Lo Sguardo
_rivista di filosofia

N. 34_2022 (I)

Algoritmo_
Genealogia, teoria, critica

_a cura di Igor Pelgreffi,
Ubaldo Fadini, Massimiliano Badino

_ISSN: 2036-6558

Lo Sguardo – Rivista di Filosofia

Aut. Tribunale di Roma n° 387/2011 – ISSN: 2036-6558
InSchibboleth Edizioni

Comitato Scientifico:

Nunzio Allocca (sapienza Università di Roma) – Antonello D'Angelo (sapienza Università di Roma) – Lisa Block de Behar (Universidad de la República) – Paolo D'Angelo (Università degli studi Roma Tre) – Mário Santiago de Carvalho (Universidade de Coimbra) – Roberto Esposito (Scuola Normale Superiore) – Miguel Angel Granada (Universidat Autònoma de Barcelona) – Richard Kearney (Boston College) – Thomas Macho (Humboldt Universität zu Berlin) – Marcello Musté (sapienza Università di Roma) – Maria Teresa Pansera (Università degli studi Roma Tre) – Fabio Polidori (Università degli studi di Trieste) – Lorena Preta (Psicanalista, membro ordinario SPI) – Paola Rodano (sapienza Università di Roma) – Wolfgang Rother (Universität Zürich) – Emanuela Scribano (Università di Venezia) – Donald Philip Verene (Emory University) – José Luis Villacañas (Universidad Complutense de Madrid) – Christoph Wulf (Freie Universität zu Berlin) ❧ † Jean-Luc Nancy (Université de Strasbourg)

Caporedattori:
Simone Guidi (Coordinamento scientifico-editoriale)
Antonio Lucci (Rapporti con stampa e editoria)

Comitato di Redazione:
Cristina Basili, Federica Buongiorno, Marco Carassai, Igor Pelgreffi, Libera Pisano, Alberto Romele.

Collaboratori Editoriali:
Luca Tonetti (Revisione Editoriale) ❧ Emanuela Carta, Eleonora De Caroli, Riccardo Fraddosio, Ambrogio Garofano, Maria Enrica Giannuzzi, Roberto Granieri, Gianluca Longa, Alessandro Nanni, Emiliano Paparazzo, Manlio Perugini, Federica Pitillo, Lorenzo Sagripanti, Laura Turano.

Sito web: www.losguardo.net
Contatti: redazione@losguardo.net

Lo Sguardo è una rivista elettronica di filosofia *Full Open Access* pubblicata da InSchibboleth Edizioni. La rivista pubblica con cadenza semestrale (quadrimestrale fino al 2017) numeri esclusivamente monotematici costituiti da articoli scientifici inediti, saggi-intervista, traduzioni di estratti da opere scientifiche significative e di recente pubblicazione o articoli rilevanti per la comunità scientifica, recensioni di libri ed eventi culturali.

Le sezioni che compongono *Lo Sguardo* sono le seguenti:

Articoli: la sezione ospita contributi scientifici prodotti e sottoposti su invito diretto della redazione. Tutti i contributi della detta sezione sono sottoposti a doppia *blind peer review*.

Interviste: la sezione ospita saggi-intervista ad autori di opere o edizioni di opere rilevanti per il panorama scientifico e luminari di chiara fama sulle questioni di competenza dei singoli numeri.

Contributi: la sezione ospita contributi scientifici prodotti e sottoposti in risposta alle *call for papers* pubblicate per ogni singolo numero dai relativi curatori. Tutti i contributi della detta sezione sono sottoposti a doppia *blind peer-review.*

Testi: la sezione ospita prime traduzioni italiane di estratti da opere scientifiche significative e di recente pubblicazione o articoli rilevanti per la comunità scientifica e/o di difficile reperibilità.

Recensioni, discussioni e note: la sezione ospita recensioni, note o discussioni che abbiano come oggetto pubblicazioni scientifiche.

Scopus® Clarivate Web of Science™ OpenAIRE DOAJ

N. 34, 2022 (I)
Algoritmo. Genealogia, teoria, critica
a cura di Igor Pelgreffi, Ubaldo Fadini, Massimilano Badino

Indice:

Algoritmo. Genealogia, teoria, critica
Editoriale di Igor Pelgreffi, Ubaldo Fadini, Massimilano Badino
. . . . . pp. 5-10

¶ Sezione Prima – *Genealogia del concetto di algoritmo*

Contributi/1: *Alan Turing and the cognitive foundation of the concept of algorithm*
di Simone Pinna, Marco Giunti . . . . . pp. 13-29

Contributi/2: *Intentionalities of the Algorithm: Historical Practices and Socio-Computing Infrastructures. A Philosophical Account*
di Javier Toscano . . . . . pp. 31-52

Contributi/3: *Produire l'inconnu. L'intelligence artificielle au prisme des ambivalences du rationalisme leibnizien*
di François Levin . . . . . pp. 53-66

¶ Sezione Seconda – *Teoria ed epistemologia dell'algoritmo*

Interviste/1: *On Algorithms: Ethical and Epistemological Questions. Interview with Igor Pelgreffi*
di Luciano Floridi . . . . . pp. 69-78

Contributi/4: *Dialettica dell'algoritmo. Per una teoria critica*
di Massimiliano Carli, Vittorio Lubrano . . . . . pp. 79-98

Contributi/5: *Dalla cognizione alla comunicazione. Epistemologia degli algoritmi secondo la teoria dei sistemi*
di Alberto Giustiniano . . . . . pp. 99-117

Articoli/1: *Notes for a political epistemology of algorithms*
di Massimiliano Badino . . . . . pp. 119-133

¶ Sezione Terza – *Algoritmo, etica e corpo*

Contributi/6: *Oltre l'algoritmo. Questioni etiche e trasformazioni antropologiche*, di Filippo Pianca . . . . . pp. 137-154

Contributi/7: *Piattaforme online e polarizzazione politica. Alcuni spunti dalla teoria spinoziana degli affetti*
di Marianna Capasso . . . . . pp. 155-175

Articoli/2: *Algorithm and Learning. Ethical and Theoretical Hypothesis on Empathic Robotics and Social Robotics*
di Igor Pelgreffi . . . . . pp. 177-203

Contributi/8: *Algoritmi viventi tra Shannon e Simondon*
di Francesca Sunseri . . . . . pp. 205-218

¶ Sezione Quarta – *Algoritmo e società*

Interviste/2: *Questioning Algorithms. Interview with Igor Pelgreffi*
di Andrew Feenberg . . . . . pp. 221-227

Contributi/9: *Governare il comportamento: le radici comportamentiste del capitalismo digitale*, di Marco Dal Pozzolo . . . . . pp. 229-249

Contributi/10: *Oltre il paradigma Smart City. Ontologia ed epistemologia dell'intelligenza urbana*
di Otello Palmini . . . . . pp. 251-271

Contributi/11: *L'algoritmo nelle trame del lavoro immateriale*
di Vincenzo Pallara . . . . . pp. 273-293

Articoli/3: *Divenire classe. Per una critica del paradigma estrattivista applicato al capitalismo digitale*
di Roberto Ciccarelli . . . . . pp. 295-318

Recensioni, discussioni e note . . . . . pp. 321-398

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7768552

Editoriale

Algoritmo
Genealogia, teoria, critica

Igor Pelgreffi 0000-0002-4490-009X

Ubaldo Fadini 0000-0002-1425-0540

Massimilano Badino 0000-0002-3521-9454

Algorithm: Genealogy, Theory, Critique

The paper presents the scientific setting of the monographic issue *Algorithm. Genealogy, theory, critique*. Actually, today's digital society poses the urgency of which could be the suitable theoretical-philosophical tools for a consistent reflection on the theme of the widespread diffusion of algorithms in our historical-social world. The approach of the articles of the issue is twofold. One concerns the analysis of the history of the concept of algorithm, and therefore fits into the vein of the history of ideas and at the same time, into the folds of epistemological questions that the algorithm poses within mathematical and information theories, understood in their historical evolution. The other is more directly related to the ethical and political consequences of the extensive application of algorithms in individual and social life. In the first aspect, continuity and discontinuity are shown with respect to references such as Turing or Babbage, but also to the origins of the universal calculus in Leibniz and in Modern Philosophy as well. In the second, the analyses place the topics within the framework of human-machine ethical dilemmas, as well as international guidelines on AI ethics. Here therefore a series of open questions are developed ranging from *artificial empathy* linked to algorithms or the future role of *Machine Learning*, up to the critique of 'platform capitalism', here with references to the most up-to-date critical thinking, such as Hardt, Zuboff, Ciccarelli, also by re-actualizing Marx's positions on the replacement of man by the intelligent machines.

***

Lo si ripete spesso: l'algoritmo è oggi un elemento pervasivo, strutturale e strutturante, nelle nostre società, un elemento in grado di condizionare in maniera crescente i nostri stili di vita (il lavoro, lo studio, la salute, il 'tempo di riposo'…): forse, un nuovo *Leitmotiv* storico-esistenziale. Il numero 34 della rivista *Lo Sguardo*, monografico su *Algoritmo. Genealogia, teoria, critica*, tenta di occuparsene in termini filosofici. La cosa è probabilmente utile, benché risulti al contempo estremamente semplice e molto complicata, per le medesime ragioni.

Infatti, esiste ormai un numero rilevante di testi di impianto filosofico sui temi dell'intelligenza artificiale, della robotica, dell'automatizzazione del nostro mondo storico-sociale, dai quali[1] emerge il valore quasi scontato, come *datum*, della assoluta centralità dell'algoritmo. Con elenco non esaustivo, si pensi solo al concetto di *infosfera* (Luciano Floridi), di *governamentalità algoritmica* (Antoinette Rouvroy), di un'etica della macchina intesa specificamente come *algor-etica* (Paolo Benanti), di *società automatica* (Bernard Stiegler). Data tale molteplicità – contemporaneamente *datum* e *doxa* – nel presente monografico si è scelto di adottare un taglio, rappresentato in realtà da due versanti.

Un versante porta alla storia del concetto di algoritmo; l'altro all'approfondimento della sua portata etico-politica. Seguendo la definizione corrente, l'algoritmo non sarebbe altro che «un procedimento di calcolo esplicito e descrivibile con un numero finito di regole che conduce al risultato dopo un numero finito di operazioni»[2]; esso rappresenta dunque un *dispositivo procedurale* con caratteristiche di non ambiguità, di ricorsività, di ottenimento costante del medesimo. In particolare, sul piano filosofico, l'algoritmo sarebbe dotato di una propria *autonomia* funzionale: la sequenza delle sue 'azioni' risulta, a un certo momento, indipendente dall'operatore (macchina o un essere umano che sia). Il primo versante prova quindi a chiarire l'origine storica e logica di tale costrutto di ripetizione automatica, che evidentemente è anche alla base dell'informatica attuale. In termini filosofici, un algoritmo è tale se ripete se stesso; ma è oggi in grado di auto-modificarsi? Il tema dell'intelligenza dell'algoritmo, cioè del *Machine Learning*, amplia decisamente le coordinate epistemologiche della questione. Qui entra in campo il secondo versante, che espande e completa i temi del primo in una chiave *interpretativa* più etica e, successivamente, socio-politica. Da questo punto di vista, i due contributi-intervista di Andrew Feenberg, *Questioning algorithms*, e di Luciano Floridi, *On algorithms: ethical and epistemological questions*, sono sintomatici di questa doppia direzione. L'uno, quello di Feenberg, ben marcato da coordinate di indagine materialistica, mostra – in coerenza con la sua lunga ricerca, almeno da *Questioning technology* del 1999 – l'assoluta necessità di mostrare, ogni volta, l'origine socio-storica delle nuove tecnologie e dei nuovi linguaggi/idee ad esse interallacciati. L'altro, quello di Floridi, pone in maniera consistente il baricentro dell'indagine sull'analisi 'logica' del tema algoritmico, partendo però da un'attenta apertura ai nodi più attuali, irrisolti o futuri, delle *Information and Communication Technologies* o delle *Digital Technologies* comunque intese. Entrambi, rappresentativi di tendenze

[1] Oltre alle ampie bibliografie in calce ai vari contributi (a cui si invita il lettore), segnaliamo, tra gli studi più recenti e con più vasta e aggiornata bibliografia: C. Bartnek et al., *An Introduction to Ethics in Robotics and AI*, Springer, 2021 (on line book: https://link.springer.com/book/10.1007/978-3-030-51110-4) e F. Fossa, V. Schiaffonati, G. Tamburrini, *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Carocci, Roma 2021. Molto utile, semrpe in chiave di esplorazione bibliografica multivariata, anche il fascicolo *Come pensa la macchina? Incognite dell'Intelligenza artificiale* (a cura di M. Pacini), «aut aut», 392, dic 2021.
[2] Cfr. la voce 'algoritmo', Vocabolario on line Treccani [www.treccani.it/vocabolario/algoritmo/ (consultato il 12.11.2022)].

in atto nel dibattito scientifico sul tema, si mostrano sensibili anche a un'altra questione chiave, e cioè che occorre oggigiorno comprendere i modi, i linguaggi e le forme storiche di come un discorso sull'algoritmo possa ri-significare un discorso – anche – sull'umano: sui suoi limiti, sui suoi superamenti, sulle sue 'ibridazioni', intesi però sempre 'in processo', cioè a dire oltre ogni dicotomia ingenuo-statica tra tecnofobìa e tecnofilìa.

