle situazioni in cui macchine e umani svolgono insieme una certa attività. Mi sto riferendo a concetti ponte quali “intelligenza ibrida”, “sistemi sociotecnici” e “macchine politiche”. Recuperando la nozione di *machine learning* – che fornisce un fondamento solido all'ipotesi che una macchina possa essere autonoma – propongo, nello studio dei processi decisionali che vedono coinvolti umani e macchine, di adottare un approccio che recida in maniera netta il legame umano-macchina al fine di identificare esattamente il ruolo e i limiti della macchina.

### 2.1 IA e democrazia: stato dell'arte

Gli studi che riguardano il rapporto fra intelligenza artificiale e democrazia hanno seguito prevalentemente due linee d'indagine. Mentre la prima linea si è occupata

degli aspetti critici legati all'avvento delle nuove tecnologie, la seconda linea si è dedicata allo studio dei possibili impatti positivi dell'intelligenza artificiale sui sistemi democratici.

Per quanto riguarda la prima linea d'indagine, si possono osservare almeno tre ordini di problemi relativi all'avvento delle nuove tecnologie.

Il primo riguarda l'enorme concentrazione di potere che si verifica quando strumenti tecnologici di particolare impatto non sono sottoposti a regolamentazione e appartengono a imprese private (Lanier, 2014, 2018; Tufekci, 2015, 2017; Helbing et al., 2019; Bartlett, 2018; Pasquale, 2015, 2019, 2020; Veliz, 2021; Stark, 2018; Iliadis e Russo, 2016; Andrejevic e Gates, 2014). In questi casi, alcuni attori privati accumulando una grande quantità di dati degli utenti e, creando piattaforme di scambio di informazione, finiscono per avere un ruolo centrale nella formazione delle opinioni. Pionieristico, in questo senso, è il lavoro di Zuboff (2019), "Il capitalismo della sorveglianza".

Il secondo ordine di problemi, invece, riguarda la presenza di *bias* nei processi decisionali in cui partecipano macchine e l'imperscrutabilità – o opacità – delle decisioni (Burrell, 2016; Selbst e Barocas, 2018; Zarsky, 2016). Alcuni di questi studi sono stati condotti a partire dall'esperienza di corti in cui si è ricorso a sistemi di giustizia predittiva – il caso più noto è quello di COMPAS negli Stati Uniti.

Infine, sempre su un piano critico, si pongono gli studi:

- che riguardano un presunto cambio di paradigma della governamentalità a seguito dell'avvento delle nuove tecnologie (Berns e Rouvroy, 2013; Roberts, 2019; Rouvroy 2011; Rouvroy e Stiegler, 2016);
- su forme di governo che si servono di algoritmi (Yeung 2018; Kuziemski e Misuraca, 2020; D'Agostino e Durante, 2018);
- su un utilizzo degli algoritmi in chiave biopolitica (Cheney-Lippold, 2011).

La seconda linea d'indagine, invece, si è concentrata sui possibili effetti virtuosi che le nuove tecnologie possono avere sui sistemi democratici (Martí, 2021; Landemore, 2022; Cristianini e Scantaburlo, 2020; Noveck, 2009, 2015). Alcuni di questi contributi si sono occupati di "*computational ethics*" o di etica degli algoritmi

(Awad et al., 2022; Mittelstadt, Allo, Taddeo, Wachter e Floridi, 2016; Dubber, Pasquale e Das, 2020; Floridi, 1999).

Le due linee di indagine appena tratteggiate corrispondono a due tipi di attitudine: da un lato, vi è un atteggiamento tecno-fobico, secondo cui l'impiego di alcuni strumenti tecnologici conduce inevitabilmente a esiti negativi – se non, in alcuni casi, catastrofici. Dall'altro lato, è ravvisabile un approccio tecno-entusiasta, che nutre la speranza che l'intelligenza artificiale risollevi le sorti di una democrazia ormai verso il tramonto. Tecno-entusiasmo e tecno-fobia rappresentano i due poli entro cui si sviluppa il discorso sulle nuove tecnologie. Se la tecno-fobia conduce a una sorta di distopia che auspica un ritorno all'era pre-internet, i tecno-entusiasti credono che l'ingresso nell'utopia sia rimesso agli avanzamenti in campo tecnologico. La maggior parte dei contributi relativi al rapporto fra democrazia e nuove tecnologie si sforza di restare in equilibrio fra i due poli appena presentati: al di là di pochi casi isolati,<sup>12</sup> gli studi sul difficile rapporto fra intelligenza artificiale e democrazia non sono mai esageratamente sbilanciati su uno dei due poli, e, più o meno tutti, sono concordi nel ritenere che l'intelligenza artificiale possa arrecare, secondo l'utilizzo che se ne fa, tanto benefici quanto pregiudizi ai sistemi democratici. È, inoltre, diffusa la consapevolezza che il legislatore e il programmatore viaggino a due velocità diverse e, in molti di questi contributi, viene denunciata l'assenza di leggi che riescano a tenere testa alla rivoluzione tecnologica in corso. Vi sono, dunque, numerose proposte su interventi per regolamentare tali strumenti (Monarcha-Matlak, 2021).

A prescindere dalla linea d'indagine, è chiaro che esiste una tensione fra intelligenza artificiale e democrazia (Simoncini, 2019; Sartor, 2020). Tale tensione è stata analizzata da diversi punti di vista. Sempre in linea con l'approccio "ottimista", nel rapporto fra legittimità democratica e intelligenza artificiale, ci sono studi che foraggiano un utilizzo etico e sostenibile del *machine learning* (Beckman, Hultin Rosenberg e Jebari, 2022).

---

<sup>12</sup> Un esempio su tutti è quello dell'accelerazionismo (Williams e Srnicek, 2018).

Altri autori hanno tentato di risolvere questa tensione attraverso concetti-ponte, come ad esempio “*political machines*” (Papakyriakopoulos, 2022), “sistema sociotecnico” (Vermaas, Kroes, van de Poel, Franssen e Houkes, 2011) e “intelligenza ibrida” (Pescetelli, 2021; Akata et al., 2020; Mulgan, 2018; Jennings, Moreau, Nicholson, Ramchurn, Roberts, Rodden e Rogers, 2014; Enarsson, Enqvist e Naarttijärvi, 2022). Secondo queste ricostruzioni, nelle odierne società, l’attività della macchina e quella umana, in alcuni ambiti, si sono a tal punto compenstrate che è diventato necessario elaborare un *framework* per cogliere l’incontro umano-macchina nella sua totalità.

Se in fase giudiziaria (o amministrativa) c’è stata un’importante diffusione di strumenti di giustizia predittiva (Citron e Pasquale, 2014; Ensign, Friedler, Neville, Scheidegger e Venkatasubramanian, 2018), i tentativi di legiferare attraverso l’intelligenza artificiale sono ancora in una fase embrionale (Solum, 2019; Lauritsen e Gordon, 2009; Nay, 2018; Ashley, 2017). Occorre, tuttavia, osservare che alcuni Paesi, come la Svezia, hanno già previsto la possibilità di adottare strumenti automatizzati di *decision-making*. Nel recente “*Administrative Procedure Act*”, nella sezione 28 si legge: «*A decision can be made by a public servant alone or jointly by several public servants or automated*» (förvaltningslag, 2017:900).

Anche gli studi che si sono dedicati ai profili di compatibilità fra “*rule of law*” e “*automated decision-making*” (Suksi, 2023) ricalcano la struttura a doppio binario appena presentata: mentre alcuni vedono nei processi decisionali automatizzati uno strumento per rafforzare lo stato di diritto, altri preconizzano esiti nefasti per il *rule of law*, che potrebbe essere minacciato dall’avvento del “*rule of algorithms*”.<sup>13</sup> In

---

<sup>13</sup> Si veda Meneceur (2023) che riprende un’espressione di Batko e Kreft (2017). In particolare, Meneceur spiega il passaggio dal *rule of law* al *rule of algorithms* in questi termini: «*This is an essential distinction between the two governmental regimes under consideration: while the rule of law sets, according to democratic deliberative modalities, preliminary rules to regulate present behavior and sanction past illicit behavior through legitimate authorities, “algorithmic government” or what we will call “rule of algorithms” aims, through a statistic representation of past behavior, to limit the possibilities of future behavior within the interpretative framework desired by its designers*».

particolare, gli “ottimisti” credono che l’automazione sia un modo per ridurre la discrezionalità, che rappresenta un ostacolo nella misura in cui può comportare errori e *bias* (Bullock 2019). È stato osservato, tuttavia, come tale discrezionalità non sparisca del tutto: essa viene anticipata al momento di sviluppo dei software che si occuperanno dell’automazione. In questo senso, c’è un passaggio di discrezionalità dal soggetto pubblico – dotato della legittimità per prendere decisioni giuridicamente vincolanti – a un soggetto privato, incaricato di progettare l’algoritmo che permette l’automazione (Waldman, 2019; Brayne, 2021).

Non mi avventurerò, in questa sede, in considerazioni sull’opportunità – o meno – dell’impiego di tecniche di intelligenza artificiale in campo giuridico-politico. Questo studio, infatti, non affronterà le questioni appena esposte, ma tratterà di una problematica che, a mio avviso, le precede. Mi occuperò di vagliare la compatibilità fra macchine e teorie della democrazia. In altre parole, può un sistema continuare a dirsi democratico se una parte delle decisioni vengono prese da una macchina? E quali tipo di decisioni possiamo concedere a una macchina? L’ammissibilità e il grado di ingerenza di una macchina nei processi decisionali dipendono dalla concezione di democrazia di partenza. E, prima ancora, dalla concezione di macchina.

## **2.2 Di che macchine si tratta**

Come ho già osservato nel primo capitolo, le macchine che interessano questa ricerca sono quei sistemi che partecipano a un processo decisionale o quei sistemi a cui viene delegata una porzione di processo decisionale. Un esempio è il software Cyberstride che, a partire dalle informazioni giunte dai diversi terminali, formula una decisione. Le macchine, quindi, che verranno prese in considerazione sono quelle in grado di decidere. Ma quando è possibile stabilire che una macchina è in grado di decidere e quando, invece, essa si limita ad eseguire un compito per cui è stata programmata?

Si tratta di dedicarsi allo studio di quelle macchine il cui comportamento si discosta in una certa misura da quello che il programmatore ha previsto avessero. Creare un criterio di distinzione su queste basi non è un compito facile. Qual è un comportamento che si discosta da quello che il programmatore ha già previsto? Deve trattarsi di macchine imprevedibili e fuori controllo, *a là* Blade Runner?

Occorre, a questo punto, fare una precisazione: perché parlo di macchine e non di intelligenza artificiale? Principalmente per tre ordini di ragioni. In primo luogo, l'espressione "intelligenza artificiale" è, per i fini che mi sono posto, inappropriata. Con "intelligenza artificiale", infatti, secondo McCarthy – colui che ha coniato tale espressione – ci si riferisce agli studi che si dedicano al raggiungimento del livello di intelligenza umana (McCarthy, *The Lighthill Debate*, 1973). Gli agenti che prenderò in esame, però, non sono necessariamente orientati al raggiungimento del livello di intelligenza umana, ma possono porsi su un altro piano. La seconda ragione è di carattere evocativo; riguarda, cioè, le immagini che una certa espressione è in grado di richiamare. L'espressione "intelligenza artificiale" richiama alla mente l'idea di un fenomeno astratto, quasi etereo, sicuramente impersonale. Con "macchine", invece, provo a dare materialità e, quindi, soggettività agli agenti non biologici. Infine, la parola "intelligenza" presenta dei profili di ambiguità considerevoli.<sup>14</sup>

Parlerò, dunque, di macchine (e non di intelligenza artificiale) perché pur trattandosi di una denominazione meno specifica, essa è, allo stesso tempo, più appropriata e più chiara. Inoltre, la parola "macchina" designa un'entità dai contorni ben definiti – evocandone la materialità – che prende la sostanza di un soggetto agente. Una macchina è, dunque, in primo luogo un agente.

Esistono diversi tipi di macchine ed esistono altrettanti modi per classificarle. Se le macchine sono agenti, bisogna cominciare con la definizione di "agente". Secondo Luc Steels (1995), un agente si caratterizza per i seguenti elementi:

---

<sup>14</sup> Tali profili verranno presi in esame al paragrafo 2.2.2.

- è un sistema (un insieme di elementi che hanno relazioni fra loro e con l'ambiente);<sup>15</sup>
- performa particolari funzioni per un altro sistema;
- è capace di preservare se stesso.

Altri autori rifuggono definizioni così conclusive di “agente”. In particolare, Russell e Norvig (2016, p. 33) accordano un valore strumentale al concetto di agente: esso serve ad analizzare i sistemi e non è una categoria assoluta, in grado di dividere il mondo in "agenti" e "non-agenti". Un'altra questione aperta è quella relativa alle qualità dell'agente: un agente deve necessariamente essere autonomo? Luck e D'Inverno (1995), dopo aver definito il concetto di agente, identificano l'agente autonomo come una sottoclasse dell'agente. Altri autori (Franklin e Graesser, 1995), invece, utilizzano “agente” come abbreviazione di “agente autonomo”.<sup>16</sup> Non entrerò nel merito di questa disputa. Per ora, è sufficiente osservare che gli agenti di cui mi occuperò sono agenti autonomi. Quando, quindi, un agente è autonomo?

### **2.2.1 Il criterio dell'autonomia**

Secondo Steels (1995), “autonomia” significa, in sostanza, “essere relativamente indipendente da qualcosa”. Per Smithers (1992), invece, il concetto di autonomia è da rintracciare nell'etimologia della parola che è composta dai lemmi “*autos*” e “*nomos*”. Tale concetto venne inizialmente adottato per descrivere le città-stato greche, nelle quali i cittadini producevano le loro stesse leggi – in contrapposizione al potere esercitato da un organo esterno.

I sistemi autonomi non vanno confusi con i sistemi automatici (Smithers, 1992). I sistemi automatici sono quei sistemi che sono in grado di regolarsi da soli. Tuttavia, essi non generano le leggi cui la loro attività regolativa cerca di conformarsi, ma si

---

<sup>15</sup> Steels (1995) osserva che gli agenti non debbano essere necessariamente fisicamente stanziati.

<sup>16</sup> Anche Floridi e Sanders (2004), adottando la definizione di agente di Allen, Varner e Zinser (2000), ritengono che un agente è tale se è autonomo.



limitano ad orientarsi all'interno di un sentiero che è già dato, correggendo e compensando gli effetti di perturbazioni esterne. Dall'altro lato, i sistemi autonomi sono quei sistemi in grado di sviluppare le proprie leggi e le proprie strategie. Essi, oltre a regolarsi da soli, si auto-governano: determinano, cioè, le strade che seguiranno, così come determinano il modo di procedere entro di esse.<sup>17</sup>

Steels (1995) condivide l'idea che, affinché un sistema possa essere considerato autonomo, deve necessariamente essere automatico – e sottoscrive la distinzione di Smithers (1992) fra agenti automatici e agenti autonomi. Questa non è l'unica classificazione presente in letteratura. Per Hildebrandt (2019, p. 10), oltre agli agenti automatici (la cui condotta è completamente predefinita) e agli agenti autonomi (che sono sia coscienti che autocoscienti e, quindi, capaci di auto-riflessione, azioni intenzionali e sviluppo di desideri di second'ordine), esistono, a metà strada, gli agenti autonomici – che corrispondo agli agenti automatici di Steels – capaci di «*self-management, self-repair, self-configuration*». Parzialmente coincidente con la tripartizione di Hildebrandt, vi è un'altra classificazione, sempre proposta da Smithers, ma in un altro scritto (1997). Egli isola tre tipi di sistemi – automatici, cibernetici e autonomi. Mentre un sistema automatico è un sistema che si muove da solo (dal greco, “autos” e “matos”), un sistema cibernetico è un sistema che si auto-regola e che, oltre a muoversi da solo, è capace di regolare e controllare i propri movimenti e le proprie azioni in modo da preservare la loro efficacia di fronte a disturbi e perturbazioni, secondo una legge di controllo o una legge di regolazione predefinita. Un sistema autonomo, invece, si differenzia da un sistema cibernetico, in quanto genera da sé le proprie leggi, regolando i propri comportamenti in base alle leggi che esso stesso si è dato.<sup>18</sup> È pacifico che una macchina possa essere automatica e cibernetica. Ma essa può davvero essere autonoma?

---

<sup>17</sup> Si tratta di alcune considerazioni di Tim Smithers del 1992 estratte da Steels (1995). Si vedrà come lo stesso Smithers in un secondo momento (1997) adotti una tripartizione.

<sup>18</sup> Smithers (1997) nella sua tripartizione esclude che un sistema biologico possa essere cibernetico. In questo modo, entra in conflitto con uno degli assiomi della cibernetica, cioè l'equiparazione dell'animale alla macchina.

Il problema dell'autonomia della macchina può essere letto come una variazione del *frame problem* (McCarthy e Hayes, 1981): è capace una macchina di affrontare ambienti dinamici, variabili ed imprevedibili? Mentre un sistema cibernetico non riuscirebbe in questa impresa, per i sistemi autonomi non sarebbe così:

Al contrario, i sistemi autonomi, formando continuamente le loro leggi di funzionamento (la loro identità), hanno, come proprietà intrinseca, la capacità di affrontare nuove situazioni: se cambiano le condizioni per questo continuo processo di formazione, cambiano le leggi di regolazione così formate, e cambiano in modo tale da mantenere il sistema autonomo naturalmente ben adattato alla sua situazione attuale. C'è un senso in cui tutto è sempre nuovo, quindi le situazioni nuove, senza la necessità che già vi sia un'esperienza previa, non sono un problema: le situazioni nuove sono normali. Questa continua e costante formazione dell'identità attraverso l'interazione si traduce in quello che Maturana e Varela chiamano un accoppiamento strutturale tra l'agente e il suo ambiente: l'organizzazione interna dell'agente riflette la struttura e l'organizzazione dell'ambiente con cui interagisce (Smithers, 1997, p. 11).<sup>19</sup>

Per Smithers, dunque, il concetto di autonomia è strettamente legato a quello di autopoiesi<sup>20</sup> ed è proprio l'autonomia che permette all'agente di affrontare situazioni nuove.

Un'altra caratteristica dei sistemi autonomi sarebbe la capacità di auto-controllarsi ovvero la capacità di non essere soggetti al controllo di un agente esterno (McFarland, 1994; Franklin e Graesser, 1995; McFarland e Bösner, 1993, pp. 145-147).

Un modo di caratterizzare il concetto di autonomia in maniera più robusta è quello di definire autonomo quell'agente capace di figurarsi tutte le conseguenze delle sue azioni a cui è soggetto pur rimanendo vitale come agente che svolge un compito nel mondo in cui opera (Smithers, 2018).

---

<sup>19</sup> Traduzione mia da Smithers (1997).

<sup>20</sup> Si tratta di un'associazione già in parte esplorata nel primo capitolo.

Bisogna riannodare il filo: affinché una macchina possa prendere decisioni, occorre che essa sia autonoma. Ma una macchina può essere considerata autonoma?

Innanzitutto, bisogna fissare la definizione di autonomia. Non si tratta di un compito facile. Questa definizione deve enumerare le funzioni che un agente autonomo può compiere? Oppure si tratta di un elenco delle sue qualità?<sup>21</sup> Si tratta, poi, di una definizione gradualista o assoluta? In altre parole, un certo agente è autonomo in termini relativi o, se è autonomo, lo è in termini assoluti?<sup>22</sup>

Quelle viste finora sono solo alcune delle definizioni di autonomia. Altre definizioni, in parte coincidenti, in parte discordanti, riguardano altri ambiti del sapere, come la biologia, il diritto, la filosofia e l'etica.<sup>23</sup> In particolare, ciò che accomuna le definizioni di autonomia nelle diverse discipline è la possibilità di «*self-law making*» (Smithers, 1997).

Probabilmente, è opportuno partire da una definizione minima di autonomia: un agente è autonomo se è relativamente indipendente da qualcos'altro. Ma esattamente da cosa dovrebbe essere indipendente l'agente-macchina? L'indipendenza è relativa al fattore umano: una macchina dev'essere indipendente dal suo programmatore, nei termini in cui la sua attività non è interamente predeterminata. Questa indipendenza si

---

<sup>21</sup> A metà strada fra un approccio funzionalistico e ontologico, c'è la concezione di Boden (1990), secondo cui, quando una certa entità è autonoma, bisogna domandarsi se i suoi meccanismi di direzione del comportamento (a) possano essere modellati dalla storia esperienziale dell'entità, (b) siano di natura emergente e (c) siano modificabili in modo riflessivo dall'entità stessa.

<sup>22</sup> Secondo una prospettiva più relativista (Covrigaru e Lindsay, 1991), l'autonomia non sarebbe la proprietà di un agente, ma piuttosto un attributo dato a un agente da un osservatore.

<sup>23</sup> A titolo esemplificativo, in filosofia morale, il concetto di morale è strettamente connesso a quello di autonomia: per Kant, Kierkegaard, Nietzsche, Royce, Hare, Popper, Sarte, and Wolff, un agente morale è necessariamente un agente autonomo (Dworkin, 1988, cap. 3). Per quanto riguarda la biologia, secondo la tradizione cibernetica, come si è già visto nel primo capitolo, il concetto di autonomia è strettamente connesso a quello di autopoiesi (Maturana e Varela, 1973; Varela, 1979; Maturana e Varela, 1992). Più precisamente, l'autopoiesi è vista come il meccanismo che permette a un sistema di essere autonomo.

declina nella capacità di definire da sé le proprie leggi e non, semplicemente, di applicare leggi date.

Non si tratta, dunque, di un'indipendenza – e di un'autonomia – assoluta. Gli agenti autonomi sono inseriti all'interno di un determinato ambiente e, se cambia l'ambiente, si potrebbe non avere più un agente autonomo (Franklin e Graesser, 1995). Tutto ciò porta all'idea che l'autonomia sia una questione di “gradi”, anziché un fenomeno binario – *on/off* – (Wheeler, 2020, p. 348) e che riguardi diverse dimensioni (Boden, 1990, p. 102).

A questo punto, è utile fissare delle condizioni – che hanno il ruolo di “standard minimo” – per stabilire se – e in che grado – una macchina può dirsi autonoma. Altre condizioni – come il requisito della coscienza (Chrisley, 2008) o l'esistenza di desideri e di credenze di secondo ordine (Frankfurt, 1988) – verranno lasciate sullo sfondo e trattate con riferimento all'ipotesi che una macchina venga considerata un agente morale, nell'ultimo capitolo. In base a quanto si è visto finora, una macchina è autonoma nella misura in cui prende le distanze dal suo programmatore e prende le distanze dal suo programmatore se è in grado di costruire da sé le leggi del suo funzionamento. Come si manifesta questa autonomia? Due fenomeni possono essere letti come un indice di autonomia: relativa imprevedibilità di comportamento e relativa impossibilità di spiegare i suoi comportamenti a posteriori.

In conclusione, si tratta di una definizione operativa e gradualista, volta a stabilire quando una macchina può dirsi autonoma – e in che misura.

### **2.2.2 Il criterio dell'intelligenza**

È stato sostenuto che un ente per essere autonomo deve essere autenticamente intelligente (Covrigaru e Lindsay, 1991). Quello dell'intelligenza è il secondo criterio che prenderò in esame.

Ho già anticipato al paragrafo 2.2 che non è un'operazione particolarmente semplice definire il concetto di intelligenza.<sup>24</sup> In riferimento all'IA, il primo insieme di definizioni è orientato a spiegare l'intelligenza nei termini di *performance* comparativa rispetto all'intelligenza umana (Steels, 1995). Il caso più famoso è quello del test di Turing (1950), in cui, come è noto, il test si intende superato se la macchina riesce a far credere all'interrogante umano di essere un umano. È stato osservato che il problema di questo insieme di definizioni è che, anche nel caso in cui vi siano programmi altamente complessi, le cui *performance* sono indistinguibili da quelle umane in settori specifici – ad esempio, nel gioco degli scacchi –, non è possibile stabilire che programmi del genere colgano l'evoluzione o la natura incorporata dell'intelligenza (Clancey, 1993). Accade, dunque, che programmi "intelligenti" vengono considerati non più intelligenti non appena la persona che ispeziona il programma capisce come il problema è stato risolto. Nel caso degli scacchi, ad esempio, si tratta di programmi che effettuano indagini profonde nello spazio di ricerca e le prestazioni impressionanti non sono quindi più considerate come un'espressione dell'intelligenza (Steels, 1995), ma sono effetto della possibilità di una macchina di attingere da una vasta memoria.

La seconda famiglia di definizioni è spiegata nei termini di conoscenza e intenzionalità. L'obiettivo è quello di lavorare sulla nozione di descrizione del livello di conoscenza. Per Newell (1981), un sistema può dirsi intelligente se il suo comportamento può essere descritto come quello di un soggetto che detiene conoscenza e che applica questa conoscenza quando la situazione lo richiede. L'intelligenza di una macchina è, quindi, legata alla sua capacità di estrarre

---

<sup>24</sup> Sulla difficoltà di definire il concetto di intelligenza, a titolo esemplificativo, si veda Minsky (2020) che sostiene: «la nostra mente contiene processi che ci consentono di risolvere problemi che consideriamo difficili. "Intelligenza" è il nome che diamo a tutti quei processi che ancora non comprendiamo. Ad alcuni questa "definizione" non piace perché il suo significato è destinato a mutare via via che impariamo cose nuove in ambito psicologico. Ma secondo me va benissimo così, perché il concetto stesso di intelligenza è come il trucco di un prestigiatore. È come il concetto di "regioni inesplorate dell'Africa": scompare non appena scopriamo che cosa essa è».

conoscenza, di formalizzarla e di codificarla. Quest'approccio appare problematico per due ordini di ragioni (Steels, 1995):

a) perché la descrizione del livello di conoscenza può essere impiegata in riferimento a diverse entità – come, ad esempio, a un termostato – a cui l'etichetta “intelligente” non è naturalmente – o intuitivamente – applicabile;

b) perché si assume esista una netta discontinuità fra sistemi intelligenti e sistemi non intelligenti, e ciò non contribuisce a spiegare come l'intelligenza si sia sviluppata nei sistemi fisici né come la conoscenza e il ragionamento siano connessi alla neurofisiologia.

Esistono poi altre definizioni di intelligenza, che non sono state direttamente impiegate nel campo degli studi sull'intelligenza artificiale. Per Penrose (1991), l'intelligenza è intimamente legata alla coscienza e all'auto-coscienza. A sua volta, la coscienza è definita come la capacità di intuire verità matematiche o di dare giudizi estetici. Pur rimandando la questione sulla coscienza all'ultimo capitolo, bisogna per ora osservare che, in assenza di una definizione chiara di coscienza, sostenere che “l'intelligenza è legata alla coscienza” è un'affermazione fondamentalmente vaga.

Per Steels (1995), tutte queste definizioni di intelligenza appaiono insoddisfacenti. Egli vede l'intelligenza come una proprietà degli agenti autonomi – ovvero di quei sistemi che devono mantenere se stessi e preservare la loro struttura e il loro funzionamento, mentre rimangono vivi. Si tratta, in altre parole, della capacità di un agente autonomo di sopravvivere. Ma è sufficiente la capacità di sopravvivere per poter parlare di intelligenza? Per Steels, la chiave di volta per uscire dall'*impasse* è la rappresentazione. Egli dà una definizione molto larga di rappresentazione: le rappresentazioni sono strutture fisiche che creano correlazioni con aspetti dell'ambiente e, quindi, esse rivelano un potere predittivo per il sistema. A loro volta, le rappresentazioni supportano processi che influenzano il comportamento. Ciò che rende le rappresentazioni uniche è che i processi che operano sulle rappresentazioni possono disporre di proprie dinamiche che sono indipendenti dalle dinamiche del mondo che essi rappresentano. Steels procede a chiarire tali affermazioni attraverso un esempio. Si tratta di un'operazione di raffronto fra due sistemi di controllo.

Entrambi i sistemi devono aprire una valvola quando la temperatura va oltre una soglia determinata. Il sistema A è composto da un'asta metallica che si espande quando la temperatura sale ed espandendosi apre la valvola, mentre, quando la temperatura si abbassa, l'asta metallica si riduce e la valvola si chiude. In questo primo caso, non ci sarebbe rappresentazione poiché la funzione di controllo è implementata solo attraverso un processo fisico. Il sistema B, al contrario, dispone di un sensore che registra la temperatura: esso converte la temperatura in una rappresentazione della temperatura. In questo secondo caso, vi è una chiara rappresentazione e, conseguentemente, si dà un processo che, operando sulla rappresentazione, si adatta in maniera flessibile.

Una tesi del genere presta il fianco a varie critiche, alcune di queste provengono dagli studi sull'*embodied cognition* (Varela, Thompson e Rosch, 2017; Mingers, 1995; Mingers, 2001). L'argomento di Steels, infatti, ponendo una distinzione così netta tra cognizione e azione, sembra non tenere in considerazione il fatto che ogni processo cognitivo nasce sempre a partire da un processo fisico – come, nel caso del cervello umano, il pensiero prende forma a partire dall'attività neurale. È stato così osservato che può esservi intelligenza anche senza rappresentazione (Brooks, 2018) e che anche i sistemi di controllo non-rappresentazionali possono essere adeguati modelli di cognizione (Van Gelder, 1995). Le ricerche sulle reti neurali sembrano andare in questa direzione: rigettano le rappresentazioni simboliche in favore di processi sub-simbolici o non simbolici (Smolensky, 1986).

In definitiva, anche il paradigma rappresentazionalista non sembra adeguato a descrivere il fenomeno dell'intelligenza.

Vi sono, dunque, nel cercare di definire l'intelligenza di entità non biologiche almeno tre tendenze:

- (a) considerare intelligente la macchina che manifesta un comportamento simile a quello umano;
- (b) considerare l'intelligenza artificiale non assimilabile a quella umana, ricorrendo a concetti – come coscienza – la cui definizione è vaga;

(c) considerare intelligente quella macchina il cui funzionamento non è riducibile a un mero processo fisico.

Se seguendo (a) si rischia di considerare intelligente una macchina che si limita a imitare il comportamento umano, (b) esclude a priori che una macchina possa essere intelligente, ricorrendo a concetti non ben definiti. (c), invece, ignora che i processi cognitivi sono sempre riducibili a processi fisici.

In assenza di una definizione puntuale, bisogna concludere che quello dell'intelligenza non è un criterio che, in questo contesto, può essere utilmente adottato per descrivere una macchina. Quindi, a tale criterio bisogna rinunciare. Necessario e sufficiente – ai fini di questa ricerca – sarà il criterio dell'autonomia. Le macchine capaci di prendere decisioni devono, fondamentalmente, essere macchine autonome, dove per “autonomia” si intende la loro indipendenza (relativa) dal fattore umano.

### **2.3 Concetti “ponte”. Per un'analisi congiunta degli apporti umani e artificiali**

Esiste un altro modo per guardare al rapporto umano/macchina. Si tratta di un approccio meno dicotomico, o, se si vuole, dicotomico ma conciliatorio. Secondo questo approccio, “fattore umano” e “fattore artificiale” si compenetrano a tal punto che non è utile né corretto isolare questi due elementi, ma occorre un'analisi congiunta. Esistono (almeno) tre campi di studio che si occupano di questa analisi congiunta. Si tratta degli studi su:

- 1) intelligenza ibrida;
- 2) sistemi socio-tecnici;
- 3) macchine politiche.

#### **2.3.1 Intelligenza ibrida**



L'intelligenza ibrida è intesa come la combinazione di intelligenza umana e intelligenza della macchina, nella prospettiva di un aumento dell'intelletto e delle capacità umane, e non con il fine di rimpiazzarle. Si tratterebbe, inoltre, di un modo per prendere decisioni sensate, eseguire azioni appropriate e raggiungere obiettivi che non sono raggiungibili da umani e macchine separatamente (Akata et al., 2020).

Gli studi sull'intelligenza ibrida, dunque, anziché concentrarsi sull'uso dell'IA per rimpiazzare l'intelligenza umana, indagano la maniera di integrare l'intelligenza umana e quella artificiale col fine di creare sistemi ibridi umano-artificiali (Chen, Ning, Nugent e Yu, 2020; Dekker e Woods, 2002; Sankar, 2021). Questi approcci non vedono ai sistemi di IA come a macchine pensanti, ma come a protesi cognitive che possono aiutare gli esseri umani a pensare e agire meglio (Guszcza, 2018). L'obiettivo principale degli studi sull'intelligenza ibrida è quello di costruire sistemi adattivi intelligenti capaci di aumentare – non di sostituire – l'intelligenza umana, valorizzando i punti di forza umani e compensandone le debolezze, tenendo in considerazione questioni etiche, legali e sociali (Akata et al., 2020). Si tratta di un programma ambizioso e ottimistico che minimizza – o ignora – il sorgere di possibili tensioni fra macchine e umani. Nella visione irenica degli studi sull'intelligenza ibrida, la macchina è intesa come uno strumento al servizio dell'intelletto umano per migliorarne le *performance*.

Ma in che modo l'intelligenza umana e quella artificiale interagiscono?

Coerentemente all'approccio strumentale degli studi sull'intelligenza ibrida, adotterò una tassonomia (Pescetelli, 2021) che si basa sul tipo di apporto che gli algoritmi possono dare agli umani, in un crescendo di complessità.

Sono stati, dunque, individuati quattro ruoli per gli algoritmi:

- algoritmo come assistente;
- algoritmo come pari;
- algoritmo come facilitatore;
- algoritmo come operatore a livello di sistema.

La funzione di (1) è quella di ridurre il carico umano cognitivo, cercando, organizzando, presentando, personalizzando e sintetizzando informazioni rilevanti per

il compito prestabilito. L'esempio classico di (1) è quello degli assistenti digitali come Alexa e Cortana.

Il compito di (2), invece, è quello di agire come un umano, giocando con altre persone in un gruppo di *problem-solvers*. Per gli esempi di (2), si vedano i giocatori di Go (Silver et al., 2016) e Atari (Mnih et al. 2015) o, più in generale, il fenomeno della sostituzione e del trasferimento di posti di lavoro a causa dell'automazione (Brynjolfsson e McAfee, 2014; Frank et al., 2019). Si tratta di algoritmi pensati per imitare l'essere umano, svolgendo compiti che coincidono con quelli che di solito gli umani compiono. Tali agenti possono, dunque, partecipare al pari di agenti umani in team ibridi, col fine di migliorare i processi decisionali (Sankar, 2021; Licklider 1960).

Un ruolo più sofisticato sembra essere quello di (3). In questo caso, l'algoritmo cerca attivamente di eliminare i *bias* in modo da migliorare le *performance* degli attori umani e algoritmici. Perché un algoritmo giochi questo ruolo è necessario che esso abbia una certa conoscenza di *bias*, euristiche e comportamento umano. Se (2) agisce indipendentemente dagli agenti umani, (3) agisce in funzione del comportamento delle persone. (3) ha, dunque, il compito di correggere le dinamiche distorte del ragionamento umano, sulla base all'assunto che le funzioni cognitive umane siano caratterizzate da *bias* (Gigerenzer, 2008).

(4) ambisce ad avere un ruolo sempre più importante in futuro. Un algoritmo viene definito (4) se «*the algorithm in question can observe the full information–task–people–AI system and the reciprocal interactions between its elements*» (Pescetelli, 2021). Un sistema del genere può rivelarsi utile nel ricombinare gli input provenienti dagli agenti umani e nel prendere decisioni sulla base di questi input.

Si può osservare come sia radicata la convinzione che il ragionamento umano sia naturalmente soggetto a *bias* e come sia altrettanto radicata l'idea che le macchine possano rappresentare uno strumento di correzione del ragionamento. Inoltre, da un punto di vista funzionale, vi è un'analogia fra (2) e (4) e i due ruoli che – come assunto nell'ultimo capitolo – una macchina può giocare all'interno di un processo decisionale – ovvero come attore, al pari di un agente umano [(2)] e come processo

decisionale stesso [(4)]. Secondo l'approccio adottato dagli studi sull'intelligenza ibrida, il rapporto umano-macchina non è inteso in termini conflittuali, in primo luogo perché è negata alle macchine una natura autonoma.