Le prime due sezioni di questo numero monografico – *Genealogia del concetto di algoritmo* (contributi di Simone Pinna e Marco Giunti, Javier Toscano e François Levin) e *Teoria ed epistemologia dell'algoritmo* (contributi di Luciano Floridi, Massimiliano Carli e Vittorio Lubrano, Alberto Giustiniano, Massimiliano Badino) – si incentrano sulla storia delle idee, sulle indagini genealogiche e epistemologiche, in senso esteso, relative all'origine dell'algoritmo, come si diceva, inteso quale calcolabilità sistematica, come ripetizione di un codice scritto in anticipo sul tempo, e dunque anche come dispositivo calcolante che è sempre, alternativamente, di previsione o di controllo. Su un piano complessivo, emergono riferimenti a una linea, consolidata negli studi di settore, che inquadra la concezione contemporanea dell'algoritmo quale frutto dei lavori di Babbage, Turing, Church e altri, ma che si spinge poi a indagare i territori greci o medievali. Senza dimenticare, tuttavia, l'eredità dei classici della storia della filosofia e delle scienze 'moderne', da Descartes a Leibniz – molto presente in queste sezioni – senza omettere, per altre ragioni, più complesse probabilmente, Spinoza. L'interesse aggiuntivo di questa sezione, specie *Teoria ed epistemologia dell'algoritmo*, è rappresentato dalla circostanza che frequentemente al tema epistemologico e di scienza dell'informazione gli autori associno una dimensione euristica maggiormente estesa, e che sembra egualmente necessaria all'organizzazione teorica del problema. Rientra in ciò la messa a fuoco di una *dialettica* dell'algoritmo, che apre a un'opzione di teoria critica; oppure l'approccio 'culturale' (emblematico il ricorso a Finn e alla sua idea dell'algoritmo come processo logico assimilabile a una «cultural machine»); oppure il ricorso alla cornice della *teoria dei sistemi* (Luhmann) e, ancora, l'idea che ogni analisi dei big data su base algoritmica sia sempre politicamente *situata*, mai neutrale, nel senso che una qualche «epistemic performance» (Badino) sarebbe strutturale all'operare stesso dell'algoritmo.

La sezione 3, *Algoritmo, etica e corpo* (contributi di Filippo Pianca, Marianna Capasso, Igor Pelgreffi e Francesca Sunseri), tocca invece le questioni etiche più classiche – rapporto uomo/macchina; esonero/sostituzione; decisione, ripetizione automatica e libertà – sulla scorta delle risultanze più aggiornate dell'antropologia filosofica e di confronti con le linee guida internazionali per l'etica dell'IA. Oltre a questo, tuttavia, i saggi tentano anche di ragionare criticamente su quali caratterizzazioni etiche, eventualmente aggiornate, siano oggi pensabili o quantomeno praticabili. Si tratta qui di esplorare alcune 'questioni limite' dell'etica stessa. Una, per esempio, è l'eventualità di integrazione reciproca di agenti umani e agenti artificiali, spesso giocabile su piani meno consueti, come quello degli affetti o della cosiddetta *empatia artificiale*. Un'altra, in continuità

con la prima, è tra le frontiere teoriche più spinose, in campo etico ma non solo: la robotica sociale quale limite delle capacità di *Machine Learning* dell'algoritmo, capacità ancora inesplorate e non sempre riconducibili ad un alveo 'razionale', con la rilevazione delle sempre maggiori lacune nella cosiddetta *explanaibility* delle procedure artificiali intelligenti.

La sezione 4, *Algoritmo e società* (contributi di Andrew Feenberg, Marco Dal Pozzolo, Otello Palmini, Vincenzo Pallara, Roberto Ciccarelli), ospita testi fra loro anche diversi, quanto a impostazioni e riferimenti, che spaziano dai motivi del capitalismo della sorveglianza (Zuboff), alle analogie tra dominio algoritmico e teorie comportamentiste (Skinner) con anche apertura al campo biopolitico (mediante Foucault o Canguilhem); da Lefebvre, Jonas e il tema degli spazi *smart* urbani, sino alle filosofie critiche di David Harvey o Roberto Ciccarelli, transitando dalle ben note analisi di Hardt e Negri. Andando a ritroso, si arriva sino a Marx, con riferimenti al tema del *General Intellect* o a quello dell'automaticità della riproduzione stessa del Capitale, ivi compreso il capitale umano. In questi saggi, tale inclinazione è misurabile anche mediante la loro riapertura a motivi, soprattutto ma non solo, presenti nell'operaismo italiano quale chiave interpretativa del *Digital Labour* e delle implicazioni dell'algoritmo in termini di *processi di soggettivazione*. In altri termini, riassumendo, le analisi qui rivelano un problema condiviso, appunto quello dell'impatto sulla vita quotidiana, sulle forme di sfruttamento 'algoritmo-centriche', sulla proposta di un pensiero minimamente critico su questa sorta mondo (digitale) totalmente amministrato, per declinare diversamente Adorno e Horkheimer.

Sembra qui compiersi, in tal modo, quella connessione tra l'algoritmo quale nozione logica, soltanto matematica, e la sua 'sostanzializzazione' materiale – connessione che mostra la concreta realtà dell'algoritmo quale elemento di una logica storica di diametro molto più ampio – a conferma forse della liceità dei 'due versanti' che si erano ipotizzati nell'impostazione.

Ciò detto, una sintesi tematica dei saggi appare piuttosto ardua. Tentando, diciamo che il primo aspetto che emerge è che trattasi, evidentemente, di sondaggi su geologie assai variabili. Troppo variabili? Come interpretare questo dato? La sensazione è che, pur nella puntualità di molti riferimenti – teoretici e epistemologici –, spesso anche molto classici, la ricerca stia ancora lavorando a una sorta di 'bibliografia generale' aggiornata, cioè a una preliminare definizione di campo che si voglia minimamente consistente. In altri termini: i sondaggi che i saggi forniscono, localmente chiarificatori di una faccia del prisma, mostrano di convesso la necessità di confronti realmente (e non retoricamente) multidisciplinari, se non transdisciplinari. Del resto, diversi saggi aprono a una vera pluralità di rapporti eterogenei tra diversi campi, che spostano continuamente l'argomentazione sull'algoritmo dal piano logico o storico non solo a quello, prevedibile, della sociologia, ma anche verso la letteratura, la critica letteraria o l'estetica. Difficile dirlo, ma tutto ciò è forse un indice del fatto che per definire l'orografia o la geologia di questo terreno comune, per costruire, cioè, questo campo di linguaggio o quantomeno di area semantica algoritmo-

specifica – sempre che qualcosa del genere sia possibile – occorra ulteriore lavoro di analisi sempre associato, tuttavia, a una maggiore disponibilità alla *plasticità* ed *estroversione* dei saperi. Conferma indiretta, tutto ciò, della grande pervasività/impatto del tema 'algoritmo'? ulteriore prova della centralità del suo operare nel piano prelogico in cui si svolgono le nostre vite, prima che divengano 'oggetto'?

Un secondo aspetto che, nell'insieme, sembra imporsi all'attenzione (strettamente connesso al primo) è il seguente: posta la conferma delle tesi evoluzioniste, per cui l'algoritmo in fondo è null'altro che il coerente sviluppo storico di un impianto tecnico del mondo occidentale, se si vuole di una *razionalità strumentale* molto interna alla Ragione (da cui si spiegano le frequenti risonanze, nel monografico, con un approccio come quello della *Dialettica dell'illuminismo*), quel che emerge è la necessità di un'ottica differente: non un'ottica convergente ma, semmai, un'ottica *rifrangente ma non totalmente disperdente* i raggi provenienti dall'oggetto di studio. Ammesso – di nuovo – che quanto passa sotto il nome di 'algoritmo' sia assimilabile all''oggetto'. In effetti, almeno la metà dei saggi – compresi quelli di Floridi e Feenberg – insiste su questa istanza, che è di domanda ma al contempo di realistica presa d'atto di una difficoltà di messa a fuoco. Da qui, ribaltando, emerge il tema della sollecitazione che l'algoritmo pone alla filosofia. E tale sollecitazione non è solo teorica o culturale, ma anche politica: dove *ci posizioniamo* nei confronti degli enormi interrogativi sull'algortimizzazione del mondo? Che tipo di 'soggetto' vorremmo essere, in questi contesti di indagine? Si noti che è a partire da queste risonanze problematiche che rientra potentemente in gioco il tema dei corpi, cioè il tema della società come *campo di intercorporeità indefinita*, in cui possono prendere forma inediti tipi di relazioni e di interazioni, giocate in quel *milieu* tra interattività e interpassività con gli algoritmi così tipico dei fenomeni in esame. E sia quando questa tensione affiora in modo molto diretto, proponendo riflessioni di critica radicale del 'capitalismo delle piattaforme', sia quando essa si coniuga a soluzioni più raffinate o 'progressiste', la spinta principale pare essere non tanto e non solo verso la definizione di una filosofa politica, quanto piuttosto verso l'affermazione di uno *studio teorico degli aspetti politici* che l'algoritmo rivela. La questione politica dell'algoritmo, nello specifico, viene presa in carico mediante ottiche di impianto poststrutturalista, attente da sempre agli sviluppi della linea *automa-programma-macchinico*; oppure mediante ottiche di impianto dialettico, legate cioè al pensiero critico materialista più tradizionale; oppure ancora, ed è forse uno dei lati più interessanti, mediante tentativi di ibridazioni di queste due macro categorie, spesso in aderenza al *coté* di studi di filosofia della tecnica (si spazia da Heidegger ad autori posteriori, quali Shannon, Simondon o Guattari). Questi tratti 'plurali' di approccio rappresentano, d'altra parte, una conferma del tenore spesso sovradeterminato, che probabilmente è tale per necessità, dei saggi qui pubblicati, per le ragioni che si è tentato di esaminare. Tale caratteristica è sia una difficoltà che una ricchezza dello sguardo, che trova conferma anche nell'eterogeneità delle tre note di lettura che chiudono il monografico, da considerarsi parte integrante dello stesso: quella al libro, del 2018, di Dominique

Cardon, *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei big data* (di Marko Ćeranić), quella al libro del 2022 di Roberto Ciccarelli, *Labour-power. Virtual and Actual in Digital Production* (di Ubaldo Fadini) e quella del volume di studi, uscito nel 2020, curato da Christian Barone, *L'algoritmo pensante. Dalla libertà dell'uomo all'autonomia delle intelligenze artificiali* (di Remo Trezza).

Igor Pelgreffi
Università degli studi di Verona
✉ igor.pelgreffi@univr.it

Ubaldo Fadini
Università degli studi di Firenze
✉ ubaldo.fadini@unifi.it

Massimiliano Badino
Università degli studi di Verona
✉ massimiliano.badino@univr.it

¶ Sezione Prima

Genealogia del concetto di algoritmo

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7768560

Contributi/1

*Alan Turing and the Cognitive Foundation of the Concept of Algorithm**

Simone Pinna 0000-0001-9049-5585

Marco Giunti 0000-0003-4182-4850

Articolo sottoposto a doppia blind peer review. Inviato il 08/12/2021. Accettato il 08/06/2022.

The work of Alan Turing (1936) set a milestone for the foundation of the concept of algorithm by grounding the notion of effective procedure on a special type of real cognitive phenomenon, namely, that of a human being performing rule-based symbolic transformations with the only aid of paper and pencil. In this work, after a brief historical overview, we show how Turing arrived at a negative solution of the decidability problem for first order logic and in which sense Turing's explication of the intuitive concept of effective procedure is sufficient to justify Church's Thesis. We then present a cognitive interpretation of Turing's theory of computation, according to which Turing machines are viewed as models of real phenomena of mind-environment interaction.

***

Introduction

During the first decades of the 20th century a vast group of mathematicians and logicians was involved in the precise definition of a series of foundational concepts for mathematics. The theoretical background of these studies was the so-called logicist project, i.e., the attempt to reduce all mathematics to logic.[1] Logicists notably recognized Cantor's set theory as a starting point for a rigorous redefinition of all mathematical theories in logical terms.

Bertrand Russell's discovery of a contradiction in Cantor's set theory (the Russell's antinomy) marked the failure of the original logicist project. However, the studies on the foundation of mathematics gained renewed impetus at the

* Supported by Fondazione Banco di Sardegna (FdS 2019, research grant n. F72F20000420007).
[1] G. Frege, *The Foundations of Arithmetic: A Logico-mathematical Enquiry into the Concept of Number*, Northwestern University Press 1968.

beginning of the 20th century thanks to David Hilbert's work. According to Hilbert's formalist project, any mathematical theory should be reduced to a pure symbolic system, such that one must be able, only through the means provided by the system, to prove the consistency of the theory. An outstanding problem in the formalist framework is the problem of decidability (*Entscheidungsproblem*), proposed in its most general version by Hilbert and Hackermann[2]: Is there any effective procedure (or any algorithm) for deciding whether an arbitrary first-order language statement is universally valid (i.e., a tautology, or a logical truth) or not?

An important side problem of the Entscheidungsproblem is the rigorous definition of the intuitive notion of effective calculability, or to find an adequate method to set apart computable numerical functions from non-computable ones, where computable numerical functions are those for which there exists an algorithm (or an effective procedure) that, for any argument, returns the value of the function in a finite number of steps. (For example, the function that returns the *n*-th decimal digit of π is computable, but for most real numbers the analogous function is not.)

In 1936,[3] Alonzo Church proposed the λ-calculus as a way to express in a rigorous formal manner the intuitive concept of effectively calculable function. He proved that the class of functions defined within his calculus is in fact equivalent to the class of recursive functions defined by Gödel.[4] He then proposed the following thesis (Church's Thesis): «We now define the notion […] of an effectively calculable function of positive integers by identifying it with the notion of a recursive function of positive integers (or of a λ-definable function of positive integers)».

According to Church, the fact that either recursiveness or λ-definability captures the notion of effective calculability rests on intuitive considerations. However, it is very difficult to find any similarity between a logical system such as the λ-calculus and the ways human beings execute algorithms.

The work of Alan Turing (1936)[5] set a milestone on this debate by grounding the intuitive notion of algorithm on a special type of real cognitive phenomenon, namely, that of a human being performing rule-based symbolic transformations with the only aid of paper and pencil. In this work, after a brief historical overview, we show how Turing arrived at a negative solution of the Entscheidungsproblem, and in which sense Turing's explication of the intuitive concept of algorithm is sufficient to justify Church's Thesis. We then propose a cognitive interpretation of Turing's theory of computation, according to which

[2] D. Hilbert, W. Ackermann , *Grundzüge der theoretischen logik*, Springer Verlag, Berlin 1928.
[3] A. Church, *An unsolvable problem of elementary number theory*, «Journal of Mathematics», 58, 1936, pp. 345-363.
[4] K. Gödel, *On Undecidable Propositions of Formal Mathematical Systems*, 1934, in Davis, M. (ed.), The Undecidable, Raven Press, Hewlett, New York, 1965, pp. 39-74.
[5] A. M. Turing, *On computable numbers, with an application to the* Entscheidungsproblem, «Proceedings of the London Mathematical Society» (pp. 230–265), London 1936, Oxford Journals.

Turing machines are viewed as models of real phenomena of mind-environment interaction.[6]

1. In Search of a Formal Notion of Algorithm

The most general formulation of the problem of decidability is the following: Is there any *effective procedure* to decide, for an arbitrary first-order language statement, whether it is *universally valid* (i.e., a tautology), or not? According to Hilbert, this was the most important problem of mathematical logic. Not only was first order logic at issue, but also any axiomatic theory based on a finite number of axioms. For this kind of theories, indeed, the solution of the problem of decidability would immediately yield an effective procedure to determine whether any statement of the theory is a logical consequence of the axioms or not.

In order to tackle the *Entscheidungsproblem*, however, a preliminary problem to be solved was the formulation of a satisfying formal definition of the intuitive concept of an *effective procedure*, or an algorithm. Intuitively, an algorithm (or a mechanical or effective procedure) is a finite set of clear-cut formal instructions for symbol manipulation. Given a finite collection of initial data, a human being must be able to carry out such instructions in a definite sequence of steps, with the exclusive aid of paper and pencil (or similar external supports or devices), and without resorting to any special insight or ingenuity. Note that, if a formal definition of algorithm were available, it would then be possible to formally define the computable functions, namely, those functions for which effective procedures to calculate their values do exist.

The solution to this preliminary problem is also clearly connected to the problem of decidability, for the latter presupposes a clear definition of the notion of an effective procedure. Moreover, once we have such a formal definition at hand, we will be able to try out the following strategy: if an effective procedure to decide the universal validity of any first order language statement *does not exist*, then we may be able to actually produce a first order language statement for which *we can prove* that there is no algorithm to decide whether it is universally valid or not (i.e., whether it is a tautology or not).

1.1 Church's Thesis

In 1936 Alonzo Church proposed λ-calculus as a formal method to rigorously express the intuitive notion of an effectively calculable function.[7] First, he proved that the class of λ-definable functions (namely, the functions

[6] M. Giunti, S. Pinna, *Toward a dynamical theory of human computation*, «Logic Journal of the IGPL», 24 (4), 2016, pp. 557-569; Wells, A., *Rethinking cognitive computation: Turing and the science of the mind* (hereafter, Wells 2005), Basingstoke, UK (2005), Palgrave Macmillan.
[7] A. Church, *An unsolvable problem of elementary number theory*, cit.

definable in his λ-calculus) is equivalent to the class of *recursive* functions, previously defined by Kurt Gödel). Then, he proposed the following thesis (known thenceforth as *Church's Thesis*):

> We now define the notion, already discussed, of an *effectively calculable* function of positive integers by identifying it with the notion of a recursive function of positive integers (or of a λ-definable function of positive integers).
>
> This definition is thought to be justified by the considerations which follow, so far as positive justification can ever be obtained for the selection of a formal definition to correspond to an intuitive notion.[8]

Church argued that an algorithm for computing a function exists if and only if that function can be expressed as a string of symbols written in his logical language or, equivalently, as a recursive function. However, he recognized that he could not prove this claim, for there is no secure way (i.e., a formal proof) to justify the identification of a formal definition with an intuitive notion. In other words, this justification can only be obtained by an intuitive stance.

Here, we cannot give an even cursory presentation of Church's λ-calculus. However, some brief remarks on the equivalent theory of the recursive functions may be useful to better understand the difference between Church's and Gödel's approach on the one hand, and Turing's one on the other. Similar to Church's λ-definability, recursivity theory provides us with a formal system to define a class of computable functions. First of all, we start from the following three basic *primitive recursive* functions:

[1] $z(n) = 0$ (*zero* function);
[2] $id(n) = n$ (*identity* function);
[3] $s(n) = n+1$ (*successor* function).

Then, all other primitive recursive functions are defined by the following rules:

Composition: a function is primitive recursive if it can be defined by the application of a primitive recursive function to others primitive recursive functions.

Recursion: a function $f(x_1, \ldots, x_n)$ of $n \geq 1$ arguments is primitive recursive if the following two conditions are met:

[i] $f(0, x_2, \ldots, x_n) = g(x_2, \ldots, x_n)$, where, if f has at least two arguments, g is a primitive recursive function and some of its arguments may be missing. If f is a single argument function, g is a zero argument function, namely, it is a natural number;

[8] *Ibid.*, p. 356.

[ii] $f(s(x_1), x_2, \ldots, x_n) = h(f(x_1, x_2, \ldots, x_n), x_1, x_2, \ldots, x_n)$, where h is a primitive recursive function and some of its arguments, except the first, may be missing.

Closure: nothing else is a primitive recursive function.

In a recursive definition the value of a function is the result of the application of the same function to smaller arguments, in a process named *recursion*. Consider the definition of the two arguments function *addition*, denoted by $a\ (x_1, x_2)$, as a primitive recursive one:

[4] $a(0, x_2) = id(x_2)$
[5] $a(s(x_1), x_2) = s(a(x_1, x_2))$

The function $a(x_1, x_2)$ is primitive recursive, for it meets both conditions of the rule of recursion. Indeed, condition [i] is met by equation [4], for *id* is primitive recursive; moreover, condition [ii] is met by equation [5] for, in the second member of the equation, s is a primitive recursive function with a single argument.

We now illustrate by an example how the recursive definition of addition is used in calculation. Let us take $x_1 = 2$ and $x_2 = 2$. By equation [3] we know that $s(2) = 3$, then, with this setting (by equation [5]), we are going to calculate the sum 3+2.

We start the computation by applying equation [5] to the chosen values of x_1 and x_2, then we decrease x_1 by 1 at each step, until $x_1 = 0$, as in the following series of equations:

$a(3, 2) = s(a(2, 2))$
$a(2, 2) = s(a(1, 2))$
$a(1, 2) = s(a(0, 2))$

Here, by [4] and [2], we arrive at the only result explicitly defined by the function, namely, the case where $x_1 = 0$. Knowing that $a(0, x_2) = x_2$, we can now substitute the result thus obtained for the case $x_1 = 0$; then, we increase x_1 by 1 at each step, until $x_1 = 2$, as in the following series of equations:

$a(1, 2) = s(2) = 3$
$a(2, 2) = s(3) = 4$
$a(3, 2) = s(4) = 5$

Here we have focused only on primitive recursive functions, but to cover all recursive functions we would have to present the more complex theory of *general* recursive functions, which is beyond the scope of the present paper. However, the previous explanation is sufficient for our purpose, that is, to provide the reader with a clear understanding of the peculiar kind of procedure by which a recursive function is computed.

We have seen above that a justification of Church's Thesis can be obtained only from an intuitive point of view. However, the fact that, *for any function computable by an effective procedure*, this function must also be definable in terms of recursiveness (or, equivalently, of λ-definability) is not intuitively implied by the notion of a recursive (or λ-definable) function. Indeed, the formal methods individuated by Church and Gödel are highly abstract and profoundly different from the calculation procedures we normally use (for example, the algorithms used for carrying out simple arithmetical operations with paper and pencil). Due to these reasons, Church's Thesis was initially deemed as *materially inadequate* and, ultimately, *not adequately justified*, because it lacked a sufficiently strong argument to ensure that all the numerical functions computable by a human being through the application of appropriate rules were captured by the class of recursive or λ-definable functions.

1.2 Turing's analysis of human computation

Turing's work[9] tackled the problem of explicating the notion of effective calculability by a totally different stance. His major concern was the definition of a theory of computation rigorously modeled on concrete phenomena of human computation, namely, on the actual procedures carried out by a human being that is calculating with the aid of paper and pencil. The analysis of this kind of phenomena needs a previous consideration of human cognitive abilities and limits.

Turing argues that his computing machines (known thereafter as *Turing machines*) capture the intuitive notion of an algorithm. This claim is justified by an *analogical argument*, namely, an inductive argument which correlates two or more phenomena in order to substantiate the hypothesis that they share some characteristic property. For example, (1) consider an object *a* whose characteristic property *p* we are interested in is known, then (2) notice that some essential features of *a* are similar to those of another object *b*; given these similarities we can claim that (3) *a* and *b* likely share *p*.