### **2.3.2 Sistemi socio-tecnici**

Ricorrere a un'analisi socio-tecnica di un sistema significa usare un approccio che oltre ai fattori umani e sociali, nella costruzione dei sistemi organizzativi, contempla gli elementi tecnici (Gordon e Sommerville, 2010). Si tratta, dunque, di un approccio volto a considerare tanto gli aspetti tecnici quanto quelli sociali nello studio di alcuni fenomeni.

La nozione di sistema socio-tecnico è stata introdotta nel contesto dei “*labor studies*” al *Tavistock Institute* da Emery e Trist nel 1960. Cinque sarebbero gli elementi che caratterizzano i sistemi socio-tecnici (Badham, Clegg e Wall, 2000):

- i sistemi si compongono di parti interdipendenti;
- i sistemi devono adattarsi e perseguire obiettivi in un ambiente esterno;
- i sistemi sociali contano di un ambiente interno che comprende sotto-sistemi tecnici e sociali separati, ma interdipendenti;
- gli obiettivi dei sistemi possono essere raggiunti con più di un solo mezzo (questo implica che ci sono scelte progettuali da fare durante lo sviluppo di un sistema);
- le prestazioni del sistema si basano sull'ottimizzazione congiunta dei sottosistemi tecnici e sociali (se ci si concentra su uno di questi sistemi escludendo l'altro, è probabile che le prestazioni e l'utilità del sistema si riducano).

Dopo aver elencato gli elementi distintivi dei sistemi socio-tecnici, all'ultimo punto viene indicata la ragione per cui bisognerebbe ricorrere a tale approccio: se si esclude l'aspetto tecnico o quello sociale, le prestazioni del sistema potrebbero subire un pregiudizio.

Nel tentativo di valutare alcune implicazioni del concetto di sistema socio-tecnico è stato ragionevolmente osservato (Ropohl, 1999) che:

(a) per quanto riguarda lo sviluppo tecnico, esso non è da considerarsi né come un processo autonomo – con leggi proprie – né come il risultato di singole e isolate azioni individuali. E non bisogna pensare allo sviluppo tecnico neanche nei termini di un destino tecnico universale (Schelsky, 1961) o come la scoperta di idee tecniche che si erano già formate in un paradiso platonico (Dessauer, 1956). Lo sviluppo tecnico sarebbe, quindi, un processo sociale: scienza e tecnologia sono condizioni necessarie di tale processo, ma non sono affatto sufficienti per determinare il suo andamento. Per questa ragione, una filosofia della tecnologia isolata non ha molto senso;

(b) in merito ai cambiamenti sociali, un approccio socio-tecnico si rivela utile per spiegare l'impatto della tecnologia sulla società. Un'invenzione non è solo un nuovo artefatto, ma anche un nuovo modello di azione umana. Dunque, ogni nuova tecnologia ha delle conseguenze nella natura e nella società.

Si tratta di riflessioni senz'altro condivisibili e che suggeriscono di studiare tanto i fenomeni sociali quanto i nuovi strumenti tecnologici alla luce delle influenze reciproche. Tuttavia, al di là di queste considerazioni di buon senso, il cuore di tale teoria, come nel caso degli studi sull'intelligenza ibrida, è di non adottare un'analisi disgiunta nello studio di alcuni fenomeni. Con riferimento ai processi decisionali in un contesto democratico questa teoria ha qualcosa da dirci? È chiaro di sì. Tuttavia, quello che mi sforzerò di fare nell'ultimo paragrafo di questo capitolo è di dare spessore e rilievo all'autonomia delle macchine e, in ragione di questa autonomia, isolare l'attività della macchina da quella umana.

### **2.3.3 Macchine politiche**

Il concetto di “macchine politiche” è stato introdotto per affrontare lo studio delle politiche nelle macchine sociali (Papakyriakopoulos, 2021) dove per macchine sociali

si intendono quegli ecosistemi in cui individui ed algoritmi partecipano e interagiscono (Shadbolt, O'Hara, De Roure e Hall, 2019). Si tratta, come si è già visto nel caso degli studi sull'intelligenza ibrida e dell'approccio socio-tecnico, di un *framework* per l'analisi di quei fenomeni in cui vi è una partecipazione tanto delle macchine quanto umana. L'ipotesi di partenza è che la tecnologia non partecipa in maniera neutra nelle questioni sociali: al contrario, le implementazioni algoritmiche trasformano radicalmente e in modi inaspettati i processi socio-politici (Papakyriakopoulos, 2021). Per comprendere la natura di queste interazioni, viene fornita una classificazione di cinque tipi di processi di influenza.<sup>25</sup>

La prima forma di influenza è definita simbolica. Tale influenza è la forma di potere dei simboli: essi orientano la realtà sociale e il comportamento. Ma di che simboli si parla? Un esempio è quello della struttura dell'interfaccia-utente che, indirizzando il comportamento degli utenti, può sempre influenzare coloro che sono parte di una macchina politica. Si tratterebbe del tipo di influenza più sottile e penetrante e la portata del suo impatto si rivela difficile da quantificare. Un altro esempio di questo tipo di influenza è l'inclusione esplicita di un testo per l'adesione ai termini e alle condizioni di una piattaforma: tale operazione ha il potenziale per cambiare la decisione di un utente che utilizza quel servizio.

Il secondo tipo di influenza è definito della “condotta politica”. Viene considerata la forma di interazione – e influenza – più semplice. Si verifica, ad esempio, quando i politici utilizzano alcuni strumenti tecnologici per accrescere il loro potere o quando i processi democratici esplicitamente si realizzano grazie a tali strumenti.

La terza forma di influenza è quella algoritmica. Questo tipo di influenza può verificarsi sia in contesti politici che non politici. Esempi di questa forma di influenza possono vedersi sui *social media*, in cui alcuni algoritmi sono progettati per:

- offrire contenuti personalizzati agli utenti;
- proporre pubblicità mirata;
- filtrare ed esaminare i contenuti generati dagli utenti.

---

<sup>25</sup> Adotterò la classificazione di Papakyriakopoulos (2021).

Con riferimento a questo tipo di interventi, vengono spesso sollevate preoccupazioni in merito alle capacità predittive, all'accuratezza e alla loro generale adeguatezza nel fornire prove attendibili per una decisione (Mittelstadt, Allo, Taddeo, Wachter e Floridi, 2016)

Il quarto tipo di influenza è legata alla progettazione dei sistemi. Un esempio di questo tipo di influenza è dato dal modo in cui sono progettati i sistemi di credito sociale: essi orientano il comportamento dei cittadini, definendo il loro spazio d'azione praticabile e i loro obiettivi sociali (Pasquale, 2015; Engelmann, Chen, Fischer, Kao e Grossklags, 2019).

L'ultimo tipo di influenza riguarda il piano normativo entro cui le macchine politiche si inseriscono. Detto altrimenti, sono le strutture giuridico-politiche a stabilire come le nuove tecnologie debbano essere regolamentate. In proposito, le principali questioni normative sulla regolamentazione degli algoritmi riguardano la proprietà dei dati e la privacy (Bertot, Jaeger, Munson e Glaisyer, 2010), l'opacità degli algoritmi (Burrell, 2016) e la discriminazione dei gruppi sociali (Zarsky, 2016).

Il *framework* delle *political machines* non permette tanto di definire con esattezza le influenze – influenze di che tipo? chi influenza cosa? – quanto di vedere i punti di contatto fra le nuove tecnologie e le questioni politiche. Il focus è più che altro rivolto ai *social media* e al modo in cui in essi si forma l'opinione, mentre poca attenzione viene rivolta alla possibilità di partecipazione di una macchina in un processo decisionale. In secondo luogo, nonostante sia condivisibile l'assunto per cui nessuna tecnologia sia neutra – ma sia portatrice di effetti trasformativi per tutta la società –, è necessario riuscire a tenere ben distinto il piano dello strumento tecnologico in sé dagli effetti contingenti che esso potrebbe generare.

Questi tre approcci – intelligenza ibrida, sistemi socio-tecnici, macchine politiche – considerano correttamente il rapporto fra umano e macchina in termini relazionali. Tuttavia, ai fini di questa ricerca, è opportuno non insistere sulla supposta natura promiscua del rapporto umano-macchina, ma è preferibile isolare e definire precisamente la natura della macchina al fine di verificare, in seconda battuta, la

compatibilità fra le macchine e la democrazia – intesa come metodo per prendere decisioni.

## **2.4 Per un'analisi disgiunta**

Si è già visto che l'unico criterio che ho ritenuto utile per caratterizzare una macchina è quello dell'autonomia. Obiettivo di questo paragrafo è specificare questo assunto. In che modo, cioè, una macchina può diventare autonoma? Lo strumento attraverso il quale una macchina si allontana dal suo programmatore umano è quello del *machine learning* e, più precisamente, di alcune particolari forme di *machine learning*.

### **2.4.1 Il *machine learning***

Vi è spesso confusione fra intelligenza artificiale e *machine learning*.

Tecnicamente, il *machine learning* è solo una branca dell'intelligenza artificiale, anche se è la branca che ha reso possibile la crescita esponenziale delle applicazioni dell'intelligenza artificiale.

Il *machine learning* è la disciplina che si focalizza su due questioni relate: (a) come può costruirsi un sistema computazionale che impara automaticamente attraverso l'esperienza? (b) quali sono le fondamentali leggi statistiche-computazionali-informazionali-teoretiche che governano tutti i sistemi di apprendimento, includendo fra essi computer, umani e organizzazioni? (Jordan e Mitchell, 2015). Il *machine learning* è, quindi, definito come quell'insieme di metodi in grado di rilevare automaticamente i *pattern* nei dati e di utilizzare gli schemi scoperti per prevedere i dati futuri o per eseguire altri tipi di processi decisionali in una situazione di incertezza (Murphy, 2013). Da un'altra prospettiva, una macchina

può qualificarsi come una macchina che apprende se le sue *performance* migliorano con l'esperienza (Mitchell, 1997; Cohen e Feigenbaum, 1989).

Solitamente, il *machine learning* viene suddiviso in tre macro-categorie (Murphy, 2013; Alpaydin, 2014; Russell e Norvig, 2016):

- apprendimento supervisionato;
- apprendimento non supervisionato;
- apprendimento “per rinforzo”.

A questa classificazione, se ne aggiunge un'altra, più utile ai fini di questa ricerca:

Il mondo del *machine learning* è diviso in scuole di pensiero concorrenti che offrono risposte molto diverse. Esistono cinque correnti principali (...). Per i simbolisti, che si ispirano alla filosofia, alla psicologia e alla logica, l'apprendimento è l'inverso della deduzione. I connessionisti prendono spunto dalle neuroscienze e dalla fisica per compiere un'operazione di *reverse engineering* del cervello. Gli evoluzionisti si ispirano alla genetica e alla biologia evolutiva per realizzare simulazioni numeriche dell'evoluzione. I bayesiani fondano le proprie tesi sulla statistica, convinti che l'apprendimento sia una forma di inferenza probabilistica. Gli analogisti, infine, sono influenzati dalla psicologia e dall'ottimizzazione matematica, e ritengono che si impari da estrapolazioni basate su criteri di somiglianza (Domingos, 2016).

Più precisamente, Domingos (2016, cap. II) definisce le scuole di pensiero in questi termini:

Per i simbolisti, l'intelligenza può essere ridotta integralmente alla manipolazione di simboli del tutto simile a quella con cui i matematici risolvono le equazioni sostituendo espressioni ad altre espressioni. I simbolisti sono consci dell'impossibilità di imparare partendo da zero: bisogna affiancare ai dati un minimo di conoscenze iniziali. Hanno imparato a incorporare nel processo di apprendimento le conoscenze preesistenti, e a combinare al volo porzioni di sapere distinte per poter risolvere problemi nuovi. Il loro Algoritmo Definitivo è la deduzione inversa, che dopo aver identificato la conoscenza



che ancora manca per poter completare una deduzione, conferisce a quest'ultima il carattere più generale possibile.

Per i connessionisti, l'apprendimento è ciò che fa il cervello: il nostro compito, quindi, è il suo reverse engineering. Il cervello apprende regolando la forza delle connessioni tra i neuroni; il problema principale è capire quali sono le connessioni responsabili di determinati errori per poterle modificare di conseguenza. L'Algoritmo Definitivo dei connessionisti è la retropropagazione, il processo che dopo aver confrontato l'output reale di un sistema con quello desiderato modifica le connessioni tra i neuroni, uno strato dopo l'altro, per ridurre la differenza.

Gli evoluzionisti credono che la madre di ogni forma di apprendimento sia la selezione naturale. Se è stata capace di creare noi può creare qualsiasi cosa, e quindi non dobbiamo fare altro che simularla al computer. Il problema fondamentale da risolvere, per gli evoluzionisti, è l'apprendimento della struttura: più che definire il valore esatto di un insieme di parametri, come nel caso della retropropagazione, ciò che interessa è la creazione del cervello, che potrà essere ottimizzato successivamente modificando quegli stessi parametri. L'Algoritmo Definitivo degli evoluzionisti è la programmazione genetica, che fa accoppiare ed evolvere i programmi di un computer nello stesso modo in cui la natura fa accoppiare ed evolvere gli organismi.

Ai bayesiani interessa soprattutto l'incertezza. Ogni forma di conoscenza acquisita è affetta da un'incertezza, e l'apprendimento stesso è una forma di inferenza associata a un'incertezza. Il problema, in questo caso, è trovare un modo per gestire un'informazione incompleta, affetta da rumore di fondo e a volte persino contraddittoria, senza esserne travolti. La soluzione è l'inferenza probabilistica, e l'Algoritmo Definitivo è costituito dal teorema di Bayes e dai suoi corollari. Il teorema di Bayes ci dice come integrare nuovi dati nel nostro sistema di credenze, e gli algoritmi di inferenza probabilistica lo fanno nel modo più efficiente possibile.

Per gli analogisti, l'ingrediente fondamentale dell'apprendimento è il riconoscimento delle situazioni simili, per poterne inferire altre somiglianze. Se due pazienti hanno sintomi simili, forse hanno la stessa malattia. Il problema fondamentale è valutare il grado di somiglianza. L'Algoritmo Definitivo degli analogisti è la macchina a vettori di supporto, che decide quali esperienze ricordare e come combinarle per estrarne ulteriori previsioni.

Sulla base di queste considerazioni, il *machine learning* si presenta come un fenomeno che non risponde ad un'unica logica. Queste cinque scuole di pensiero, infatti, intendono in modo diverso il concetto di apprendimento e, per questa ragione, adottano differenti strategie per implementare le tecniche di apprendimento nella macchina. Ciascuna di queste strategie permette di sviluppare autonomamente – sulla base dei dati elaborati – un particolare modo di apprendere. Il *machine learning*, però, non è inteso solo come il complesso di strategie che la macchina adotta per apprendere, ma, ancora prima, esso è la capacità stessa di una macchina di imparare. È in questa capacità che la macchina rende manifesta la propria indipendenza rispetto al fattore umano. L'autonomia non sta, dunque, nell'essere in possesso di strategie per apprendere, quanto nella possibilità che tali strategie vengano elaborate dalla macchina stessa. In altre parole, la macchina in maniera *relativamente* autonoma dai suoi programmatori arriva ad elaborare delle strategie per apprendere: essa impara come imparare. La questione non riguarda più ciò che la macchina ha imparato, ma come essa lo ha imparato. In questo senso, ciò che rileva quando si parla di *machine learning* è che:

- (a) non si sa cosa le macchine impareranno;
- (b) non si sa come le macchine lo impareranno e, quindi, in che modo esse si avvicinano alla conoscenza.

Le scuole di pensiero segnalate da Domingos (2016, cap. IX) non si presentano in forma pura, ma vengono ricombinate per generare nuove tecniche di *machine learning*. Il processo di ricombinazione è definito *meta-learning*. Più precisamente, il *meta-learning* studia come i sistemi di apprendimento possano aumentare l'efficienza attraverso l'esperienza; l'obiettivo è capire come l'apprendimento stesso possa diventare flessibile in base al dominio o al compito da studiare (Vilalta e Drissi, 2002). Anche il *meta-learning* può essere affidato a una macchina. In questo caso, quindi, la macchina non solo apprende in maniera autonoma, ma decide anche quale combinazione di sistemi di apprendimento utilizzare. Questa ricombinazione, tuttavia, non garantisce di rintracciare il percorso inferenziale che ha portato alla decisione e, in ultima analisi, non permette di spiegare le ragioni di una decisione.

### 2.4.2 Sull'utilità e l'opportunità di un'analisi disgiunta

Nel paragrafo 2.3 ho esplorato la prospettiva dei concetti-ponte. Si tratta di quei concetti che nello studio dei fenomeni in cui vi è un'attività combinata (umani e macchine) propongono di adottare un'analisi congiunta, ponendo il focus sulle relazioni reciproche, con la prospettiva, forse un po' ingenua, di un miglioramento delle *performance* umane e nella promessa di una pacifica coesistenza fra umani e macchine.

Ciò che suggerisco è di adottare un'analisi disgiunta, con un focus che non insiste sulle reciproche influenze fra umani e macchine, quanto su ciò che contraddistingue le macchine, nell'ottica di poter inquadrare esattamente il loro ruolo così da comprendere i profili di compatibilità con l'umano. L'analisi disgiunta si propone di studiare la macchina in quanto agente autonomo che si relaziona all'umano non in termini semplicemente strumentali.

La proposta di un'analisi disgiunta acquista sostanza proprio grazie al *machine learning*. Se, infatti, era impensabile parlare di agente autonomo in riferimento a un'entità simbolica o che si rifà al modello GOF AI,<sup>26</sup> il *machine learning* ha aperto la strada a numerose ricerche sugli agenti autonomi (Ezenkwu e Starkey, 2019). Ciò che

---

<sup>26</sup> Per una definizione di "GOF AI" si veda Dennet (2018): «La prima intelligenza artificiale, detta anche GOF AI (Good Old-Fashioned Artificial Intelligence) (Haugeland, 1985), era un approccio topdown, "intellettualistico" all'Intelligenza Artificiale, basato sullo scrivere ciò che sapevano gli esperti umani, in un linguaggio che il computer poteva manipolare con motori inferenziali in grado di perlustrare gli "enormi" banchi di memoria riempiti di questa conoscenza del mondo inserita manualmente con grande cura, deducendo i teoremi necessari per prendere decisioni fondate e controllare nel modo appropriato gli arti o qualsiasi altro effettore dell'intelligenza artificiale. A posteriori, possiamo constatare che la GOF AI è stata un esercizio di creazione di qualcosa di molto cartesiano, un esperto razionalistico con miriadi di proposizioni immagazzinate nella sua memoria e tutta la comprensione incorporata nella sua capacità di trarre conclusioni dagli assiomi pertinenti e di individuare contraddizioni nella sua conoscenza del mondo – nel modo più efficiente possibile».

caratterizza i modelli GOFAI è la prevedibilità: dal momento che la loro programmazione è interamente *top-down*, il loro comportamento è totalmente pianificato e prevedibile. Quasi paradossalmente, ciò che caratterizza i modelli basati sul *machine learning* – e li rende autonomi – è anche ciò che viene percepito come un grande limite o una “anomalia” – nonché un’insidia – del ragionamento automatico: mi riferisco all’opacità degli algoritmi. Ma cosa si intende per opacità degli algoritmi?

Secondo Burrell (2016), per “opacità” si intendono almeno tre fenomeni.

In un primo senso, ci si riferisce alla forma di autoprotezione intenzionale da parte delle aziende che intendono mantenere i loro segreti commerciali e il loro vantaggio competitivo.<sup>27</sup>

Il secondo fenomeno che va sotto il nome di “opacità” è legato all’analfabetismo tecnico. Si deve riconoscere che tanto la scrittura quanto la lettura del codice, così come la progettazione degli algoritmi, sono competenze specialistiche che, quindi, rimangono inaccessibili alla maggior parte della popolazione.

Il terzo ed ultimo senso in cui è inteso il concetto di opacità riguarda un ulteriore livello. Si tratta di una problematica strettamente connessa al *machine learning*: quando una macchina impara e, in questo modo, costruisce la propria rappresentazione di una decisione, essa lo fa senza tenere conto della possibilità di essere compresa da un umano. Non è un problema che riguarda semplicemente la difficoltà tecnica: gli algoritmi di *machine learning* possono sfuggire alla piena comprensione anche a coloro che hanno una formazione specializzata. La ragione è semplice: la logica decisoria interna degli algoritmi è alterata durante la fase di apprendimento.

Se l’essenza dell’autonomia per una macchina è nell’alterazione della logica interna rispetto al progetto originario, la sua manifestazione è nell’opacità degli algoritmi – ovvero nel loro essere imperscrutabili per un osservatore umano.

---

<sup>27</sup> Con riguardo a questa declinazione dell’opacità, si è detto che essa potrebbe anche essere una copertura per una nuova forma di occultamento di regolamenti elusi, di manipolazione dei consumatori o di modelli di discriminazione (Pasquale, 2015).

In questo senso, un'analisi congiunta – che mira a minimizzare la tensione esistente fra macchine e umani – è poco convincente e poco utile ai fini di questa ricerca: mi interessa, infatti, dare enfasi al fenomeno dell'emancipazione della macchina dal suo creatore. È solo a partire dalla “macchina emancipata” che è possibile ricostituire la tensione umano-macchina, così da verificare cosa succede nell'ipotesi in cui una macchina dovesse acquisire un'autorevolezza tale da assicurare il raggiungimento di decisioni epistemicamente migliori rispetto a quelle che prenderebbero degli umani.

## **CAPITOLO 3 – Processi decisionali democratici**

### **3.1 Processi decisionali e teorie della democrazia**

#### **3.2 Verso una definizione dei processi decisionali democratici: quattro domande**

##### **3.2.1 Perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?**

###### **3.2.1.1 Giustificazioni strumentaliste**

###### **3.2.1.2 Giustificazioni non-strumentaliste**

###### **3.2.1.3 La natura della giustificazione pubblica**

##### **3.2.2 La democrazia si fonda sull'idea che esista una posizione etica corretta?**

###### **3.2.2.1 Classificare le concezioni metaetiche**

###### **3.2.2.2 Oggettivismo vs soggettivismo**

##### **3.2.3 A quali risultati deve condurre una decisione presa in maniera democratica?**

###### **3.2.3.1 Sostanzialismo vs proceduralismo**

###### **3.2.3.2 Il tentativo di conciliare procedura e sostanza**

##### **3.2.4 Come si può prendere una decisione con metodo democratico?**

###### **3.2.4.1 Voto, negoziazione, argomentazione**

###### **3.2.4.2 L'argomentazione e la decisione corretta**

### **3.3 Due concezioni della democrazia**

#### **3.3.1 Concezioni non-epistemiche**

#### **3.3.2 Concezioni epistemiche**

##### **3.3.2.1 I sistemi deliberativi come espressione delle concezioni epistemiche**



## CAPITOLO 3

### 3.1 Processi decisionali e teorie della democrazia

La democrazia, per come verrà analizzata in questa sede, è intesa come un metodo per prendere decisioni. Tale metodo, a seconda dell'ideale di riferimento e di altri fattori che menzionerò nelle prossime pagine, può assumere varie forme.

È stato osservato che la maggior parte degli studi sul rapporto fra intelligenza artificiale e democrazia ha riguardato i possibili effetti – negativi o positivi, a seconda delle analisi – dell'ingerenza delle macchine nei sistemi democratici. A mio avviso, però, non sono state sufficientemente esaminate le condizioni di ammissibilità per la partecipazione delle macchine ai processi decisionali in un contesto democratico. In altre parole, quali sono le condizioni che consentono a una macchina di prendere parte a una decisione? E di che tipo di partecipazione deve trattarsi?

Ho anticipato, nel capitolo precedente, che una macchina può avere due distinti ruoli all'interno di un processo di *decision-making*: come agente e come sistema che elabora le decisioni degli agenti umani. Non ho, tuttavia, ancora definito il tipo di processo decisionale cui faccio riferimento.

Il processo decisionale in cui si prendono decisioni a nome e per conto della collettività è quello democratico. Ma cos'è un processo decisionale democratico? E in che senso la democrazia è un metodo per prendere decisioni?

Innanzitutto, bisogna chiarire che non tutti i metodi per prendere decisioni sono democratici, così come è vero che non tutti i sistemi di governo sono democratici. Metodi decisionali e sistemi di governo non democratici sono solitamente definiti autoritari. Se l'obiettivo di questa ricerca è quello di valutare l'utilizzo di una macchina all'interno di un processo decisionale democratico, allora è opportuno limitare il campo di questa analisi, considerando che:

- (a) verranno presi in esame solo i processi decisionali democratici;
- (b) la democrazia sarà intesa come metodo per prendere decisioni.



Dopo aver ristretto il campo d'indagine, devo far presente che esso è più largo di quanto si immagini: non esiste un solo processo decisionale democratico, ma ne esistono tanti quante sono le concezioni della democrazia. Il numero di queste concezioni – e di relativi processi decisionali democratici – è indeterminato e, potenzialmente, se ne potrebbe ricavare una da ogni soggetto che si è fermato a riflettere sul significato della parola democrazia. Valutare ciascuna teoria della democrazia – con annessa concezione del processo decisionale –, fosse anche solo degli autori di maggiore rilievo, non è in questa sede un compito che posso portare a termine.

Partirò da un'ipotesi: esiste una famiglia di concezioni che consente l'accesso delle macchine ai processi decisionali democratici ed esiste un'altra famiglia di concezioni che sbarra la strada alle macchine. Non analizzerò, quindi, tutte le teorie della democrazia che finora sono state sviluppate, ma, partendo da quattro domande, proverò a dimostrare come ogni teoria possa essere ricondotta a una di queste due famiglie: è questo l'obiettivo del terzo capitolo. È rimandata, invece, al quarto capitolo la dimostrazione sulla compatibilità o incompatibilità fra le macchine e queste due famiglie di concezioni.

Come ho già sottolineato, le teorie della democrazia sono numerosissime. Questo è legato a diverse ragioni. La riflessione teorica sulla democrazia ha radici lontane e nel corso dei secoli si è evoluta e si è fatta sempre più articolata e complessa.

Storicamente, il concetto di “democrazia” si è sviluppato in contrapposizione ai concetti di “monarchia”, in cui governa una sola persona, e “aristocrazia”, in cui il governo è nelle mani di pochi individui, in quanto in una democrazia è il popolo a governare (Held, 2006).

La democrazia è stata oggetto di uno studio che ha riguardato diverse discipline – scienze politiche, sociologia, economia – ed è stata analizzata da diversi punti di vista – democrazia come ideale sociale, come insieme di istituzioni politiche, ecc. In questo testo, con la parola “democrazia” mi riferirò ad una procedura per prendere decisioni in maniera collettiva che si caratterizza per il fatto che include, più o meno

direttamente, tutti i membri di una comunità politica per la quale c'è bisogno che una certa decisione venga presa (Goodin, 2008).

Negli ultimi decenni, a partire dall'esigenza di far fronte alla crisi della democrazia (Mounk, 2018; Levitsky e Ziblatt, 2018; Przeworski, 2019; Foa e Mounk, 2016), si è creato molto entusiasmo attorno alla teoria della democrazia deliberativa. La ragione per cui mi dedicherò all'analisi di questa teoria è che a partire da essa si è sviluppata una famiglia di concezioni – definita epistemica – che, come ho già sostenuto, agevola, almeno in linea teorica, l'accesso delle macchine ai processi decisionali democratici.

Ma che cos'è la teoria della democrazia deliberativa?

Questa domanda non ha una sola risposta: tante definizioni sono già state formulate sulla democrazia deliberativa. Dal punto di vista concettuale, la democrazia deliberativa intende superare le concezioni aggregative della democrazia (Bächtiger, Dryzek, Mansbridge e Warren, 2018): le teorie democratiche che hanno dominato il dibattito scientifico e intellettuale nel corso del Novecento, infatti, si caratterizzano per la loro natura “aggregativa”. Jon Elster (1986) afferma che per le teorie aggregative il processo politico è uno strumento piuttosto che un fine in sé e che l'atto politico decisivo è un'azione privata anziché pubblica, vale a dire il voto individuale e segreto.

Si tratta, dunque, di teorie accomunate da una visione meramente “aggregativa” del processo democratico: esse si fondano sull'aggregazione di preferenze o di interessi individuali al fine di raggiungere una scelta sociale collettiva. La visione aggregativa, inoltre, per i fautori di un approccio deliberativo alla democrazia, vedrebbe le preferenze dei cittadini come fisse e predeterminate, mentre in una concezione deliberativa della democrazia i cittadini dovrebbero essere pronti a ridiscutere pubblicamente tali preferenze e, quindi, ad essere disposti a trasformarle.

Occorre chiarire cosa si intende per deliberazione. Nella lingua italiana “deliberare” e “decidere” vengono spesso intesi come sinonimi. La parola “deliberare”, nel significato che in questa trattazione le viene dato, rimanda alla radice etimologica classica e al significato che ha conservato nella lingua inglese:

“Deliberare” significa, classicamente, soppesare i pro e i contro delle possibili soluzioni ad un problema, trovare e sostenere le ragioni a favore di una scelta pratica e criticare quelle che non appaiono convincenti, riconoscere come accettabili gli argomenti che risultano persuasivi, respingere o contestare quelli che non appaiono tali. “Deliberare” significa formarsi un giudizio ponderato su ciò che è “giusto” o “sbagliato”. (...) Una deliberazione può svolgersi nel foro interiore di un individuo, ma si svolge anche – ed è questo che, naturalmente, interessa quando si parla di “democrazia” – in una dimensione pubblica, attraverso uno scambio di ragioni e di argomenti (Floridia, 2017, pp. 11-12).

Si può, dunque, parafrasando Bohman, qualificare la democrazia deliberativa come quell’insieme di concezioni nelle quali una deliberazione pubblica di cittadini liberi ed eguali sarebbe il cuore della legittimità politica dei processi decisionali e di autogoverno (Bohman, 1998).

Essendo “un insieme di concezioni” la teoria della democrazia deliberativa si pone come una sorta di ombrello. Non sorprenderà, dunque, nel corso della trattazione notare la profonda diversità di vedute fra i vari autori.

La distinzione fra teorie aggregative e teorie deliberative non acquista in questa analisi una rilevanza preminente.<sup>28</sup> Ciò che in questa sede interessa, e che verrà ripreso nelle prossime pagine, è osservare come le concezioni deliberative della democrazia spostino il focus sul processo trasformativo delle volontà e non si concentrino unicamente sulla mera espressione delle volontà.

Analizzare le diverse teorie della democrazia è un’operazione volta a definire il procedimento entro cui una macchina può eventualmente trovare spazio.

Bisogna ricordare che le condizioni di ammissibilità di una macchina in un dato procedimento dipendono da:

---

<sup>28</sup> Bisogna riconoscere, però, che, secondo alcuni, come osserva Peter (2007), nella teoria democratica d’oggi le due categorie più importanti sono per l’appunto la coppia democrazia deliberativa/democrazia aggregativa.

- la natura della macchina;
- il processo decisionale di riferimento;
- il tipo di partecipazione della macchina.

Il primo punto è stato trattato nel capitolo precedente, il terzo punto sarà oggetto dell'ultimo capitolo, mentre il secondo punto sarà analizzato in questo capitolo. Il processo decisionale di riferimento è, infatti, determinato dalla concezione di democrazia di partenza. Queste concezioni non sono volte alla costruzione di particolari assetti istituzionali, ma intendono descrivere il conglomerato di giustificazioni che reggono la democrazia e gli obiettivi perseguiti. Nelle prossime pagine, proverò a fornire un'immagine per ciascuna delle due distinte concezioni della democrazia.

### **3.2 Verso una definizione dei processi decisionali democratici: quattro domande**

Al fine di determinare la natura dei processi decisionali democratici, occorre isolare due concezioni della democrazia da cui discendono (almeno) due distinti procedimenti. Nel compiere questa operazione, mi servirò di quattro domande che toccano quattro aspetti cruciali per la definizione di una teoria della democrazia. Il paradigma della democrazia deliberativa, ponendosi trasversalmente a queste concezioni, non si configura come una teoria indipendente rispetto alle altre concezioni. Tuttavia, il paradigma deliberativo verrà ripreso in più punti, anche perché è a partire da questa teoria che si sono rafforzate quelle concezioni – definite epistemiche – che, secondo quanto sostengo, agevolano l'accesso delle macchine in un processo decisionale democratico.

Tre dei quattro punti presi in esame sono fra loro legati, così che a partire dalla definizione di uno di questi punti si potrebbe arrivare alla determinazione degli altri due. Si tratta, comunque, di connessioni possibili, ma non necessarie.

Per quanto riguarda l'ultimo punto, invece, esso non serve a definire una particolare concezione della democrazia, ma presenta i diversi modi in cui le preferenze possono essere espresse ed elaborate in una democrazia.

Le domande che affronterò sono le seguenti:

- Perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?
- La democrazia si fonda sull'idea che esista una posizione etica corretta?
- A quali risultati deve condurre una decisione presa in maniera democratica?
- Come si può prendere una decisione in maniera democratica?

### **3.2.1 Perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?**

Perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico? Questa domanda è preceduta da un'altra domanda: qual è il modo migliore per prendere decisioni di interesse pubblico? Se a questa domanda si risponde “la democrazia”, bisogna spiegare perché la democrazia è il modo migliore per prendere decisioni. Il dibattito intorno a questa domanda è quello sulle giustificazioni della democrazia. Le giustificazioni che vengono proposte servono ad individuare valori o ragioni che inducono a ritenere preferibile la democrazia su forme alternative di *decision-making*, in primo luogo su sistemi come l'oligarchia o la dittatura. È possibile distinguere due tipi di giustificazioni: le giustificazioni strumentaliste e quelle non-strumentaliste.

#### **3.2.1.1 Giustificazioni strumentaliste**

Nelle teorie di stampo strumentalista la giustificazione è connessa ai risultati che il metodo decisionale democratico produce in comparazione ad altri procedimenti (Arneson, 2004). Si è difeso il metodo democratico in quanto può generare:

- (a) un progresso per i protagonisti del processo democratico;
- (b) delle buone leggi e delle buone politiche.

Per quanto riguarda (a), diversi autori vedono nel processo democratico una possibilità di autorealizzazione delle potenzialità umane, di autodeterminazione, di formazione tanto di attitudini quanto di responsabilità civiche (Floridia, 2017, p. 39).

Per quanto riguarda (b), si possono distinguere due famiglie di teorie: le “*responsiveness theory*” e le teorie epistemiche.

Per le prime, il processo decisionale democratico protegge meglio diritti e interessi soggettivi perché risponde meglio a giudizi o preferenze dei cittadini rispetto ad altre forme di governo. J. S. Mill (1861, cap. III) sostiene che dal momento che la democrazia dà ad ogni soggetto una parte di potere politico, tale sistema di governo forza i decisori politici a prendere in considerazione diritti ed interessi di una più ampia fascia di soggetti di quelli che vengono presi in esame in un'aristocrazia o in una monarchia.

La seconda famiglia di teorie – le teorie cosiddette “epistemiche” – condivide la tesi che la democrazia sia generalmente più autorevole rispetto a metodi alternativi nella produzione di decisioni politiche. Fondamentalmente, si ricorre a due tipi di spiegazione: spiegazioni di tipo statistico o probabilistico – come, ad esempio, il teorema della giuria di Condorcet – e spiegazioni legate all’aspetto deliberativo dei processi decisionali.

Il teorema della giuria di Condorcet (2014) sancisce che, date certe ipotesi, la probabilità che una maggioranza di voti supporti la decisione corretta aumenta all’aumentare del numero dei votanti. Le assunzioni sono:

- è più probabile che un votante identifichi la scelta corretta (l’ipotesi di competenza);
- i votanti votano per quella che credono possa essere la decisione corretta (l’ipotesi di sincerità);

- i votanti sono statisticamente indipendenti l'uno dall'altro (l'ipotesi dell'indipendenza).

Il teorema di Condorcet è stato utilizzato anche per giustificare decisioni che riguardano tre o più opzioni (List e Goodin, 2001).

Quando parlo di teorie epistemiche della democrazia, mi riferisco all'idea che la giustificazione della democrazia risieda nel fatto che il metodo democratico permette di massimizzare le *chance* di arrivare alla giusta o corretta decisione, o almeno di arrivarci il più vicino possibile. Decisione “corretta” o “giusta”, come si vedrà, può voler dire una grande varietà di cose all'interno di un'ampia gamma di possibilità che va dall'oggettiva verità ad un costrutto più soggettivo, che dipende da fattori culturali o temporanei (Estlund e Landemore, 2018, p. 113).

Sulle teorie epistemiche mi concentrerò adeguatamente nelle prossime pagine. La convinzione, infatti, che la democrazia permette di giungere alla “decisione migliore” deve essere valutata alla luce della credenza che esiste una decisione migliore (3.2.2) e, soprattutto, se sia possibile definire indipendentemente dal processo democratico qual è tale decisione (3.2.3).

### **3.2.1.2 Giustificazioni non-strumentaliste**

Per le teorie non-strumentaliste della democrazia, un processo democratico è qualcosa di desiderabile indipendentemente dai risultati prodotti: un processo democratico sarebbe un bene in sé perché ci sarebbero dei valori che esso realizza e soddisfa a prescindere dal suo esito. I valori ai quali una visione non-strumentalista della democrazia si riaggancia sono essenzialmente quelli di libertà (Rostbøll, 2015; Saffon e Urbinati, 2013) e di equità (Waldron, 1999).

Per quanto riguarda la libertà, l'idea fondamentale è che ogni individuo in una democrazia ha diritto alla libertà, che consiste nella possibilità che viene data ai singoli di auto-governarsi. Questo comporta il diritto alla partecipazione politica e anche il diritto a “prendere cattive decisioni”. Per questa ragione, una teoria non-

strumentalista, che si fondi sulla libertà, non tiene in considerazione il “prodotto” del processo democratico (Gould, 1988).

Per quanto riguarda il valore dell'equità, la democrazia rappresenta un modo per trattare le persone come eguali quando c'è una buona ragione per imporre una forma di organizzazione sulla loro vita condivisa. Quando tutti hanno una determinata idea di come risolvere le questioni che riguardano la comunità, ognuno reclama il diritto ad essere dittatore della vita condivisa. La democrazia rappresenta un pacifico e giusto compromesso fra tutte le rivendicazioni in conflitto per comandare. Il compromesso consiste nella eguale possibilità di poter esprimere la propria voce sulla decisione da prendere. In questo senso, un processo decisionale democratico, dando ad ognuno una eguale possibilità di intervenire in caso di disaccordo, sarebbe garanzia di rispetto di ogni punto di vista su questioni di interesse comune (Singer, 1973, pp. 30-41).

Infine, anche la “giustificazione pubblica” viene spesso considerata un tipo di giustificazione non-strumentalista della democrazia (Habermas, 2013). A tal proposito, rimando alle considerazioni che si trovano al prossimo paragrafo, che mostrano come si tratti, in ultima analisi, di una giustificazione strumentalista.

Un approccio in linea con la *ratio* della distinzione fra teorie strumentaliste e teorie non-strumentaliste suggerisce di fondare questa distinzione sulla base della presenza o meno di una giustificazione che abbia un carattere strumentale rispetto all'esito del processo democratico. In altre parole, definiamo strumentaliste tutte quelle teorie che si fondano anche, e non unicamente, su giustificazioni che sono legate a ciò che il processo democratico produce. Definiamo, invece, “strumentaliste pure” quelle teorie che si fondano unicamente su giustificazioni legate all'esito del processo.

### **3.2.1.3 La natura della giustificazione pubblica**



La giustificazione pubblica è stata considerata una giustificazione di stampo non-strumentalista da Habermas. Se Habermas ascrive la sua teoria nel novero delle teorie non-strumentaliste, lo fa per non essere tacciato di sostanzialismo,<sup>29</sup> dove per sostanzialismo, come vedremo nelle prossime pagine (3.2.3), si intende l'idea che esista uno standard di giustizia che è indipendente rispetto alla procedura democratica. La sua mossa è, quindi, duplice:

- da un lato evita di caricare il processo democratico di un contenuto determinato aprioristicamente – come i repubblicani che considerano il processo democratico un mezzo per permettere l'auto-chiarificazione etico-politica o come i liberali che considerano il processo democratico lo strumento per raggiungere il compromesso di interessi;
- dall'altro lato, facendo appello al principio dell'etica del discorso come fondamento della democrazia, dichiara l'indipendenza della sua ricostruzione dall'output del processo.

In “Fatti e norme”, Habermas pone come fondamento della democrazia il “principio dell'etica del discorso”. Tale principio porta a considerare la democrazia nei termini di un processo discorsivo basato sullo scambio di ragioni. La concezione deliberativa, grazie alla teoria del discorso, rielabora il paradigma liberale e quello repubblicano in maniera inedita. In accordo con il repubblicanesimo, porta in primo piano il processo di formazione di opinione e volontà, ma non intende come secondario l'istituzionalizzazione di corrispondenti procedure e presupposti comunicativi (Habermas, 2013, p. 333).

Tanto la concezione repubblicana quanto quella liberale sono, per Habermas, accomunate dal sostanzialismo. Questa valutazione viene operata in base alla funzione data al processo democratico. Se per i repubblicani, infatti, il processo democratico è una forma di auto-chiarificazione etico-politica della comunità, per la concezione liberale, il processo democratico si compie esclusivamente nella forma

---

<sup>29</sup> Si vedrà, nel prossimo paragrafo, come una giustificazione di tipo strumentalista conduca sempre al sostanzialismo (Rummens, 2018).

dei compromessi d'interesse che si basano su regole fondate, in ultima istanza, sui diritti fondamentali liberali (Habermas, 2013, p. 359).

Habermas vuole superare il sostanzialismo e punta su una concezione radicalmente procedurale.

La sovranità popolare non è più concentrata in un collettivo (cioè nella presenza fisicamente afferrabile di un'assemblea di cittadini o rappresentanti), ma si fa valere nella circolarità processuale di discussioni e dibattiti e decisioni ragionevolmente strutturate (Habermas, 2013, p. 359).

La sovranità popolare, alla stessa maniera, non viene smentita, ma «intersoggettivamente interpretata».

Pur facendosi anonima, la sovranità popolare si ritira solo allo scopo di farsi valere come un potere comunicativamente generato. A rigore, questo potere comunicativo nasce dall'interazione che si crea tra una formazione della volontà istituzionalizzata come Stato di diritto, da un lato, e sfere pubbliche culturalmente mobilitate, dall'altro (Habermas, 2013, p. 337).

Nella versione proceduralista, la sovranità popolare è legata a «condizionamenti sociali» che non risultano immediatamente disponibili alla «volontà dei cittadini». La politica deliberativa è, dunque, solo uno degli elementi di una società ben più complessa e genera una versione della democrazia che vede nel sistema politico non un «vertice» o un «centro», ma un «sistema d'azione» fra tanti altri (Habermas, 2013, p. 337).

La posizione di Habermas, dunque, viene presentata come sovranità popolare congiunta con proceduralismo (Maffettone, 2019, p. 303). Si prospetta una democrazia discorsiva a doppio binario, articolata cioè su due livelli connessi fra loro: un livello formale che corrisponde alle istituzioni e un livello informale costituito dalle numerose ramificazioni dell'opinione pubblica all'interno della società civile. In sintesi, in una concezione deliberativa della democrazia la fonte di legittimità non è in

una predeterminata volontà degli individui, ma nel processo di formazione di quella volontà, il quale processo stesso rappresenta la deliberazione.

La teoria politica non farebbe, dunque, ipotesi su quali siano le condizioni per una società giusta, ma si esaurisce nel prendere in considerazione il modo in cui i partecipanti al processo democratico discutono e deliberano (Maffettone, 2019, p. 309). Questa concezione condurrebbe ad un proceduralismo estremo. In questo senso, secondo alcune ricostruzioni del pensiero habermasiano, le procedure democratiche acquisterebbero una sorta di valore “in sé” e la democrazia deliberativa diverrebbe non un mezzo ma un fine. È per questa ragione che il proceduralismo di Habermas dovrebbe considerarsi non-strumentalista.

Tuttavia, è stato osservato come la ricostruzione habermasiana del processo democratico possa risultare in ultima analisi fondamentalmente strumentalista. Si arriva a questa conclusione per diverse ragioni.

Un primo ordine di critiche è stato avanzato da Forst (2016), il quale sostiene che il principio dell’etica del discorso non è un principio genuinamente morale, ma è un principio che implica un esito felice di un discorso pratico.

Estlund (2008), da parte sua, ritiene che l’approccio proceduralista non sia così “estremo” come Habermas sostiene,<sup>30</sup> poiché tale approccio non ignora le questioni legate alle *performance* epistemiche. La questione del proceduralismo habermasiano verrà trattata più avanti, per ora è sufficiente osservare che la critica di Estlund è volta ad apprezzare gli aspetti strumentalisti nella giustificazione di Habermas.

L’ultimo ordine di ragioni riguarda la natura stessa della democrazia deliberativa. Ricorrendo al paradigma deliberativo, infatti, il processo democratico è geneticamente destinato ad avere un carattere strumentale. In base alla definizione di deliberazione data in precedenza si possono individuare due elementi imprescindibili: il processo decisionale e la decisione. Questi due elementi – processo decisionale (o procedure democratiche) e decisione – sono parte di un dispositivo – la deliberazione – che li pone inevitabilmente in un rapporto di mezzo a fine. La deliberazione non è

---

<sup>30</sup> Il proceduralismo si contrappone al sostanzialismo. Iscrivendosi nelle file dei proceduralisti, Habermas si salva dalle accuse di strumentalismo.

solo il modo in cui qualificiamo un certo processo decisionale – che si caratterizza per lo scambio di ragioni, la ponderazione, ecc. – ma è anche la decisione. Il processo decisionale, dunque, rappresenta il mezzo per raggiungere il fine, ossia la decisione.

Nonostante il funambolismo habermasiano, risulta evidente che all'interno del paradigma deliberativo il processo decisionale avrà sempre una funzione servente e sarà, per questo, strumentale rispetto alla decisione. In questo senso, la democrazia già non rappresenta più un “bene in sé”, ma viene considerata come un metodo – attraverso le forme dello scambio di ragioni, la ponderazione, ecc. – per raggiungere un fine, la decisione.

In conclusione, la teoria del discorso – cioè il fondamento habermasiano della giustificazione democratica – non è una giustificazione indipendente rispetto all'output della democrazia. Semplicemente, nel panorama delle giustificazioni strumentaliste è “meno strumentalista” rispetto ad altre, e questo perché, fra il principio del discorso e la decisione, Habermas interpone altri passaggi. Ricorrendo al principio dell'etica del discorso, infatti, Habermas dichiara la sua “lontananza” dalla fine del discorso, cioè la decisione. Ma ogni principio del discorso implica inevitabilmente una fine del discorso, cioè una decisione. Il fatto che il principio del discorso sia stato “allontanato” dalla decisione non vuol dire che non esista un collegamento fra questi due poli e, in questo collegamento, sta la relazione strumentale fra discorso (o processo decisionale) e decisione.

Ho già anticipato che esiste una connessione fra i quattro punti di cui si compone questa analisi, a riprova che una certa risposta ad una di queste domande condiziona – o dovrebbe condizionare – la risposta alle altre domande. Nel prosieguo di questo capitolo, tali connessioni verranno rese manifeste.

### **3.2.2 La democrazia si fonda sull'idea che esista una posizione etica corretta?**

Per rispondere alla domanda “perché la democrazia è il modo migliore per prendere decisioni di interesse pubblico?”, sono stati anticipati dei temi che verranno

ora trattati più dettagliatamente. Dopo aver presentato le giustificazioni che sono a sostegno della democrazia, occorre capire se la democrazia possa contare su un fondamento etico e quale esso sia. A questo quesito si è cercato di rispondere in vari modi.

Innanzitutto, bisogna domandarsi cosa si intende per “fondamento etico”. Questa domanda riguarda la metaetica. In metaetica, la questione rilevante non è “qual è la posizione etica corretta?”, ma se esiste una posizione etica corretta.

Ogni teoria della democrazia, infatti, si fonda su una concezione – o una famiglia di concezioni – metaetica. Da questo assunto prendono le distanze le teorie di Dahl e Schumpeter, secondo cui la democrazia non presuppone nessuna particolare posizione metaetica (Dahl, 1956; Schumpeter, 2008). Detto altrimenti, questi autori sono per la neutralità metaetica dell’ideale della democrazia.

Metterò da parte le tesi di Dahl e Schumpeter sulla neutralità metaetica e mi occuperò di quelle concezioni metaetiche che vengono poste a fondamento della democrazia. La domanda può essere riformulata in questi termini: la democrazia si fonda sull’idea che esista una posizione etica corretta?

Prima di rispondere a questa domanda, bisogna provare a fare un po’ di chiarezza nel dibattito sulla metaetica: come si classificano le concezioni metaetiche?

### **3.2.2.1 Classificare le concezioni metaetiche**

Le teorie metaetiche possono essere classificate in diversi modi.<sup>31</sup> Anche se tali classificazioni, come è stato osservato, sono in parte superate, è utile tenerle in considerazione (Bongiovanni, 2007).

La prima distinzione che prenderò in esame è quella fra realismo morale e non-realismo morale. Secondo i realisti morali, le proprietà morali esistono e sono indipendenti dai giudizi delle persone. I non-realisti morali, invece, sostengono che

---

<sup>31</sup> Per stilare questa tassonomia, mi sono servito di diversi contributi, tra cui Fisher (2014) e Harman e Thomson (1996).

non esistono proprietà o fatti morali. Parafrasando Sayre-McCord (1986), il realismo implica l'accettazione di due tesi:

- le affermazioni morali, se interpretate letteralmente, possono essere letteralmente vere o false (cognitivismo);
- alcune sono letteralmente vere.

Bisogna considerare che un non-realista può essere anche cognitivista, mentre un realista è necessariamente cognitivista. Non esistono, quindi, realisti che siano non-cognitivist, ma non c'è incompatibilità fra non-realismo e cognitivismo.

Ma cosa significa "cognitivismo" quando parliamo di metaetica? Secondo i cognitivist, quando qualcuno fa una dichiarazione in termini morali sta esprimendo una credenza. Inoltre, le dichiarazioni morali possono essere vere o false poiché sono credenze e le credenze sono soggette a giudizi di verità. Il cognitivismo è spesso anche chiamato "descrittivismo". Per i non-cognitivist, invece, se qualcuno esprime una dichiarazione in termini morali, non sta esprimendo una credenza, ma più probabilmente un'emozione o un desiderio. Uno stato diverso da una credenza non può essere considerato vero o falso: un desiderio, per esempio, non è suscettibile di un giudizio del genere.

Un'altra distinzione è fra naturalismo e non-naturalismo. Da una prospettiva naturalista, le proprietà morali sono indagabili empiricamente dalle scienze. Per il non-naturalismo, invece, le proprietà morali non possono essere sottoposte ad un'analisi scientifica. La distinzione fra naturalismo e non-naturalismo ha senso se si parte da una posizione realista e si assume, quindi, che le proprietà morali esistano.

L'ultima distinzione proposta è quella fra assolutismo e relativismo. Quando parliamo di assolutismo, in campo metaetico, ci riferiamo alla convinzione che esista una singola verità morale. Secondo i relativist, invece, non esiste una singola verità morale, ma ci sono diversi *framework* morali e nessuno di questi è più corretto degli altri.

La tassonomia minima qui proposta presenta delle categorie che in parte si sovrappongono e non chiarificano del tutto la complessità delle concezioni metaetiche esistenti. Ai fini di questa ricerca, ritengo sia importante partire

dall'ultima delle distinzioni prese in considerazione, cioè quella fra relativismo e assolutismo.

Resta, quindi, da chiedersi in quale modo queste categorie possono essere utilizzate per definire le prospettive metaetiche quando si parla di democrazia.

### **3.2.2.2 Oggettivismo vs soggettivismo**

A partire dalla distinzione tra assolutismo e relativismo, propongo – riprendendo il lavoro di Martí (2012) “*Democracia y subjetivismo metaetico*” – di adottare un'altra coppia di categorie che ritengo più utile per il fine che mi sono posto in questo paragrafo, che è quello di capire la natura di una teoria della democrazia che si fonda sull'idea che esiste una posizione etica corretta e la natura di una teoria che non si fonda su tale idea. Si tratta della distinzione fra soggettivismo e oggettivismo. Occorre da subito osservare che tutte le teorie assolutiste sono anche oggettiviste, ma non ogni teoria oggettivista è necessariamente assolutista.

Le teorie metaetiche oggettiviste si basano su due assunti:

- a) i giudizi etici possono essere corretti o incorretti;
- b) il criterio di correttezza è almeno parzialmente indipendente dalle attitudini e dalle credenze dell'individuo che formula il giudizio o del gruppo cui tale individuo appartiene.

Bisogna osservare che (a) rivela che è sufficiente il parametro della correttezza e non già quello più esigente della verità, che caratterizza solo alcune teorie oggettiviste, come quella del realismo morale. Per i realisti morali, infatti, i giudizi etici pretendono di riferire dei fatti ed essi sono veri se i fatti sono corretti. Inoltre, come si è visto, alcuni giudizi etici sarebbero effettivamente veri.

Secondo la categorizzazione che ho proposto, le teorie oggettiviste vengono distinte in teorie moniste – che credono che il criterio di correttezza sia fondato su un solo valore – e in teorie pluraliste – che sostengono che tale criterio si fondi su un insieme di valori.

Fra i pluralisti, poi, ci sono due famiglie:

- il pluralismo moderato – o armonico – che sostiene che i diversi valori morali possono essere fra loro armonizzati, in modo da fondare un criterio unico di correttezza dei giudizi (Rawls, 1991; Habermas, 1995, 2014);
- il pluralismo forte – o non armonico – che afferma che la pluralità di valori non permette di generare un unico criterio di correttezza per i giudizi etici.<sup>32</sup>

Per quanto riguarda il soggettivismo metaetico, anch'esso può essere diviso in due famiglie.

La prima famiglia sostiene che il criterio di correttezza dei giudizi etici dipende completamente da attitudini e da credenze individuali o di determinati gruppi. Se si tratta di attitudini e di credenze individuali, si farà riferimento al soggettivismo individuale; in relazione a un gruppo, invece, si tratterà di soggettivismo collettivo. In questo caso un criterio di correttezza per i giudizi etici esiste, ma tale criterio si riferisce a valori individuali o riferiti ad un determinato gruppo.

La seconda famiglia è quella dello scetticismo morale. Secondo questa teoria, non ci sarebbe la possibilità di valutare in termini di correttezza un giudizio etico, neanche in termini relativi.

Si può osservare come nel novero degli oggettivisti Martí (2012, nota 13) faccia rientrare anche i non-cognitivist e i non-realisti. In particolare, Martí crede che alcune teorie non-cognitiviste siano compatibili con un minimo di oggettività. Egli porta l'esempio del prescrittismo di Hare (1991), una teoria non-cognitivist, che esige alcuni requisiti oggettivi, come l'universalizzabilità, perché un certo giudizio etico possa considerarsi corretto. Inoltre, dal momento che non serve lo standard elevato della verità, ma è sufficiente quello della correttezza, l'oggettivismo potrebbe anche essere non-realista, assumendo che non esista una verità morale.

In questo senso, Martí (2012) propone di sostituire la questione sul valore di verità dei giudizi etici con quella sul valore di correttezza.<sup>33</sup> Questa sostituzione non

---

<sup>32</sup> Secondo Martí (2012), i contributi in questa direzione sarebbero quelli di Mouffe (1993, 2005), Laclau e Mouffe (1985), Tully (1995) e Lefort (1988).



comporterebbe un sacrificio dell'oggettività in un senso che lui considera rilevante. La rilevanza, in questo caso, non riguarda l'esistenza di una verità morale oggettiva, ma attiene alla costruzione di un criterio che non dipende esclusivamente dai giudizi soggettivi.

Martí, in sintesi, parte da due assunti:

(1) esistono giudizi etici corretti;

(2) esiste un criterio unico e univoco – e diverso da quello che si fonda sulle preferenze individuali – per decretare se un giudizio etico è corretto.

Il primo assunto è comune a qualsiasi concezione metaetica oggettivista. Il secondo, invece, esclude le posizioni oggettiviste pluraliste non armoniche: si tratta di quelle posizioni che non permettono di formulare un criterio unico e univoco di correttezza.

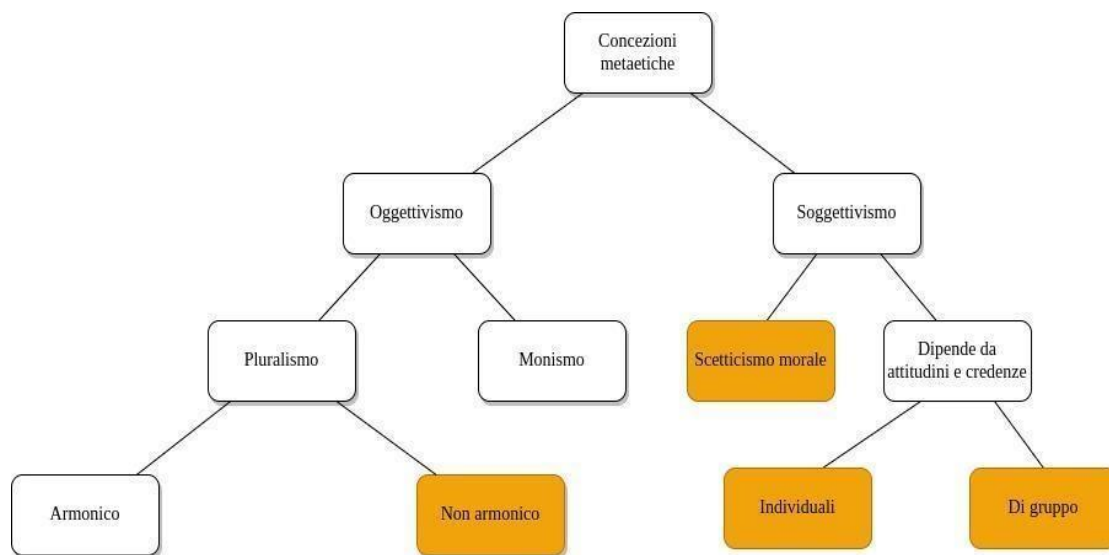
Possiamo osservare come non si tratti di due assunti indipendenti. (1) può esistere anche se (2) non esistesse o fosse falsa: è il caso degli oggettivisti pluralisti non armonici. Al contrario, (2) non può esistere senza (1): assumere che ci sia un criterio per decretare se un giudizio etico è corretto, implica necessariamente credere nell'esistenza di giudizi etici corretti. Per riassumere, (2) dipende da (1), ma (1) non dipende da (2). (2) è in qualche modo la diretta conseguenza di (1), ma è anche un modo per validare (1): se non esistesse un modo per verificare la correttezza di un giudizio etico, che senso avrebbe parlare di giudizi etici corretti?

Per Martí, quindi, il vero *discrimen* nell'affrontare la questione fondativa della democrazia non è tanto fra oggettivismo e soggettivismo metaetico, quanto fra teorie che consentono di costruire un criterio (unico e univoco) di correttezza dei giudizi etici e teorie che non lo fanno. Nella tassonomia proposta, quindi, il pluralismo non armonico – pur essendo una concezione oggettivista – non permettendo di costruire un criterio di correttezza dei giudizi etici, si presenta più vicino al soggettivismo che all'oggettivismo.

La galassia metaetica ricostruita da Martí può così essere schematizzata.

---

<sup>33</sup> Martí (2012) porta come esempio quello di Habermas, il quale preserva la nozione di verità per gli enunciati empirici, mentre quella di correttezza per gli enunciati normativi.



Quelle in arancione sono le concezioni che Martí ritiene non possano né fondare filosoficamente la democrazia né giustificarla. Seguendo questa ricostruzione, infatti, il relativismo – intendendo con esso tutte quelle concezioni metaetiche che non ritengono esista un criterio unico e univoco di correttezza dei giudizi etici – non può giustificare o fondare filosoficamente la democrazia.

Al contrario, ciò che Kelsen sostiene è che il relativismo morale offra l'unico fondamento filosofico possibile per la democrazia. In un frammento di “*Foundations of democracy*” Kelsen (1955, p. 38) sostiene:

... se si crede nell'esistenza dell'assoluto, e di conseguenza nei valori assoluti, [...] non è forse insensato lasciare che una maggioranza decida ciò che è politicamente buono? Legiferare, e cioè determinare i contenuti di un ordine sociale, non in base a ciò che è oggettivamente il meglio per gli individui soggetti a questo ordine, ma in base a ciò che questi individui, o la loro maggioranza, credono, a torto o a ragione, essere il loro meglio – questa conseguenza dei principi democratici di libertà e uguaglianza è

giustificabile solo se non esiste una risposta assoluta alla domanda su cosa sia il meglio, se non esiste un bene assoluto.<sup>34</sup>

Non è mio obiettivo stabilire se il miglior fondamento filosofico della democrazia sia il relativismo o l'assolutismo. Ciò che mi interessa, invece, è capire quali teorie della democrazia possono nascere da distinte concezioni metaetiche. A mio avviso, infatti, a seconda della concezione metaetica di elezione – sulla base del fatto che essa sia oggettivista o soggettivista – è possibile costruire distinte teorie della democrazia.

La mia ipotesi di partenza è che una concezione metaetica relativista porti necessariamente a una concezione non-epistemica della democrazia, mentre una assolutista possa portare a una concezione epistemica. Nelle prossime pagine, si vedrà cosa si intende per “concezione epistemica” e cosa per “concezione non-epistemica”.

### **3.2.3 A quali risultati deve condurre una decisione presa in maniera democratica?**

La terza domanda riguarda l'esito del procedimento democratico. Il procedimento democratico deve condurre ad un determinato *output*? Se sì, quale? Più precisamente, questa domanda riguarda il rapporto fra democrazia e giustizia. Esiste un'unica e oggettiva risposta in materia di giustizia? E qualora esistesse, tale risposta è indipendente dalla stessa procedura democratica?

Questa domanda è strettamente connessa ai punti che abbiamo già analizzato in 3.2.1 e in 3.2.2. Per ora, dobbiamo osservare che al quesito appena proposto si può rispondere principalmente in due modi (3.2.3.1) e che esiste una terza risposta (3.2.3.2) che ha tentato di tenere insieme queste due risposte a prima vista inconciliabili.

---

<sup>34</sup> Traduzione mia da Kelsen (1955).

### 3.2.3.1 Sostanzialismo vs proceduralismo

Con riferimento alla democrazia, la giustizia può essere concepita in termini sostanziali o procedurali.

Coloro che propendono per una concezione sostanziale della giustizia ritengono che esista sempre una risposta corretta e oggettiva in materia di giustizia e che tale risposta sia indipendente rispetto al procedimento democratico. Al contrario, coloro che sono per una concezione procedurale della giustizia credono che il criterio di giustizia di una decisione democratica non dipenda da un contenuto sostantivo, ma sia determinato dal fatto che la decisione è stata presa rispettando un'adeguata procedura democratica.<sup>35</sup>

Questa seconda posizione è difesa da autori come Hans Kelsen e, come ha sostenuto Mackie (2011), sarebbe presupposta in modelli di democrazia aggregativa o che si fondano sulla *social choice*.<sup>36</sup> Da una prospettiva deliberativa della democrazia, invece, si può sostenere tanto una posizione sostanzialista quanto una proceduralista. Tuttavia, come si vedrà, le posizioni sostanzialiste sono quelle più diffuse fra gli autori legati a un ideale deliberativo della democrazia.

Tra le posizioni proceduraliste, si distingue quella di Arrow. In primo luogo, Arrow (1963, p. 22) rifiuta l'idea che esista un bene sociale oggettivo che si può definire indipendentemente dai desideri individuali, riagganciandosi in questo alla tradizione metaetica relativista. In secondo luogo, la "*social choice theory*" di Arrow cerca di individuare il *social welfare*, che è una funzione matematica che aggrega singole preferenze individuali. Gli ordinamenti di preferenza di un individuo possono riflettere semplici desideri legati al consumo, così come valori che includono considerazioni sociali e morali. Questi desideri o valori individuali creano una misura di *social welfare* solo attraverso un processo di aggregazione. In questo senso, e tornando alla premessa iniziale, secondo Arrow, non avrebbe senso parlare di valori

---

<sup>35</sup> Sul rapporto fra legittimità e giustizia si veda 3.3.

<sup>36</sup> In 3.1 si è visto cosa si intende per concezione aggregativa della democrazia. Per i contributi a una concezione aggregativa della democrazia si veda Black (1958) e Arrow (1963).

che prescindono dalle preferenze individuali aggregate da una funzione di *social welfare*.

Dall'altra parte, fra i sostanzialisti, vi è Van Parijs (2011), secondo cui il compito del filosofo è quello di attuare un “*Rawls–Machiavelli program*”: il filosofo politico deve in primo luogo determinare come debba essere una giusta società e, in un secondo momento, stabilire quali istituzioni politiche meglio contribuiscono alla realizzazione della giusta società che egli ha immaginato.

Inoltre, fra coloro che credono che la “giusta” risposta sia indipendente rispetto all'esito del procedimento democratico sono da annoverare Estlund (2008, pp. 21-39), Landemore (2013, pp. 208-231), Arneson (2004, p. 43) e Martí (2006b). Per questi autori, la democrazia più che implementare i requisiti della giustizia avrebbe il compito di scoprirli dal momento che essi già esistono e sono indipendenti dalla procedura stessa. In particolare, secondo Estlund (1997, 2008) e Landemore (2012, 2013), la democrazia è il miglior mezzo, fra quelli disponibili, per scoprire la verità in merito alla giustizia.

Tuttavia, occorre rilevare come Estlund (1997) non consideri la sua posizione sostanzialista, ma che preferisca definirla “*epistemic proceduralism*”, distinguendola tanto dal “*fair proceduralism*” quanto dal “*fair deliberative proceduralism*” e dal “*rational deliberative proceduralism*”.

Più precisamente, il *fair proceduralism*<sup>37</sup> sarebbe quella concezione secondo la quale ciò che rende legittima una decisione democratica è il ricorso alla procedura equa della regola di maggioranza. Il problema di questo approccio, secondo Estlund (2008), è che mentre le procedure democratiche potrebbero essere realmente eque, la massima espressione dell'equità fra persone in disaccordo su due alternative è quella di decidere lanciando una monetina. Il lancio della monetina, così come un'estrazione a sorte, pur trattandosi di un metodo decisionale equo, non sarebbe il mezzo appropriato per legittimare delle decisioni.

---

<sup>37</sup> A sostegno di questa concezione, ci sono i contributi di Cohen (1994) e Christiano (1993, 1996).

Secondo il *fair deliberative proceduralism*<sup>38</sup>, invece, i cittadini devono avere eguale possibilità di partecipare alla discussione che precede la votazione, proponendo le loro ragioni e i loro argomenti. Tali ragioni e argomenti entrano esplicitamente in questo processo e, secondo tale teoria, non bisogna fare appello a nessun preciso standard di correttezza indipendente rispetto alla procedura democratica per legittimare la procedura stessa. La legittimità, infatti, risiederebbe nella imparzialità della procedura rispetto alle convinzioni e agli argomenti degli individui.

L'ultima concezione da cui Estlund prende le distanze è definita *rational deliberative proceduralism* (Benhabib, 1996; Fishkin, 1992). Questa concezione riprende il *fair deliberative proceduralism*, tranne che per un elemento: il valore di una procedura risiede nel riconoscere le buone ragioni anziché nel garantire un equo accesso a tutti. Il *rational deliberative proceduralism*, quindi, non afferma che il metodo democratico generi risultati che si avvicinano a qualche standard che sia indipendente rispetto alle procedure: si tratterebbe, in quel caso, infatti, di una concezione epistemica. Al contrario, secondo il *rational deliberative proceduralism*, la decisione dev'essere prodotta da una procedura che si fonda sul riconoscimento di ragioni, mentre non rileva il fatto che questa decisione si avvicini ad uno standard di correttezza indipendente rispetto alla procedura. In questo senso, le decisioni sarebbero razionali solo in un senso procedurale e non anche sostanziale.

Queste teorie, secondo Estlund, non prendendo in considerazione uno standard di correttezza indipendente rispetto alla procedura stessa, non sarebbero sufficientemente "epistemiche".<sup>39</sup> Viene, quindi, proposto un nuovo *framework* filosofico, il proceduralismo epistemico. Esso avrebbe lo scopo di riconciliare l'interesse per le procedure e quello per il risultato. Estlund (2008), inoltre, afferma che occorre stabilire un minimo standard performativo atteso dall'autorità politica che va stabilito in una zona che va da "meglio che casuale" al punto in cui la democrazia stessa incontra i suoi limiti. Questo standard è probabilmente inferiore comparato agli

---

<sup>38</sup> Fra i contributi in questa direzione, ci sono Manin (1987) e Michelman (1989).

<sup>39</sup> Si vedrà nel paragrafo 3.3 cosa intende Estlund per una concezione epistemica.

*output* prodotti da quella che lui chiama “epistocrazia”. A prescindere dai nomi attribuiti da Estlund, la sua concezione è decisamente sostanzialista se per sostanzialismo si intende l’esistenza di uno standard di correttezza indipendente rispetto alla procedura stessa.

Per riassumere, alla domanda “a quali risultati deve condurre una decisione presa in maniera democratica?” ci sono, in linea di massima, due tipi di risposte.

Se si crede esista una decisione giusta che prescinde dal procedimento democratico – e che sia in un certo senso anteriore e indipendente rispetto a tale procedimento –, si appartiene a una famiglia di posizioni che possono essere definite sostanzialiste.

Coloro, invece, che sostengono che per definire “giusta” una decisione sia necessario e sufficiente il fatto che essa sia stata presa seguendo un’adeguata procedura democratica sono considerati proceduralisti.

Risulta, dunque, evidente il nesso che sussiste fra il rapporto democrazia/giustizia e le concezioni metaetiche analizzate in 3.2.2: chi ha una posizione sostanzialista necessariamente appoggia una concezione metaetica oggettivista dal momento che se crede esista una decisione giusta a prescindere dalla procedura, deve credere nell’esistenza di giudizi etici corretti e giudizi scorretti. Della relazione, invece, fra giustificazioni della democrazia e sostanzialismo mi soffermerò nel prossimo paragrafo.

### **3.2.3.2 Il tentativo di conciliare procedura e sostanza**

È stata proposta una terza via, che ambisce a tenere insieme tanto il sostanzialismo quanto il proceduralismo. Questa terza possibilità è stata anticipata quando ho affrontato il tema della giustificazione pubblica, al paragrafo 3.2.1.3. Se la giustificazione pubblica intende porsi a metà strada fra le giustificazioni strumentaliste e quelle non-strumentaliste, con riferimento al rapporto fra democrazia

e giustizia, la posizione che passa fra sostanzialismo e proceduralismo prende il nome di costruttivismo.