Turing's argument starts from the consideration of a *type* of real cognitive phenomenon, namely, the fundamental operations that we can observe in a human being carrying out a computation with paper and pencil. Turing explicitly refers to this kind of phenomena, as in the following quotation:

> Computing is normally done by writing certain symbols on paper. We may suppose this paper is divided into squares like a child's arithmetic book. In elementary arithmetic the two-dimensional character of the paper is sometimes used. But such a use is always avoidable, and I think that it will be agreed that the two-dimensional character of paper is no essential of computation. I assume then that the computation is carried out on one-dimensional paper, i.e., on a tape divided into squares. I shall also suppose that the number of symbols which may be printed is finite. If we were to allow

[9] A. M. Turing, *On computable numbers*, cit.

an infinity of symbols, then there would be symbols differing to an arbitrarily small extent.[10]

In this passage, Turing focuses on the individuation of the fundamental elements of human computation, by eliminating all those features that are *inessential* to the calculations performed by a human being. This is to be intended in the sense that they *complicate the algorithm* without yielding any gain with respect to the *results* that can be computed. The two-dimensional paper sheet is thus considered inessential in this sense, for it is always replaceable by a unidimensional tape divided into squares. By contrast, the assumption that the number of symbols that a 'computer' (i.e., according to Turing's terminology, a *human being* who computes) is able to recognize at a glance must be finite is essential, for it directly depends on human cognitive limits. Another inessential feature is the compositional character of the mental states used while computing.

> We may suppose that there is a bound B to the number of symbols or squares which the computer can observe at one moment. If he wishes to observe more, he must use successive observations. We will also suppose that the number of states of mind which need be taken into account is finite. […] Again, the restriction is not one which seriously affects computation, since the use of more complicated states of mind can be avoided by writing more symbols on the tape.[11]

It is hard to avoid the impression that Turing is considering the physical limitations we need to impose on a computing *machine* rather than the cognitive limitations of a computing *human being*. This is partly due to our present usage of the word 'computer', as a synonym of 'digital computer'. However, the concept of a digital computer as the physical machine we are used to was unknown in the mid-1930s. A thorough reading of Turing's argumentation should convince even the most skeptical reader. Consider, for example, the following passage on the identifiability of symbols:

> In most mathematical papers the equations and theorems are numbered. Normally the numbers do not go beyond (say) 1000. It is, therefore, possible to recognize a theorem at a glance by its number. But if the paper was very long, we might reach Theorem 157767733443477; then, further on in the paper, we might find "... hence (applying Theorem 157767733443477) we have […]". In order to make sure which was the relevant theorem we should have to compare the two numbers figure by figure, possibly ticking the figures off in pencil to make sure of their not being counted twice.[12]

It is obvious that this argument makes sense only if we consider human cognitive limits rather than general physical limits, we can impose on a computing machine. After the identification of the essential elements of computation,

[10] *Ibid.,* pp. 249.
[11] *Ibid.,* p. 250.
[12] *Ibid.,* p. 251.

Turing analyzes the behavior of a computer that conforms to all the imposed limitations.

> The behaviour of the computer at any moment is determined by the symbols which he is observing, and his 'state of mind' at that moment. [...]
>
> Let us imagine the operations performed by the computer to be split up into 'simple operations' which are so elementary that it is not easy to imagine them further divided. Every such operation consists of some change of the physical system consisting of the computer and his tape. We know the state of the system if we know the sequence of symbols on the tape, which of these are observed by the computer (possibly with a special order), and the state of mind of the computer. We may suppose that in a simple operation not more than one symbol is altered. Any other changes can be split up into simple changes of this kind.[13]

Now Turing describes how to build into a computational model all the elements needed to describe the essential work of a human computer, namely, the simple operations drawn from the analysis of the real cognitive phenomenon he considers.

> The simple operations must therefore include:
>
> (a) Changes of the symbol on one of the observed squares.
>
> (b) Changes of one of the squares observed to another square within L squares of one of the previously observed squares.
>
> It may be that some of these changes necessarily involve a change of state of mind. The most general single operation must therefore be taken to be one of the following:
>
> (A) A possible change (a) of symbol together with a possible change of state of mind.
>
> (B) A possible change (b) of observed squares, together with a possible change of state of mind.
>
> The operation actually performed is determined [...] by the state of mind of the computer and the observed symbols. In particular, they determine the state of mind of the computer after the operation is carried out.
>
> We may now construct a machine to do the work of this computer.[14]

All the functional features of Turing's computational models listed in this quotation are drawn from his analysis of human computation. The intuitive force of his definition of an effective procedure, in fact, arises from the analogy between the description of a class of real phenomena and the operations carried out by his ideal models. The (negative) solution to the problem of decidability

[13] *Ibid.,* p. 250.
[14] *Ibid.,* p. 251.

he proposes (see §1.3) is based on the same analogy. Indeed, by this analogical argument it is possible to justify the general thesis according to which Turing machines can execute all and only the essential computational procedures that a human being can, in principle, carry out in a finite amount of time with the only aid of paper and pencil or similar external resources, and without resorting to any special insight or ingenuity. In a nutshell, *Turing's Thesis* can be formulated as follows: *Turing machines are adequate ideal models of human computation.*

It is worth noting that the class of computable functions individuated by Turing is identical to that individuated by Church. Turing himself[15] proved that the class of all numerical functions computable by his machines (henceforth, *Turing-computable* functions) is identical to the class of λ-definable ones. Turing's argument, then, also supplies a strong intuitive justification to Church's Thesis, overcoming the theoretical impasse described above (§1.1).

2. Turing's Negative Solution of the *Entscheidungsproblem*

Turing's strategy for the negative solution of the decidability problem consists of two main moves. First, he identifies the intuitive notion of an effective procedure with the formal operations carried out by his computing machines (Turing's Thesis), and intuitively justifies this identification through an analogical argument (§1.2). Second, he employs this identification to justify Church's Thesis, which can be stated as the claim that the class of all computable functions (in the intuitive sense) is a subclass of the class of all Turing-computable functions. This finally allows him to prove by *reductio* that there is *no algorithm* that decides whether an arbitrary first order sentence is a logical truth (or a tautology). Turing remarks[16] that his result is sensibly different from the incompleteness result obtained by Gödel.[17] While the latter proved that any consistent minimally expressive axiomatic theory[18] is incomplete – for there is always a particular sentence of the theory that cannot be proved or disproved within the theory itself – Turing shows that there is no *algorithm* capable to decide, for any sentence of a first order language, whether it is universally valid.

2.1 Essential elements of Turing machines

To fully understand Turing's proof, it is necessary to briefly introduce the notion of a Turing machine (*TM*). A *TM* is made up of the following components:

[15] *Ibid., Appendix*.
[16] *Ibid.,* p. 259.
[17] K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, «Monatshefte für Mathematik und Physik», 38 (1931), 173-198.
[18] For a 'minimally expressive axiomatic theory' we mean a theory that is able to express and prove at least what can be proved in elementary arithmetic (see M. L. Dalla Chiara, *La logica*, Milano 1974, §1.5).

1. A finite automaton consisting of

- a simple input-output device that implements a specific set of instructions (machine table);

- an internal memory that holds only one discrete element at each step (internal state) and

- an internal device (read/write head) that can scan and change the content of the internal memory.

2. An external memory consisting of a tape divided into squares, potentially extendible in both directions *ad libitum*;

3. an external device (read/write/move head) that scans the content of a cell at a time and allows the finite automaton to work on the memory tape.

At each step of a computation, the external device scans one cell of the tape and changes its content in accordance with the symbol written on the cell and its present internal state (as scanned by the read/write head in the internal memory). The rules for symbol transformations are specified in the machine table, which is a non-empty set of *quintuples* of the form $<q_i, s_j : s_k, M, q_r>$. No two quintuples of the machine table can begin with the same pair of symbols $<q_i, s_j>$. The pair $<q_i, s_j>$ in the first part of a quintuple represents the input, where the first symbol is taken from a finite set $Q = \{ q_1, q_2, \ldots, q_n \}$ of internal states and the second from another finite set $S = \{ s_1, s_2, \ldots, s_m, b \}$ of symbols, which represents the vocabulary of the tape and includes at least two symbols, one of which, i.e., b (the *blank*), means that the content of the cell is null. The number of tape cells is always finite, but it can be increased as needed by adding blank cells. The triple $< s_k, M, q_r >$ in the second part of a quintuple represents the output, where the first and the third symbol are taken, respectively, from the above mentioned sets $\{ S \}$ and $\{ Q \}$, while M represents one of the possible movements of the external head on an adjacent cell of the tape with respect to the presently scanned one (L for 'move left' and for 'move right') or a null movement (H for 'halt'). Each single quintuple of a machine table, then, can be interpreted as a conditional instruction that tells the machine: 'if you read q_i in your internal memory and s_j on the tape, then substitute s_k for s_j on the tape, q_r for q_i in the internal memory, and move the external head according to M'. The machine stops computing if and only if the current internal state q_i and the scanned symbol s_j do not correspond to the initial pair of any quintuple in the machine table, or the corresponding quintuple has the form $< q_i, s_j : s_j, H, q_i >$.

Turing introduces some important conventions in his notation. He distinguishes between two types of symbols that form the vocabulary of the tape: 'symbols of the first type', or *figures*, are only 0 and 1 – they are used to encode the binary representation of an arbitrary real number. All the others are 'symbols of the second type'. Let an arbitrary *TM* start computing from a given initial state on a blank tape. The sequence of all figures that the machine writes on the tape while it is computing is called the *sequence computed by the machine*. Any

such sequence can be interpreted as the binary representation of a real number,[19] which can thus be said to be computed by the machine as well. A *TM* is called *circular* if, when started on a blank tape, it eventually stops computing or it writes at most a finite number of figures. Otherwise, it is said to be *circle free*. An infinite sequence of figures (or the corresponding real number) is *computable* if there is a circle free *TM* such that, when started on a completely blank tape, produces that sequence.

2.2 Turing's proof

Consider that the set of all *TMs* is enumerable (namely, it can be put in one-one correspondence with the set of natural numbers). In the first place, any *TM* can be identified with its machine table. A machine table, indeed, consists of a set of quintuples drawn from a finite vocabulary, and the set of all symbol strings which can be built on this vocabulary is enumerable. In the second place, any machine table may be seen as a single symbol string made up by concatenating its quintuples. Hence, the set of all machine tables is enumerable as well, because it is a decidable subset of the set of all symbol strings.

Once we have the list of all possible *TMs* at hand, we can assign each of them a *description number*, namely a number, obtained through a process named *Gödelization*,[20] that completely represents a specific machine table.

Now we have all the notions needed to understand Turing's negative solution of the decision problem. Assuming Turing's Thesis, as specified at the end of §1.2, his negative solution of the decision problem derives from the preliminary solution of another more specific problem, known as the *satisfactoriness problem*, which is a type of halting problem: Is it possible to find a *TM* that decides whether an arbitrary *TM* is circle free?

Once he has demonstrated that the satisfactoriness problem is unsolvable, Turing considers a second problem, called *printing problem:*[21] Is it possible to find a *TM* that decides, in a finite number of steps, whether an arbitrary *TM* will eventually print a 0 on the tape?

Turing then proves that, if the printing problem is solvable, the satisfactoriness problem is solvable as well. But we already know that the latter is unsolvable. Then we conclude (for *modus tollens*), that the printing problem is unsolvable as well.