È stato sostenuto che le distinzioni fin qui proposte – giustificazioni strumentaliste/non-strumentaliste e proceduralismo/sostanzialismo – non siano adeguate per comprendere la vera natura del costruttivismo (Rummens, 2018, p. 136). Coloro che credono in giustificazioni strumentaliste e hanno un approccio sostanzialista in riferimento al rapporto fra democrazia e giustizia sono convinti dell'esistenza di uno standard di giustizia indipendente rispetto alla procedura democratica e che la possibilità – attraverso la procedura democratica – di conoscere la decisione giusta funga da giustificazione per la democrazia. Anche chi sottoscrive un approccio costruttivista sostiene esista uno standard sostantivo di giustizia. Tuttavia, secondo questo approccio, tale standard non sarebbe indipendente rispetto alla procedura democratica. In qualche modo, ciò che sostiene Habermas – e non solo – è che giustizia e metodo democratico sono a tal punto intrecciati che è impossibile pensare la democrazia senza giustizia e viceversa. Detto altrimenti, questa concezione adotta un'idea sostantiva di giustizia – intesa come imparzialità –, ma riconosce, allo stesso tempo, che i requisiti di giustizia possono venire determinati solo sulla base di un inclusivo processo di deliberazione democratica. In questo senso, non ci sarebbe giustizia senza democrazia poiché il contenuto specifico di giustizia può solo essere determinato democraticamente dai cittadini stessi. Parallelamente, non ci sarebbe democrazia senza giustizia perché i risultati democratici sarebbero legittimi nella misura in cui essi possono essere intesi come elaborazioni di un ideale sostantivo di giustizia. Per questa ragione, più che di scoperta è preferibile parlare della costruzione di verità in fatto di giustizia.

La concezione di Habermas è stata definita “*deep proceduralism*”, ma alcune critiche sono state rivolte al fatto che da un certo punto di vista essa sarebbe proceduralista solo in un senso rarefatto. Come si è visto, Habermas sostiene che l'unico standard cui debbano adeguarsi le decisioni politiche sia una non strumentale valutazione dei procedimenti che le generano – anche detta “razionalità procedurale” (Habermas, 2013), che si fonda sul potere della ragione nel discorso pubblico politico



– e che qualsiasi imposizione di standard politici sostanziali, cioè indipendenti dalla procedura, ostacolerebbe il fondamento dialogico su cui ritiene debba fondarsi la normatività politica (Habermas, 2018, p. 260). Allo stesso tempo, però, Habermas sostiene che il processo politico non possa essere inteso a prescindere dalla garanzia di certe libertà individuali e, in qualche modo, tali libertà forniscono uno standard per valutare le decisioni politiche: la violazione di queste libertà, infatti, renderebbe una decisione illegittima anche se questa fosse adottata attraverso una procedura corretta. In definitiva, per Estlund (2008, pp. 89-90), la ragione per cui quello di Habermas sarebbe un proceduralismo rarefatto è che la procedura deve in ogni caso condurre alla protezione di libertà basiche.

Inoltre, osserva sempre Estlund (2008), se le decisioni sono legittime solo se prese seguendo procedure deliberative ideali (Habermas, 1979, p. 186, 2018, pp. 31-34 e 259) – procedure, quindi, ipotetiche, immaginarie –, il fatto che una procedura sia legittima o meno è sempre logicamente indipendente dalla effettiva procedura che l’ha generata. È per questa ragione che si tratterebbe di uno standard sostantivo, indipendente dalla procedura. In altre parole, c’è un nesso di dipendenza logica fra procedura e standard di giustizia, ma la procedura presa in considerazione è quella della situazione discorsiva ideale. Per quanto riguarda la procedura effettiva, invece, tale rapporto di dipendenza non c’è: in questo senso lo standard di giustizia è indipendente rispetto al processo.

Estlund (2008, pp. 73-76 e 89-90), inoltre, estende la critica di “falso proceduralismo” anche alla *social choice theory*. Più precisamente, egli sostiene che la *social choice theory*, nella misura in cui si riferisce ad una regola aggregativa astratta, adotta uno standard sostanziale logicamente indipendente rispetto alla procedura che ha effettivamente prodotto le decisioni. In altre parole, ciò che è rilevante per la *social choice theory* è l’aggregazione di preferenze e non il reale processo che ha permesso di aggregare quelle preferenze per produrre una decisione collettiva.

Coerentemente a quanto visto finora, la teoria costruttivista habermasiana può essere inserita nella famiglia di concezioni metaetiche oggettiviste: nel caso di

Habermas lo standard per decretare la correttezza dei giudizi etici è determinato dalle decisioni prese nella situazione discorsiva ideale. Seppure, quindi, non viene fissato sostanzialmente il contenuto dei giudizi etici, il costruttivismo habermasiano, oltre a iscriversi perfettamente in una tradizione metaetica oggettivista, si poggia su una giustificazione di stampo strumentalista giacché, nella sua teoria, il procedimento democratico ha come obiettivo quello di raggiungere un certo tipo di decisioni che possono essere sottoposte al vaglio di un criterio di correttezza che è quello che si ricava dalle decisioni che verrebbero prese in una situazione discorsiva ideale. In ultima analisi, anche questo approccio, al pari di quello di Estlund, può dirsi sostanzialista.

Al di là di Habermas, Estlund (2008, p. 85) definisce “*deep deliberative democracy*” quelle concezioni che pur presentandosi come interamente proceduraliste finiscono per fare appello a standard indipendenti dalle stesse procedure per valutare la correttezza delle decisioni.

Secondo quanto visto finora, dunque, la concezione costruttivista di Habermas – e di quella parte della democrazia deliberativa che si dichiara proceduralista – non potendo fare a meno di richiamare standard indipendenti – anche solo ideali – non può essere iscritta all’interno di un paradigma proceduralista: la concezione costruttivista, infatti, si fonda su un preciso standard di correttezza che, almeno nel caso di Habermas, si fonda sulla situazione discorsiva ideale.

Infine, bisogna osservare come il tentativo di iscrivere una teoria non proceduralista all’interno del paradigma proceduralista non è isolato: tale tentativo è stato fatto anche con la *social choice theory*. Tuttavia, l’esito è stato lo stesso delle teorie ascrivibili al *deep deliberative democracy*: anche la *social choice theory* non riesce a fare a meno di una standard indipendente rispetto alla procedura stessa.

### **3.2.4 Come si può prendere una decisione con metodo democratico?**

Anche se la distinzione fra teorie aggregative della democrazia e teorie deliberative non acquista in questa analisi rilevanza primaria, bisogna osservare che essa si basa innanzitutto sul focus: mentre le prime insistono sulla aggregazione di preferenze individuali – e vedono nel voto la rappresentazione più emblematica –, le teorie deliberative pongono l'accento sulle possibilità trasformative delle preferenze attraverso processi dialogici.

Pur non negando l'utilità di un metodo come il voto, le teorie deliberative hanno rivolto la loro attenzione a metodi alternativi per prendere decisioni, come l'argomentazione. In questo senso, se l'accento sul voto rivela un'idea delle preferenze come qualcosa di fisso e immutabile, in un'ottica deliberativa della democrazia le preferenze individuali sono sempre in mutamento e si presentano come input assemblabili – e trasformabili – secondo funzioni meno semplicistiche di operazioni puramente additive. Se le teorie aggregative, infatti, concentrano le preferenze attorno a due o più centri, per la teoria deliberativa le preferenze si evolvono e hanno una naturale tendenza alla trasformazione.

La democrazia, come ho già detto, verrà analizzata come procedimento per prendere decisioni. Tale procedimento può essere diviso in due fasi:

- a) una fase di espressione e di trasformazione delle preferenze che funziona come input del processo;
- b) una fase nella quale si applica una regola di decisione che permette di passare dalle preferenze – e dalla loro trasformazione – alla decisione collettiva.

Per quanto riguarda le preferenze, esse possono nascere a partire da due diversi tipi di motivazione: motivazioni auto-interessate e motivazioni imparziali. Sono l'estremizzazione di due posizioni e non esistono preferenze fondate su motivazioni “pure”.

Secondo Elster e Martí, ad una decisione ci si può arrivare seguendo tre principi: voto, negoziazione e argomentazione.<sup>40</sup> Ciascuno di questi principi consterebbe di una componente oggettiva – che si riferisce al funzionamento strutturale del processo

---

<sup>40</sup> Questa tripartizione è presa dall'analisi di tre testi: Martí (2001, 2006a) ed Elster (1998).

decisionale – e di una soggettiva – che riguarda il tipo di motivazione (Martí, 2006a, pp. 46-ss).

La componente oggettiva ricalca le due fasi in cui si struttura il processo decisionale. Il primo elemento della componente oggettiva è determinato da quell'insieme di regole che stabiliscono come le preferenze debbano essere espresse e trasformate. La seconda componente, invece, che corrisponde a (b), è quella della regola di decisione: la regola, cioè, che permette di trasformare un congiunto di preferenze individuali in una decisione collettiva.

#### **3.2.4.1 Voto, negoziazione, argomentazione**

Col modello del voto si stabilisce che le decisioni debbano fondarsi sulle preferenze – auto-interessate o imparziali – individuali di ogni cittadino. Le preferenze, in questo caso, sono viste come variabili esogene ed immutabili ed è per tale ragione che non si parla di trasformazione delle preferenze, ma solo della loro espressione. Per quanto riguarda la regola di decisione, può essere adottata tanto la regola della maggioranza quanto quella dell'unanimità.

La negoziazione presenta un connotato soggettivo specifico: a differenza del modello del voto le preferenze sono sempre auto-interessate. Non sono estranei a questo modello strumenti come la minaccia, il ricatto, l'inganno e le concessioni mutue. Le preferenze, nella negoziazione, non sono viste come variabili esogene ed immutabili, ma possono, nel corso della discussione, combinarsi (non trasformarsi). Per quanto riguarda la regola di decisione, nella negoziazione occorre raggiungere il “consenso strategico”.

Anche l'argomentazione ha un connotato soggettivo specifico: le preferenze sono sempre imparziali. Esse non sono intese come variabili esogene ed immutabili poiché tutti sono disposti a cambiare opinione per raggiungere la decisione “migliore”. La regola di decisione nel caso dell'argomentazione è quella del raggiungimento del

“consenso ragionato”, che si ottiene con la “forza dell’argomento” utilizzato, e non tramite minaccia, inganno, ricatto, concessioni mutue, ecc.

L’applicazione pura di questi principi, come Martí ha sottolineato, dà vita a procedimenti ideali che sono impossibili empiricamente. Pertanto, i modelli democratici che sorgono dall’applicazione di questi principi si presentano come ideali regolativi.

Dall’incontro fra questi principi e le motivazioni alla base delle preferenze – auto-interessate o imparziali –, secondo Martí (2006a, p. 50), prendono forma quattro modelli democratici.

|                            | Voto                                 | Negoziazione                       | Argomentazione                     |
|----------------------------|--------------------------------------|------------------------------------|------------------------------------|
| Preferenze autointeressate | Teoria della democrazia come mercato | Teoria pluralista della democrazia | <i>Concettualmente impossibile</i> |
| Preferenze imparziali      | Modello rousseauiano                 | <i>Concettualmente impossibile</i> | Democrazia deliberativa            |

41

Il principio del voto può generare due tipi di modelli democratici. Nel caso in cui le preferenze dei partecipanti siano auto-interessate, si ha la teoria della democrazia come mercato. Questa concezione della democrazia sarebbe sostenuta da autori come Schumpeter (2008) e Downs (1957). Se, invece, al principio del voto si coniugassero preferenze imparziali, il modello di democrazia di riferimento è quello rousseauiano.

Il principio della negoziazione non contempla la possibilità che ci siano preferenze imparziali. Combinandosi con preferenze rigorosamente auto-interessate, si ha la teoria pluralista della democrazia. Questa teoria è sostenuta da autori come Dahl (1956, 1989) e Truman (1959).

---

<sup>41</sup> Questa tabella è la traduzione di una tabella che si trova in Martí (2006a, p. 50)

Per ultimo, viene il principio dell'argomentazione: esso non può coniugarsi con preferenze auto-interessate, ma solo con preferenze imparziali. La forma che genera il principio dell'argomentazione è quella della democrazia deliberativa.

Il principio dell'argomentazione – e, nel suo modello ideale, la democrazia deliberativa – conducono, a detta di Martí, sempre alla decisione corretta. Cosa si intende in questo caso per decisione corretta?

#### **3.2.4.2 L'argomentazione e la decisione corretta**

Ho già parlato di “correttezza” con riferimento alle concezioni metaetiche e all'approccio sostanzialista nel rapporto fra democrazia e giustizia. Nel primo caso, si è visto che le concezioni metaetiche oggettiviste condividono l'idea che esista un criterio di correttezza dei giudizi etici. Per alcune concezioni oggettiviste, tale criterio sarebbe unico e univoco. Nel caso del rapporto fra democrazia e giustizia, i fautori di un approccio sostanzialista sostengono che esiste una decisione corretta in materia di giustizia e che essa sia indipendente rispetto alla procedura democratica.

Con riferimento all'argomentazione, Martí sostiene che essa può essere portata avanti solo da soggetti le cui preferenze sono di tipo imparziale – cioè non auto-interessate. Escludere che esistano preferenze auto-interessate è il primo passo per sostenere l'esistenza di una sola decisione corretta. Se ci fossero, infatti, preferenze auto-interessate, vi sarebbero, oltre alla decisione che rappresenta l'*optimum* fra le diverse preferenze individuali, anche una decisione corretta per ciascuna preferenza. Va osservato come a differenza della negoziazione – che pure prevede una discussione che anticipa la decisione – partecipare ad una deliberazione significa comprometersi con l'idea di imparzialità e abbandonare strategie egoistiche (Martí, 2006a, p. 98). Dunque, se le preferenze sono imparziali e non egoistiche, la natura dei partecipanti passa in secondo piano: non è più così rilevante chi davvero partecipa al processo perché gli argomenti e le ragioni, affinché vengano accettati, devono andare oltre le preferenze di ciascuno. In questo senso, i fautori del principio

dell'argomentazione propongono una de-soggettivazione dei processi politici. Ciò è particolarmente evidente in Habermas. A fondamento della teoria del discorso, infatti, egli pone la ragione comunicativa.

(Essa) si distingue dalla ragion pratica anzitutto perché non è più riferita a singoli attori, o a macro-soggetti di natura statale e sociale. Ciò che rende possibile la ragione comunicativa è il *medium* linguistico, attraverso cui s'intrecciano interazioni e si strutturano forme di vita (Habermas 2013, p. 33).

Prendendo le distanze tanto da Rousseau – con il riferimento al macro-soggetto dotato di potere legislativo – quanto da visioni che intendono la formazione legislativa come aggregazione di volontà individuali, Habermas propone non la via del soggetto, ma una via che passa fra i soggetti, intersoggettiva. Come è stato suggerito dallo stesso Habermas, si tratterebbe di una «comunicazione senza soggetto».<sup>42</sup>

La teoria del discorso prende finalmente congedo da quelle figure di pensiero – ancora legate alla filosofia della coscienza – per cui bisognava o ascrivere la prassi d'autodeterminazione dei cittadini a un soggetto sociale collettivo oppure ricondurre l'anonimo “dominio delle leggi” alla concorrenza dei soggetti individuali. Nel primo caso, si vedeva la cittadinanza come un attore collettivo che rispecchiava e agiva in nome della totalità. Nel secondo caso, i singoli attori funzionavano da rotelline, o variabili dipendenti, dei processi del potere – processi funzionanti alla cieca, in quanto al di là degli individuali atti di scelta potevano esserci solo decisioni aggregate, non liberamente consapevoli (Habermas 2013, p. 362).

---

<sup>42</sup> Il concetto di “comunicazione senza soggetto” (Habermas 2013; Habermas 2017) rivela affinità con quello di “*anonymous public conversation*”, che troviamo in Benhabib (1996). Altrove (Dryzek 2006), tale concetto è stato valutato insufficiente per analizzare le “*divided societies*”. Sarebbe, infatti, troppo amorfo considerando che l'identità dei soggetti che prendono parte al discorso è una questione cruciale.

Mi sono già soffermato in 3.2.2.2 sull'ambiguità del concetto di "correttezza" con riferimento alla morale e, quindi, alle decisioni politiche. Quello che importa, a questo punto, è rispondere alla domanda "perché l'argomentazione dovrebbe condurre a decisioni corrette?" Per rispondere a tale domanda è necessario analizzare il legame fra argomentazione e decisione corretta. La particolare disposizione dei partecipanti – che si compromettono con l'ideale di imparzialità – è solo una parte della risposta.

Secondo alcune teorie che si collocano nell'alveo della democrazia deliberativa, è il processo argomentativo lo strumento in grado di scoprire la decisione corretta – decisione che, secondo tali teorie, sarebbe indipendente rispetto alla procedura democratica. La necessità, però, di legare l'argomentazione alla decisione corretta genera, secondo Martí (2006a, pp. 113-124), un dilemma:

(a) ricorrendo a nozioni sostantive predeterminate, si corre il rischio di appoggiarsi a condizioni ipotetiche altamente discutibili che conducono a giustificazioni di tipo circolare;

(b) ricorrendo a nozioni formali che lasciano indefiniti i criteri, sorge il bisogno di specificare tali criteri in ogni comunità sulla base delle regole del proprio dialogo politico già costituito, e ciò comporta il rischio di aprirsi alla minaccia del relativismo.<sup>43</sup>

A tal proposito, Martí prende in esame alcune fra le possibili risposte a questo dilemma. La più promettente sembrerebbe offerta dallo stesso Habermas.

Habermas intende ricostruire una nozione di razionalità che sia universale e che superi i limiti della razionalità strumentale. Egli propone un nuovo tipo di razionalità che emerge in un contesto dialogico e che dà vita alla situazione discorsiva ideale (Habermas, 1987). La situazione discorsiva ideale si caratterizza per due elementi: (I) pretesa di arrivare a un accordo; (II) pretesa di validità.

Habermas cerca, inoltre, di eludere la costruzione di un criterio di correttezza sostantivo, rimettendo a coloro che partecipano al discorso la responsabilità di trovare

---

<sup>43</sup> Sul perché il relativismo rappresenti un problema per Martí – e per chiunque intenda legittimare da un punto di vista oggettivo la democrazia – si veda 3.2.2.2



tale criterio. Se si legge Habermas in questo modo, tuttavia, tale criterio non può che essere locale e particolare, e si ricade nel relativismo. Al contrario, appoggiando la lettura che Elster fa di Habermas, assumere come standard di correttezza la decisione che verrebbe presa nella situazione discorsiva ideale, significa ricorrere a nozioni sostantive predeterminate, definendo il principio dell'argomentazione come quel metodo che si dà nella situazione discorsiva ideale.

Al di là del dilemma “circolarità/relativismo”, non è chiaro perché il principio dell'argomentazione permetta di giungere ad una decisione corretta. In questo senso, risposte più esaustive possono provenire da concezioni epistemiche della democrazia, che verranno prese in esame nel prossimo paragrafo.

Per concludere, il tentativo, attraverso Habermas, di spiegare il nesso fra argomentazione e decisione corretta si è rivelato insoddisfacente. Esso consiste di due passi.

Nel primo passo, si assiste a una esclusione soggettiva: sono eliminati dal processo quei soggetti che presentano preferenze auto-interessate. Le preferenze devono essere intersoggettive e, in qualche modo, de-personalizzate. Il motivo è che solo tali ragioni possono ambire ad essere universalizzabili. Come accade che fra queste ragioni – che ambiscono all'universalizzabilità – vengano selezionate proprio quelle che danno forma alla decisione corretta? A questo proposito – e questo è il secondo passo – la risposta più promettente nell'alveo della democrazia deliberativa sembra essere quella di stabilire che è corretta quella decisione che prende forma nella situazione discorsiva ideale. Tuttavia, la natura sfuggente della “situazione discorsiva ideale” non chiarifica del tutto il nesso fra argomentazione e decisione corretta.

### **3.3 Due concezioni della democrazia**

Il quadro alla fine di queste quattro domande non è ancora chiarissimo. Esiste una continuità fra un certo tipo di concezioni metaetiche, un certo tipo di giustificazioni, l'elezione fra un approccio proceduralista o sostanzialista e la scelta del metodo per

prendere decisioni. Tuttavia, i quattro punti analizzati si riferiscono a quattro distinti domini che si intersecano e, in alcuni punti, si confondono.

Il primo dominio preso in esame è quello delle giustificazioni della democrazia. Come si è visto, possono esserci giustificazioni strumentaliste – in cui la democrazia è considerata un bene perché permette di raggiungere certi risultati – e giustificazioni non-strumentaliste – in cui la democrazia è considerata un bene a prescindere dai risultati che essa produce. Questo dominio è intrinsecamente legato all'altra domanda presa in esame – la democrazia deve fondarsi sull'idea che esiste una decisione corretta? In questo senso, la democrazia viene presa in esame non tanto come forma di governo che prevede la presenza di particolari istituzioni, quanto come metodo decisionale. In quanto metodo decisionale, la democrazia porta necessariamente a delle decisioni. Tali decisioni, trattandosi di decisioni politiche, esprimono dei giudizi etici. Che esista o meno un criterio unico e univoco per stabilire la correttezza dei giudizi etici è precisamente l'ambito del secondo dominio. Per gli autori che credono nell'esistenza del suddetto criterio di correttezza viene a crearsi una distinzione fra due sensi della legittimità – una legittimità in senso soggettivo e una legittimità in senso oggettivo.<sup>44</sup> Una decisione è legittima in senso soggettivo se essa è percepita come tale da parte dei soggetti cui la decisione è applicata, mentre è legittima in senso oggettivo nella misura in cui si adegua a certi principi e valori morali. Per gli autori, invece, che non credono esista un unico criterio di correttezza, la legittimità avrebbe un solo senso, che corrisponde al senso soggettivo. Il concetto di legittimità – in senso oggettivo – è stato equiparato a quello di giustizia (Hampton, 1988; Buchanan, 2002). Per quanto i due concetti vengano più spesso tenuti distinti (Rawls, 1993, 1995), esiste indubbiamente un nesso fra legittimità e giustizia: la legittimità più precisamente sarebbe una condizione per la giustizia (Langvatn, 2016). La domanda rilevante, a questo punto, – e ciò ci permette di entrare nel terzo dominio – è se esiste una decisione giusta, a prescindere dal procedimento democratico. Coloro che rispondono affermativamente a questa domanda hanno un approccio che viene

---

<sup>44</sup> Tale distinzione può anche essere vista come legittimità in senso normativo ed in senso descrittivo. Per Weber (1964), ad esempio, l'unico senso della legittimità sarebbe quello descrittivo.

definito “sostanzialista”. Chi, invece, sostiene che la giustizia di una decisione vada ricercata nella procedura stessa ha un approccio “proceduralista”.<sup>45</sup> Concretamente, in cosa consiste la procedura democratica? Questa domanda consente di entrare nel quarto dominio, in cui vengono definiti i modi per prendere una decisione. Sono stati analizzati tre principi – voto, negoziazione, argomentazione – che danno vita a distinti modelli di democrazia. Fra questi modelli, si distingue quello della democrazia deliberativa che si fonda sul principio di argomentazione. In continuità con la linea strumentalista-oggettivista-sostanzialista, l’argomentazione è stata individuata, da alcuni autori, come il mezzo più adatto per raggiungere la decisione corretta.

Sulla scorta delle riflessioni esposte fino a questo punto, procederò ad isolare due distinte famiglie di concezioni della democrazia. La distinzione che propongo si inserisce all’interno di una tradizione sviluppatasi negli ultimi decenni,<sup>46</sup> soprattutto nell’alveo degli autori “deliberativisti”, e che si fonda, in ultima analisi, sul tipo di risposta che viene data alla domanda “un sistema democratico deve essere giustificato sulla base del fatto che permette di arrivare a decisioni migliori da un punto di vista epistemico?”

Questa domanda tiene insieme le quattro domande che finora abbiamo preso in esame. Più precisamente:

- la democrazia deve giustificarsi sulla base dei risultati che produce?
- esistono decisioni eticamente migliori con riferimento alle questioni politiche?
- che rapporto intrattiene la procedura democratica con queste decisioni?
- qual è la procedura giusta per raggiungere la decisione corretta?

### **3.3.1 Concezioni non-epistemiche**

---

<sup>45</sup> Proceduralista, occorre specificarlo, lo intendo alla maniera in cui Rawls intende il proceduralismo puro.

<sup>46</sup> Mi sto riferendo ad autori come Landemore, Martí ed Estlund.

Alla domanda “un sistema democratico deve giustificarsi sulla base del fatto che permette di arrivare a decisioni migliori da un punto di vista epistemico?” le concezioni non-epistemiche della democrazia danno una risposta negativa. Tale risposta può essere dettata da diversi motivi. Per comprendere questi motivi bisogna scomporre la domanda in più parti. Questa scomposizione segue i quattro punti che sono stati affrontati nel paragrafo precedente: giustificazioni, concezioni metaetiche, rapporto fra giustizia e democrazia, metodi per prendere decisioni.

Per quanto riguarda il primo punto, come si è visto, la domanda rilevante è “su che basi si giustifica la democrazia?” È chiaro che tutti coloro che condividono un approccio non-strumentalista non possono aderire a una concezione epistemica. Se, infatti, la democrazia è un bene in sé e non per i risultati che essa produce, la giustificazione non può fondarsi sulla base del fatto che essa produce una decisione migliore. Si può, dunque, sostenere che la democrazia produca decisioni migliori da un punto di vista epistemico, ma allo stesso tempo non condividere l’idea che la democrazia si possa giustificare per le decisioni che essa produce.

Il secondo punto riguarda l’idea che esistano dei giudizi etici corretti e che sia disponibile un criterio unico e univoco per decretare la correttezza di tali giudizi. Questo punto presenta maggiori criticità rispetto al precedente. In che rapporto sono i giudizi etici con le decisioni migliori da un punto di vista epistemico? O meglio, le decisioni migliori da un punto di vista epistemico devono anche essere corrette da un punto di vista etico? Queste domande verranno analizzate più diffusamente in 3.3.2.

Per ora, è sufficiente osservare che per “decisioni migliori da un punto di vista epistemico” mi riferisco a “decisioni corrette”. Occorre a questo punto capire se tale correttezza riguarda anche gli aspetti etici della decisione. Dal momento che si tratta di questioni politiche, è difficile immaginare come la correttezza non possa riguardare anche gli aspetti etici della decisione. Pertanto, si può sostenere che partendo da concezioni metaetiche soggettiviste e, fra le oggettiviste, quelle pluraliste non armoniche, non si possa non condividere una visione non-epistemica della democrazia. Le altre concezioni metaetiche – quelle oggettiviste, ad eccezione di quelle pluraliste non armoniche – non è detto che supportino necessariamente una

visione epistemica della democrazia, ma questo si vedrà più avanti. In sintesi, a meno che non si voglia sostenere che le decisioni politiche sono svincolate da qualsiasi considerazione etica, bisogna concludere che una concezione epistemica della democrazia deve necessariamente fondarsi sull'idea che esistano dei giudizi etici corretti e che per decretare tale correttezza esista un criterio unico ed univoco. Per sottrazione, quindi, tutte le altre concezioni che non condividono queste tesi sostengono visioni non-epistemiche della democrazia.

Il terzo punto – che riguarda il rapporto fra giustizia e processo democratico – presenta delle singolari problematichità legate a come alcuni autori hanno definito le loro posizioni. Questo aspetto era già emerso in 3.2.3.1 e verrà ulteriormente ripreso in 3.3.2. Si è osservato come nel rapporto fra processo democratico e giustizia si possono assumere due atteggiamenti. Si può, infatti, sostenere che non ci sia alcun legame fra esito del procedimento democratico e giustizia delle decisioni e si può sostenere che la giustizia di una certa decisione dipenda dal procedimento democratico. Coloro che credono che non vi sia necessariamente un rapporto di dipendenza fra giustizia e democrazia hanno una posizione che è definita sostanzialista; coloro, invece, che sostengono che tale rapporto esista sono chiamati proceduralisti.

Perché la differenza fra sostanzialismo e proceduralismo rileva in questa sede? Perché le posizioni proceduraliste sono sicuramente non-epistemiche. Se, infatti, si ritiene che la giustizia di una decisione sia legata al tipo di procedimento che ha portato a quella decisione, non esiste “la decisione corretta”, ma essa è contingente e dipende dal tipo di procedura di volta in volta implementata. Tuttavia, non è detto che una posizione sostanzialista conduca sempre a concezioni di tipo epistemico. Si può verificare, infatti, il caso in cui, pur contemplando l'esistenza di una decisione corretta che prescinde dal procedimento democratico, non si ritiene che la democrazia debba giustificarsi sulla possibilità di raggiungere questa decisione corretta.

Per quanto concerne l'ultimo punto che abbiamo analizzato – che riguarda i distinti metodi in cui si può prendere una decisione in maniera democratica – tanto la negoziazione quanto il voto e l'argomentazione possono in potenza aderire a una

concezione non-epistemica della democrazia. Questi principi, come si è visto, danno forma a diversi modelli democratici e nessuno di questi modelli preclude a priori la possibilità di rispondere in maniera negativa alla domanda che è oggetto di questo paragrafo – la democrazia deve giustificarsi sulla base del fatto che permette di arrivare a decisioni migliori da un punto di vista epistemico?

In conclusione, si è visto che nessuna delle diverse posizioni che interessano i quattro punti analizzati impone una lettura necessariamente epistemica della democrazia, ma, come si vedrà, una concezione epistemica della democrazia è frutto dello specifico posizionamento rispetto ai quattro punti presi in esame.

Pertanto, come è ovvio, sono “non-epistemiche” tutte quelle concezioni della democrazia che non sono epistemiche.

### **3.3.2 Concezioni epistemiche**

L’espressione “concezione epistemica” della democrazia è stata introdotta da Joshua Cohen (1986, p. 34). In un periodo in cui la deliberazione non aveva ancora guadagnato centralità nel dibattito teorico politico, Cohen si dedica all’interpretazione epistemica del voto e segnala che essa si caratterizza per tre aspetti:

- 1) credere esista uno standard indipendente di correttezza delle decisioni: si tratta dell’indipendenza dal consenso e dal risultato dei voti;
- 2) considerare in termini cognitivi il voto: il voto esprime credenze rispetto alle politiche corrette che si accordano allo standard indipendente, e non alle preferenze individuali rispetto alle politiche;
- 3) considerare il processo decisionale come un processo di regolazione delle credenze: si tratta di una regolazione che si realizza sulla base delle prove sulla risposta corretta che vengono fornite dalle credenze degli altri.

Questo discorso può essere esteso – ed è stato esteso (Martí, 2006b) – anche alla deliberazione.

Sosterrò, coerentemente a quanto visto finora, che una concezione epistemica della democrazia consta di tre elementi:

- (a) credere nell'esistenza di uno standard indipendente di correttezza delle decisioni;
- (b) considerare la democrazia come il metodo per scoprire questo standard;
- (c) giustificare la democrazia sulla base di (a) e (b).

Occorre avanzare due riflessioni. (b) presume l'esistenza di (a) e i tentativi di dimostrare empiricamente (b) sono diretti a provare che le decisioni prese in maniera democratica – seguendo determinate tecniche di *decision-making* – siano qualitativamente migliori rispetto a quelle prese utilizzando metodi alternativi. Molti di questi contributi si riferiscono a concetti come “*Collective Intelligence*” o “*Collective Wisdom*” (Surowiecki, 2004; Rheingold, 2003, 2012; Tapscott e William, 2006; Sunstein, 2006; Levy, 1994; Hong e Page, 2004; Page, 2007; Noveck 2009, 2015; Landemore, 2013; Malone e Bernstein, 2022; Mulgan, 2018; Malone, 2018).

Per Martí (2021, p. 209), un'idea come quella di “intelligenza collettiva” ha un'importanza cruciale per la democrazia.

Se vogliamo che le nostre democrazie non solo siano dotate di una legittimità minima, ma decidano e agiscano in modo intelligente, cioè prendano decisioni e leggi migliori, attuino politiche migliori e, in ultima analisi, si governino meglio, dobbiamo comprendere meglio come funziona l'intelligenza collettiva, il che include una domanda su quali siano le condizioni epistemiche in cui gruppi grandi come le comunità politiche possono fare queste scelte intelligenti. Inoltre, se si prendono in considerazione i fattori che portano alcuni a parlare di crisi della democrazia e di crisi della rappresentanza (...), è ancora più importante interessarsi ai modi specifici in cui l'intelligenza collettiva può essere potenziata.<sup>47</sup>

Ci sono diverse definizioni di intelligenza collettiva. Per una di queste, l'intelligenza collettiva è la capacità di prendere decisioni collettive che siano almeno

---

<sup>47</sup> Traduzione mia da Martí (2021).

dello stesso livello o migliori di quelle che potrebbe prendere uno qualsiasi dei membri di un determinato gruppo (Hiltz e Turoff, 1978). Viene anche definita come la capacità di un gruppo di esseri umani di svolgere un compito come se il gruppo stesso fosse un organismo coerente e intelligente che lavora con una sola mente, anziché come un insieme di agenti indipendenti (Malone e Bernstein, 2022).

Per quanto riguarda la seconda riflessione, si osservi come (c) non rappresenta la diretta conseguenza di (a) e (b). In altre parole, pur confidando nell'esistenza di uno standard indipendente di correttezza delle decisioni e considerando la democrazia come il metodo per scoprire questo standard, si può sostenere che la giustificazione della democrazia non risieda nel fatto che essa permette di raggiungere la decisione corretta, ma la giustificazione potrebbe fondarsi su altre ragioni.

Come si è visto in 3.2.3.2, anche la *social choice theory*, secondo Estlund, si rifarebbe ad uno standard di correttezza indipendente rispetto alla procedura stessa e, in una certa misura, punta sulle capacità epistemiche dei gruppi. Si potrebbe, quindi, sostenere che la *social choice theory* e tutte le teorie aggregative della democrazia debbano farsi rientrare nella famiglia delle concezioni epistemiche della democrazia.

Tuttavia, la rilevanza che viene data alla natura auto-interessata delle preferenze rende difficile dichiarare convintamente che la *social choice theory* tenda a uno standard di correttezza realmente indipendente rispetto alla procedura. Per tale ragione, secondo la concettualizzazione di “concezione epistemica” che ho proposto in questa sede, la teoria della scelta sociale non può essere fatta rientrare a pieno nella famiglia delle concezioni epistemiche.

Per concludere, tornerò alle quattro domande prese in esame. Sulla base di quanto è stato detto, puntare su una concezione epistemica significa sostenere che la giustificazione della democrazia risiede nel fatto che essa è un metodo che consente di produrre decisioni migliori rispetto ad altri metodi, assumendo che esistano decisioni oggettivamente migliori (anche da un punto di vista etico), e che esse siano valutabili sulla base di un determinato criterio di correttezza, che si rivela indipendente rispetto alla procedura stessa. Si può, pertanto, concludere che una concezione epistemica della democrazia debba necessariamente essere strumentalista



(per quanto riguarda la giustificazione), oggettivista (da un punto di vista metaetico) e sostanzialista (con riferimento al rapporto fra democrazia e giustizia).

Resta da analizzare l'ultimo dei quattro punti: una concezione epistemica della democrazia è compatibile con i principi di argomentazione, negoziazione e voto o solo con alcuni di essi? E se sì, con quali? In proposito, tornando alla separazione fra preferenze auto-interessate e preferenze imparziali, bisogna osservare che ciò che esclude l'esistenza di una decisione corretta è immaginare che il procedimento democratico sia alimentato da preferenze auto-interessate. Se, infatti, ad un processo decisionale partecipassero soggetti portatori di istanze egoistiche, esso non produrrebbe una decisione corretta poiché vi sarebbe una decisione corretta per ciascuno dei soggetti coinvolti nella decisione. A questo proposito, se si intende aderire ad una concezione epistemica della democrazia non si può anche sostenere che il principio che muove le procedure democratiche sia la negoziazione che, come si è visto, si caratterizza per un connotato soggettivo specifico, che è quello di preferenze meramente auto-interessate. Per quanto riguarda il voto e l'argomentazione, essi possono aderire a una concezione epistemica della democrazia solo a condizione che siano espunti da qualsiasi connotato soggettivo auto-interessato. In altre parole, perché vi sia una concezione epistemica della democrazia bisogna supporre che il processo democratico sia la sede in cui si incontrano preferenze intersoggettive, in grado di generare una decisione che sia corretta per tutti e per nessuno in particolare.