The last step towards the solution of the decision problem concerns the translation of any *TM* in formulae of a first-order language. This way, it is possible

[19] More precisely, Turing takes the sequence computed by a machine to represent *any* real number whose decimal part, expressed in binary, is equal to that sequence. To avoid confusion, then, he focuses in his 1936 paper on computable sequences, rather than computable numbers (A. M. Turing, *On computable numbers*, cit., p. 233).
[20] See G. S. Boolos, J. P. Burgess, R. C. Jeffrey, Computability and logic, Cambridge University Press, Cambridge, England (2002), § 4.1.
[21] The following explanation is drawn from A. Wells, *Rethinking cognitive computation: Turing and the science of the mind*, Palgrave McMillan, Basingstoke, 2005, Ch. 15.

to interpret the computational steps of a *TM* as subsequent steps in a first order logic proof. Turing is then able to produce a first order formula, *Un*(*M*), which expresses the following proposition: 'In some complete configuration of *M*, 0 is written on the tape'. It is thus possible to prove that the following implications are true:

(a) if, in some complete configuration of *M*, 0 is written on the tape, then *Un*(*M*) is provable in first order logic;
(b) if *Un*(*M*) is provable in first order logic, then, in some complete configuration of *M*, 0 is written on the tape.

The conjunction of a and b tells us that *Un*(*M*) is provable in first order logic if and only if *M* will write 0 on the tape; but we already know that the printing problem is unsolvable. The conclusion is that no algorithm can determine whether a specific first-order statement, namely *Un*(*M*), is provable in first order logic. But this entails that no general algorithm exists that can decide, for any first order statement, whether it is universally valid or not. Therefore, the decision problem is unsolvable.

It is important to note that, even though Turing's negative solution of the *Entscheidungsproblem* is the result of a formal proof, his argument ultimately relies on the intuitive correspondence between the computational processes of a *TM* and his analysis of human computation, that is to say, on the intuitive justification of Turing's Thesis.

3. The Cognitive Relevance of Turing's Work

We have seen in the last section a rigorous formal proof which is ultimately based on the intuitive correctness of the underlying analysis of human cognitive powers and limits. The fact that a mathematical proof could be based on intuitions, however, should not surprise the reader. The link between mathematical proofs and cognitive performances has been highlighted, *e.g.*, by Giuseppe Longo,[22] who notes that the notion of *order*, on which a large part of mathematical reasoning relies, is founded on cognitive notions such as the *mental number line*, through which we can internally 'observe' the geometrical relations between numbers.

Turing explicitly recognizes the importance of a serious cognitive analysis for the faithful formal characterization of an intuitive notion. However, this point represents only an indirect contribution to the study of cognition. We propose that Turing's work may be thought of as an attempt to characterize a specific class of cognitive phenomena, namely, the *phenomena of human computation*. In this section we present, first, Wells' interpretation of Turing's theory of computation, which considers Turing machines as models of mind-

[22] G. Longo, *Reflections on Concrete Incompleteness*, «Philosophia Mathematica», 19, 3, 2011, pp. 255-280 (p. 256).

environment interaction.[23] Second, we propose a methodological interpretation of Turing's Thesis that suggests a precise line of research in cognitive science, according to which *algorithmic skills* are seen as a set of cognitive capacities that deserve to be studied on their own.

3.1 TMs as models of mind-environment interaction

A long established tradition in cognitive science considers Turing as a precursor of *classic computationalism*, according to which cognition is seen as the result of *purely internal* rule-based symbolic transformations. The major reasons for ascribing Turing's thought to this theoretical framework are fundamentally two: 1) classic computationalism considers the *digital computer* as a model of the mind, and the ideal prototype of a digital computer is the *universal* Turing Machine;[24] 2) Turing implicitly advocated the *computational theory of mind*,[25] a basic tenet of classic computationalism, in his famous 1950 article, published in the philosophical journal *Mind*, where he proposed the *imitation game* as a test for machine intelligence.[26] There, he pointed at digital computers as good candidates for playing this game, and was confident that machines of this kind would eventually pass the test, thus showing intelligence and thought. But, as Turing claimed, a digital machine would pass the test in force of the right kind of algorithms it would execute. Thus, intelligent thought processes would be nothing over and above algorithmic activities.

Nevertheless, the ascription of Turing's thought to classic computationalism is marked by a fundamental misunderstanding, which consists in considering the external tape of a *TM* as an *internal* memory.[27] This misinterpretation leads to a view of cognition in which all cognitive activities are made up of rule based transformations of *internal symbols*, which encode either the mental or the environmental features relevant to a given cognitive task. But we have seen, on

[23] *Op. cit.*; A. Wells, *Turing's analysis of computation and theories of cognitive architecture*, «Cognitive Science», 22, 1998, 269–294.
[24] A. M. Turing, *On computable numbers*, §6. A universal *TM* is a *TM* that can emulate the work of any other *TM*. To this purpose, the tape of the universal *TM* is split into two parts by a symbol that is never canceled or overwritten during the computation. On one part of the tape (say the right one), the machine table of the *object TM* (i.e., the *TM* to be emulated) is encoded, while on the other part the universal *TM* keeps a representation of the other components of the object *TM* (the internal and the external memory with the respective contents, and the position of the external head). The analogy with a digital computer is straightforward: the machine table of the universal *TM* corresponds to the *hardware* of a digital computer (specifically, to its CPU), while the quintuples of the object *TM,* encoded on the right part of the tape, correspond to the *software*, namely, to the program in use.
[25] The computational al theory of mind claims that all thought processes have an algorithmic nature.
[26] A. Turing, *Computing machinery and intelligence,* «Mind», 59(236), 1950, 433-460.
[27] See A. Wells, *Rethinking cognitive computation: Turing and the science of the mind*, Basingstoke 2005, for a theoretical analysis of this misunderstanding and its consequences for the *status* of Turing's theory of computation with respect to the philosophy of mind and the philosophy of cognitive science.

the contrary (§1.2), that the tape of a *TM* is an abstraction of 'a child's arithmetic book', so that it is to be considered as a part of the *external environment* used to carry out the cognitive task. By a cognitive point of view, then, a computation is not just a sequence of transformations of symbols of the internal vocabulary of the machine, but it consists in operations on *external symbols*, namely, on the symbols of the vocabulary *of the tape*. These symbols represent environmental features that need not be internalized in order to carry out the cognitive task, because in most cases their internalization would only result in wasting cognitive resources.

If we consider the *mind as a computer* analogy from this perspective, and take a universal *TM*, instead of a digital computer, as a model of cognition, the resulting view turns out to be very different from classic computationalism. For, now, not only symbols (as shown above), but also *programs* need not be internalized in order to be executed. This is tantamount to saying that a cognitive system may not keep the rules of behavior it needs to survive in an internalized library of programs. On the contrary, the external world itself may supply the cognitive system with the symbolic structures needed to adapt its behavior to a continuously changing environment. From this perspective, then, Turing's cognitive view can no longer be seen as a precursor of classic computationalism, but it is more akin to recent approaches to cognitive science such as embodied/extended cognition and the dynamical approach to cognition.

3.2 The methodological interpretation of Turing's Thesis

Wells' interpretation of Turing's cognitive view, on which the previous section is based, expresses an original position in the philosophy of mind (named *ecological functionalism*) which is relevant to the debates on classic *vs* new cognitive science[28] and on the *status* of the computational theory of mind.[29] However, it is not clear how to use this interpretation of a *TM* as a general model of cognition to actually study real cognitive phenomena. To this purpose, we need a step back to the motivations underlying Turing's choices for the design of his computational models.

As we have seen, Turing's proof (§2.2) is based on an analogical argument, namely, it presupposes that the simple operations carried out by a *TM* are an idealization of those performed by a human being that executes an algorithm with the exclusive aid of paper and pencil. We then proposed the following formulation of Turing's Thesis: *Turing machines are adequate ideal models of human computation.* But how could we define the *type* of cognitive phenomenon of which *TMs* are ideal models? Let us call it *phenomenon of human computation*. By this expression we mean any activity of a human being that consists in

[28] A. Newen, L. De Bruin, S. Gallagher, (Eds.) *The Oxford handbook of 4E cognition*, Oxford University Press, New York (NY) 2018.
[29] R. A. Wilson, *Wide computationalism*, «Mind», 103(411), 1994, 351-372; G. Piccinini, *Computationalism in the philosophy of mind*, «Philosophy Compass», 4(3), 2009, 515-532.

the deliberate and controlled execution of a given algorithm, i.e., a purely mechanical or effective procedure. Any specific instance of a phenomenon of human computation is thus uniquely individuated by its particular task, which consists in the request of executing a specified algorithm. An algorithm (or a mechanical or effective procedure) is a finite set of clear-cut formal instructions for symbol manipulation. Given a finite collection of initial data, a human being must be able to carry out such instructions in a definite sequence of steps, with the exclusive aid of paper and pencil (or similar external supports or devices), and without resorting to any special insight or ingenuity.

We have also seen that a *TM* is a *highly idealized* model of a human computer that, given these idealization, is not *per se* fit to model the *whole* variety of phenomena of human computation. In fact, standard *TMs* are able to model only a small class of this kind of cognitive phenomena, namely, those that can be carried out on a unidimensional tape. This means that all the algorithms that need a bidimensional paper (like the standard column algorithms for solving the four basic arithmetical operations) could not be modeled by *TMs*. And we would even need one more dimension to solve, for example, a Rubick's cube, or any other cognitive task involving the execution of some algorithm on a three-dimensional support. If we want to design a model of the full set of phenomena of human computation, we then need to somehow soften these idealizations by retaining the basic *cognitive* insight of Turing's analysis.

To sum up, we can thus formulate the following *methodological interpretation* of Turing's Thesis: *The kind of model adequate to describe and explain all phenomena of human computation is to be searched as an appropriate generalization of Turing's machines, which preserve their structure and basic design.*

3.3 Dynamical models of human computation

In §3.1 we have seen an interpretation of *TMs* as models of mind-environment interaction, where the tape of a *TM* is seen as the external environment on which symbol transformation is performed. Taking a cue from this view, we propose an interpretation of *TMs* as idealized *dynamical models* of human computation. Turing himself, indeed, individuated the three main components of these dynamical models:[30] the internal state, the position of the external head and the content of the tape. In order to fulfill the methodological interpretation of Turing's Thesis, then, we propose the definition of an *Enhanced Turing Machine* (*ETM*) as a dynamical model of the full variety of the phenomena of human computation.

An *ETM* is a dynamical model which satisfies the following three requirements:

[30] «We know the state of the system if we know the sequence of symbols on the tape, which of these are observed by the computer (possibly with a special order), and the state of mind of the computer». (A. M. Turing, *On computable numbers*, p. 250).

1. Its instantaneous state has the three components individuated by Turing: *q* = internal state, *p* = head position, *c* =content of the external support;

2. its basic actions should be the three ones also individuated by Turing:[31] (i) the head reads the whole content of a finite neighborhood (*read/write neighborhood*) of its present location, but then changes the symbol of just one cell of this neighborhood; (ii) the head moves from its present location to a new one within a fixed finite neighborhood (*move neighborhood*) of the present location; (iii) the machine changes its internal state;

3. finally, the global transition function of an *ETM* should satisfy the constraint that Turing pointed out:[32] At each time step, the next instantaneous state of the machine should be completely determined by the present internal state and by the present content of the read/write neighborhood (*scanned content*), by means of the three basic actions. More precisely, this means that the present internal state and the scanned content should determine (i.a) the cell of the read/write neighborhood whose present symbol is to be changed and (i.b) the new symbol to be substituted for the present one in such a cell; (ii) the cell of the move neighborhood where the head is going to relocate; (iii) the new internal state.

These three requirements represent, in our view, the distillation of the cognitive insight on which Turing's theory of computation is based. A model satisfying these requirements is thus a good candidate for an adequate description and explanation of the phenomena of human computation in all their variety. We proposed elsewhere[33] a dynamical theory of human computation based on a type of dynamical model (named *Algorithmically enhanced Turing Machine – ATM*) that satisfies the above requirements, but we cannot give here the details of these models. In the present paper, our aim has been to highlight the cognitive foundation of Turing's theory of computation and how it is relevant not only to the history of the concept of algorithm, but also to the philosophy of cognitive science and to the scientific study of a *specific type* of cognitive phenomena.

Simone Pinna
Università degli studi di Cagliari
✉ simonepinna@hotmail.it

Marco Giunti
Università degli studi di Cagliari
✉ giunti@unica.it

[31] See footnote 13 of the present paper.
[32] «The operation actually performed is determined [...] by the state of mind of the computer and the observed symbols. In particular, they determine the state of mind of the computer after the operation is carried out» (A. M. Turing, *On computable numbers*, cit., p. 251).
[33] M. Giunti, S. Pinna, *Toward a dynamical theory of human computation*, «Logic Journal of IGPL», 24 (4), 557-569.

References

Primary Sources

Church, A. 1936. *An unsolvable problem of elementary number theory*, «Journal of Mathematics», 58, pp. 345-363.

Giunti, M., Pinna, S. 2016. *Toward a dynamical theory of human computation*, «Logic Journal of IGPL», 24 (4), 557-569.

Gödel, K. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, «Monatshefte für Mathematik und Physik», 38, pp. 173-198.

Gödel, K. 1934, *On Undecidable Propositions of Formal Mathematical Systems*, in Davis, M. (ed.) 1965. *The Undecidable*, Raven Press, Hewlett, New York, pp. 39-74.

Turing, A. M. 1936. *On computable numbers, with an application to the* Entscheidungsproblem, «Proceedings of the London Mathematical Society» (pp. 230–265), London, Oxford Journals.

Turing, A. M. 1950. *Computing machinery and intelligence,* «Mind», 236)59), pp. 433-460.

Wells, A. 1998. *Turing's analysis of computation and theories of cognitive architecture*, «Cognitive Science», 22, 269–294.

Wells, A. 2005. *Rethinking cognitive computation: Turing and the science of the mind*, Palgrave McMillan, Basingstoke.

Secondary Literature

Boolos, G. S., Burgess, J. P., Jeffrey, R. C. 2002. *Computability and logic*, Cambridge University Press, Cambridge, England.

Dalla Chiara, M. L. 1974. *La logica*, Isedi, Milano.

Longo, G. 2011. *Reflections on Concrete Incompleteness*, «Philosophia Mathematica», 19, 3, pp. 255–280.

Newen A., De Bruin L., Gallagher S. (Eds.) 2018. *The Oxford handbook of 4E cognition*, Oxford University Press, New York (NY).

Piccinini, G. 2009. *Computationalism in the philosophy of mind*, «Philosophy Compass», 4(3), pp. 515-532.

Wilson, R. A. 1994. *Wide computationalism*, «Mind», 103(411), pp. 351-372.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7768562

Contributi/2

Intentionalities of the Algorithm: Historical Practices and Socio-Computing Infrastructures

A Philosophical Account

Javier Toscano 0000-0001-8317-8311

Articolo sottoposto a doppia blind peer review. Inviato il 29/11/2021. Accettato il 08/06/2022.

Computing does not only imply a logical interaction with and through machines, but also – maybe more poignantly – a way of thinking. As historians of technology acknowledge, computing meant in the past so much as counting, or even reasoning. But in this sense, the history of computing has a much earlier beginning than what is popularly thought. The first machines that we can recognize as abstract computers were imagined by Charles Babbage in the 19th Century, but the first algorithms were assembled centuries before, to be performed by social machineries. Drawing on an expanded understanding of Tomasello's hypothesis on the nature of collective thinking, this article argues that collective intentionalities – which can be thought of as pre-conceptual or non-conscious states of collective organization – are inherent settings embedded in social algorithms, and that they play an important role in the socio-computing infrastructure in which these algorithms perform. The article then investigates this path by exploring historical samples of social algorithms and collective intentionalities, from Antiquity and the Middle Ages to contemporary socio-algorithmic practices. It then attempts to reconstruct the functional patterns and the complex relations they afford, while it looks at some of the cognitive articulations they conform. In the end, the article explains an algorithm as a collective social technology that emerges as a cultural script, where we find recurring process of signifying, inscribing and interpreting, and in which a given form of social thinking, involving a characteristic worldview, is always mobilized.

***

Introduction

An algorithm entails a piece of coding within a given language where a specific solution for an observed problem is proposed. Articulating a notion of code in the early days of computing history, pioneer logicians Newell, Simon and Shaw wrote that «the appropriate way to describe a piece of problem solving behavior is in terms of a program […]. Computers come into the picture only

because they can, by appropriate programming, be induced to execute the same sequences of information processes that humans execute when they are solving problems».[1] In this sense, even if we have become used to the idea of algorithms and codes for computers, the machine is contingent, and only the processes of defining, framing and solving a problem become relevant. This allows to shift our attention from a computer-based mechanism to the interrelated, cognitive and normative cultural practices sustained through socio-computing infrastructures, which have a history of their own.

Furthermore, if an algorithm (a code, a program) implies an instruction to achieve a particular end (a solution), this can only be implemented if the encoding of that instruction is readable, i.e. if it can be interpreted. As such, the interpretative demand inscribes the algorithm under a hermeneutic domain where two further premises follow: on the one hand, an interpretation implies that the implementation of an algorithm cannot be thought of as the repetition of an exact procedure, but rather entails an adaptation, an adjustment, transformation – even if of a very slightly different kind each time.[2] This is what, in a lateral but rather fruitful development of their own agenda, pioneer mathematicians Goldstine and von Neumann hinted at when they wrote in a groundbreaking paper that «coding is not a static process of translation, but rather the technique of providing a dynamic background to control the automatic evolution of a meaning».[3] In this sense, an algorithm opens itself beyond the idea of an inert repetition, towards creative developments where a social machinery flows in rhythms and resourceful iterations.

On the other hand, the necessary encoding/decoding process supposes already a common ground or shared context between the individual or group that proposes the procedure and those that receive and apply it. If we look deeply into this feature, we can see that the encoding/decoding transaction not only requires a linguistic understanding of the procedure itself, but a whole immersive comprehension of the particular task, i.e. a pragmatic know-how associated with a worldview, which stems out of and at the same time allows for social coordination. This is what the linguist and cognitive scientist Michael Tomasello

[1] A. Newell, J.-C. Shaw and H. Simon, *Elements of a Theory of Human Problem Solving*, «Psychological Review» LXV, 3, 1958, pp. 151-166.

[2] As we will see, this amounts to a re-centering of the critical tradition of hermeneutics (Heidegger, Gadamer) that wanted to distance itself from a positivist approach to computing. In the argument presented here, computing is instead thought of as a complex cognitive interaction with the social and material environment. And in this view, as the social cognitivist Shaun Gallagher writes, «cognition is an enactive and emotionally embedded engagement with the world by which we are able to solve problems, control behavior, understand, judge, explain, and generally do certain kinds of things – much of that constitutionally shaped by tools, environmental factors, social practices, etc.» S. Gallagher, *The socially extended mind*. «Cognitive Systems Research» 25-26, 2013, pp. 4-12, here p. 8. http://dx.doi.org/10.1016/j.cogsys.2013.03.008

[3] H. Goldstine, and J. von Neumann, *Planning and Coding of Problems for an Electronic Computing Instrument*, *Report on the Mathematical and Logical aspects of an Electronic Computing Instrument Part II*, Princeton 1947, vol I-3, p. 2.

has termed *collective intentionalities*. For Tomasello, collective intentionalities are necessary pre-conditions through which shared goals and social cognitive processes are inherently organized and anonymously systematized. He links them directly to the development of human thinking when he states: «important aspects of human thinking emanate not from culture and language per se but, rather, from some deeper and more primitive forms of uniquely human social engagement».[4] Tomasello argues that human thinking has evolved not only through the organization of cultural conventions, norms and institutions, but also, on an equal weight but seldom acknowledged, through joint goals and joint forms of attention – a common ground – which have created the possibility of individual roles and perspectives within «an ad hoc shared world or 'forms of life»[5].

Tomasello cunningly leans his hypothesis on the phenomenological notion of intentionality, which should be distinguished from that of intention. An intention implies a purpose, a causal antecedent of an action; an intentionality, by contrast, is the property of a mental state.[6] In a way, intentional states may be pre-conceptual and sometimes non-conscious, but do not need to be visceral or irreflexive. Intentionality is a way of being-in-the-world, a form of orienting oneself in one's environment, and not a mere psychological state. According to the philosopher Alfred Schutz, intentional descriptions classify over situations and actions, as typifications clear enough to imply pattern recognition, but with a certain inherent indeterminacy, since they invariably include an «open horizon of unexplored content».[7] Therefore, they become the framework for both conscious and unconscious intentions and actions. Intentionality occurs as a background mental activity, implying a collective pragmatic know-how, intuition, belief, affect and habituation, rather than a conscious state of the mind.

But how exactly can a collective intentionality ensue? The link is set through a socially-derived knowledge transmitted through customs, social habits, routines and particularly the syntax of everyday language. As Schutz put it: «This includes ways of life, methods of coming to terms with the environment, efficient recipes for the use of typical means for bringing about typical ends in typical situations».[8] Now, rather than presenting here the different epistemological models that discuss

[4] M. Tomasello, *A Natural History of Human Thinking*. Cambridge and London 2014, p. 2.
[5] *Ibid.,* p. 5.
[6] We could also remind of the discussion in Thomas Aquinas, who established that the object which is thought is intentionally in the thinking subject, the object which is loved is intentionally in the person who loves, the object which is desired in the person desiring, and so forth. In that sense, intentionality is clearly something that can be predicated of inexistent phenomena, but which has an effect on our own conceptions, desires and beliefs. The difference between intentionality and intention is further discussed in J. R. Searle, *The intentionality of intention in action*, «Cognitive Science» IV, 1980, pp. 47–70; or in M. Chapman, *Intention, intentionality, and the constructive character of scientific knowledge,* «Psychological Inquiry» I, 1990, pp. 251–253.
[7] A. Schutz, *Collected papers I: The problem of social reality*, The Hague 1962, p. 14.
[8] *Ídem.*

the issue[9], we can think of a working example. If a community engages in a given public ritual, individuals might participate with the intention of strengthening their bond with the collective, but for the community at large the ritual might be the occasion, encompassing a very clear intentionality, of performing a political act where roles and hierarchies are set or distributed. A collective intentionality is expected for social coordination or cooperation to occur, but this does not mean that every individual pertaining to that collective might share, or even be conscious of, the intentions behind the actions that this entail.[10] In any case, going back to Tomasello's hypothesis, collective intentionalities are pre-conditions of social thinking; in that sense, they can arguably be used to explain and articulate forms of social programming where a social algorithm performs a specific task. Therefore, an expansion of Tomasello's hypothesis is put here to the test. In the first section of this article, some historical practices will be presented as prospects of such collective intentionalities guiding characteristic social algorithms. The descriptions seek to explain likely assemblages, but also to shed light on typical traits of what can be understood as a social algorithm from different perspectives. In a second step, some of these practices will be further analyzed, outlining specific cases, in order to yield analytical insights, with the aim of understanding how intentionalities may function as part of a socio-computing infrastructure in which a social algorithm performs.

1. Historical sampling

Collective intentionalities are not always easy to locate for an external observer, as one usually lacks the necessary contextual insights that could make them entirely readable, which would in turn make a social procedure transparent and a cultural device wholly interpretable. Nonetheless, under Tomasello's expanded hypothesis, intentionalities are assumed to form an ensemble with social algorithms. So if we look for the type-form of an algorithm – an instruction or a sequenced procedure – we may come up with a set of joint articulations that may be further scrutinized.

[9] Different models on this issue are contrasted in S. R. Chant, F. Hindriks, and G. Preyer (eds.), *From Individual to Collective Intentionality*, Oxford 2014.

[10] As Julian Nida-Rûmelin writes: «The kind of collective intentionality that is constitutive of cooperation does not necessarily require a community or a group that is bound together by a specific ethos. Think of a society in which a strong individualism prevents the development of any sense of cultural, social, or ethnic community. If the members of this society are asked which community they belong to, they would answer 'none'. If communities are understood as being held together by a common ethos, that is, a common system of rules and norms and a specific feeling of belonging, we still can imagine a society in which there are no communities of this kind but in which cooperation among individuals nevertheless exists. The basis of this cooperation may be trust, interest, and some common normative beliefs regarding autonomy and fairness.» J. Nida-Rûmelin, *Structural Rationality and Collective Intentions*, S. R. Chant et al. (eds.) *From Individual…* cit., pp. 207-222.