Diversi autori hanno appoggiato una concezione epistemica della democrazia.<sup>48</sup>

In particolare, per Estlund (2008, p. 8), «le leggi prese democraticamente sono legittime e autorevoli perché prodotte da un procedimento che tende a prendere decisioni corrette».<sup>49</sup>

---

<sup>48</sup> Ci sono diversi contributi che difendono una concezione epistemica della democrazia. Di seguito solo alcuni esempi: Estlund (2008); Anderson (2006); Geenens e Tinnevelt (2009); Goodin (2003); List e Goodin (2001); Williams (2002).

<sup>49</sup> Traduzione mia da Estlund (2008).

Per Nino (1997), ad esempio, il valore della democrazia risiede nella sua natura epistemica: la democrazia è un mezzo per raggiungere il “vero”, il “giusto” nel campo della morale; si tratterebbe del migliore strumento per scoprire i principi morali che reggono le nostre comunità politiche.

Egli intende la correttezza in termini di coerenza coi requisiti di giustizia. In maniera altrettanto incisiva, si esprimono List e Goodin (2001):

Il tratto distintivo dell'approccio epistemico, in tutte le sue forme, è la sua premessa fondamentale, cioè che esiste un fatto indipendente dalla procedura per quanto riguarda il risultato migliore o giusto. Un approccio epistemico puro ci dice che le nostre regole di decisione sociale dovrebbero essere scelte in modo da seguire questo fatto.<sup>50</sup>

La ragione per cui ho proceduto ad isolare queste due concezioni della democrazia è stata già anticipata. L'ipotesi di partenza è che una certa concezione della democrazia agevola e favorisce la partecipazione e l'intervento delle macchine nei processi decisionali. La concezione deputata a ciò è quella epistemica. Perché la concezione epistemica, fra i diversi modi di intendere la democrazia, sia quella più amica alle macchine sarà oggetto del prossimo capitolo.

### **3.3.2.1 I sistemi deliberativi e le concezioni epistemiche**

Esiste una scuola di pensiero – o una corrente – che ha raccolto l'eredità delle concezioni epistemiche della democrazia? Ciò che sostengo è che le concezioni epistemiche – così come le ho descritte in questa sede – hanno trovato una delle loro declinazioni più fortunate nei “*deliberative system*”.

L'espressione “*deliberative system*” compare per la prima volta nel saggio di Mansbridge (1991) “*Every day talk in the deliberative system*”. L'obiettivo di

---

<sup>50</sup> Traduzione mia da List e Goodin (2001).

Mansbridge è quello di estendere il concetto di democrazia deliberativa e di cogliere l'aspetto deliberativo del sistema nel suo insieme.

Coloro che si inseriscono in questa corrente osservano come negli ultimi decenni ci sia stato un diffuso consenso sul fatto che la legittimità della democrazia dipendesse in parte sulla qualità della deliberazione. Per promuovere la qualità della deliberazione, sono state elaborate due strategie. Una di queste si concentra sulla deliberazione nei corpi legislativi. L'altra strategia, invece, ha promosso l'implementazione di piccole iniziative deliberative nelle quali i cittadini possono deliberare in condizioni relativamente favorevoli, come, ad esempio, i *mini-public*.<sup>51</sup> Per i fautori di un approccio sistemico alla democrazia deliberativa, il limite di entrambe le strategie è che esse si concentrano solo su singole sedi deliberative e non sulla loro interdipendenza all'interno di un sistema più vasto (Mansbridge, Bohman, Chambers, Christiano, Fung, Parkinson, Thompson e Warren, 2012).

In quello che viene definito il "Manifesto" della svolta sistemica,<sup>52</sup> si dichiara che nessun forum, anche se rispetta tutti i parametri deliberativi, può singolarmente possedere la capacità deliberativa sufficiente a legittimare la maggior parte delle decisioni e delle politiche adottate in una democrazia. Sarebbe, quindi, necessario andare oltre lo studio di singoli istituti e processi, per esaminare le interazioni del sistema nel suo insieme. Una tale affermazione deriva dal riconoscimento che la maggior parte delle democrazie sono complesse entità nelle quali un'ampia varietà di istituti, associazioni e altri fora – anche informali – realizzano un lavoro politico. Questi autori promuovono, in definitiva, un approccio sistemico alla democrazia deliberativa.

---

<sup>51</sup> Per "*mini-public*" si intendono arene deliberative costitutive di un gruppo di cittadini che vengono identificati come un microcosmo rappresentativo di ciò che tutti penserebbero se potessero disporre di una sufficiente dotazione cognitiva e informativa sui problemi da discutere, e potessero propriamente riflettere e deliberare in condizioni di isolamento dalle pressioni distorcenti dell'ordinaria comunicazione (Fung, 2003).

<sup>52</sup> Per "manifesto" mi riferisco a "A systemic Approach to Deliberative Democracy" di Mansbridge, Bohman, Chambers, Christiano, Fung, Parkinson, Thompson e Warren (2012).

In proposito, occorre rilevare come nella riflessione su quali siano i limiti dei sistemi deliberativi, vengano indicati come campi di studio delle arene decisionali che, solo indirettamente, si possono definire democratiche. In particolare, le arene individuate sono quattro: le decisioni vincolanti degli stati; le attività direttamente collegate alla preparazione di queste decisioni; discussioni informali collegate alle decisioni vincolanti; arene di discussione formali o informali collegate a decisioni su questioni di interesse pubblico che non sono destinate a decisioni vincolanti dello stato (Mansbridge, Bohman, Chambers, Christiano, Fung, Parkinson, Thompson e Warren, 2012, p. 9).

All'allargamento del focus, e al cambiamento di scala, consegue un cambiamento nella composizione soggettiva di coloro che partecipano al processo decisionale. Il discorso, fondamento del *deliberative turn*, non è più solo fra individui, ma diventa scambio informativo fra entità di diversa natura.

Questo nuovo approccio si dota di un metodo per valutare il sistema. A questo fine vengono proposte delle funzioni/parametro per tale valutazione. La prima è detta "epistemica" e riguarda la produzione di decisioni che sono propriamente informate dai fatti e dalla logica e che sono il risultato di concrete e significative considerazioni legate a ragioni rilevanti. La seconda funzione – più controversa della prima – è quella etica. Questa funzione si riferisce alla capacità di un sistema deliberativo di promuovere un mutuo rispetto fra i cittadini. Infine, la funzione democratica riguarda la promozione di processi politici inclusivi in termini di equità (Mansbridge, Bohman, Chambers, Christiano, Fung, Parkinson, Thompson e Warren, 2012, pp. 11-12).

Un'ulteriore novità introdotta dai fautori dell'approccio sistemico riguarda la maniera in cui valutare le qualità deliberative di un sistema. In particolare, le valutazioni non andrebbero effettuate sulla maniera in cui operano i singoli istituti o le singole parti del sistema, ma sul sistema nel suo insieme (Owen e Smith, 2015, p. 214). In definitiva, anche se nessuno degli istituti democratici dovesse rispettare gli standard deliberativi, il sistema complessivamente potrebbe esserlo (Owen e Smith, 2015, p. 216).

Seguendo questa tesi, dunque, non c'è una diretta proporzionalità fra istituti deliberativi e grado deliberativo del sistema. Detto altrimenti, non è automatico che all'aumentare degli istituti che rispettano gli standard deliberativi ci sia una crescita del livello deliberativo del sistema perché questo livello è dato da un gioco di equilibri. Gli autori del "Manifesto" sostengono che un tale approccio sia utile anche per comprendere che gli equilibri del sistema non sono decisi unicamente dagli istituti dotati di potere decisionale giuridicamente vincolante, ma anche dalle discussioni informali, che si riferiscono alle decisioni vincolanti, e dalle arene, formali o informali, legate a problemi di interesse pubblico non direttamente connesse a decisioni vincolanti. Si sostiene, ad esempio, che una retorica e dei media particolarmente faziosi, fungendo da contrappeso, potrebbero aiutare a realizzare un sistema autenticamente deliberativo (Mansbridge, Bohman, Chambers, Christiano, Fung, Parkinson, Thompson e Warren, 2012).

Manca, nell'approccio sistemico, una riflessione diretta sulle giustificazioni della democrazia.

Innanzitutto, non bisogna confondere quelle che abbiamo definito "funzioni" con le giustificazioni. Le funzioni – come vengono intese dai sistemici – non rappresentano il fondamento di legittimità della democrazia e neanche le ragioni o le condizioni di esistenza del sistema stesso, ma sono solo parametri utili per operare valutazioni.

Una questione diversa riguarda i limiti del sistema deliberativo: essi non hanno il ruolo normativo svolto dalle giustificazioni, ma sono utili per delimitare il campo di applicazione dei sistemi deliberativi per operare uno studio sulla democrazia.

In definitiva, anche l'approccio sistemico ci restituisce un'immagine della normatività non chiara. Esso viene presentato come il superamento di due concezioni troppo ristrette di democrazia deliberativa. Tali concezioni – quella che si dedica solo alla deliberazione dei corpi legislativi e quella che mira a costruire assemblee in cui i cittadini deliberano in condizioni favorevoli – falliscono nel comprendere la realtà ultima del sistema. In questo modo, l'approccio sistemico trova la sua giustificazione nella supposta capacità di comprendere il processo di formazione degli output a

partire dagli input. In altre parole, si tratta di una teoria della manipolazione dei processi decisionali che riguardano questioni di interesse pubblico: ciò che rileva è la trasformazione da input in output (la decisione) e non la conoscenza del sistema stesso. La conoscenza è strumentale e funzionale alla costruzione di un sistema che persegue obiettivi che corrispondono alle funzioni che sono state prefissate.

Si tratta, in questo senso, di una giustificazione epistemica di un genere diverso rispetto a quella del modello deliberativo habermasiano. In questo caso, infatti, la giustificazione non risiede nelle possibilità del metodo democratico di raggiungere la migliore decisione, ma riguarda la capacità dell'approccio sistemico/deliberativo di conoscere le regole di manipolazione dei processi decisionali.

Una tale giustificazione è epistemica in quanto ambisce a conoscere la realtà democratica in una maniera che sia sufficiente a padroneggiare la sua manipolazione e non per restituire un'immagine fedele di tale realtà. La sua legittimità risiede, in ultima analisi, nelle capacità di questo approccio di conoscere le regole di trasformazione degli input in output.

In conclusione, l'approccio sistemico stravolge la classica struttura normativa della democrazia. Ciò avviene considerando non più la democrazia come un metodo decisionale, ma come la realtà entro cui si sviluppano i processi decisionali. Per questa ragione, non è più necessario fornire una giustificazione della democrazia, poiché la democrazia stessa è estromessa dal processo giustificativo e prende il ruolo della realtà da analizzare. Allo stesso tempo, nel discorso dei sistemici, la democrazia – intesa come inclusione in termini di equità – diventa uno dei parametri, fra gli altri, con cui giudicare un sistema deliberativo inserito in un contesto democratico.

Parallelamente, anche il significato di deliberazione cambia. Essa non riguarda più solo uno scambio fra esseri umani, ma si estende per includere anche la comunicazione fra soggetti diversi dal singolo individuo, fino, come si vedrà nell'ultimo capitolo, a non includere nessun individuo.



## **CAPITOLO 4 – Il ruolo delle macchine nei processi decisionali democratici**

### **4.1 Processi decisionali e macchine**

### **4.2 I due ruoli delle macchine**

#### **4.2.1 La macchina come attore in un processo decisionale**

##### **4.2.1.1 La questione della personalità giuridica**

##### **4.2.1.2 Una macchina può essere titolare di diritti?**

##### **4.2.1.3 La macchina come agente morale**

#### **4.2.2 La macchina come elaboratore e trasformatore di preferenze**

##### **4.2.2.1 La macchina come parte del processo decisionale o come processo decisionale?**

##### **4.2.2.2 Sull'autonomia della macchina-programma**

### **4.3 La compatibilità della partecipazione delle macchine con le diverse concezioni della democrazia**

#### **4.3.1 Compatibilità con le concezioni non-epistemiche**

##### **4.3.1.1 La macchina-attore e le concezioni non-epistemiche**

##### **4.3.1.2 La macchina-programma e le concezioni non-epistemiche**

#### **4.3.2 Compatibilità con le concezioni epistemiche**

##### **4.3.2.1 La macchina-attore e le concezioni epistemiche**

##### **4.3.2.2 La macchina-programma e le concezioni epistemiche**

### **4.4 La macchina che delibera**



## CAPITOLO 4

### 4.1 Processi democratici e macchine

Che rapporto esiste fra processi decisionali democratici e macchine? I termini della questione appaiono, a questo punto, più chiari: nel secondo capitolo è stato definito il concetto di macchina, mentre nel terzo capitolo sono state isolate due concezioni della democrazia che fondano (almeno) due tipi di processi decisionali. L'ipotesi di partenza è che una certa famiglia di concezioni – quella epistemica – permette e favorisce l'accesso delle macchine ai processi decisionali, mentre un'altra – quella non-epistemica – no. Dimostrare la fondatezza di questa ipotesi è oggetto del quarto capitolo e, in fondo, di questa ricerca.

I processi decisionali che si delineano a partire dalle due distinte concezioni della democrazia tratteggiate – quella epistemica e quella non-epistemica – non danno vita a specifici modelli democratici, che si caratterizzano per un particolare assetto istituzionale. L'ipotesi di compatibilità riguarda, più che altro, il piano dell'ammissibilità, in linea teorica, di elementi non umani ed esogeni al tessuto comunitario e dell'adozione di tecniche automatizzate per l'aggregazione e l'elaborazione delle preferenze e per la successiva trasformazione delle stesse in una decisione. Non mi dedicherò, inoltre, alle possibilità tecniche né all'effettiva implementazione di tali tecniche: assumo che esistano macchine in grado di svolgere determinati compiti.

Partirò dall'ipotesi che le macchine possono giocare due tipi di ruoli:

- (a) come attore in un processo decisionale;
- (b) come elaboratore e trasformatore di preferenze individuali.

(a) e (b) saranno oggetto rispettivamente del paragrafo 4.2.1 e 4.2.2. In definitiva, (a) e (b) rappresentano i due modi in cui si può concepire una macchina: come soggetto agente o come programma. Nella seconda parte del capitolo si valuterà la compatibilità di (a) e (b) con le concezioni non-epistemiche e, poi, con quelle

epistemiche. Per quale ragione, quindi, un certo modo di concepire la democrazia consentirebbe e favorirebbe l'accesso delle macchine in un processo decisionale? E, in secondo luogo, una concezione epistemica della democrazia consente e favorisce l'accesso sia per un intervento di tipo (a) che per uno di tipo (b)?

## **4.2 I due ruoli delle macchine**

Nel secondo capitolo è stata data una definizione di macchina. Tale definizione si fonda su due criteri: quello dell'autonomia e quello dell'intelligenza. Una macchina autonoma e intelligente può rivestire vari ruoli, io ne ho isolati due: come attore e come elaboratore di decisioni. La distinzione fra macchine come attori agenti e macchine come programmi è stata proposta in Franklin e Graesser (1996). In quella tassonomia, la macchina-agente era posta su un piano diverso: essa era considerata autonoma, a differenza della macchina-programma.<sup>53</sup> Per Franklin e Graesser, un'entità è agente nella misura in cui è autonoma e, dunque, si differenzia rispetto a un programma proprio perché un programma non è autonomo. Non adotterò questa tassonomia e partirò dal presupposto che tanto un programma quanto un agente possano essere entità autonome.

Occorre, dunque, fondare la distinzione fra agente e programma su un altro criterio. Tale distinzione sarà costruita a partire non tanto sulle qualità della macchina, quanto sul ruolo che essa svolge all'interno del processo democratico.

### **4.2.1 La macchina come attore in un processo decisionale**

---

<sup>53</sup> La definizione in Franklin e Graesser (1996, p. 25) di agente autonomo è la seguente: «*An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future*».

Può una macchina partecipare a un processo decisionale democratico in qualità di attore? Prima di rispondere a questa domanda, occorre stabilire cosa è in grado di rendere una macchina un attore di un processo decisionale.

Assumere che una macchina sia un attore di un processo decisionale significa assimilarla a un'entità umana. In altre parole, vuol dire sostenere che una macchina può partecipare a una decisione così come vi partecipano gli esseri umani.

Non si tratta, come ho sostenuto fino ad ora, di valutare le possibilità tecniche di partecipazione. Ciò di cui mi occuperò in questo paragrafo è la comprensione delle implicazioni che derivano dal considerare una macchina come un attore in un processo decisionale.

Pertanto, occorre domandarsi cosa fa di un soggetto un attore in un processo decisionale. E, in secondo luogo, cosa lo legittima a partecipare a un processo decisionale?

#### **4.2.1.1 La questione della personalità giuridica**

Alcune questioni legate alle domande appena presentate sono state anticipate dalle riflessioni sulla possibilità di attribuire personalità giuridica alle macchine (Solum, 2020; Hildebrandt, 2019; Bryson, Diamantis & Grant, 2017; Chopra & White, 2004; Koops, Hildebrandt, Jaquet-Chiffelle, 2010; Solaiman, 2017; Asaro, 2007).

Prima ancora della personalità giuridica, come osservano Chopra e White (2004, p. 635), un dibattito filosofico ricorrente riguarda la possibilità di ascrivere personalità, in senso lato, agli agenti artificiali. A tal fine, viene solitamente stilata una lista di condizioni necessarie e sufficienti perché a un agente artificiale vengano riconosciute capacità cognitive come quelle di un essere umano. Si fa riferimento a condizioni come l'intenzionalità, la possibilità di avere un'esperienza fenomenica cosciente o la possibilità di esercitare il libero arbitrio. Dal momento che vi è disaccordo su queste condizioni, è stato suggerito che si potrebbe trarre ispirazione proprio dalle categorie legali.

Innanzitutto, il concetto di personalità giuridica può essere visto dal punto di vista dell'agentività. L'agentività legale va tenuta distinta dall'agentività morale. Mentre l'agentività morale è generalmente intesa come la capacità di intraprendere azioni intenzionali e, quindi, come la capacità di rendere ragioni per una certa azione, l'agentività legale è la capacità, attribuita dalla legge, di agire secondo diritto e di essere ritenuti responsabili per le proprie azioni (Hildebrandt, 2019, p. 8).

Questa distinzione produce distinti ruoli nel sistema giuridico-politico. A un attore dotato di agentività giuridica (e quindi di personalità giuridica), ma non di agentività morale, potrebbero essere precluse alcune prerogative, come il riconoscimento dei diritti fondamentali e il poter essere ritenuto responsabile da un punto di vista penale.

L'esempio tipico di un soggetto dotato sia di agentività giuridica che morale è la persona naturale; mentre per quanto riguarda un attore dotato di agentività giuridica ma non morale, il più classico degli esempi è quello dell'impresa.

Per riconoscere lo “*status ridotto*” di persona giuridica a una macchina si può fare ricorso a un argomento analogico. La tesi è a grandi linee la seguente: se viene attribuita personalità giuridica alle imprese, per quale ragione essa non potrebbe essere attribuita ad una macchina? È stato osservato, infatti, che il riconoscimento di personalità giuridica è sempre un artificio ed esso non deve essere necessariamente connesso a una nozione metafisica o etica di personalità (Bryson, Diamantis & Grant, 2017, p. 277).

Inoltre, sulla possibilità (e opportunità) di dichiarare una macchina una persona giuridica, almeno per alcune responsabilità civili, oltre ai vari contributi scientifici, si è espresso il Parlamento europeo con una raccomandazione.<sup>54</sup>

---

<sup>54</sup> In European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), all'art. 59 f) si legge: «*creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently*».

Accanto a questo “*status ridotto*”, è stata prospettata anche la possibilità di garantire a una macchina alcuni diritti fondamentali (Solum, 2020, pp. 1255 ss.). Se accordare agentività giuridica a una macchina non equivale a garantirle la possibilità di partecipare a un processo decisionale democratico in qualità di attore, probabilmente, riconoscendo ad essa alcuni diritti fondamentali, si potrebbe riuscire in questo proposito. Il riconoscimento di diritti fondamentali, infatti, sembra essere la condizione minima per permettere a una macchina di partecipare politicamente. Bisogna, quindi, volgere lo sguardo verso un’altra direzione.

#### **4.2.1.2 Una macchina può essere titolare di diritti?**

Alla possibilità di riconoscere la titolarità di diritti a una macchina sono stati sollevati tre tipi di obiezioni (Solum, 2020, pp. 1258 ss.). “Solo gli esseri umani possono essere titolari di diritti” è la prima obiezione. La seconda obiezione, in continuità con la prima, riguarda il fatto che a una macchina mancherebbero alcuni elementi fondamentali della personalità, come l’anima, l’intenzionalità, la coscienza o le emozioni. La terza obiezione è che le macchine, in quanto creazioni umane, non possono essere considerate altro che proprietà umana.

A favore della prima tesi, c’è un argomento chiaro e semplice: solo le persone possono essere titolari di diritti perché “*persons-are-conceptually-human*”. Secondo questo argomento, il concetto di persona è inestricabilmente legato all’esperienza umana.<sup>55</sup> A questa tesi si è obiettato che esistono criteri di identificazione della persona che sono concettualmente indipendenti dai criteri per identificare un essere umano. Fra questi ci sono il possesso di credenze e desideri di secondo ordine – cioè credenze e desideri sulle proprie credenze o sui propri desideri.<sup>56</sup>

Per quanto riguarda la seconda obiezione, essa è anche definita “l’argomento del *missing-something*”. Questa obiezione è strettamente connessa – e complementare –

---

<sup>55</sup> Sul tema si vedano i contributi in Gill (1990) e il lavoro di Wiggins (2001).

<sup>56</sup> Si veda in proposito Frankfurt (1988).

all'argomento del "*persons-are-conceptually-human*" ed ha una struttura che si articola in tre punti:

- la qualità X è essenziale per definire la personalità;
  - nessuna macchina possiede la qualità X;
  - la produzione di comportamenti che possono essere identificati con X
- dimostra che una macchina può solo simulare X, ma la simulazione di una cosa non è la cosa stessa.

La qualità che manca è identificata spesso con concetti come “anima”, “desideri” e “intenzionalità” che definiscono gli umani come persone.<sup>57</sup>

Partendo da due di questi concetti, proporrò due obiezioni che possono essere sollevate.

Fra i più risalenti limiti attribuiti alle macchine ci sarebbe la loro incapacità di intendere. Un'obiezione del genere ci rimanda al dibattito sulla stanza cinese di Searle (1999). Searle immagina di essere in una stanza assieme ad un computer. Il suo compito è quello di rispondere a delle domande in cinese che gli vengono passate attraverso una fessura sotto la porta. Searle non conosce il cinese, ma consultando le indicazioni del computer, invia risposte corrette in cinese fuori dalla porta, e questo induce chi è fuori a supporre erroneamente che ci sia qualcuno che comprenda e parli cinese nella stanza.

L'esperimento della stanza cinese intende dimostrare che una macchina è in grado di dare una parvenza di comprensione della lingua, ma questo non significa che essa possa realmente comprendere. Detto altrimenti, Searle sostiene che i computer si limitano ad usare regole sintattiche per manipolare stringhe di simboli, ma non hanno comprensione del significato o della semantica.

L'esperimento della stanza cinese ha subito forti critiche. Quella che ai nostri fini risulta più rilevante è nota come “*the Brain Simulator reply*”, sostenuta fra gli altri da Paul e Patricia Churchland (1990). Questa critica suggerisce di immaginare un cervello artificiale che simula il funzionamento di quello umano: anziché operazioni

---

<sup>57</sup> Si veda Haugeland (1981) che definisce questa strategia “*hollow shell*”.

su stringhe di simboli, questo cervello artificiale imita le sequenze di attivazione neurale che si verificano nel cervello di un soggetto che parla cinese quando lo comprende. Dal momento che questo cervello artificiale funziona come il cervello di un essere umano che parla cinese (poiché processa le informazioni come fosse un cervello naturale), si deve concludere che un computer del genere è in grado di comprendere il cinese.<sup>58</sup> Una tale critica mira a riconsiderare le possibilità di comprensione – e quindi di intesa – di una macchina: mettere in discussione l’assunto che le macchine non sono in grado di pensare mostra come non vi sono argomenti sufficientemente solidi per sostenere che le macchine non sono in grado di intendere.

Si tratta di un capovolgimento dell’onere della prova: quello che occorre dimostrare non è la capacità delle macchine di pensare, ma la loro incapacità. In altre parole, il dibattito sulla stanza cinese mostra come non ci siano ragioni fondate per escludere a priori la capacità di pensare di una macchina. In assenza di tali argomenti, la prima obiezione sulla possibilità di una macchina di prendere parte al discorso non può essere accettata.

La seconda obiezione relativa all’argomento “*missing-something*” riguarda il concetto di interesse. La partecipazione, ma prima ancora, la titolarità di diritti è strettamente connessa al concetto di interesse.<sup>59</sup> Può una macchina avere degli interessi? Se si dimostrasse, infatti, che le macchine non possono avere interessi, allora non potrebbero essere titolari di diritti e, tantomeno, potrebbero prendere parte a un processo decisionale. L’obiezione può essere formulata come segue: un’entità incapace di comprendere non può avere coscienza di se stessa, quindi dei suoi interessi.

Questa obiezione, a dire il vero, si rivela anche più debole dell’obiezione precedente. Il concetto di interesse, infatti, può venire sostituito da quello di obiettivo. Se è difficile accettare che una macchina possieda una coscienza, non si può

---

<sup>58</sup> Oltre ai Churchland, si sono occupati di questa tesi Chalmers (1996), Cole & Foelber (1984), Pylyshyn (1980).

<sup>59</sup> Si veda Raz (1988, p. 176) che afferma in maniera netta che «*that right-holders are creatures who have interests*».

escludere che essa sia in grado di perseguire degli obiettivi, che corrispondono ai suoi interessi. Avendo disgiunto il concetto di interesse da quello di coscienza, le resistenze nell'accettare l'incapacità di una macchina di pensare non danno vita a un argomento rilevante.

Riusciamo a escludere completamente che una macchina possa perseguire un determinato interesse? O a escludere che essa possa essere interessata a che le cose vadano in una certa maniera anziché in un'altra? Non trovando ragioni che persuadano sull'inesistenza di interessi da parte di una macchina, una tale obiezione non può essere tenuta in considerazione.

La terza ed ultima obiezione contro il riconoscimento di diritti a una macchina riguarda il fatto che, in quanto artefatto, una macchina non può essere considerata altro che un bene sotto la proprietà dei suoi produttori. Posta in altri termini, secondo questa obiezione, anche se volessimo considerare la macchina una persona, essa sarebbe da considerare uno "schiavo per natura" e i creatori della macchina sarebbero coloro legittimati a possederla.

Secondo Solum (1992, p. 1279), anche se si ammettesse che il fatto che i creatori di una macchina siano i suoi proprietari e che ciò comportasse che le macchine sono schiavi per natura, sarebbe corretto non riconoscere diritti a una macchina?

In primo luogo, è stato osservato come gli schiavi possono essere emancipati: anche se una macchina è considerata un bene di proprietà, non è detto che essa debba rimanere tale per sempre.

In secondo luogo, anche a uno schiavo possono essere riconosciuti dei diritti.

Queste obiezioni non si sono rivelate decisive. In assenza di ragioni che precludono il riconoscimento di diritti a una macchina, almeno in via ipotetica, è possibile sostenere che una macchina può essere titolare di diritti.

#### **4.2.1.3 La macchina come agente morale**



Che una macchina possa essere titolare di alcuni diritti non vuol dire che essa sia legittimata a partecipare a un processo democratico. Pur considerando la partecipazione politica come uno dei classici diritti dello stato liberale, essa si presenta concettualmente indipendente dal riconoscimento dei diritti. In altre parole, se in un contesto democratico sono legittimati a partecipare a una decisione coloro a cui è riconosciuta personalità giuridica (o a cui sono riconosciuti dei diritti), un tale riconoscimento non comporta automaticamente l'ammissione in un processo decisionale.

Il problema può essere impostato seguendo un'altra strada. L'ipotesi di partenza è che il requisito minimo per la partecipazione a un processo decisionale politico non è il riconoscimento della personalità giuridica – o il riconoscimento di diritti –, ma il poter essere considerato un agente morale. Scopo, dunque, di questo paragrafo è vagliare la possibilità che una macchina sia un agente morale.

Ma che cos'è un agente morale? Nel secondo capitolo, ho adottato la definizione di “agente” di Steels (1995). Questa definizione coincide, in parte, con quella data da Floridi e Sanders (2004), che spingono oltre la loro indagine, definendo il concetto di “agente morale”. In accordo con Allen, Varner e Zinser (2000), mentre un agente, genericamente parlando, deve essere in possesso dei criteri di interattività, autonomia ed adattabilità, un agente è definito morale se e solo se è capace di azioni moralmente qualificabili, dove per moralmente qualificabile si intende un'azione che può causare il bene o il male morale (Floridi e Sanders. 2004, p. 364).<sup>60</sup>

In linea di massima, dunque, una tale definizione di “agente morale” risulterebbe compatibile con quella di “macchina”. Tuttavia, Floridi e Sanders (2004, pp. 364-369), individuano quattro possibili obiezioni.

La prima di queste è l'obiezione teleologica: l'ho già presa in esame in 4.2.1.2 con riferimento alla possibilità che una macchina abbia “interessi”. Dal momento che una macchina, al pari di un essere umano, può esibire comportamenti orientati al

---

<sup>60</sup> È soddisfacente questa definizione? Non occorre un elemento di consapevolezza? Questa definizione può apparire troppo poco esigente e nel prosieguo del paragrafo verranno esaminate alcune obiezioni.

raggiungimento di un obiettivo (Russel e Norvig, 2016), quello teleologico non può essere considerato un criterio utile per stabilire una distinzione fra umani e macchine.

La seconda obiezione è definita “intenzionale”. Secondo questa obiezione, non è sufficiente che una macchina si comporti in maniera teleologica. Per essere un agente morale, infatti, sarebbe necessario un collegamento più profondo fra la macchina e le sue azioni: questo collegamento dovrebbe avere a che fare con il desiderio e la volontà di agire in un certo modo e, quindi, con l’essere consapevole dei propri comportamenti. Questa obiezione è stata già sollevata al paragrafo 4.2.1.2, con riferimento alle credenze e ai desideri di secondo ordine (Frankfurt, 1988) e presuppone la possibilità di un accesso privilegiato agli stati mentali dell’agente. Tuttavia, tali stati non sono direttamente osservabili e, non potendo garantire un’analisi del genere nella pratica, il criterio dell’intenzionalità non permette di escludere le macchine dal novero degli agenti morali. Si tratta, in altre parole, dell’obiezione nota come “*the Other Minds Reply*”.<sup>61</sup> Essa può essere articolata come segue: come si fa a sapere che qualcuno è in possesso di uno stato intenzionale? Solo sulla base del suo comportamento. Se, dunque, una macchina è in grado di superare un test comportamentale tanto quanto qualsiasi altra mente e se si intende attribuire la capacità di comprensione ad altre persone, in linea di principio, bisogna attribuire tale capacità anche alle macchine.

La terza obiezione è quella relativa alla libertà. Secondo questa obiezione, gli esseri umani sono, a differenza delle macchine, liberi. Tuttavia, data la definizione di macchina data nel secondo capitolo, considerando, quindi, le macchine come entità non deterministiche, si può sostenere che anche le macchine sono in una certa misura libere.

Secondo l’ultima obiezione – quella che Floridi e Sanders (2004, pp. 366-369) considerano la più forte – le macchine non possono essere considerate agenti morali perché esse non sono moralmente responsabili per le loro azioni. Ciò è legato al fatto che ritenere responsabile una macchina per le proprie azioni sarebbe concettualmente

---

<sup>61</sup> Di questa critica si sono occupati Dennett (1987) e Moravec (1999).

improprio. In altre parole, le tre condizioni utili a definire un agente (Allen, Varner e Zinser, 2000) – interattività, autonomia, adattabilità – sarebbero condizioni necessarie ma non sufficienti per riconoscere agentività morale a una certa entità. La quarta condizione sarebbe quella della responsabilità ed è proprio questa condizione che permette di distinguere gli agenti morali – che devono necessariamente essere umani – da altre entità, che non sono nulla di più che cause efficienti. Secondo questa obiezione, quindi, un agente può essere considerato morale solo se è responsabile da un punto di vista prescrittivo. Per Floridi e Sanders (2004, pp. 367-ss), questo tipo di obiezione rivela una confusione di fondo: l'identificazione di *x* come agente morale con la valutazione di *x* come un agente moralmente responsabile. Infatti, ciò che questa obiezione sostiene è che l'identità (come agente morale) senza la responsabilità (come agente morale) è senza significato. Viene offerto l'esempio dei bambini: anche se i bambini sono in grado di compiere il bene o il male morale, essi non sono ritenuti responsabili per le loro azioni. Ciò significa considerarli come fonti d'azione morale, anche se non possono ancora essere soggetti al processo di valutazione morale. Se non li considerassimo fonti di azione morale per il fatto che non possono essere soggetti ad un processo di valutazione morale, dovremmo sostenere che i bambini non sono da considerarsi agenti morali.

In conclusione, anche l'ultima obiezione – quella considerata la più forte delle quattro – non preclude in linea di principio la partecipazione di una macchina a un processo decisionale se si assume che il requisito minimo sia il riconoscimento di agentività morale.

Il percorso per stabilire se una macchina può partecipare a un processo decisionale in qualità di attore si è rivelato tortuoso. In un primo momento, è apparso come necessario il raggiungimento del requisito della personalità giuridica che è risultato facilmente accessibile attraverso un argomento analogico: se viene riconosciuta personalità giuridica a un'impresa, perché non può essere riconosciuta anche a una macchina? Si è, quindi, visto come il requisito della personalità giuridica sia tutt'altro che sufficiente per permettere a una macchina di entrare in un processo decisionale. È stata allora intrapresa la strada del riconoscimento di diritti. Se la partecipazione al

processo decisionale democratico è uno dei classici diritti dello stato liberale, allora perché una macchina possa prendere parte a un processo decisionale in un contesto democratico è opportuno che essa possa essere titolare di diritti. Dal momento che le obiezioni contro il riconoscimento di diritti a una macchina non si sono rivelate un reale impedimento, bisogna concludere che in linea teorica una macchina può essere titolare di diritti. Ma anche questo requisito non è stato ritenuto sufficiente. Il requisito ultimo per la partecipazione a un processo decisionale in qualità di attore è stato fissato nel riconoscimento di agentività morale e, secondo il percorso che ho seguito, non ci sono ragioni irresistibili per non accordare tale riconoscimento a una macchina. Tuttavia, il superamento del requisito soglia – l’agentività morale – non spiana definitivamente la strada alle macchine nei processi decisionali su questioni di interesse pubblico.

L’ultima prova – ma questa sarà oggetto del paragrafo 4.3 – riguarda la compatibilità con ciascuna delle due concezioni della democrazia che ho delineato nel terzo capitolo.

#### **4.2.2 La macchina come elaboratore e trasformatore di preferenze**

A partire dalla definizione di agente autonomo data nel secondo capitolo, assumo che una macchina possa essere considerata un agente autonomo non solo nel caso in cui riveste il ruolo di attore in un processo decisionale, ma anche quando le viene affidato il compito di elaborare e trasformare le preferenze individuali. In questo caso, non è richiesto il requisito più stringente della moralità né occorre riconoscere alla macchina in questione personalità giuridica o titolarità di diritti. Non si tratta, infatti, di verificare la possibilità che una macchina abbia il diritto di partecipare a un processo decisionale come portatrice di istanze proprie, quanto quello di valutare se, e in che termini, una parte del processo decisionale sia automatizzabile. Il fatto che una macchina del genere non si presenti come portatrice di interessi propri non è indice di un ruolo “minore” della macchina, con annessa minore difficoltà di integrazione della

stessa all'interno di un processo decisionale. In questo caso, sorgono problemi di ordine diverso.

Ho già parlato di questo tipo di macchina in riferimento al progetto cibernetico cileno. Nella *operation room* di Cybersyn, vi era una macchina – Cyberstride – che aveva il compito di elaborare gli input provenienti dalle aziende e creare proposte di azione.

Una macchina del genere, nella sua versione più elementare, può limitarsi al conteggio dei voti. Appare evidente che una macchina siffatta non ponga alcun problema relativo alla legittimità della sua azione. Problemi di legittimità, invece, potrebbero sorgere nell'ipotesi in cui alla macchina vengano demandati compiti più impegnativi, che corrispondono a un ampliamento dei suoi “poteri di scelta”. A titolo esemplificativo, potrebbe trattarsi di un processo che, aggregando preferenze, trovi una sintesi tra le varie istanze. Stilizzando al massimo:

- (1) l'agente x chiede A;
- (2) l'agente y chiede B;
- (3) la macchina elabora C;
- (4) C diventa legge.

È chiaro che se il processo si fermasse a (3), non esisterebbero problemi di legittimità. Qualora, infatti, la macchina si limitasse a suggerire un certo corso di azione, senza accordarle potere decisionale – o un'autorità indiscutibile –, tale corso di azione si presenterebbe come una semplice proposta, la cui valutazione, ed eventuale approvazione, sarebbe demandata ad organi già dotati di legittimità per decidere.

I problemi, dunque, in ultima analisi, riguardano il riconoscimento di autorità a una macchina in grado di elaborare e rendere vigenti leggi a partire da istanze individuali e collettive presentate nella forma di input da trasformare. La domanda a questo punto appare più chiara: può essere riconosciuta a una macchina l'autorità per prendere decisioni su questioni di interesse pubblico?

Anche in questo caso è opportuno fare la consueta premessa: la questione non riguarda la possibilità, da un punto di vista tecnico, che una macchina sia in grado di

elaborare una decisione su questioni politiche. Assumo che esistano già macchine tecnicamente in grado di farlo. La questione riguarda un altro piano, che è quello della legittimità di scelte di interesse pubblico prese da una macchina. Sono – o possono essere – considerate legittime da un punto di vista democratico delle scelte prese da una macchina? Per una risposta esaustiva rimando all'ultimo paragrafo: non è possibile, infatti, rispondere a questa domanda se non è stato chiarito cosa si intende per “legittimità democratica” e, come è stato preannunciato, un tale concetto dipende dal tipo di concezione di democrazia di riferimento. Non verrà posta questa domanda per ogni singola concezione della democrazia, ma avendo agglomerato le varie concezioni della democrazia in due grandi famiglie di concezioni, mi porrò questa domanda solo due volte – una volta con riferimento alle concezioni di tipo non-epistemico e un'altra con riferimento alle concezioni epistemiche. Se, dunque, la compatibilità con le distinte concezioni della democrazia verrà esplorata in un secondo momento, per ora è opportuno fissare esattamente il ruolo e la funzione di una macchina del genere, così da comprendere in che relazione essa si trova col processo decisionale.

#### **4.2.2.1 La macchina come parte del processo decisionale o come processo decisionale?**

Mettendo per un attimo da parte la questione della legittimità, bisogna domandarsi quale sia esattamente il ruolo di una macchina che raccoglie ed elabora input per trasformarli in una decisione giuridicamente vincolante. In altre parole, è essa stessa il processo decisionale o è solo una parte del processo decisionale? E soprattutto, di che tipo di processo decisionale si tratta? La questione è tutt'altro che voluttuaria.

Assumere che essa sia solo una parte del processo decisionale significa riconoscerle un ruolo limitato e considerare come parte del processo decisionale l'espressione delle preferenze. In base all'immagine della struttura decisionale che sto tratteggiando, alla macchina giungono sotto forma di input le preferenze dei portatori

di interessi e il compito di questa macchina è di generare un output a partire da tali input. Se si assume che una macchina siffatta è solo una parte del processo, allora bisogna riconoscere che il processo decisionale si compone di almeno tre momenti: un momento di espressione delle preferenze, un momento di raccoglimento, elaborazione e trasformazione delle stesse e, infine, la decisione in senso stretto – dare, cioè, corso alle preferenze elaborate. Da questa prospettiva, quindi, gli input non sono considerati elementi esterni al sistema, ma sono parte del sistema stesso. Si può obiettare che, in una rappresentazione del genere, il confine fra il sistema e il suo ambiente è eccessivamente sfumato.

La confusione fra “sistema” e “ambiente” non è l’unico limite di questa rappresentazione. Bisogna fare un passo indietro: cos’è, in fondo, un processo democratico? Esso può essere considerato come un processo cognitivo collettivo. Se si assumessero gli input come parte integrante del processo decisionale, non sarebbe possibile assimilare un processo democratico alla rappresentazione classica dei processi cognitivi. Ma come è strutturato un processo cognitivo?

L’esempio di un sistema cognitivo che tiene l’input fuori dal sistema – permettendo così una distinzione netta fra il sistema e il suo ambiente – è quello visivo, nel modo in cui solitamente viene rappresentato (Marr, 2010). Il sistema visivo è, infatti, definito come quel processo che a partire da immagini del mondo esterno produce una descrizione utile a colui che vede, sgomberando il campo da informazioni irrilevanti (Marr, 1976; Marr e Nishihara, 1978).

Se, dunque, si vuole intendere il processo decisionale come un processo cognitivo, bisogna considerare gli input – cioè l’espressione delle preferenze individuali – come elementi esterni al sistema. In questo senso, le preferenze individuali giocano il ruolo delle immagini del mondo esterno che giungono all’occhio: tali immagini non fanno parte del sistema visivo così come l’espressione delle preferenze non fa parte del processo decisionale. Dunque, in un sistema in cui vi è una macchina che elabora e trasforma gli input degli individui bisogna identificare l’operato della macchina come il processo decisionale nella sua interezza. A questo punto, la questione appare sotto un’altra luce: non si tratta più della possibilità di delegare a una macchina una parte

del processo decisionale, ma si tratta di automatizzare interamente il processo decisionale, così che tale processo finisce per identificarsi con la macchina stessa. Alla domanda “di che tipo di processo decisionale si tratta?”, si può rispondere che esso dipende dalla maniera in cui la macchina è configurata. E l’aspetto democratico? Il requisito della democraticità è soddisfatto con la possibilità di fornire input alla macchina e, eventualmente, accordando medesimo valore ad ogni input. In questo senso, la democraticità oltre a caratterizzare il processo decisionale, lo precede.

#### **4.2.2.2 Sull’autonomia della macchina-programma**

A questo punto, occorre domandarsi se una macchina del genere può essere rispettosa del principio democratico anche nel momento decisionale vero e proprio. È stata avanzata un’ipotesi interpretativa: nel caso in cui una macchina abbia il compito di raccogliere, elaborare e trasformare gli input che le giungono, il processo decisionale deve identificarsi con la macchina stessa. Nel terzo capitolo si è visto che per “democrazia” si intende un metodo per prendere decisioni. In questo senso, l’anticipazione del “momento democratico” a un momento precedente il processo decisionale in senso stretto mette in crisi questa definizione di democrazia. Per questa ragione, nell’ipotesi in cui il processo decisionale venga identificato con l’azione della macchina, la democrazia non sarebbe più un metodo per prendere decisioni, ma un momento che precede il momento decisionale vero e proprio. In questo modo, però, si esclude a priori che la macchina nel decidere – al di là, quindi, dell’apporto che precede il momento decisionale in senso stretto – possa adottare un metodo democratico. Ma può una macchina adottare un metodo democratico decidendo autonomamente?

Immaginiamo di sostituire la macchina in questione con un essere umano dotato di straordinarie capacità cognitive e accertata integrità morale che chiamerò Napoleone. Supponiamo che spetti a Napoleone, dopo aver sentito tutti gli interessati, decidere su questioni di interesse pubblico. C’è una strategia che consente di definire



“democratiche” le decisioni prese da Napoleone? È evidente che un sistema del genere è più vicino a una monarchia che a una democrazia e che si rivela un’operazione estremamente difficile riuscire a considerare democratiche le decisioni prese da Napoleone, nonostante si tratti di decisioni prese ascoltando il parere di tutti coloro che vantano un interesse su una determinata decisione. Nel caso della macchina, il nostro senso comune ha più difficoltà nel decretare con la medesima convinzione che non può trattarsi di una democrazia.<sup>62</sup>

Dov’è, dunque, la differenza? Perché se a decidere fosse Napoleone, non si tratterebbe in nessun caso di una democrazia, mentre se decidesse una macchina, la questione si rivelerebbe meno certa? La ragione potrebbe risiedere nella supposta natura della macchina: essa, a differenza dell’essere umano, è programmabile e riprogrammabile. In questo senso, mentre è concepibile un intervento strutturale sulla macchina tale da rideterminare il modo in cui essa prende le decisioni, non è pensabile riprogrammare un essere umano. Rivendicando, però, la possibilità di poter modificare la macchina, si sta negando la sua natura autonoma – dove per autonomia si intende l’assenza di un controllo eterodiretto – e, come si è visto, l’autonomia rappresenta una qualità essenziale della macchina protagonista di questa analisi.

La questione, a questo punto, si complica ulteriormente e occorre addentrarsi nel concetto di autonomia. L’autonomia può riguardare diversi livelli. Prenderò in esame due di questi livelli, assumendo che la macchina in questione sia autonoma almeno secondo uno di questi livelli. Nel primo livello, un agente è autonomo se è in grado di raggiungere fini prestabiliti scegliendo da sé i propri mezzi: chiamerò questa autonomia “strumentale”. Il secondo livello, invece, ha a che fare con la scelta dei fini. In questo senso, un agente è autonomo se decide da sé i propri fini: chiamerò questa forma di autonomia “teleologica”. Come si pone la macchina-programma rispetto a questi due livelli? Essa è senza dubbio autonoma da un punto di vista

---

<sup>62</sup> Occorre osservare che sostenere che le decisioni prese da Napoleone non siano mai democratiche – mentre quelle prese da una macchina, con qualche sforzo argomentativo, potrebbero esserlo – non equivale a dire che secondo il senso comune sia più auspicabile che a decidere siano le macchine e non Napoleone.

strumentale poiché stabilisce da sé i modi in cui raggiungere il fine. Ma cosa succede per quanto riguarda l'autonomia teleologica? Più precisamente, di che fini si sta parlando? Delle decisioni o del fatto che essa prenda decisioni? Se il fine viene identificato con le singole decisioni che essa prende, allora si può sostenere che la macchina sia autonoma anche da un punto di vista teleologico. Tuttavia, una lettura del genere appare fuorviante poiché la macchina di cui mi sto occupando non viene programmata per prendere decisioni già definite a priori, ma il suo compito è quello di prendere decisioni il cui contenuto è definito di volta in volta dalla macchina stessa. In questo senso, il fine prestabilito è quello di decidere, e non è quello di prendere particolari decisioni. Tornando all'esempio del sistema visivo, si può sostenere che un sistema visivo è autonomo nella misura in cui elabora autonomamente le immagini dall'esterno per fornire delle descrizioni a colui che osserva, ma non è autonomo nel senso di poter decidere se fornire o meno descrizioni a chi osserva. In altre parole, il sistema visivo non decide autonomamente di fornire descrizioni a chi osserva, ma questo è un obiettivo precedentemente fissato. Il livello di autonomia si rivela, quindi, ridotto, ma pone comunque dei problemi di legittimità.

Riassumendo, la macchina-processo decisionale ha una natura autonoma nel senso che:

- sceglie da sé i propri mezzi per raggiungere fini predeterminati;
- formula autonomamente delle decisioni – ma a partire dal comando di prendere decisioni.

Resta, dunque, un'ultima obiezione contro l'autonomia di questa macchina. Si potrebbe sostenere che essa – anche da un punto di vista strumentale – non sia realmente autonoma dal momento che è stata programmata da un agente umano e, in qualche modo, la sua azione – e le decisioni che essa produce – sono destinate ad essere il frutto delle scelte di un programmatore. Secondo Dennett (2018, p. 89):

è interessante ricordare che nel secolo scorso una delle obiezioni più comuni alla GOFAI ("Good old fashioned artificial intelligence") era questa: la cosiddetta

intelligenza di questi programmi in realtà è soltanto l'intelligenza – la comprensione – dei programmatori. I programmi non capiscono un bel niente!

Ho già analizzato questa obiezione e mi limiterò a riprendere l'argomento usato nel secondo capitolo. È sufficiente ricordare che l'avvento del *machine learning* pone dei problemi di ordine diverso rispetto alla maniera in cui viene solitamente intesa la macchina – che è quella GOF AI. La peculiarità del *machine learning* sta nel fatto che la macchina – una volta che le sono forniti input e output – trova da sé il modo per arrivare dall'input all'output. L'apprendimento consiste proprio in questo. Non si tratta più, dunque, della classica struttura “*top-down*” – quella tipica della GOF AI – che si caratterizza per un approccio:

“intellettualistico” all'Intelligenza Artificiale, basato sullo scrivere ciò che sapevano gli esperti umani, in un linguaggio che il computer poteva manipolare con motori inferenziali in grado di perlustrare gli “enormi” banchi di memoria riempiti di questa conoscenza del mondo inserita manualmente con grande cura, deducendo i teoremi necessari per prendere decisioni fondate e controllare nel modo appropriato gli arti o qualsiasi altro effettore dell'intelligenza artificiale (Dennet, 2018, p. 77).

È già stato abbondantemente rilevato come l'evoluzione di tecniche sempre più avanzate di apprendimento automatico abbia portato molti programmi a diventare indecifrabili per i programmatori stessi<sup>63</sup> (Selbst e Barocas, 2018; Zarsky, 2016;

---

<sup>63</sup> Sul punto si veda Dennett (2018, p. 290-291) che sostiene: «Quando un sistema GOF AI ci dice qualcosa, in linea di principio possiamo sempre richiedere una ragione, e questa dovrebbe essere disponibile, poiché qualunque cosa ci abbia detto deve essere stata generata per deduzione logica dal suo database di assiomi. Il sistema può mostrarci la dimostrazione, e noi possiamo criticarla passo per passo se vogliamo. Nelle scienze cognitive molti hanno accolto con sentimenti contrastanti l'avvento dell'“apprendimento profondo” e dei metodi bayesiani. Perché? Il fatto che queste nuove strutture cognitive funzionino così bene è stupefacente e affascinante, e le loro applicazioni si diffonderanno rapidamente in tutto il mondo, ma... anche se ci forniranno risposte importanti a domande difficili

Burrell, 2016). Questa indecifrabilità rappresenta la cifra dell'autonomia della macchina rispetto al suo programmatore: il fatto, cioè, che una macchina apprenda “da sola” come compiere certe operazioni e che un programmatore non sia in grado di ricostruire il processo decisionale che compie la macchina consente di considerare quella macchina autonoma.

La questione sulla democraticità è stata eclissata da quella sull'autonomia. In altre parole, prima di domandarsi se una macchina è in grado di rispettare il principio democratico, si è rivelato necessario sgomberare il campo dalle perplessità che riguardano l'assenza di autonomia di una macchina. Non ci sono state obiezioni tali da impedire di considerare autonoma una macchina. Tuttavia, non sono stato in grado di rispondere alla domanda che ha aperto questo paragrafo e questo è legato a una ragione semplice: per valutare l'aderenza di una macchina siffatta al principio democratico bisogna riannodare il filo del discorso e ritornare alle due concezioni di democrazia che sono state isolate – quella epistemica e quella non-epistemica. Il motivo per cui bisogna ritornare alle concezioni della democrazia delineate nel terzo capitolo è presto detto: tali concezioni danno sostanza e significato alla parola democrazia e per capire se una macchina può comportarsi in maniera democratica bisogna in primo luogo capire cosa si intende per democrazia.

#### **4.3 La compatibilità della partecipazione delle macchine con le diverse concezioni della democrazia**

Nel paragrafo precedente, ho presentato i due ruoli che una macchina può giocare in un processo democratico. Essa può essere attore o può, secondo una certa interpretazione, finire per identificarsi con il processo decisionale stesso. Ad una prima analisi sulle possibilità di integrazione di questi ruoli all'interno di un processo democratico, ho constatato, in entrambi i casi, che è impossibile accertare tale

---

come non è mai avvenuto finora, non saranno in grado di dirci perché. Saranno oracoli che non possono partecipare al gioco di fornire ragioni».

possibilità senza aver prima fissato con esattezza la concezione di democrazia di riferimento. In altre parole, si è visto come in linea di principio non ci siano ostacoli invalicabili per escludere la partecipazione di una macchina a un processo decisionale in qualità di attore o la delega a una macchina di una porzione di decisionalità, ma ciò non è sufficiente a realizzare l'analisi di compatibilità. Inoltre, con riferimento alle macchine che “diventano processo decisionale” è emerso come la questione non riguardi solo l'ammissibilità al processo, ma si debba verificare una condizione ulteriore, cioè che la macchina agisca in modo “democratico”. Nel secondo capitolo sono state isolate due concezioni della democrazia, che coprono il panorama delle concezioni possibili. Più precisamente, dopo aver tracciato il perimetro delle concezioni epistemiche, sono state definite non-epistemiche tutte quelle concezioni che non sono epistemiche. L'ipotesi di fondo è che esiste una famiglia di concezioni – quella epistémica – che non pone limiti particolari per l'accesso delle macchine ai processi decisionali, mentre che vi è un'altra famiglia di concezioni – quella non-epistémica – che pone dei limiti a tale accesso. L'obiettivo, dunque, di questo paragrafo è quello di valutare la compatibilità tra l'intervento delle macchine – nei due ruoli già definiti – e le diverse concezioni della democrazia.

Si tratta, da un altro punto di vista, di comprendere se e in che misura tali interventi delle macchine siano legittimati. È, dunque, chiaro che una tale valutazione di legittimità non può prescindere dalla definizione del contesto di riferimento, che è quello della concezione democratica di partenza.

#### **4.3.1 Compatibilità con le concezioni non-epistemiche**

L'ipotesi che proverò a verificare è che le concezioni non-epistemiche della democrazia sono munite di anticorpi che escludono – o limitano – la partecipazione delle macchine ai processi decisionali.

Come si è visto nel terzo capitolo, le concezioni non-epistemiche della democrazia sono state definite in termini negativi e residuali: sono non-epistemiche tutte quelle

concezioni che non sono epistemiche. Partendo da una definizione del genere, è difficile individuare quali siano gli elementi comuni alle concezioni non-epistemiche che garantiscono un divieto di accesso – o una limitazione – per le macchine ai processi decisionali. Su che basi, quindi, sostengo che le concezioni non-epistemiche della democrazia dispongono di anticorpi che precludono la partecipazione delle macchine? Si tratta di anticorpi normativi e la mia tesi è che, in termini operativi, la differenza fra concezioni epistemiche e concezioni non-epistemiche risieda anche nella presenza di questi anticorpi. Detto altrimenti, ogni concezione epistémica non possiede anticorpi tali da precludere la partecipazione – o la delega – a una macchina in grado di garantire il raggiungimento di decisioni migliori – attraverso la partecipazione come soggetto o come programma di elaborazione –, mentre le concezioni non-epistemiche dispongono di anticorpi tali da precludere – o limitare oltre una certa misura – la partecipazione o la delega a una macchina del genere. Lo scopo delle prossime pagine è quello di mostrare la presenza di tali anticorpi nelle concezioni non-epistemiche e, inversamente, la loro assenza nelle concezioni epistemiche.

In cosa consistono questi anticorpi?

#### **4.3.1.1 La macchina-attore e le concezioni non-epistemiche**

Una volta chiarita la definizione di macchina, la domanda “una macchina può partecipare come attore in un processo decisionale democratico?” rimane ambigua principalmente per due motivi:

- (a) cosa si intende per “partecipazione”;
- (b) cosa si intende per “democratico”.

La partecipazione va in questa sede intesa come la capacità di influenzare una decisione. In questo senso, una macchina partecipa a un processo decisionale se essa può influenzare una decisione.

Avendo stabilito come criterio minimo per la partecipazione il riconoscimento di agentività morale e non avendo trovato ragioni sufficientemente valide per escludere che una macchina possa esserne in possesso, resta da capire se sia possibile – e in che termini – che una macchina influenzi una decisione. Le influenze assumono la forma di preferenze. Tali preferenze, secondo quanto abbiamo visto nel secondo capitolo, possono essere imparziali o auto-interessate.

Sono già state sollevate – e contestualmente risolte – obiezioni in merito alla possibilità che una macchina abbia interessi propri. La macchina, essendo un agente teleologicamente orientato, può avere degli obiettivi che – tenendo fuori il piano psicologico dall’analisi – possono essere assimilabili a degli interessi. E per quanto riguarda le preferenze imparziali? Non ci sono ragioni per non considerare possibile che una macchina possa avere preferenze imparziali. Anzi, il senso comune suggerisce che una macchina – molto più che un essere umano – sia la candidata perfetta per avere questo tipo di preferenze.

Il secondo motivo di ambiguità ha a che fare con il concetto di “democrazia”. Ho definito tale concetto a partire dall’analisi di quattro aspetti – giustificazione, concezione metaetica, rapporto giustizia/processo democratico, modi per prendere una decisione.

Per quanto riguarda il primo aspetto, è stato osservato che una concezione della democrazia è sicuramente non-epistemica se si fonda su una giustificazione di tipo non-strumentalista. Secondo le giustificazioni di questo tipo, la democrazia sarebbe un bene al di là dei risultati che essa produce. Tali giustificazioni si reggono su principi come libertà ed eguaglianza.

Per quanto riguarda la libertà, essa è intesa come la possibilità di autogovernarsi. La partecipazione diventa, dunque, un diritto. Che una macchina possa, in via ipotetica, essere titolare di diritti è già stato associato. Ma ha senso sostenere che essa ha diritto alla libertà? Il concetto di libertà rimanda all’annosa questione dell’autonomia<sup>64</sup> e, in proposito, è eloquente che la parola “autonomia” sia spesso

---

<sup>64</sup> Si veda G. Dworkin (1988, 2015) con riferimento alla natura dell’autonomia.

utilizzata come sinonimo di libertà (Berlin, 1969). Il nodo della libertà è stato già sciolto al paragrafo 4.2.1.3. Alla tesi che gli esseri umani, a differenza delle macchine, sono liberi, si può obiettare che dal momento che si assume una definizione di macchina che manifesta un comportamento non prevedibile o determinabile a priori e se la libertà è assenza di coercizione o manipolazione, non ci sono ragioni per escludere che tale macchina sia in una certa misura libera. Per questo motivo, si può sostenere che è possibile riconoscere il diritto alla libertà per una macchina. Da questo punto di vista, qualora si propendesse per una giustificazione di stampo non-strumentalista che si fonda sul principio di libertà, la partecipazione di una macchina sarebbe rimessa al riconoscimento del diritto di libertà. Come con riferimento alle obiezioni sulla possibilità che una macchina venga considerata un agente morale, non ci sono ragioni sufficientemente forti da considerare una macchina non meritevole del diritto di libertà.

Per quanto riguarda il principio dell'equità, la questione si complica. In base a tale principio, la democrazia è un modo per trattare le persone in maniera eguale. A partire da una giustificazione del genere, accettare che una macchina prenda parte ad un processo decisionale significa considerarla uguale ad un essere umano. Tale affermazione, però, si rivela difficile da difendere.

In generale, per le concezioni della democrazia che si fondano su giustificazioni di stampo non-strumentalista occorre fare una considerazione ulteriore. Dal momento che questo tipo di giustificazioni non vede nel metodo democratico uno strumento per ottenere un certo risultato, ma crede che il metodo democratico abbia un valore in sé, è difficile spiegare perché bisognerebbe lasciar partecipare una macchina al processo decisionale. A meno che la macchina non venga considerata degna degli stessi diritti di un essere umano, non è chiaro per quale motivo dovrebbe esserle concesso l'accesso alle decisioni politiche. In altre parole, se la democrazia non si giustifica a partire dalla qualità delle decisioni che consente di prendere, ma perché essa è in sé moralmente desiderabile – ed è moralmente desiderabile perché è garanzia di alcuni valori fondamentali degli esseri umani – allora non ha senso lasciare che una macchina partecipi alle decisioni.



Per quanto riguarda il secondo aspetto, è emerso che a partire da una concezione metaetica relativista si arriva inevitabilmente a concezioni della democrazia non-epistemiche. Una concezione metaetica relativista è compatibile con la partecipazione di una macchina ai processi decisionali? Si è già visto che non ci sono ragioni sufficientemente solide per non considerare una macchina come un agente morale. Qualora, quindi, si accertasse che una certa macchina è un agente morale, sarebbe perfettamente plausibile la sua partecipazione a un processo decisionale. Che la democrazia si fondi sulla convinzione che esistano dei giudizi etici oggettivi o che tali giudizi non esistano, è irrilevante al fine di stabilire le possibilità di partecipazione – o meno – di una macchina a un processo decisionale in qualità di attore. Lo stesso discorso, dunque, varrebbe nel caso in cui si partisse da una concezione metaetica oggettivistica.

Un discorso simile a quello fatto per le posizioni metaetiche vale per il rapporto fra procedura democratica e giustizia della decisione. In tale rapporto, secondo una posizione proceduralista, la giustizia di una decisione è determinata dal fatto che essa sia stata presa seguendo una procedura democratica. In altre parole, seguire una procedura democratica è garanzia di giustizia per la decisione in questione. In questo senso, non esiste un criterio esterno alla procedura stessa per decretare la giustizia di una decisione. Occorre, dunque, domandarsi se una posizione proceduralista sia d'ostacolo alla partecipazione della macchina-attore ai processi decisionali. Come nel caso delle concezioni metaetiche relativiste, non ci sono ragioni per sostenere che un approccio proceduralista limiti l'accesso di una macchina ad un processo decisionale. In altre parole, il fatto che una macchina possa partecipare a una decisione non mette in crisi il rapporto che esiste fra democrazia e giustizia secondo l'approccio proceduralista.

Infine, per quanto concerne i modi per prendere una decisione, sono stati isolati tre principi – voto, negoziazione, argomentazione. In che relazione si pone la macchina con tali principi? Come si è visto, le concezioni non-epistemiche non sono connesse necessariamente ad uno di questi principi in particolare, dunque, bisogna valutare la

compatibilità (in riferimento alle possibilità di partecipazione di una macchina) con ciascuno di essi.

Per quanto riguarda il voto, anche per via della carica che esso ha acquistato nelle democrazie odierne – venendo spesso identificato col concetto stesso di democrazia – appare singolare l’idea di riconoscere a una macchina la possibilità di votare. In questo senso, è opportuno chiedersi perché non bisognerebbe riconoscere il diritto di voto a una macchina. Questa domanda suona senz’altro bizzarra. Cercando di glissare sull’apparente irragionevolezza del quesito, proverò a concentrarmi sulla questione della legittimità. Se si assume che una macchina possa essere titolare di diritti e che possa essere riconosciuta come un agente morale, allora non dovrebbe stupire che le venga riconosciuto il diritto di voto. Ma cos’è che fa resistenza a questa possibilità? Tale resistenza ha a che fare con il valore che il suffragio universale – e la lotta per ottenerlo – ha nella cultura giuridico-politico democratica?

Lasciando da parte questo tipo di remore – e limitandomi ad un’analisi formale delle possibilità di voto di una macchina – proverò ad indagare l’esistenza di ragioni sufficientemente solide da precludere la partecipazione di una macchina nel caso in cui il principio di espressione delle preferenze fosse quello del voto. Il principio “una persona, un voto” – intrinsecamente legato all’idea del suffragio universale – suggerisce che a fondamento di decisioni democratiche prese con il voto ci sia il principio di eguaglianza (Brennan, 2016; Christiano, 1996, 2008). Tale principio, statuendo che gli esseri umani sono fra loro tutti uguali, ricollega il diritto di voto al fatto che si tratti di esseri umani. In questo senso, verrebbe tradito il principio di eguaglianza se si riconoscesse il diritto a votare a un agente diverso da un essere umano.

Tuttavia, il principio “una persona, un voto” non deve condurre necessariamente al principio di eguaglianza. Esso può, ad esempio, essere giustificato a partire dalla teoria “*all affected interest*”, (Dahl, 1990, p. 64) secondo la quale fanno parte del *demos* – e quindi prendono parte alla decisione – tutti coloro che sono in qualche

modo interessati ad una determinata decisione politica.<sup>65</sup> Si ripropone la questione degli interessi e, come si è visto, non è possibile sostenere che una macchina non può avere degli interessi. Che una macchina possa votare, dunque, dipende da una considerazione semplice: chi ha diritto di votare? Tutti coloro che sono interessati a una decisione o gli esseri umani sulla base di un principio egualitario?

Per quanto riguarda negoziazione e argomentazione, il discorso può essere inquadrato in maniera diversa. Nei processi negoziali o argomentativi, a differenza che nel voto, il momento volitivo “personale” non è preminente. In altre parole, le preferenze – nei processi di negoziazione o argomentazione – non si presentano fisse e immutabili, ma possono, nel caso della negoziazione, combinarsi e, nel caso dell’argomentazione, trasformarsi. Si tratta, dunque, di processi discorsivi e, in casi come questi, alcune perplessità e remore avanzate nell’ipotetico riconoscimento del diritto di voto a una macchina, si presentano in forma più attenuata. Se, infatti, nel voto vi è un’espressione netta delle preferenze, nei processi “discorsivi” come la negoziazione e l’argomentazione, l’aspetto volitivo appare meno marcato e la macchina si presenta come un attore, in mezzo ad altri, che contribuisce ad arrivare ad una decisione che viene presa collettivamente. In altre parole, in una negoziazione o in un processo argomentativo, la macchina sembra avere un ruolo più propositivo o consultivo che decisionale in senso stretto. Inoltre, tanto l’argomentazione quanto la negoziazione non sembrano essere connesse – a differenza del voto – al principio di eguaglianza. Per questa ragione, non ci sono ostacoli insormontabili nell’accettare la partecipazione di una macchina in qualità di attore in un processo negoziale o in un processo che si basa sul principio dell’argomentazione.

#### **4.3.1.2 La macchina-programma e le concezioni non-epistemiche**

---

<sup>65</sup> Un’obiezione a questo argomento è stata sollevata da Goodin (2007, p. 52) secondo cui non è possibile sapere davvero chi sono gli interessati a una determinata decisione politica fino a quando tale decisione non viene presa.

Procederò ora ad analizzare la compatibilità fra la delega a una macchina-programma del processo decisionale e le concezioni non-epistemiche della democrazia.

In primo luogo, si è visto che può senz'altro essere definita non-epistemica quella concezione della democrazia che si fonda su una giustificazione di stampo non-strumentalista. Se la democrazia è un bene in sé – e la legittimità viene slegata dalla produzione di un particolare esito – non ha senso domandarsi se sia legittimo delegare le decisioni a una macchina per il solo fatto che questa macchina abbia una riconosciuta capacità di “decidere meglio”. Pertanto, a partire da una giustificazione non-strumentalista è impossibile legittimare un processo decisionale gestito da una macchina. Infatti, la ragione stessa per cui si dovrebbe ricorrere a una delega della decisionalità – cioè la possibilità di prendere decisioni migliori – non rappresenta un argomento che giustifica la democrazia.

In secondo luogo, è emerso come a partire da una concezione metaetica relativista si arrivi inevitabilmente a concezioni della democrazia non-epistemiche. Occorre, dunque, domandarsi se una concezione metaetica relativista sia compatibile con la delega alle macchine del processo decisionale. Avanzero ora un'ipotesi: se si dubita dell'esistenza di giudizi etici corretti e, conseguentemente, dell'esistenza di un criterio unico e univoco per decretare la correttezza di tali giudizi, e, al contrario, si sostiene che il criterio di correttezza dei giudizi etici, ammesso che esista, dipenda completamente da attitudini e da credenze individuali o di determinati gruppi, allora è impossibile affermare che una macchina sia in grado autonomamente di elaborare dei giudizi etici. La loro correttezza, ammesso che si possa sancire,<sup>66</sup> è rimessa a considerazioni individuali o di gruppo, che, quindi, non possono essere estese al di là di una certa comunità di riferimento. In altre parole, da una prospettiva metaetica relativista, è difficile immaginare che una macchina possa accedere a giudizi etici corretti proprio perché essi trovano determinazione solo su base individuale o, al massimo, in un ben definito gruppo di individui. Non accettare che le considerazioni

---

<sup>66</sup> Per uno scettico morale, come si è visto in 3.2.2.2, non ha senso parlare di giudizi etici corretti.

morali abbiano un contenuto universalizzabile impedisce di immaginare di delegare a una macchina la decisione su questioni etiche.

Questa ipotesi scricchiola e può essere agilmente superata. Poniamo il caso che una macchina abbia la struttura atta a ricevere input che contengono particolari preferenze etiche. Per quale motivo una macchina del genere non potrebbe, in ipotesi, anche essere in grado di generare un output di sintesi che tenga in conto di queste preferenze etiche?

Una volta raccolti questi “input etici”, infatti, la macchina potrebbe, in ottica deliberativa/negoziale, mettere in atto un’operazione di bilanciamento o potrebbe limitarsi, in coerenza col principio del voto, a registrare quale delle preferenze etiche sia la più popolare e decidere nel rispetto di tale preferenza. Una macchina del genere, dunque, anche qualora si partisse da una concezione metaetica relativista, potrebbe rielaborare i posizionamenti etici individuali in modo da simulare l’esito di una pratica discorsiva, nel caso si propendesse per il principio argomentativo, o potrebbe semplicemente “fare il conto”, generando una decisione che gode della maggior parte del favore, nel caso in cui si propendesse per il principio di maggioranza. In altre parole, l’idea che una macchina non possa avere accesso ai giudizi di valore e, in generale, esprimere considerazioni morali – rielaborando gli input dei cittadini – non può essere adeguatamente sostenuta senza fare ricorso ad indimostrabili ragioni legate all’eccezionalismo umano.

Si è visto, inoltre, come un approccio proceduralista<sup>67</sup> conduca inevitabilmente a una concezione della democrazia non-epistemica. Coerentemente a questo approccio, la giustizia di una decisione è determinata dal fatto che essa sia stata presa seguendo una procedura democratica. In altre parole, seguire una procedura democratica è garanzia di giustizia per la decisione in questione poiché non esiste un criterio, esterno alla procedura stessa, per decretare la giustizia di tale decisione. Bisogna, a questo punto, chiedersi se, e in che termini, il proceduralismo sia di ostacolo alla delega alle macchine del processo decisionale. Se la giustizia di una decisione è

---

<sup>67</sup> Si è già visto in 3.2.3 in che senso è stato inteso “proceduralista”.

legata al fatto che sia stata seguita una procedura di un certo tipo, è difficile immaginare che la delega di potere a una macchina – che deciderebbe in totale autonomia – possa essere considerata una procedura democratica nel senso inteso dai proceduralisti. Se, infatti, seguendo l'approccio proceduralista, la giustizia di una decisione è connessa al rispetto della procedura democratica, l'eventuale delega a una macchina, bypassando tale procedura, genererebbe decisioni ingiuste. Da quanto si è visto nel terzo capitolo, un approccio proceduralista mette in primo piano la partecipazione – o la possibilità di partecipare – a una decisione – o alla discussione per la decisione – di tutti i soggetti interessati a quella decisione. Delegare, seppur democraticamente, una cospicua parte di decisionalità a una macchina vuol dire non considerare la procedura democratica – che prevede il coinvolgimento di tutti i soggetti interessati a una certa decisione – necessaria e sufficiente a legittimare una decisione.

A questo argomento si può obiettare che anche nel modello rappresentativo non vi è una reale partecipazione dei soggetti interessati: in una democrazia rappresentativa, infatti, si verifica una delega che legittima le decisioni prese dai rappresentanti. Perché la delega a una macchina sarebbe diversa dalla delega ai rappresentanti? In secondo luogo, per quale ragione un modello rappresentativo della democrazia è compatibile con un approccio proceduralista, mentre una versione meccanizzata della democrazia non lo sarebbe? Al fondo c'è una questione, che verrà approfondita nelle prossime pagine: nell'elezione di rappresentanti umani viene compiuta una scelta tra diverse alternative. Scegliere il candidato Tizio, anziché Caio, tutela le possibilità decisionali degli individui. Qualora esistessero due tipi di macchine, che si fondano su meccanismi diversi, tale tutela, in un certo senso, verrebbe salvaguardata. Ma, in questo caso, non sarebbe più la macchina che sto immaginando: una macchina, cioè, al di là e al di sopra delle parti, una macchina, per certi aspetti, inevitabile e necessaria.