Furthermore, intentionalities do not behave as stable categories nor conceptual tags, and the precise pre-figural mental state that they could eventually denote would most certainly be historically determined too. Given such difficulties, this section tries to group together typical cultural practices and devices as they are enacted in their algorithmic form, in order to organize them and pierce through their performativity. The resulting sets refer then to collective intentionalities as semantic clusters that involve more or less evident historical procedures, where a social algorithm or an algorithmic device is seen to function in a specific social dynamic. The samples were selected out of archeological or ethnographic accounts; these usually provide more detail and depth, but they are here synthetized so they can fit an argumentative discussion.

Moreover, due to space limits, the groups referred here are briefly sketched and reduced to five sets, even if their number could be easily enlarged on an expanded quest. Section two will then provide a closer analysis of some examples stemming from these divisions, to show how these can contribute to a deeper understanding of their intrinsic articulation.

1.1 To command / to control / to order

From ancient times, the most evident drive embedded in an algorithm is to achieve control over the world and over matter. Many of the subsequent intentionalities will be later subsumed to this original dynamic, which denotes the algorithm as an instrument of power. For instance, in a broad sense, the act of praying in any religion is a way of invoking a non-human force to affect the world. A prayer is a call for the order of the world to be regulated, used in favor of the speaker uttering the plea[11]. In the most fundamental prayer of Christianity, the *Lord's Prayer*, that is the function of the words *fiat voluntas tua*: through a speech act – reinforced through the embodied gesture of joining the palms together – the orders of heaven and earth are expected to coincide. This is the same principle ascribed in magic to an enchantment, a spell, or a ritual. The examples in which an invocation operates in this sense are multiple, but we cannot follow here further categorizations. In any case, we can explore another dimension, since it is possible to distinguish this aspect of command over a situation or a state of things in one type of algorithm that operates in a similar way, as a set of instructions that aims at ordering a course of action: the operation of the law.

[11] As Frazer writes, «From the earliest times man has been engaged in a search for general rules whereby to turn the order of natural phenomena to his own advantage.» Religion is one of those techniques in which man, driven by faith, seeks «a propitiation or conciliation of powers superior to man which are believed to direct and control the course of nature and of human life.» J. G. Frazer. *The Golden Bough. A Study in Magic and Religion*. London 1983, p. 65.

The rule of law enacts an order over a given territory (a *Raumordnung*, according to Carl Schmitt[12]), a jurisdiction. One of the earliest known sets of laws is the Code of Hammurabi, sixth king of Babylon. That code itself, composed between 1792 and 1750 BCE, was a compilation of different commandments, a unification of regulations from different cities of the Babylonian empire, but it was also a guiding device that generated at least three things: firstly, a sense of order over the territory it aspired to control; secondly, an instrument that symbolically established and reinforced sovereign power – in this case through evoking a direct connection between the monarchs and their god Marduk – and finally, it also sanctioned the vassalage of the subject, since subjects were all those individuals for whom the code applied, and for whom the stone with the inscribed code was placed in the center of public squares. Furthermore, this code was one of the first devices of civilization written entirely in conditional instructions *If… then…*[13], where punishments and fines were administered according to precise parameters.

1.2 To calculate / to compute / to reason

It is difficult to separate the boundaries in which an algorithm stops being an executable instruction that involves a command and becomes a way of calculating and generating knowledge. We doubtlessly enter here the terrain of a will to know that articulates knowledge and power in a straightforward way. In any case, it is worth exploring the operation and scope of some devices that can be collected under this modality.

In ancient Greece, Pythagorean thought was founded on the idea that the world is organized in numerical proportions; encoded, for instance, in music and mathematics. The golden section and the Pythagorean theorem – one of the foundations of Euclidean geometry and classical trigonometry – were elements of an aesthetic that tried to prop itself around an idea of beauty of mathematical proportions. The Pythagorean project consisted then of «the extraction and application of a numerical code that organizes both art and nature.»[14] This code allowed for the formation of correspondences between the micro and the macrocosm and aimed at describing the harmonic principle of the world. In this sense, Pythagorean thought found in the notion of harmony the possibility of defining and anticipating – that is, of computing – the exchanges between nature and humankind.

Another tradition where we find a source of authority based in calculation is the Kabbalah. Kabbalah is a mystical system that was developed during the

[12] C. Schmitt, *Der Nomos der Erde im Völkerrecht des Jus Publicum Europaeum*, Berlin 1950.
[13] The code was made up of 282 laws written under this format (although a number of them, those between 66-99 and 110-111, have been lost). An early commentary on the code can be read in R. F. Harper, *The Code of Hammurabi, King of Babylon about 2250 B.C.*, Chicago and London 1904.
[14] F. Cramer, *Words made Flesh. Code, Culture, Imagination*, Rotterdam 2005, p. 25.

European High Middle Ages and the Renaissance. One of the oldest books, the *Sepher Yetzirah* or *Book of Creation*, was published towards the end of the 9th century CE. Here, the letters of the Hebrew alphabet appear, just as Arabic numbers in other systems, as building blocks of the cosmos.[15] The language of God is said to have created the universe, and the language of mankind would be the only way to arrive at a reconstruction of its original intention. That is why human language itself is thought to be of divine inspiration. In an interesting analysis, Florian Cramer makes a translation of the idea of divine creation expressed in words to a logical system that shows its consistency and reaffirms the notion of the Kabbalistic code as an algorithm. Cramer quotes the fourth chapter of the *Sepher Yetzirah*:

> 1. There were formed seven double letters, Beth, Gimel, Daleth, Kaph, Pe, Resh, Tau, each has two voices, either aspirated or softened. […]
> 3. These seven double letters He formed, designed, created, and combined into the Stars of the Universe, the days of the week, the orifices of perception in man; and from them he made seven heavens, and seven planets, all from nothingness […].
> 4. From two letters or forms He composed two dwellings; from three, six; from four, twenty-four; from five, one hundred and twenty; from six, seven hundred and twenty; from seven, five thousand and forty; and from thence their numbers increase in a manner beyond counting; and are incomprehensible.[16]

Then Cramer draws attention to the mathematics involved in this permutation, such that 2! = 2 * 1 = 2; 3! = 3 * 2 * 1 = 6, and 7! = 7 * 6 * 5 * 4 * 3 * 2 * 1 = 5040. In the *Sepher Yetzirah*, the combination of letters grants divine creation a strictly formal and verifiable character; the alphabet transcends its metaphoric value to be inscribed as a mechanical and metonymic computing device. Consequently, the formalization of this and other passages scattered throughout the text favored, through different analogies, readings of the Torah that took it as an acrostic (*notaricon*), a permutation of letters (*temurah*) or a numerical code (*gematria*) of the name YHWE. In this way, the primitive Kabbalah can be counted as one of the first comprehensive speculative sciences that took linguistic computation tools as its fulcrum, with the aim of unveiling the enigma of divine creation.

A final example worth mentioning in this path is the mathematical system constructed by Leibniz, comparatively closer to our day. For his is one of the most complex notational developments on record. Leibniz builds his system on a combinatorial of signs and characters with which he hoped to order mankind's thinking and the possibilities of knowing the world. This system was already shown in its time as a historical descendant, although structured

[15] The Kabbalists shared this belief with the Pythagoreans, for which the letters – *stojeion* – were elements of discourse (*logos*) and at the same time foundations that were used in the construction of the universe.
[16] *Sepher Yetzirah or The Book of Creation*, trans. by W. W. Wescott 1887, available at: http://www.sacred-texts.com/jud/yetzirah.htm

with a visible logic, of those Renaissance efforts put forward by Ramon Lull and other philosophers. What Leibniz was looking for with his set of characters – as all mathematicians had done before him – was certainty for a system, a parameter of objectivity. Leibniz thus elaborates a vision in which the secrets of the world could be opened to whoever could read the projection of geometric lines and the laws of calculation that govern them. Leibniz called this language a *characteristica universalis*, and it was supposed to accurately translate human thoughts, making them objective and free from error. «This very writing – the philosopher argues – would be a kind of general algebra and would provide a means of reasoning through calculus, so that instead of arguing, one could say: let's count».[17] However, due to its scope and sheer ambition, the Leibnizian project was left as an unfinished attempt.[18]

1.3 To contemplate / to meditate / to worship

The intentionalities inscribed on technical devices have not always been oriented to control or to count, nor to produce efficiently under an economic logic. Many times the encoded objects have emerged as mere pastimes, like ancient board games and automata referred to in Greek and Medieval texts. Along this line of inquiry, there have been technological developments that have simply sought to serve as reflections of the ephemeral elements of the natural world, as aids in contemplation, so to speak. Heron of Alexandria, for example,

[17] G. W. Leibniz, *Scientia Generalis. Characteristica*, Gerhardt, K. I. (ed.) *Die Philosophische Schriften*, Berlin 1890, p. 26.

[18] Even if Leibniz did not achieve what he expected to be a universal language, his work in mathematics, set towards that aim, led to the invention of modern calculus (an accomplishment he shares with Newton, who derived a similar mathematical system out of different metaphysical assumptions). Leibniz is of course a very well known and estimated figure among contemporary historians of mathematics and science. Norbert Wiener even named him the saint patron of cybernetics. According to Wiener, «The philosophy of Leibniz centers about two closely related concepts – that of a universal symbolism and that of a calculus of reasoning. From these are descended the mathematical notation and the symbolic logic of the present day. Now, just as the calculus of arithmetic lends itself to a mechanization progressing through the abacus and the desk computing machine to the ultra-rapid computing machines of the present day, so the calculus ratiocinator of Leibniz contains the germs of the machina ratiocinatrix, the reasoning machine» (N Wiener, *Cybernetics. Or Control and Communication in the Animal and the Machine*, Cambridge 1985, p. 12). This quote is full of allusions that need clarification and context. It is evident, for example, that the 'current' desktop computer Wiener mentions does not refer to those existing in our time, although it prefigures them. For this reason, the concretion of a *machina ratiocinatrix*, which Wiener takes as fact, must remain as an assumption. Wiener also assumes that cybernetics fulfills Leibniz's desire to grant order and a deep knowledge of the world, but if that were the case, it would mean that cybernetics fulfills a Kabbalistic and a Pythagorean desire as well. On the other end of this speculation, it could also be stated that today's desktop computer reproduces the world to its own image, a claim that most probably would have horrified Leibniz. In other words, these can only count as anachronisms. This article proposes precisely that similar intentionalities might be at work, but given that they all function within their own historical context, further provisions need to be taken before they can be contrasted and analyzed, as we will see.

was the inventor of the first steam engine, the aeolipile, which was an apparatus made up of a chamber of air or water (generally a half sphere or a cylinder that could be heated) with tubes protruding out of it, by which the steam was expelled, moving thus the whole metallic body. However, with this mechanism Heron merely designed fountains and other mobiles (*Pneumatics* II, XI), that is, his invention was not intended to engage in mobility or productive projects, but was proposed instead for the embellishment of public places.

Along the same lines, Lynn White Jr. suggests that the technology behind the windmills used in medieval Europe for grinding grains was actually used as early as the 8th century BCE in China and Tibet for religious purposes. Thus, the Buddhist cylinders activated by the wind, which revolved around their vertical axis like anemometers, were part of a complex technology of prayer in which all the technical knowledge of an era came together in an organized way.[19] White also reminds us that the most complex machinery of the European Middle Ages was the wind organ, a monumental instrument built for Christian churches of different confessions and located in the most diverse latitudes.[20] In the 10th century, a 400-flute specimen was built inside the Winchester Cathedral in England; it required two people to play it and 70 men to pump air into its air retainer.

These devices remind us that the spiritual aspect linked to technology cannot be underestimated. After all, the so-called information age was born when the printing press set up by Gutenberg, turning an old wine press into a rotary press, managed to print a Bible around 1455. In that sense, the socio-computing infrastructure implied in these early devices worked under a completely different intentionality than the one that emerged later with the aim of maximizing productivity and profit, even if the visible mechanisms may perform similar functions.

1.4 To structure / to coordinate / to enable

By structuring an order (access, sequence, coordination, alternation, etc.), an algorithm not only regulates, but also enables a process. When this is the case, the algorithm behaves as a protocol that assigns and distributes tasks, while it effectively manages the sequences that comprises a generative chain. Thus, an algorithm *qua* protocol is an ordering method. It is productive since the steps it includes were articulated at a given moment with a certain logic, to be used at different points in an automatic, decentralized way, and without the need to reflect thoroughly upon the consequences of each of the steps during its application. There are different types of protocols. As Alexander Galloway writes:

[19] L. White, *Medieval Technology and Social Change*, London and Oxford 1974, pp. 85-6.
[20] *Ibid.*, p. 117.