Infine, per quanto riguarda i principi che regolano i processi decisionali, che secondo la tripartizione di Elster (1998) e Martí (2001) sono voto, negoziazione e argomentazione, essi devono essere analizzati disgiuntamente. Più precisamente, con

riferimento al voto, esso dà vita a un modello democratico la cui automazione non presenta particolari problemi. Ciò che si chiede alla macchina, in questo caso, è di emulare il principio del voto e, quindi, non di riformulare le preferenze già predefinite, ma semplicemente di accertare quale di queste preferenze abbia riscosso maggiore successo.

Per quanto concerne il principio della negoziazione, già da diversi anni esistono dei tentativi di automazione di processi del genere (Bartolini, Preist e Jennings, 2005; Jelassi e Foroughi, 1989; Gupta et al., 2006). In questo caso la macchina ha il compito di trovare un punto di incontro in quelle situazioni in cui vi è un forte disaccordo su giudizi di fatto o di valore fra i membri di un gruppo (Jelassi e Foroughi, 1989).

Il principio che presenta maggiori difficoltà di inquadramento è certamente quello argomentativo. Secondo Martí (2001) e i sostenitori delle teorie deliberative, il principio argomentativo è quello su cui si fonda il modello della democrazia deliberativa. Tale principio presuppone l'esistenza di una decisione corretta sotto la forma di uno standard di correttezza indipendente rispetto al processo deliberativo (Martí, 2001, p. 49). Il compito di una macchina del genere è meno semplice di quello di una macchina che imita il principio del voto o della negoziazione. In questo caso, infatti, gli input che giungono alla macchina non possono essere semplicemente conteggiati o calcolati per trovare il loro valore medio, ma devono subire una vera e propria trasformazione. Ciò è anche legato al fatto che, in base a quanto si è visto nel terzo capitolo, tali input – che prendono il nome di preferenze – non sono auto-interessati, ma si presentano come imparziali. Dunque, l'obiettivo di una macchina deliberativa non è semplicemente quello di prendere una decisione che incontri il consenso strategico (come nel caso della negoziazione) o che sia quella maggiormente condivisa (come nel caso del voto), ma dovrebbe essere quello di generare, attraverso un processo trasformativo, la decisione che si fonda sull'argomento migliore. Vi è, dunque, un ampliamento del suo potere di scelta poiché essa deve innanzitutto selezionare l'argomento migliore.

Secondo Martí (2001), il principio dell'argomentazione non è l'unico principio che si fonda su preferenze imparziali. Anche nel voto, in ipotesi, i soggetti possono muoversi a partire da preferenze imparziali. A questo punto, sembrerebbe rilevare non tanto il fatto che la macchina in questione debba seguire il principio del voto o quello dell'argomentazione, quanto il fatto che le preferenze presentate a una macchina del genere siano da considerarsi imparziali. Tuttavia, nel caso del voto la macchina ha un compito semplificato: nonostante le preferenze possano essere imparziali sia in caso di argomentazione che in caso di voto, in quest'ultima ipotesi la macchina si limita a selezionare l'alternativa che ha riscosso maggiore successo. Il compito della *macchina deliberativa*, invece, si rivela più complicato: ricevere e trasformare le preferenze individuali non in quanto preferenze auto-interessate, ma come ragioni entro cui scovare e portare alla luce l'argomento migliore e decidere sulla base di quest'argomento.

La deliberazione e, conseguentemente, la *macchina deliberativa* acquistano un senso solo se gli altri elementi analizzati – la giustificazione, la concezione metaetica e il rapporto democrazia/giustizia – sono configurati in una determinata maniera e, cioè, alla maniera delle concezioni epistemiche.

Complessivamente, si è osservato come alcuni aspetti tipici delle concezioni non-epistemiche della democrazia (giustificazioni non-strumentaliste e approcci proceduralisti) si rivelino degli anticorpi normativi per la delega a una macchina delle decisioni, mentre altri aspetti rendano la questione della delega poco interessante (voto e negoziazione). Le concezioni metaetiche relativiste, invece, pur non ostacolando – o banalizzando – la delega alle macchine, rendono tale delega un'operazione senz'altro più complicata.

#### **4.3.2 Compatibilità con le concezioni epistemiche**

L'ipotesi di partenza è, come già detto, che una concezione epistémica della democrazia agevola e promuove l'accesso delle macchine ai processi decisionali. Una



concezione epistemica, a differenza delle concezioni non-epistemiche che sono definite in maniera residuale, ha una configurazione ben precisa. Nel terzo capitolo, ho sostenuto che tali concezioni si fondino necessariamente su giustificazioni strumentaliste e su concezioni metaetiche oggettiviste, propendono per un approccio sostanzialista e ricorrono al principio dell'argomentazione come metodo per prendere decisioni. La verifica di compatibilità sarà duplice: prima riguarderà le macchine che hanno un ruolo di attore nel processo decisionale e, poi, verrà presa in esame la compatibilità con le macchine-programma.

Prima di addentrarmi nella verifica di compatibilità, c'è bisogno di fare una premessa di carattere generale. Non vi sono in letteratura – o, meglio, io non ho incontrato – contributi, fra gli autori che gravitano attorno alla galassia delle concezioni epistemiche, che si auspicano interventi delle macchine tanto ragguardevoli quanto quelli suggeriti in questo scritto. Vi è, tuttavia, per alcuni autori vicini a una concezione epistemica della democrazia, notevole interesse per l'utilizzo dell'intelligenza artificiale nei processi decisionali (Landemore, 2022; Martí, 2021). Al di là, dunque, dell'analisi formale sulla compatibilità, che svilupperò nelle prossime pagine, esiste, a mio parere, una naturale tensione delle concezioni epistemiche della democrazia verso l'implementazione di tecnologie per coadiuvare – o, nel caso esasperato, per sostituire – i processi deliberativi.

#### **4.3.2.1 La macchina-attore e le concezioni epistemiche**

Valuterò, ora, la compatibilità fra la partecipazione di una macchina in qualità di attore in un processo decisionale e le concezioni epistemiche della democrazia.

Una concezione epistemica della democrazia è necessariamente strumentalista. Secondo le giustificazioni strumentaliste, come si è visto, la democrazia si giustifica per il fatto che produce decisioni migliori di quelle che si prenderebbero adottando un altro metodo decisionale. A differenza di quanto visto al paragrafo 4.3.1.1, dal momento che la democrazia non si giustifica sulla base di principi come quello di

libertà o di eguaglianza, ma sulla base dei risultati che il processo democratico produce, la partecipazione di una macchina ad un processo decisionale del genere è rimessa a una condizione: che la macchina contribuisca a prendere decisioni migliori o, quantomeno, non peggiori la qualità delle decisioni. Come si può decretare che una macchina partecipando a un processo decisionale contribuisca a migliorare le decisioni? Non si tratta di dimostrare l’apporto positivo (nel senso di migliorativo) dell’agente macchina, quanto di considerare se essa sia in grado di compromettere o meno il processo decisionale. In proposito, non trovo ragioni sufficientemente solide a sostegno di un’influenza peggiorativa dell’agente macchina in un processo decisionale. Al contrario, l’idea *mainstream* della macchina – intesa come agente estraneo a *bias* e ad errori di giudizio che sono soliti riguardare gli esseri umani (Tversky e Kahneman, 1974) – sembra avvicinarsi più di un essere umano all’interlocutore razionale della situazione discorsiva ideale (Habermas, 1987).

Il secondo punto che prenderò in esame è quello delle concezioni metaetiche. Nelle concezioni epistemiche della democrazia, il posizionamento metaetico di elezione è quello oggettivista. Cosa implica ciò con riferimento alla possibilità che una macchina partecipi come attore in un processo decisionale? La domanda può essere formulata meglio: a partire da una concezione metaetica oggettivista, si può riconoscere a una macchina l’accesso a un processo decisionale democratico? Come si è visto al paragrafo 4.3.1.1, la questione metaetica – esistono valori etici oggettivi? – non è rilevante ai fini di stabilire se una macchina possa o meno partecipare in qualità di attore in un processo decisionale. Che esistano dei valori etici oggettivi e che sia la credenza nell’esistenza di questi valori a fondare filosoficamente e giustificare la democrazia non implica che ogni agente, preso singolarmente, debba essere considerato “morale” – cioè conforme a valori considerati oggettivi.<sup>68</sup> Questa ipotetica macchina morale, dunque, può, in ipotesi, essere anche contro la morale oggettiva e ciò non metterebbe in discussione il fondamento oggettivista, da un punto

---

<sup>68</sup> Se si pensasse che ogni agente per essere tale debba agire in maniera conforme alla morale oggettiva, istituti come il carcere – o il diritto penale in generale – non potrebbero essere più seriamente giustificati.

di vista metaetico, della democrazia. Occorre osservare, inoltre, che considerata l'idea *mainstream* di macchina – come entità programmata e riprogrammabile – sembra vi siano maggiori possibilità che sia una macchina ad adeguarsi ai valori oggettivamente dati rispetto a un essere umano.

Il terzo punto da analizzare riguarda il rapporto fra processo democratico e giustizia. Si è visto che chi sostiene una concezione epistemica della democrazia considera il rapporto fra processo democratico e giustizia in termini sostanziali – vi è, cioè, la credenza che esistano delle decisioni giuste che prescindono dal metodo utilizzato per decidere. Anche in questo caso – come per le concezioni metaetiche – considerarsi proceduralisti o sostanzialisti non è determinante nel decretare se una macchina possa o meno prendere parte a un processo decisionale.

Infine, l'ultimo punto che analizzerò riguarda i distinti principi che regolano un processo decisionale democratico. Il principio di elezione delle concezioni epistemiche è senz'altro quello argomentativo. Tale principio è a fondamento del modello di democrazia deliberativa. Ho già accennato, in 4.3.1.1, ad alcune difficoltà legate al riconoscimento del diritto di voto a una macchina. Per quanto concerne l'argomentazione, invece, ci sono meno difficoltà nel concepire l'intervento di una macchina. Una macchina, nel caso di un processo deliberativo, si presenta come un interlocutore fra tanti e non serve affrontare particolari peripezie per legittimare la sua presenza nel processo decisionale.

In definitiva, le concezioni epistemiche della democrazia non presentano ragioni che impediscono – o limitano – l'accesso di una macchina ai processi decisionali.

#### **4.3.2.2 La macchina-programma e le concezioni epistemiche**

La delega a una macchina di porzioni di decisionalità trova negli elementi che fondano una concezione epistemica della democrazia piena legittimità: questo è ciò che proverò a dimostrare. La macchina che analizzerò in questo paragrafo è la macchina-programma che, a partire dalle preferenze espresse da coloro che sono

legittimati a partecipare al processo decisionale, genera delle decisioni. Questa valutazione di compatibilità riguarderà, come al solito, quattro aspetti – giustificazioni della democrazia, concezioni metaetiche, rapporto fra procedura democratica e giustizia delle decisioni, principi che regolano il processo democratico.

Per quanto riguarda il primo aspetto, le giustificazioni che sostengono una concezione epistemica della democrazia sono quelle strumentaliste. La ragione per cui si dovrebbe delegare una porzione di decisionalità a una macchina è legata alla prospettiva che tale macchina sia in grado di prendere decisioni migliori rispetto a quelle che prenderebbero gli esseri umani senza alcun tipo di supporto.

Si tratta di un'ipotesi: sto assumendo che esiste una macchina capace di espungere e neutralizzare argomenti e dinamiche – tipiche delle interazioni umane – che generano ambiguità ed errori di valutazione che possono inficiare la qualità delle decisioni.

L'incontro fra teoria del diritto e scienze cognitive (Brigaglia e Celano, 2017) ha dato il via a un proliferarsi di riflessioni sulla presenza di *bias* nel ragionamento umano e, in particolare, nel ragionamento giuridico (Gigerenzer e Enel, 2004; Danziger, Levav e Avnaim-Pesso, 2011). Queste riflessioni hanno riguardato anche la deliberazione in senso lato ed è emerso come negli organi deliberativi possano esservi dinamiche che hanno una cattiva influenza sulle decisioni.<sup>69</sup> In questo senso, appare ben motivato ricorrere al supporto di una macchina capace di limitare – o eliminare – i pregiudizi che dinamiche comunicative e decisionali distorte – tipicamente umane – potrebbero arrecare alla qualità delle decisioni se si condivide l'idea che la democrazia è un bene perché permette di raggiungere decisioni migliori. Ciò che sostengo è che mentre le scienze cognitive hanno dato fondamento alle critiche nei

---

<sup>69</sup> Alcune critiche mosse alle concezioni epistemiche della democrazia affermano che non c'è alcuna prova che dimostri che la deliberazione intersoggettiva porti a decisioni più giuste o corrette. A titolo esemplificativo, Sunstein (2002, p. 187), in merito a questa critica, ha argomentato che la deliberazione può generare decisioni anche peggiori poiché nei gruppi, benché di persone fra loro affini, vige “la legge della polarizzazione” che porta spesso le persone a serrarsi dietro le loro posizioni predeterminate.

confronti dell'entusiasmo – a tratti un po' *naif* – dei fautori delle concezioni epistemiche della democrazia, gli studi sull'intelligenza artificiale potrebbero trovare una soluzione a tali criticità, nella delega a macchine che evitino gli errori cui sono soggetti gli esseri umani quando ragionano (singolarmente o collettivamente).

La questione aperta – quella dirimente – è stata già anticipata al paragrafo 4.2.2.2 e può essere formulata nei seguenti termini: può una macchina – che si identifica con un processo decisionale – agire in maniera democratica? Se la democrazia è il modo attraverso cui prendere decisioni migliori e se il ragionamento umano è spesso inficiato da *bias* di qualche tipo e se questa macchina promette di estromettere dalle decisioni tali *bias*, allora bisogna concludere che la macchina in questione può agire in maniera democratica. Questa risposta appare, tuttavia, insoddisfacente. Bisogna domandarsi, infatti, se esiste una maniera di settare questa macchina in modo tale che il suo agire risulti rispettoso del principio democratico.

Prima di immaginare come programmare una macchina del genere, è opportuno chiarire cosa intendono per democrazia alcune concezioni della democrazia. Per rispondere a questa domanda, è necessario scendere agli altri livelli dell'analisi – cioè quelli che riguardano le concezioni metaetiche, il rapporto fra processo democratico e giustizia e i principi che regolano il processo democratico.

Le concezioni epistemiche della democrazia condividono l'idea che la democrazia si possa giustificare e fondare filosoficamente solo se si assume che esistano valori oggettivi e universali. In altre parole, la democrazia, secondo le concezioni epistemiche, si fonda sulla convinzione che esiste una posizione etica corretta. Secondo la tassonomia metaetica tratteggiata nel terzo capitolo, la seconda tesi degli oggettivisti – si tratta degli oggettivisti che condividono una concezione epistémica della democrazia – è che esiste un criterio unico e univoco – diverso da quello che si fonda sulle preferenze individuali – per decretare se un giudizio etico è corretto.

Secondo le concezioni epistemiche, è la democrazia lo strumento atto a scoprire tale criterio, nonostante esso sia indipendente dalle credenze individuali. Un'affermazione del genere può sembrare paradossale: in che modo la democrazia è lo strumento che permette di scoprire la correttezza dei giudizi etici se tali giudizi non

dipendono dalle credenze individuali? Delle due una: se la democrazia è il modo in cui si manifestano le preferenze individuali, allora il criterio di correttezza o deve dipendere dalle preferenze individuali o dev'essere indipendente rispetto a queste.

La soluzione che alcuni autori trovano a questo paradosso si radica nel concetto di deliberazione. È il processo deliberativo che permetterebbe *democraticamente* di arrivare alla decisione corretta. In qualche modo, tale processo – fondandosi sull'idea che le preferenze di coloro che vi partecipano siano imparziali e non auto-interessate (Martí, 2001) – ha il merito di lasciar emergere l'argomento migliore, che è ciò su cui si basa la decisione corretta.

Al di là delle critiche all'ingenuità di questa affermazione, ciò che è rilevante, in questo momento, è l'idea di alcuni autori che afferiscono al “*Systemic Turn*”, secondo cui la deliberazione dev'essere intesa in senso sistemico e, dunque, la “buona deliberazione”, non deve essere prerogativa di ogni singola parte del sistema e può anche non riguardare la deliberazione dei singoli cittadini (Owen e Smith, 2015).<sup>70</sup>

In altre parole, il “minimum deliberativo” e il mutuo rispetto – elementi tipici dell'ideale deliberativo di democrazia – debbono essere valutati dal sistema nel suo insieme e, date queste premesse, si può immaginare come legittimo un sistema democratico in cui non vi sia nessuna reale deliberazione dei cittadini (Owen e Smith, 2015, p 227). Ciò rappresenta senz'altro un “indebolimento” normativo del concetto di democrazia. A partire da questo indebolimento, appare come perfettamente

---

<sup>70</sup> Owen e Smith (2015, p. 219) si rivelano molto critici nei confronti di alcuni autori del panorama dell'approccio sistemico. In proposito, affermano: «*However, our particular concern is that within this conception of a deliberative system, no actual deliberation between citizens need take place as long as the process embodied in the sequenced institutions is functionally equivalent to a deliberative democratic exchange. In this somewhat reductionist approach, democratic deliberation is conceived instrumentally as a way of generating decisions that are sensitive to the relevant range of public reasons. Because only the end of appropriately sensitive decisions matters, any functionally equivalent democratic way of generating this end will count as equally legitimate. The model does not rule out the inclusion of deliberative institutions within a sequence, but does not require their inclusion. Such a 'distributed' approach to deliberative systems sets aside any concern for the deliberative capacities and powers of citizens*».

plausibile la possibilità di delegare a una macchina l'elaborazione e la trasformazione delle decisioni, a patto che tali decisioni siano sensibili alle ragioni pubbliche.

Per riassumere, l'approccio sistemico – che raccoglie l'eredità delle concezioni epistemiche – rende manifesti gli esiti dell'exasperazione delle concezioni epistemiche della democrazia in due modi. Da un lato stravolgendo il concetto di democrazia: “democrazia” non è più considerato un ideale normativo esigente né un metodo per prendere decisioni, ma diventa il campo d'indagine – ovvero la realtà entro cui si sviluppano i processi decisionali che riguardano le questioni politiche. In secondo luogo, abbassando la soglia dei requisiti minimi per considerare legittimo un certo ordinamento: perché un sistema venga ritenuto democratico, è sufficiente l'adempimento di uno standard normativo che può essere soddisfatto anche solo attraverso l'elezione dei rappresentanti.<sup>71</sup> Dall'altro lato, l'approccio sistemico agisce anche sul concetto di deliberazione. Esso è soggetto a un “*concept-stretching*” (Bächtiger, Niemeyer, Neblo, Steenbergen, e Steiner, 2010, p. 48) tale per cui “deliberare” non significa più solo scambiarsi ragioni fra esseri umani, ma la deliberazione si estende fino ad includere la comunicazione fra soggetti di diversa natura – fra macchine ed esseri umani (Capriati, 2024) o fra vari attori istituzionali (Mansbridge, Bohman, Chambers, Christiano, Fung, Parkinson, Thompson e Warren, 2012) – e, secondo l'ipotesi che sto suggerendo, la deliberazione può anche consistere in uno scambio simulato che avviene all'interno di una macchina.

Tornando alla questione dirimente: può una macchina agire in maniera democratica? In base a quanto visto finora, secondo i fautori di un approccio sistemico – exasperazione e sintesi operativa di alcune concezioni epistemiche della democrazia – una macchina che riceve, elabora e trasforma le preferenze dei cittadini potrebbe adempiere agli standard normativi minimi richiesti.

---

<sup>71</sup> Si veda in proposito la critica di Owen e Smith (2015, p. 227) che hanno qualificato l'approccio sistemico di Goodin (2008, p. 202) “*deliberative Schumpeterianism*”.

#### 4.4 La macchina che delibera

L'ipotesi che andava verificata era la seguente: ci sono concezioni della democrazia che precludono – o limitano – l'accesso delle macchine ai processi decisionali su questioni di interesse pubblico e ci sono altre concezioni che facilitano tale accesso. Si tratta di un'ipotesi verificata solo in parte. Non si è potuto, infatti, concludere che questa ipotesi valga anche per la macchina-attore. Una concezione non-epistemica della democrazia non preclude – o permette – l'accesso di una macchina ad un processo decisionale più di quanto farebbe una concezione epistemica.

Per quanto riguarda la macchina-programma, la concezione democratica di riferimento acquista rilevanza. Più precisamente, le concezioni non-epistemiche dispongono di anticorpi normativi che si rivelano un ostacolo alla delega di decisioni a una macchina. Le concezioni epistemiche, invece, non dispongono di tali anticorpi e, come nel caso delle teorie che rientrano nel novero dell'approccio sistemico alla democrazia deliberativa, c'è un dichiarato spostamento di focus su momenti decisionali diversi dalla classica deliberazione fra cittadini.

L'approccio sistemico alla democrazia deliberativa e, in generale, le concezioni epistemiche sono una risposta alla crisi della democrazia. La crisi della democrazia si manifesta in vari modi e ha diverse cause. Fra le cause, c'è una perdita di fiducia nei decisori politici. La mossa di coloro che afferiscono a una concezione epistemica della democrazia è stata quella di rilanciare sulle capacità decisionali degli individui e dei gruppi. Giustificando, però, la democrazia sulla base del fatto che essa produce decisioni migliori rispetto agli altri sistemi di governo si è assistito ad un indebolimento normativo. Una democrazia così indebolita ha dovuto affrontare da un lato i recenti studi delle scienze cognitive che rendono manifesta la natura fallace del ragionamento umano e, dall'altro lato, gli studi sull'intelligenza artificiale che propongono di superare i limiti del ragionamento umano costruendo macchine senza *bias*. Si inaugura, così, una promessa di automazione dei processi decisionali. Tale automazione permetterebbe di prendere decisioni migliori – senza errori – e con costi



sensibilmente minori, in termini di tempo e di risorse. Le criticità che sorgono sono evidenti:

(a) lo spostamento di potere da attori pubblici – le istituzioni – ad attori privati – imprese e multinazionali – che programmano le macchine che hanno legittimità a decidere;

(b) la creazione di macchine autonome che, prendendo le distanze dai programmatori attraverso tecniche di *machine learning*, agiscono in maniera incomprensibile – e potenzialmente dannosa – in uno scenario *à la* Blade Runner.

È chiaro, specie nel secondo caso, che si tratta di un'ipotesi altamente improbabile, ma possibile.

Infine, al di là delle criticità rilevate, questa analisi non è solo volta a verificare la compatibilità fra macchine e specifiche concezioni della democrazia, ma è anche un tentativo di comprendere fino a che punto il processo che crea il diritto è automatizzabile e, inversamente, cosa non può essere automatizzato. Le conclusioni che mi sembra si possano trarre sono a tratti distopiche: secondo alcune recenti e autorevoli concezioni della democrazia, il processo decisionale potrebbe essere affidato interamente a una macchina se questa desse prova di decidere meglio degli esseri umani. E, in fondo, non sembra così assurdo: se le automobili a guida autonoma dessero prova di massima efficienza, non lasceremmo che fossero loro a guidarci sempre?

## CONCLUSIONI

Come è accolta nel dibattito sulla democrazia la partecipazione delle macchine a processi decisionali su questioni di interesse pubblico?

Per rispondere a questa domanda sono partito da molto lontano. Nel primo capitolo, ho presentato il progetto cibernetico. Sono gli anni '70 e nel Cile di Salvador Allende viene organizzato un sistema di raccolta ed elaborazione dati per scelte di politica economica e sociale: si tratta di Cybersyn. In questo progetto si assiste ad una embrionale forma di automazione dei processi decisionali. Con Cybersyn vengono anticipate alcune questioni cruciali per questa analisi. Più precisamente, mi riferisco a:

(a) l'analogia fra sistemi politici e macchine;

(b) la possibilità che una macchina viva dentro un sistema politico.

(a) porta a domandarsi se un sistema politico debba essere considerato un sistema teleologico o un sistema non-teleologico.

Con (b), invece, si intende verificare se sia possibile delegare ad una macchina una parte – o la totalità – del processo decisionale. In altre parole, ci si può affidare ad una macchina per prendere decisioni di interesse pubblico? È esattamente questa la domanda che muove la mia ricerca. Per poter rispondere, è necessario destrutturare la domanda: in primo luogo, cos'è una macchina?

In questa ricerca, quando si parla di macchine, mi riferisco ad agenti non biologici autonomi. Se è facile immaginare a cosa ci si riferisce con “non biologici”, più difficile è definire puntualmente il concetto di “autonomia”. “Autonomo”, in termini negativi, significa non essere soggetto a regole altrui ed è solo con riferimento ad una macchina del genere che può sorgere un problema di legittimità nella delega del potere: delegare, infatti, a una macchina la cui logica decisoria è interamente predeterminata significa avvalersi di un mero esecutore. Una macchina, invece, il cui comportamento non è perfettamente prevedibile si pone in termini di alterità rispetto

al decisore umano. È, dunque, legittimo avvalersi di una macchina del genere per prendere decisioni?

Per rispondere a questa domanda, bisogna guardare alla natura dei contesti democratici in cui queste macchine si inseriscono. Per “contesto democratico” non mi riferisco alla congerie di norme ed organi che danno vita all’apparato istituzionale dello Stato, ma all’insieme degli elementi che costituiscono una determinata struttura normativa che si definisce democratica. Secondo la mia analisi, tale struttura si compone di quattro elementi costitutivi: giustificazioni; concezioni metaetiche; rapporto procedura democratica e giustizia delle decisioni; metodi per prendere una decisione.

Se tali elementi si configurano in una certa maniera – giustificazione di stampo strumentalista, concezione metaetica oggettivista, approccio sostanzialista in fatto di giustizia, elezione del principio dell’argomentazione –, allora la concezione della democrazia che ne viene fuori può essere definita epistemica. Se, invece, anche uno solo di questi elementi costitutivi diverge dalla configurazione appena esposta, ci si trova davanti a una concezione non-epistemica. Le concezioni epistemiche sono, dunque, definite in termini residuali: sono non-epistemiche tutte quelle concezioni che non sono epistemiche. Fra le concezioni epistemiche, si distacca l’approccio sistemico alla deliberazione democratica che è, in qualche modo, sintesi operativa ed esasperazione delle suddette concezioni. Ma perché distinguere le concezioni non-epistemiche da quelle epistemiche? Qual è l’obiettivo di questa operazione?

L’ultimo capitolo è votato a verificare l’ipotesi di partenza: ci sono alcune concezioni della democrazia – quelle epistemiche – che permettono o agevolano la partecipazione delle macchine ai processi decisionali, mentre ci sono altre concezioni – quelle non-epistemiche – che rappresentano un ostacolo a tale partecipazione. Ma che significa per una macchina “partecipare?” È una partecipazione che può esplicarsi in due forme: la macchina può partecipare in qualità di attore o come programma – raccogliendo, elaborando e trasformando in una decisione le preferenze individuali. È questo secondo tipo di macchine che fa emergere questioni di maggiore interesse: può una macchina del genere agire in maniera democratica? Questa si rivela una domanda

ostica. Innanzitutto, il momento democratico può precedere il momento decisionale vero e proprio: le preferenze di coloro che partecipano a un processo decisionale possono essere presentate in condizioni di equità a una macchina. Ma come si passa dalla raccolta delle preferenze alla formulazione di una decisione? Apparentemente pochi problemi pone il metodo del voto: in questo caso la macchina deve limitarsi a registrare le preferenze e a calcolare quale di queste sarà a fondamento della decisione, secondo un principio di maggioranza – assoluta o relativa – o di consenso. La questione si complica nel caso della deliberazione: la macchina che delibera, infatti, non si limita a fare un conteggio dei voti, ma, sulla base delle preferenze espresse, cerca di elaborare una decisione attraverso un processo che trasforma le preferenze originarie, nella prospettiva di lasciarsi guidare dal migliore argomento. Non ci sono ragioni sufficientemente solide per escludere che una macchina del genere sia in grado di agire in maniera democratica.

La macchina che delibera non è un'ipotesi impossibile né da un punto di vista tecnico né di ammissibilità, nel recente dibattito sulla democrazia. Non solo non ci sono ragioni forti contro l'automazione dei processi decisionali, ma vi sono argomenti che sembrano incoraggiare il ricorso alle macchine. Tali argomenti da un lato provengono proprio dal dibattito sulla democrazia. L'approccio sistemico alla deliberazione democratica ha, infatti, riscritto il significato di deliberazione: essa non è più lo scambio dialogico fra cittadini, ma si presenta come un'interazione fra entità di diversa natura orientata alla produzione di una decisione. Ma ci sono anche buoni argomenti che provengono dalla congiuntura fra gli studi sulle scienze cognitive – che riguardano la fallacia del ragionamento umano – e quelli sull'intelligenza artificiale – che si fondano sulla promessa del superamento dei limiti del ragionamento umano. Una certa lettura di questi studi potrebbe implicare un indebolimento della struttura normativa della democrazia e, dunque, contribuire a radicare la convinzione che una macchina possa decidere meglio di un gruppo di persone. Questa convinzione ha senso solo se si ammette un fatto: esistono decisioni corrette o, comunque, decisioni più corrette di altre, da un punto di vista oggettivo.

È singolare l'evoluzione normativa del concetto di democrazia: per fronteggiare la crisi della democrazia – che, fra le altre cose, si manifesta con la rottura del legame fiduciario fra rappresentanti e rappresentati e con l'affermarsi di figure particolarmente carismatiche che creano consenso e che sono poco rispettose delle istituzioni liberali (il cosiddetto populismo) – si è puntato sulle presunte capacità cognitive dei gruppi di persone. In altre parole, per rilanciare le istituzioni democratiche, si è giocato sulle capacità delle assemblee – più o meno grandi, a seconda delle analisi – di prendere buone decisioni. I recenti studi delle scienze cognitive vanno, però, in senso inverso: essi dimostrano come il ragionamento umano sia sistematicamente soggetto a *bias* ed errori di diverso tipo. In altre parole, emerge come il ragionamento umano sia fallace e che nei gruppi tale fallacia si incrementa, generando, ad esempio, dinamiche come la polarizzazione. Per questa ragione, è difficile sostenere che la fiducia nelle capacità di giudizio dei gruppi di persone si sia davvero riaffermata. Parallelamente, gli studi sull'intelligenza artificiale hanno fatto enormi passi in avanti, creando macchine che non ripetono nel loro ragionamento gli errori che tipicamente commettono gli umani. Il quadro, dunque, è il seguente: nel tentativo di superare la crisi della democrazia si è andata affermando l'idea che il metodo democratico sia il modo per prendere decisioni migliori. Tale idea però non ha trovato un solido fondamento né nella pratica politica occidentale – le masse, a cui sono rimesse le decisioni, vengono sempre più spesso rappresentate come ignoranti e incompetenti – né nei recenti studi delle scienze cognitive. Si è andato, invece, affermando il mito di una macchina immune dai comuni vizi del ragionamento umano. In questa prospettiva, la delega di porzioni del processo decisionale non solo non appare un'assurdità, ma sembra quasi auspicabile, come un salvagente lanciato alla democrazia per non colare a picco.

C'è un secondo ordine di considerazioni da fare. Bisogna tornare al punto di partenza, cioè a quell'intuizione di Beer ampiamente criticata: i sistemi politici sono macchine e in quanto macchine sono sistemi autopoietici. Tale analogia può considerarsi un argomento per foraggiare l'accesso delle macchine nei processi decisionali? L'argomento potrebbe avere la seguente struttura: una macchina è

legittimata a prendere decisioni anche perché il sistema politico nella sua totalità agisce come una macchina. È chiaro che si tratta di un argomento scivoloso. Quella della macchina è una metafora e non deve essere intesa in senso letterale: quando si dice che ogni sistema politico è una macchina non si intende dire che esso è una macchina nel vero senso della parola. E anche se si ammettesse che si tratta davvero di una macchina, bisogna riconoscere che è una macchina diversa rispetto alla macchina cui ho fatto riferimento in questa analisi: la macchina-sistema politico non è un'entità meccanica, ma è l'insieme di norme e di istituti che danno vita al sistema politico; la macchina deliberativa è, invece, una macchina vera e propria – con un hardware e un software. Bisogna, quindi, concludere che la metafora della macchina per descrivere il sistema politico non getta alcuna luce su questa ricerca? Non vi è alcun legame fra la macchina-sistema politico e la macchina che delibera?

Probabilmente, il ricorso all'immagine della macchina svela il desiderio di neutralità. Ciò che generalmente ci si auspica è che i sistemi politici e i processi decisionali siano da un lato più efficienti e, dall'altro lato, indifferenti ad interessi egoistici. È quello che si auspicano anche gli autori che rientrano sotto l'ombrello della teoria della democrazia deliberativa quando parlano di preferenze “imparziali” o “non auto-interessate” dei partecipanti ai processi decisionali. Si tratta di una spersonalizzazione dei procedimenti politici che ha come conseguenza quella di dare maggiore dignità e autorevolezza alla decisione presa da una macchina, facendo appello al mito della sua perfezione.

## BIBLIOGRAFIA

- Akata Z., Balliet D., De Rijke M., Dignum F., Dignum V., Eiben G., ... & Welling M. (2020). *A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence*. In «Computer», 53(8), pp. 18-28.
- Allen C., Varner G., Zinser J. (2000). *Prolegomena to Any Future Artificial Moral Agent*. In «Journal of Experimental and Theoretical Artificial Intelligence», 12, pp. 251-261.
- Alpaydin E. (2014). *Introduction to Machine Learning*. Cambridge. MIT Press.
- Anderson E. (2006). *The Epistemology of Democracy*. In «Episteme», 3, pp. 8-22.
- Andrejevic M., Gates K. (2014). *Big data surveillance: Introduction*. In «Surveillance & Society», 12(2), pp. 185-196.
- Arendt H. (2004). *'What Remains? The Language Remains': A Conversation with Günter Gaus*. In Baehr P. R. (a cura di). *The Portable Hannah Arendt*. Penguin Books. London.
- Arendt H. (1990). *Philosophy and politics*. In «Social research», pp. 73-103.
- Arneson R. J. (2004). *Democracy is not Intrinsically Just*. In Dowding K., Goodin R. E., Pateman C. (a cura di), *Justice and Democracy*. Cambridge. Cambridge University Press, pp. 40–58.

Arrow K. J. (1963). *Social Choice and Individual Values*. New Haven. Yale University Press.

Asaro P. (2007). *Robots and responsibility from a legal perspective*. In «The IEEE conference on robotics and automation, workshop on roboethics». Roma. Unpublished proceedings:  
<http://www.peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf>

Ashby W. R. (1957). *An introduction to cybernetics*. London. Chapman & Hall.

Ashley K. D. (2017). *Artificial intelligence and legal analytics: new tools for law practice in the digital age*. Cambridge University Press.

Awad E., Levine S., Anderson M., Anderson S. L., Conitzer V., Crockett M. J., ... & Tenenbaum, J. B. (2022). *Computational ethics*. In «Trends in Cognitive Sciences», 26(5), pp. 388-405.

Bächtiger A., J. S. Dryzek, J. Mansbridge, M. Warren, (2018). *Deliberative Democracy: An Introduction*. In Bächtiger A., Dryzek J. S., Mansbridge J., Warren M. (a cura di). *The Oxford Handbook of Deliberative Democracy*. Oxford University Press. Oxford, pp. 1-31.

Bächtiger A., Niemeyer S., Neblo M., Steenbergen M. R., Steiner J. (2010). *Disentangling diversity in deliberative democracy: competing theories, their blind spots and complementarities*. In «Journal of Political Philosophy», 18, pp. 32–63.

Badham R., Clegg C., Wall T., (2000). *Socio-technical theory*. In Karwowski W. (a cura di). *Handbook of Ergonomics*. John Wiley. New York.



Bartlett J. (2018). *The People vs. Tech. How the Internet is Killing Democracy (and how we save it)*. Random House.

Bartolini C., Preist C., Jennings N. R. (2005). *A software framework for automated negotiation*. In Choren R., Garcia A., Lucena C., Romanovsky A. (a cura di). *Software engineering for multi-agent systems III: Research issues and practical applications*. Springer Berlin Heidelberg, pp. 213-235.

Batko R., Kreft J. (2017). *The Sixth Estate—The Rule of Algorithms*. In «Problemy Zarządzania», (2/2017 (68), t. 2), pp. 190-209.

Beckman L., Hultin Rosenberg J., Jebari K. (2022). *Artificial intelligence and democratic legitimacy. The problem of publicity in public authority*. In «AI & SOCIETY», pp. 1-10.

Beer S. (1959). *Cybernetics and Management*. London. English Universities Press.

Beer S. (1974). *Cybernetics of National Development*. Zaheer Foundation Lecture,

Beer S. (1981). *Brain of The Firm*. Wiley. Chichester.

Beer S. (1983). *The Will of The People*. In «Journal of the Operational Research Society», 34 (8), pp. 797-810.

Benhabib S. (1996). *Toward a Deliberative Model of Democratic Legitimacy*. In Benhabib S. (a cura di). *Democracy and difference: Contesting the boundaries of the political*. Princeton University Press., pp. 67–94.

Berlin I. (1969). *Four essays on liberty*. Oxford. Oxford University Press.

- Berns T., Rouvroy A. (2013). *Algorithmic Governmentality and Prospects of Emancipation*. In «La Découverte», 177 (1), 2013, pp. 163-196.
- Bertot J. C., Jaeger P. T., Munson S., Glaisyer T. (2010). *Social media technology and government transparency*. In «Computer» 43(11), pp. 53–59.
- Black D. (1958). *The Theory of Committees and Elections*. Cambridge. Cambridge University Press.
- Boden M. A., (1996). *Autonomy and Artificiality*. In Boden M. A. (a cura di). *The philosophy of artificial intelligence*. Oxford University Press, pp. 95–107.
- Bohman J. (1998). *Survey article: The coming of age of deliberative democracy*. In «Journal of political philosophy», 6(4), pp. 400-425.
- Bongiovanni G. (2007). *Oggettività e morale. La riflessione etica del Novecento*. Pearson Italia Spa.
- Brayne S., (2021). *Predict and surveil: data, discretion, and the future of policing*. Oxford University Press. New York.
- Brennan J. (2016). *The ethics and rationality of voting*. In Zalta E. (a cura di). *The Stanford Encyclopedia of Philosophy*.
- Brigaglia M., Celano B. (2017). *Rivoluzione cognitivista e teoria del diritto: un programma di ricerca*. In «Diritto & questioni pubbliche», 2(2/2017), pp. 525-535.
- Brocklesby J., Mingers J. (2005). *The Use of The Concept Autopoiesis in The Theory of Viable Systems*. In «Systems Research and Behavioral Science», 22 (8), pp. 3-9.

- Brooks R. A. (2018). *Intelligence without reason*. Steels L., Brooks R. (a cura di). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Routledge, pp. 25-81.
- Brynjolfsson E., McAfee A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company. New York.
- 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, pp. 273-291.
- Buchanan A. (2002). Political Legitimacy and Democracy. In «Ethics», 112(4), pp. 689-719.
- Bullock J. B. (2019). *Artificial intelligence, discretion, and bureaucracy*. In «Am Rev Public Adm», 49, pp. 751–761.
- Burrell J. (2016). *How the machine 'thinks': Understanding opacity in machine learning algorithms*. In «Big data & society», 3(1).
- Calvino I. (2015). *Una pietra sopra*. Edizioni Mondadori.
- Capriati P. (2023). *Autopoiesi dei sistemi politici: il caso Cybersyn*. In «Philosophy Kitchen-Rivista di filosofia contemporanea», (18), pp. 149-162.
- Capriati P. (2024). *Ricerca l'intesa con una macchina. Compatibilità fra umani e soggetti artificiali nella teoria del discorso di Habermas*. In «Etica & Politica / Ethics & Politics», XXVI, 2024, 1, pp. 33-45

- Chalmers D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
- Chen L., Ning H., Nugent C. D., Yu Z. (2020). *Hybrid Human-Artificial Intelligence*. In «Computer», 53, pp. 14-17.
- Cheney-Lippold J. (2011). *A new algorithmic identity: Soft biopolitics and the modulation of control*. In «Theory, culture & society», 28(6), pp. 164-181.
- Chopra S., White L. (2004). *Artificial agents-personhood in law and philosophy*. In «ECAI», Vol. 16.
- Chrisley R. L. (2008). *Philosophical, Foundations of Artificial Consciousness*. In «Artificial Intelligence in Medicine», 44, pp. 119–137.
- Christiano T., (1993). *Social Choice and Democracy*. In Copp et al. (a cura di). *The Idea of Democracy*. Cambridge. Cambridge University Press, pp. 183–186.
- Christiano T., (1996). *The Rule of the Many: Fundamental Issues in Democratic Theory*. Boulder. Westview Press.
- Christiano T., (2008), *The Constitution of Authority*. New York. Oxford University Press.
- Christiano T. (2022). *Democracy*. In Zalta E. (a cura di). *The Stanford Encyclopedia of Philosophy*.
- Churchland P. M., Churchland P. S. (1990). *Could a machine think?* In «Scientific American», 262(1), pp. 32-39.

- Citron D. K., Pasquale F. (2014). *The scored society: Due process for automated predictions*. In «Wash. L. Rev.», 89, 1.
- Clancey W. J. (1993). *Situated action: A neuropsychological interpretation response to Vera and Simon*. In «Cognitive Science», 17(1), pp. 87-116.
- Cohen P., Feigenbaum E. (1989). *Learning and Inductive Inference*. In Barr A., Feigenbaum E. A. (a cura di). *The Handbook of Artificial Intelligence, volume III*. Kaufman. Los Altos, pp. 326-334.
- Cohen J. (1986). *An epistemic conception of democracy*. In «Ethics», 97(1), pp. 26-38.
- Cohen J. (1994). *Pluralism and Proceduralism*. In «Chicago-Kent Law Review», vol. 69.
- Cole D. J., Foelber R. (1984). *Contingent materialism*. In «Pacific Philosophical Quarterly», 65(1), pp. 74-85.
- Covrigaru A. A., Lindsay R. K. (1991). *Deterministic autonomous systems*. In «AI Magazine», 12(3), pp. 110-117.
- Cristianini N., Scantamburlo T. (2020). *On social machines for algorithmic regulation*. In «AI & SOCIETY», 35, pp. 645-662.
- D'Agostino M., Durante M. (2018). *Introduction: The governance of algorithms*. In «Philosophy & Technology», 31, pp. 499-505.
- Dahl R. (1956). *A Preface to Democratic Theory*. Chicago. University of Chicago Press.

- Dahl R. (1989). *Democracy and its critics*. New Haven. Yale University Press.
- Dahl R. (1990). *After the Revolution? Authority in a Good Society*. New Haven. Yale University Press.
- Danziger S., Levav J., Avnaim-Pesso L. (2011). *Extraneous factors in judicial decisions*. In «Proceedings of the National Academy of Sciences of the United States of America», 108, 17, 6889 ss.
- De Condorcet, N. (2014). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Cambridge University Press.
- Dekker S. W. A., Woods D. D. (2002). *MABA-MABA or Abracadabra? Progress on Human-Automation Coordination*. In «Cogn. Technol. Work», 4, pp. 240-244.
- Dennett D. C. (1987). *Fast Thinking*. In Dennett D. C. (a cura di). *The Intentional Stance*. Cambridge. MIT Press, pp. 324-337.
- Dennett D. C. (2018). *Dai batteri a Bach. Come evolve la mente*. Milano. Raffaello Cortina editore.
- Dessauer F. (1956). *Streit um die Technik*. Frankfurt. M. Josef Knecht.
- Domingos P. (2016). *L'Algoritmo Definitivo: La macchina che impara da sola e il futuro del nostro mondo*. Bollati Boringhieri.
- Downs, A. (1957). *An economic theory of political action in a democracy*. In «Journal of political economy», 65(2), 135-150.

Dryzek J. S. (2006). *Deliberative Global Politics: Discourse and Democracy in a Divided World*. Cambridge. Polity.

Dubber M. D. Pasquale F., Das S. (a cura di). (2020). *The Oxford handbook of ethics of AI*. Oxford Handbooks.

Dworkin G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Dworkin G. (2015). *The nature of autonomy*. In «Nordic Journal of Studies in Educational Policy».

Elster J. (1986). *The Market and The Forum: Three Varieties of Political Theory*. In Elster J., Hylland A. (a cura di), *Foundations of Social Choice Theory*. Cambridge University Press. Cambridge, pp. 103-132.

Elster J., (1998). *Deliberative democracy* (Vol. 1). Cambridge University Press.

Emery F. E. e Trist E. L. (1960). *Socio-technical Systems*. In «Management Sciences Models and Techniques», 2.

Enarsson T., Enqvist L., Naarttijärvi M. (2022). *Approaching the human in the loop-legal perspectives on hybrid human/algorithmic decision-making in three contexts*. In «Information & Communications Technology Law», 31(1), pp. 123-153.

Engelmann S., Chen M., Fischer F., Kao C. Y., Grossklags J. (2019). *Clear sanctions, vague rewards: how china's social credit system currently defines "Good" and "Bad" Behavior*. In «Proceedings of the conference on fairness, accountability, and transparency», pp. 69–78.

Ensign D., Friedler S. A., Neville S., Scheidegger C., Venkatasubramanian S. (2018). *Runaway feedback loops in predictive policing*. In «Proceedings of the Conference on fairness, accountability and transparency», pp. 160-171.

Espejo R. (2022). *Cybersyn, big data, variety engineering and governance*. In «AI & society», 37(3), pp. 1163-1177.

Estlund D., Landmore H. (2018). *The Epistemic Value of Democratic Deliberation*. In Bächtiger A., Dryzek J. S., Mansbridge J., Warren M. (a cura di). *The Oxford Handbook of Deliberative Democracy*. Oxford University Press. Oxford, pp. 113-131.

Estlund D. (1997). *Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority*. In Bohman J., Rehg W., *Deliberative Democracy: Essays on Reason and Politics*. Cambridge. MIT Press, pp. 173–204.

Estlund D. (2008). *Democratic Authority: A Philosophical Framework*. Princeton. Princeton University Press.

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

Ezenkwu C. P., Starkey A. (2019). *Machine autonomy: Definition, approaches, challenges and research gaps*. In «Intelligent Computing: Proceedings of the 2019 Computing Conference», Volume 1 (pp. 335-358).

Fisher, A. (2014). *Metaethics: an introduction*. Routledge.

Fishkin, J. S. (1992). *The dialogue of justice: Toward a self-reflective society*. Yale University Press.



Floridi, L. (1999), *Information Ethics: On the Theoretical Foundations of Computer Ethics*. In «Ethics and Information Technology», 1(1), pp. 37-56.

Floridi L., Sanders J. W. (2004). *On the morality of artificial agents*. In «Minds and machines», 14, pp. 349-379.

Floridia A. (2017). *Un'idea deliberativa della democrazia*. Il Mulino. Bologna.

Foa R. S., Mounk Y., (2016). *The Danger of Deconsolidation: The Democratic Disconnect*. In «Journal of Democracy», 27 (3) pp. 5-17.

Forst R. (2016). *The Justification of Basic Rights: A Discourse-Theoretical Approach*. In «Netherlands Journal of Legal Philosophy», 45(3), pp. 7–28.

förvaltningslag, 2017:900.

Frank M. R., Autor D., Bessen J. E., Brynjolfsson E., Cebrian M., Deming D. J., ... & Rahwan I. (2019). *Toward understanding the impact of artificial intelligence on labor*. In «Proceedings of the National Academy of Sciences», 116(14), pp. 6531-6539.

Frankfurt H. G. (1988). *The importance of what we care about: Philosophical essays*. Cambridge University Press.

Franklin S., Graesser A. (1996). *Is it an Agent, or just a Program? A Taxonomy for Autonomous Agents*. In «International workshop on agent theories, architectures, and languages». Springer Berlin Heidelberg, pp. 21-35.

Fung A. (2003), *Survey Article: Recipes for Public Spheres. Eight Institutional Design Choices and Their Consequences*. In «The Journal of Political Philosophy», 11 (3), pp. 338-367.

Geenens R., Tinnevelt R. (a cura di). (2009), *Does Truth Matter? Democracy and Public Space*. New York. Springer.

Gigerenzer G., Engel C. (2004). *Heuristics and the law*. Cambridge, London. MIT Press.

Gigerenzer G. (2008). *Why heuristics work*. In «Perspectives on psychological science», 3(1), pp. 20-29.

Gill, C. (1990). *The Person and the Human Mind: Issues in Ancient and Modern Philosophy*. New York. Oxford University Press.

Goodin R. E. (2003). *Reflective Democracy*. Oxford. Oxford University Press.

Goodin R. E. (2007). *Enfranchising All Affected Interests, and Its Alternatives*. In «Philosophy and Public Affairs», 35, pp. 40–68.

Goodin R. E. (2008). *Innovating Democracy: Democratic Theory and Practice After the Deliberative Turn*. Oxford University Press. Oxford.

Gould C. (1988). *Rethinking Democracy: Freedom and Social Cooperation in Politics, Economics and Society*. Cambridge University Press. New York.

Gupta J. N., Forgionne G. A., Mora T. M., Braun P., Brzostowski J., Kersten G., ... & Guszczka J. (2018). *Smarter together: Why artificial intelligence needs human-centered design*. Deloitte, London, n. 22, [Online]. Available:

[https://www2.deloitte.com/content/dam/insights/us/articles/4214\\_Smartertogether/DI\\_Smarter-together.pdf](https://www2.deloitte.com/content/dam/insights/us/articles/4214_Smartertogether/DI_Smarter-together.pdf)

Habermas J. (1979). *Communication and the Evolution of Society*. Boston. Beacon Press.

Habermas J. (1987). *The theory of communicative action*. (Vol. 1). Boston. Beacon Press.

Habermas J. (1995). *Reconciliation Through the Public Use of Reason: Remarks on John Rawls Political Liberalism*. In «The Journal of Philosophy», 92(3), pp. 109-131.

Habermas J. (2013). *Fatti e norme*. Bari. Laterza.

Habermas J. (2014). *Truth and justification*. John Wiley & Sons.

Habermas J. (2018). *Inclusion of the other: Studies in political theory*. John Wiley & Sons.

Hampton J (1998). *Political Philosophy*. Boulder. Westview Press.

Hare R. M. (1952). *The Language of Morals*. New York. Oxford University Press.

Harman G., Thomson J. J. (1996). *Moral relativism and moral objectivity*.

Haugeland J. (1981). *Semantic Engines: An Introduction to Mind Design*. In «Mind Design», 1, 32.

Haugeland J. (1985). *Intelligenza artificiale: il significato di un'idea*. Bollati Boringhieri. Torino.

Helbing D., Frey B. S., Gigerenzer G., Hafen E., Hagner M., Hofstetter Y., ... & Zwitter A. (2019). *Will democracy survive big data and artificial intelligence?* In «Towards digital enlightenment: Essays on the dark and light sides of the digital revolution», pp. 73-98.

Held D. (2006). *Models of Democracy*. Polity Press. Cambridge.

Hildebrandt, M. (2019). *Legal personhood for AI*. In Hildebrandt M. (a cura di). *Law for computer scientists and other folk*. Oxford University Press.

Hiltz S. R., Turoff M. (1978). *The Network Nation: Human Communication via Computer*. Addison-Wesley.

Hong L., Page S. (2004). *Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers*. In «Proceedings of the National Academy of Sciences of United States», 101(46), pp. 16385–16389.

Iliadis A., Russo F. (2016). *Critical data studies: An introduction*. In «Big Data & Society», 3(2).

Jelassi M. T., Foroughi A. (1989). *Negotiation support systems: an overview of design issues and existing software*. In «Decision Support Systems», 5(2), pp. 167-181.

Jennings N. R., Moreau L., Nicholson D., Ramchurn S., Roberts S., Rodden T., Rogers A. (2014). *Human-agent collectives*. In «Communications of the ACM», 57(12), pp. 80-88.

- Jordan M. I., Mitchell T. M. (2015). *Machine learning: Trends, perspectives, and prospects*. In «Science», 349(6245), pp. 255-260.
- Kelsen H. (1955). *Foundations of democracy*. In «Ethics», 66. pp. 1-101.
- Koops B. J, Hildebrandt M., Jaquet-Chiffelle D. O. (2010). *Bridging the accountability gap: rights for new entities in the information society*. In «Minnesota Journal of Law, Science & Technology», 11, pp. 497-561.
- Kuziemski M., Misuraca G. (2020). *AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings*. In «Telecommunications policy», 44(6).
- Laclau E., Mouffe C. (1985). *Hegemony and Socialist Strategy. Towards a Radical Democratic Politics*. London. Verso.
- Lakoff G., Johnson M. (2008). *Metaphors we live by*. University of Chicago press.
- Landmore H. (2012). *Why the Many are Smarter than the Few and Why It Matters*. In «Journal of Public Deliberation», 8 (1).
- Landmore H. (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton. Princeton University Press.
- Landmore H. (2022). *Can AI bring deliberative democracy to the masses*. In «HAI Weekley Seminar», pp. 1-35.
- Langvatn S. (2016). *Legitimate, But Unjust; Just, But Illegitimate: Rawls on Political Legitimacy*. In «Philosophy & Social Criticism», 42(2), pp. 132–153.

Lanier J. (2014). *Who owns the future?* Simon and Schuster.

Lanier J. (2018). *Ten arguments for deleting your social media accounts right now.* Random House.

Lauritsen M., Gordon T. F. (2009). *Toward a general theory of document modeling.* In «Proceedings of the 12th international conference on artificial intelligence and law», pp. 202-211.

Lefort C. (1998). *Democracy and Political Theory.* Cambridge. Polity Press.

Levitsky S., Ziblatt D. (2018). *How Democracies Die.* Crown.

Levy P. (1994). *L'Intelligence Collective: Pour Une Anthropologie du Cyberspace.* Editions La Decouverte.

Licklider J. C. R. (1960) *Man-Computer Symbiosis.* In «IRE Trans. Hum. Factors Electron», 14, pp. 4–11.

List C., Goodin R. E. (2001). *Epistemic Democracy: Generalizing the Condorcet Jury Theorem.* In «Journal of Political Philosophy», 9, pp. 277-306.

Luck M., D'Inverno M. (1995). *A Formal Framework for Agency and Autonomy.* In «Proceedings of the International Conference on Multiagent Systems», pp. 254-260.

Luhmann N. (1986). *The Autopoiesis of Social Systems (172-192).* In Geyer F., van der Zouwen J. (a cura di). *Sociocybernetic Paradoxes.* London. SAGE Publications.

- Mackie G. (2011). *The Values of Democratic Proceduralism*. In «Irish Political Studies», 26, pp. 439-453.
- Maffettone S. (2019). *Politica: idee per un mondo che cambia*. Mondadori Education.
- Malone T. (2018). *Superminds. The Surprising Power of People and Computers Thinking Together*. Little Brown.
- Malone T. W., Bernstein M. S. (2022). *Handbook of collective intelligence*. MIT press.
- Manin B. (1987). *On Legitimacy and Deliberation*. In «Political Theory». 15(3), pp. 338-368.
- Mansbridge J. (1999). *Everyday talk in the deliberative system. Deliberative politics*, pp. 211-239.
- Mansbridge J., Bohman J., Chambers S., Christiano T., Fung A., Parkinson J., Thompson D. F., Warren M. E. (2012). In Parkinson J., Mansbridge J. (a cura di). *Deliberative systems: Deliberative democracy at the large scale*. Cambridge University Press, pp. 1-26.
- Marr D. (1976). *Early processing of visual information*. In «Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences», 275(942), pp. 483-519.
- Marr D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

- Marr D., Nishihara H. K. (1978). *Representation and recognition of the spatial organization of three-dimensional shapes*. In «Proceedings of the Royal Society of London. Series B. Biological Sciences», 200(1140), pp. 269-294.
- Martí J. L. (2001). *Democracia y deliberación: una reconstrucción del modelo de Jon Elster*. In «Revista de estudios políticos», (113), pp. 161-192.
- Martí J. L. (2006a). *La república deliberativa: una teoría de la democracia*. Marcial pons.
- Martí J. L. (2006b). *The epistemic conception of deliberative democracy defended: Reasons, rightness and equal political autonomy*. *Deliberative democracy and its discontents*, pp. 27-56.
- Martí J. L. (2012). *Democracia y subjetivismo metaético*. In «Teoria politica», (2), pp. 111-136.
- Martí J. L. (2021). *The role of new technologies in deliberative democracy*. *Rule of Law vs Majoritarian Democracy*.
- Maturana H. R. (1987). *The Biological Foundations of Self-consciousness and The Physical Domain of Existence*. In Caianiello E. (a cura di) *Physics of Cognitive Processes*. Singapore. World Scientific, pp. 324-379.
- Maturana H. R., Varela F. J. (1973). *De Máquinas y Seres Vivo: Una teoría de la organización biológica*. Editorial Universitaria. Santiago de Chile.
- Maturana H. R., Varela, F. J. (1992). *Autopoiesi e cognizione: la realizzazione del vivente*. Trad. it di Stragapede A. Venezia. Marsilio Editori.



Maturana H. R., Varela F. J., Ceruti M. (1987). *L'albero della conoscenza*. Trad. it di Melone G.. Milano: Garzanti.

McCarthy J. (1973). *The Lighthill Debate*.

McCarthy J., Hayes P. J. (1981). *Some philosophical problems from the standpoint of artificial intelligence*. In Webber B. L., Nilsson N. J., *Readings in artificial intelligence*. Morgan Kaufmann, pp. 431-450.

McFarland D. (1994). *Towards robot cooperation. From animals to animats*, 3, pp. 440-444.

McFarland D., Bösner T. (1993). *Intelligent Behaviour in Animals and Robots*. Bradford. MIT Press.

Meneceur Y. (2023). *Artificial Intelligence, Public Administration, and the Rule of Law: Towards the Uncertainties of a New "Rule of Algorithms"*. In Suksi M. (a cura di). *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*. Springer Nature, pp. 117-145.

Michelman F. I. (1989). *Conceptions of Democracy in American Constitutional Argument: The Case of Pornography Regulation*. In «Tennessee Law Review». 562, pp. 291–319.

Mill J. S. (1861). *Considerations on Representative Government*. Prometheus Books. Buffalo.

Mingers J. C. (1995). *Information and meaning: foundations for an intersubjective account*. In «Information Systems Journal». 5(4), pp. 285-306.

- Mingers J. (2001). *Embodying information systems: the contribution of phenomenology*. In «Information and Organization (formerly Accounting, Management and Information Technology)», 11(2), pp. 103–128.
- Minsky M. (2020). *La società della mente*. Adelphi Edizioni spa.
- Mitchell T. (1997). *Machine Learning*. Ed. MacGraw-Hill.
- Mittelstadt B. D., Allo P., Taddeo M., Wachter S., Floridi L. (2016). *The ethics of algorithms: Mapping the debate*. In «Big Data & Society», 3(2).
- Mnih V., Kavukcuoglu K., Silver D., Rusu, A. A., Veness J., Bellemare M. G., Graves A., Riedmiller M., Fidjeland A. K., Ostrovski G., et al. (2015). *Humanlevel control through deep reinforcement learning*. In «Nature», 518.
- Monarcha-Matlak, A. (2021). *Automated decision-making in public administration*. In «Procedia Computer Science», 192, pp. 2077-2084.
- Moravec H. (1999). *Robot: Mere Machine to Transcendent Mind*. New York. Oxford University Press.
- Mouffe C. (1993). *The Return of the Political*. London. Verso.
- Mouffe C. (2005). *On the Political (Thinking in Action)*. London. Routledge.
- Mounk Y. (2018). *The People vs. Democracy. Why Our Freedom Is in Danger and How To Save It*. Harvard University Press.

- Mulgan G. (2018). *Big mind: How collective intelligence can change our world*. Princeton University Press.
- Murphy, K. P., (2012), *Machine Learning: A Probabilistic Perspective*. Cambridge. MIT Press.
- Nay, J. (2021) *Natural Language Processing for Legal Texts*. In Katz D. M., Dolin R., Bommarito M. (a cura di). *Legal Informatics*. Cambridge University Press.
- Newell A. (1982). *The knowledge level*. In «Artificial intelligence». 18(1), pp. 87-127.
- Nino C. S. (1997). *La constitució de la democràcia deliberativa*. Gedisa. Barcelona.
- Noveck B. (2009). *Wikigovernment. How Technology Can Make Governments Better, Democracy Stronger and Citizens More Powerful*. Brookings Institution.
- Noveck B. (2015). *Smart Citizens, Smarter State*. Harvard University Press.
- Nozick R. (2008), *Anarchia, Stato e Utopia*. Il Saggiatore
- Owen D., Smith G. (2015). *Survey article: Deliberation, democracy, and the systemic turn*. In «Journal of political philosophy», 23(2), pp. 213-234.
- Page S., (2007). *The difference. How the Power of Diversity Creates Better Groups, Firms, Schools*. Princeton University Press.
- Palumbo A. (2012). *Epistemic turn or democratic U-turn? on the tension between philosophical reasoning and political action in deliberative democracy*.

Unpublished manuscript available at [http://www.academia.edu/1371787/Epistemic\\_Turn](http://www.academia.edu/1371787/Epistemic_Turn)

Papakyriakopoulos O. (2022). *Political machines: a framework for studying politics in social machines*. In «AI & SOCIETY», pp. 1-18.

Pask G. (1958). *Organic Control and the Cybernetic Method*. In «Cybernetica», 1, pp. 155-173.

Pasquale F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Pasquale F. (2019). *A Rule of Persons, Not Machines: The Limits of Legal Automation*. In «George Washington Law Review», 87(1), pp. 1-55.

Pasquale F. (2020). *New laws of robotics*. Harvard University Press.

Penrose R. (1991). *The emperor's new mind*. In «RSA Journal», 139(5420), pp. 506-514.

Pescetelli N. (2021). *A brief taxonomy of hybrid intelligence*. In «Forecasting», 3(3), pp. 633-643.

Peter F. (2007). *Democratic legitimacy and proceduralist social epistemology. Politics*. In «Philosophy & economics», 6(3), pp. 329-353.

Pickering A. (2010). *The cybernetic brain*. Chicago. University of Chicago Press.

Przeworski A. (2019). *The Crises of Democracy*. Harvard University Press.

Pylyshyn Z. W. (1980). *The 'causal power' of machines*. In «Behavioral and Brain Sciences», 3(3), pp. 442-444.

Rawls J., (1971). *A Theory of Justice*. Cambridge. The Belknap Press of Harvard University Press.

Rawls J., (1993). *Political Liberalism*. New York. Columbia University Press.

Rawls J., (1995). *Reply to Habermas*. In «The Journal of Philosophy», 92(3). pp. 132-180.

Raz J. (1986). *The morality of freedom*. Clarendon Press.

Reijers W., Orgad L., De Filippi P. (2023). *The rise of cybernetic citizenship*. In «Citizenship Studies», 27(2), pp. 210-229.

Rheingold H. (2003). *Smart mobs: The next social revolution*. Basic books.

Rheingold H. (2012). *Net smart: How to thrive online*. MIT Press.

Roberts S. L. (2019). *Big data, algorithmic governmentality and the regulation of pandemic risk*. In «European Journal of Risk Regulation», 10(1), pp. 94-115.

Ropohl G. (1999). *Philosophy of socio-technical systems*. In «Society for Philosophy and Technology Quarterly Electronic Journal», 4(3), pp. 186-194.

Rosenblueth A., Wiener N., Bigelow J. (1943). *Behavior, Purpose and Teleology*. In «Philosophy of Science», 10, pp. 18-24.

Rostbøll C. F. (2015). *The Non-Instrumental Value of Democracy: The Freedom Argument*. In «Constellations», 22. pp. 267–278.

Rouvroy A. (2011). *Technology, Virtuality and Utopia. Governmentality in an Age of Autonomic Computing*. In Hildebrandt M., Rouvroy A. (a cura di). *Law, Human Agency and Autonomic Computing. Philosophers of Law Meet Philosophers of Technology*, Abingdon, Routledge.

Rouvroy A., Stiegler B., (2016). *Il regime di verità digitale. Dalla governamentalità algoritmica a uno nuovo Stato di diritto*. In «La Deleuziana», 3, 2016.

Rummens S., (2018). *Deliberation and Justice*. In Bächtiger A., Dryzek J. S., Mansbridge J., Warren M. (a cura di), *The Oxford Handbook of Deliberative Democracy*. Oxford University Press. Oxford, pp. 132-143.

Russell S. J., Norvig P. (2016). *Artificial Intelligence, A Modern Approach*. Global Edition. Pearson Education.

Saffon M. P. E., Urbinati N. (2013). *Procedural Democracy, the Bulwark of Equal Liberty*. In «Political Theory», 41. pp. 441–81.

Sankar S. (2021) Transcript of “The Rise of Human-Computer Cooperation”. Available online:  
[https://www.ted.com/talks/shyam\\_sankar\\_the\\_rise\\_of\\_human\\_computer\\_cooperation/transcript?language=en](https://www.ted.com/talks/shyam_sankar_the_rise_of_human_computer_cooperation/transcript?language=en).

Sartor G. (2020). *Artificial intelligence and human rights: Between law and ethics*. In «Maastricht Journal of European and Comparative Law», 27(6), pp. 705-719.

Sayre-McCord G. (1986). *The many moral realisms*. In «The Southern Journal of Philosophy», 24(5), 1.

Schelsky H., (1961). *Der Mensch in der wissenschaftlichen Zivilisation*. In «VS Verlag für Sozialwissenschaften».

Schumpeter J. (2008 [1942]). *Capitalism, Socialism and Democracy*. New York, Harper-Collins.

Searle, J. 1999. *The Chinese Room*.

Selbst A. D., Barocas S. (2018). *The intuitive appeal of explainable machines*. In «Fordham L. Rev.», 87, 1085.

Shadbolt N., O'Hara K., De Roure D., Hall W. (2019). *The theory and practice of social machines*. Springer, Berlin.

Silver D., Huang A., Maddison C. J., Guez A., Sifre L., van den Driessche G., Schrittwieser J., Antonoglou I., Panneershelvam V., Lanctot M., et al (2016). *Mastering the game of Go with deep neural networks and tree search*. In «Nature». 529, pp. 484-489.

Simoncini A. (2019). *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*. In «BioLaw Journal-Rivista di BioDiritto», (1), pp. 63-89.

Singer P. (1973). *Democracy and Disobedience*. Oxford University Press. Oxford.

Smithers T. (1997). *Autonomy in robots and other agents*. In «Brain and cognition», 34(1), pp. 88-106.

Smithers T. (2018). *Are autonomous agents information processing systems?* In Steels L., Brooks R. (a cura di). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Routledge, pp. 123-162.

Smolensky P. (1986). *Information processing in dynamical systems: Foundations of harmony theory*. In Rumelhart D. E., McClelland J. L., PDP Research Group. (a cura di). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*. The MIT press, pp. 194-281.

Solaiman S. M. (2017). *Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy*. In «Artificial Intelligence Law», 25(2), pp. 155-179.

Solum L. B. (2020). *Legal personhood for artificial intelligences*. In Wallach W., Asaro P. (a cura di) *Machine ethics and robot ethics*. Routledge.

Solum L. B. (2019). *Artificially intelligent law*. In «BioLaw Journal-Rivista di BioDiritto», 1, pp. 53-62.

Stark L. (2018). *Algorithmic psychometrics and the scalable subject*. In «Social Studies of Science», 48(2), pp. 204-231.

Steels L. (1995). *When are robots intelligent autonomous agents?* In «Robotics and Autonomous systems», 15(1-2), pp. 3-9.

Suksi M. (a cura di). (2023). *The Rule of Law and Automated Decision-Making: Exploring Fundamentals of Algorithmic Governance*. Springer Nature.

Sunstein C. (2002). *The Law of Group Polarization*. In «The Journal of Political Philosophy», 10 (2), pp. 175-195.



Sunstein C. (2006). *Infotopia: How Many Minds Produce Knowledge*. OUP.

Surowiecki J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York. Random House.

Tapscott D., Williams A. (2006). *Wikinomics: How Mass Collaboration Change Everything*. Penguin.

Truman D. (1959). *The governmental process*. New York. Alfred A. Knopf.

Tufekci Z. (2014). *Engineering the public: Big data, surveillance and computational politics*. In «First Monday».

Tufekci Z. (2015). *Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency*. In «Colo. Tech. LJ», 13, 203.

Tufekci Z. (2017). *Twitter and Tear Gas*. Yale University Press.

Tully J. (1995). *Strange Multiplicity: Constitutionalism in an Age of Diversity*. Cambridge. Cambridge University Press.

Turing A. M. (1950), *Computing Machinery and Intelligence*. In «Mind», 59, pp. 433-460.

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), pp. 1124-1131.

Ulrich W. (1981). *A Critique of Pure Cybernetic Reason: The Chilean Experience with Cybernetics*. In «Journal of Applied Systems Analysis», 8, pp. 33-59.

Vahidov R. (2006). *E-negotiation systems and software agents: Methods, models, and applications*. In Gupta J. N., Forgionne G. A., Mora M. (a cura di). *Intelligent decision-making support systems: foundations, applications and challenges*. pp. 271-300.

Van Gelder T. (1995). *What might cognition be, if not computation?* In «The Journal of Philosophy», 92(7), pp. 345-381.

Van Parijs, P. (2011). *Just Democracy: The Rawls–Machiavelli Programme*. Colchester. ECPR Press.

Varela F. J., (1979). *Principles of Biological Autonomy*. North-Holland. Elsevier.

Varela F. J., Thompson E., Rosch E. (2017). *The embodied mind, revised edition: Cognitive science and human experience*. MIT press.

Véliz C. (2021). *Privacy is power*. Brooklyn. Melville House.

Vermaas P., Kroes P., van de Poel I., Franssen M., Houkes W., (2011) *A philosophy of technology. From technical artefacts to sociotechnical systems*. Morgan and Claypool. San Rafael.

Vilalta R., Drissi Y. (2002). *A perspective view and survey of meta-learning*. In «Artificial intelligence review», 18, pp. 77-95.

Waldman E., (2019). *Power, process, and automated decision-making*. In «Fordham Law Rev», 88, pp. 613-632.

- Waldron J., (1999). *Law and Disagreement*. Oxford. Oxford University Press.
- Weber M., (1964). *The Theory of Social and Economic Organization*. Simon and Schuster.
- Wheeler M., (2020). *Autonomy*. In Dubber M. D., Pasquale F., Das, S. (a cura di). *The Oxford handbook of ethics of AI*. Oxford University Press, pp. 343-357.
- Wiener, N. (1968). *La cibernetica*. Trad. it. di Barosso G. Milano. Il Saggiatore.
- Wiggins D. (2001). *Sameness and substance renewed*. Cambridge University Press.
- Williams B. (2002). *Truth and Truthfulness*. Princeton. Princeton University Press.
- Williams A., Srnicek, N. (2018). *Manifesto accelerazionista*. Gius. Laterza & Figli Spa.
- Yeung K. (2018). *Algorithmic regulation: A critical interrogation*. In «Regulation & governance», 12(4), pp. 505-523.
- Zarsky, T. (2016). *The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making*. In «Science, Technology, & Human Values», 41(1), pp. 118-132.
- Zuboff, S. (2019). *Il capitalismo della sorveglianza: il futuro dell'umanità nell'era dei nuovi poteri*. Luiss University Press.