Prior to its usage in computing, protocol referred to any type of correct or proper behavior within a specific system of conventions. It is an important concept in the area of social etiquette as well as in the fields of diplomacy and international relations. Etymologically it refers to a fly-leaf glued to the beginning of a document, but in familiar usage the word came to mean any introductory paper summarizing the key points of a diplomatic agreement or treaty.[21]

In a world obsessed with efficiency and the higher aim of productivity, protocols are multiplying. These include procedures, formalities, or specificities that must be fulfilled in a course of events. A protocol is a waiting line at the super market or in a government office, or its substitution for another sequencing mechanism that implies giving a number to each person when they arrive. This does not mean that protocols are always effective, as anyone who has become entangled in bureaucratic procedures can testify. Getting on the public transport by paying the ticket to the driver or to a collector involves a suboptimal process compared to getting on board with a pre-paid card, but the protocol also obeys specific social conditionings and tenacious habits that sometimes defy common sense.

In any case, protocols can be understood as standards that govern the implementation of characteristic sequential processes. Technical protocols establish then the essential and agreed points for a standard of action with which the exchanges and relations of a large number of users can then be managed. By these means, technical protocols disseminate peculiar ideologies and forms of control; in a first layer, these can be decentralized and pre-agreed between institutions and directing agencies, then they get to be coordinated as a framework that regulates flows, systematizes exchanges, encodes relationships, and connects agents.

In that sense, a protocol cannot be seen as an imposition. Nobody would claim that a traffic light placed in a street junction to regulate vehicular traffic tries to impose its power; what it achieves instead is the regulation of sequences that operate in intersections or opposing trajectories. In other words, a protocol coordinates flows and needs, with a distributive logic that makes it easier for everyone to achieve their particular goal. Hence, the protocol is not simply a normative framework or a set of rules to be submitted to, but an order that enables a circuitry. And as Galloway argues here again: «protocol is like the trace of footprints left in snow, or a mountain trail whose route becomes fixed only after years of constant wear. One is always free to pick a different route. But protocol makes one instantly aware of the best route – and why wouldn't one want to follow it?».[22] A protocol is a form of organization that raises pre-formed solutions to recurring problems; it is a successful formula because the benefits of submitting to that order clearly outweigh the inconveniences it can

[21] A. Galloway, *Protocol. How Control Exists After Decentralization*, Cambridge 2004, p. 7.
[22] *Ibid.*, p. 244.

generate. In that sense, the protocol is a rational operative device, which outlines an agreement to whose jurisdiction it is assumed advantageous to adhere to.

1.5. To disorder / to deconstruct / to regenerate

We have already seen how a social algorithm is an important device to try to order and structure matter and the world under particular conceptions. But there are algorithms that seek exactly the opposite: to reverse or deconstruct a given state of affairs or a social structure, or at least to call it into question. Clearly, this intention was not a recurrent search during Antiquity, in civilizations that endured stable cultural forms as natural facts of the world. In any case, towards the Middle Ages in Europe, there were already social devices that allowed for the generation of temporary spaces of symbolic rearrangement, such as a 'carnival'. Now, the concrete structure, the sources and the imminent effects of that phenomenon constitute a particular case in the history of culture. For as long as it was linked to popular customs, the carnavalesque occurred in informal layers of the cultural production of the time, away from intellectual interpretations and analyses, so its emergence and cultural networks are difficult to reassemble and follow. In this regard, the critic Mikhail Bakhtin found in the work of the French writer François Rabelais (1494-1553) a communicating vessel that informed of a 'secret' world, insofar as it was nurtured through and thrived precisely on it. Bakhtin located thus the key to a social algorithm in a specific literary work. As the critic writes:

> Rabelais is difficult. But his work, correctly understood, casts a retrospective light on this thousand-year-old development of the folk culture of humor, which has found in his works its greatest literary expression. Rabelais' illuminative role in this respect is of the greatest importance. His novel must serve as a key to the immense treasury of folk humor which as yet has been scarcely understood or analyzed. But first of all it is necessary to take possession of this key.[23]

The Rabelaisian corpus and all the very interesting relationships that it illuminates are certainly not part of the scope of this article. However, it is worth taking into account the way in which the social algorithm that Bakhtin uncovers, rearticulates and uses in Rabelais, manages to reconstruct a particular, invisible structure for an entire era, in the complex operation of restoring, in a sort of archaeological approach, the social algorithm with its specific means of transmission – i.e. the cultural ecosystem as a socio-computing infrastructure – that gives it its characteristic fertility. In a way, Bakhtin manages to locate the central ritual act of the carnival in the false coronation and deposition of the carnival king. This is the opening act and the culminating point of that event: an inverted world is here denoted. The coronation and deposition thus constitute a dual and ambivalent ritual that expresses change, the relativity of structure

[23] M. Bakhtin, *Rabelais and his World*, trans. H. Iswolsky, Bloomington 1984, pp. 3-4.

and order, and the contingency of authority and hierarchic orderings. In that sense, an inversion is not merely a disorder, but a restructuring that dismantles social structures and habits, and thus shows the ultimate meaning of this social ceremony: the pathos of change and renewal, the cycles of death and rebirth. The carnival is for Bakhtin then a festival that brackets time (a *chronotopos*, in his own terminology) in order to end and renew everything. The source code in Rabelais allows Bakhtin to read not only aspects of an archetypal subjectivity, but also a whole cluster of life forms, of ways of being and of a set of values and inquiries for a determined historical moment.

What Bakhtin found in Rabelais was experienced by other cultural agents in the wake of the 20th century. Through different means, literary or artistic, encoding operations became flexible and variable. During the early decades of the century, these agents were already trying to leave behind an aesthetic and epistemological model to try to inaugurate an alternative cultural order. For the Dadaist poet Tristan Tzara, for example, who challenged an artistic tradition that valued abstractions such as the work of 'genius', 'talent', the 'uniqueness' of the artwork or the importance of passive contemplation, a poem should not only affect the life of an individual, but it must also pierce through the very heart of tradition itself. The generation of an algorithm allowed him to establish computational processes from arbitrary inputs that not only made every reader a creator in him or herself, but also aimed at deconstructing the tradition of the individual 'genius' in literature. His text *To Make a Dadaist Poem* from 1923 states:

> Take a newspaper.
> Take a pair of scissors.
> Choose an article as long as you are planning to make your poem.
> Cut out the article.
> Then cut out each of the words that make up this article and put them in a bag.
> Shake it gently.
> Then take out the scraps one after the other in the order in which they left the bag.
> Copy conscientiously.
> The poem will be like you.
> And here are you a writer, infinitely original and endowed with a sensibility that is charming,
> though beyond the understanding of the vulgar.[24]

This algorithmic process for the production of a literary piece turns repeatability into a subversive trait. That is, while for Christian mystics or Jewish kabbalists repeatability promised the attainment of an order and the formation of a totality, Tzara mechanizes a process that is assumed to be of personal inspiration, and with that move he reinserts chaos and dispersion into the production of meaning. Tzara's algorithm prefigures the post-structuralist conception of the world in which there are no essences or deep truths, where

[24] T. Tzara, *Pour faire une poéme dadaïste*, *Œuvres completes*, Paris 1975.

an author is a function and an effect of a work and not the expression of a predefined subjectivity.

Other artists throughout the 20th century also generated pieces as instructions that could be repeated by anyone. That was the case, for example, of the artists associated with the *Fluxus* movement. John Cage premiered in 1952 his well-known piece *4'33"*, a work in which the notation asks an interpreter to generate – or rather: to open – a frame to listen to the accidental sounds at a concert hall while the interpreter remains still. Cage's piece is a search to underline what exists as a product of a previous program (that of classical music) and to depict reality as generated and generative, rather than an attempt to introduce a new aesthetic for an unintentional and random sound, as it has sometimes been interpreted[25].

These algorithms contain expressive intentionalities, points of view, subjective inflections and subcultural interrelationships. But they are only effective when understood under the parameter of a specific social frame – the world of art experimentation – where they are seen as contained nodes of relations and exchanges. In any case, all these examples denote how a work of literature or an artwork may condense in itself, abbreviated, the character of its time. This gives us an invaluable opportunity to understand an algorithm as a cultural symptom of a specific era.

2. Filtering, Contrasting, Reassembling

The previous section aimed at showing feasible assemblages of social algorithms working on historical socio-computing practices according to common inherent structures and intentionalities. On the one hand, that procedure helps to observe certain affinities and possible genealogies. On the other, it also allows grasping fundamental differences, so that one can understand, for example, why the steam engine from the 19th century stands basically on a completely different path of development than the early aeolipiles invented by the Greeks[26]. In that sense, the focus of these emerging sets has been to shift away from the machine-based functioning of an algorithm, and to concentrate instead on the socio-computing infrastructures on which social algorithms rely, and over which collective intentionalities play a major role. Clearly there is, for every historical

[25] Cage uses the rhetoric that surrounds the meaning of 'a concert' to highlight the same concert-form, thereby underlining the fact that a sound piece can be a simple 'frame' in which 'anyone' can participate, which brings him closer to the Dadaist spirit of Tzara. On the other hand, the randomness involved in the piece is not a totally open accident, as might be believed at first, so we are not in front of a silent piece, linked to Japanese Zen, or to the mystical contemplation of the concrete. *4'33"* is a code that commands the opening of a sound receptacle in which a controlled, stochastic chance comes into play, by means of an algorithm that is also random in itself.

[26] This claim could be further investigated, to allow for example a discussion with the thesis of Gilbert Simondon and what he observed as an autarkic evolution of technical realities. Cf. G. Simondon, *Du mode d'existence des objets techniques*, Paris 1958, esp. pp. 50-70.

practice, a wider socioeconomic and cultural context that needs to come to the fore for a deeper archeo-historical analysis, but that would require a thorough study, which is not here intended. What is here pursued instead is to understand what working with the notion of intentionalities might yield, and the reach these may have when reflecting on the overall structure of a socio-computing complex.

To continue our quest, this section will now inquire deeper into specific practices. The aim is to observe what social algorithms might entail, and to remark how comparing practices ascribed to given intentionality sets can actually illuminate one another, for the intrinsic similarities and differences can be telling of the respective socio-computing infrastructure that is at work in each of them. For this end, the groups presented in the previous section need not be taken as definitive and closed sets, but only as guidelines or proxies; high porosity between categories and the slipping of meaning are assets in a cultural dynamic, although they might make difficult an analytical task that pretends to make observations across extended time periods. For this same reason, the practices here analyzed are taken from recent decades, to avoid unnecessary assumptions or just slip unwise anachronisms. That this is needed will be clear when we see that, even if the time period is restricted, shifts in technological habits and collective behaviors can be already distinctly perceived.

2.1 Completing a task: social algorithms as scripts

For this first case, we will start by laying forth another artwork that could have been easily added to the last set from the previous section. George Brecht, an artists associated with the *Fluxus* movement, took the process of algorithmic art to the extreme, showing everyday activities as programmed procedures. Stated as operational processes, that is, habits as computational loops, the aesthetic components of the everyday resurface in his works as an unintentional coded style in which individuals interact with each other. A classic Brechtian piece from 1961 formally reinterprets the simplest act of answering the phone:

> Three telephone events
>
> *When the telephone rings, it is allowed to continue ringing, until it stops.
> *When the telephone rings, the receiver is lifted, then replaced.
> *When the telephone rings, it is answered.[27]

More than commanding an action, Brecht's piece suggests that in our daily life every subject participates in some work of art (maybe his own?) without fully being aware of it. Of course, that would only be clear to a person that knows how

[27] G. Brecht, *Three telephone events,* 1961. Retrieved from: *Galeria d'arte moderna gammm* https://gammm.org/2021/03/21/three-telephone-events-george-brecht-1961/

conceptual art insinuates its meaning, not to the lay individual. For in the end, Brecht is of course not the author of this particular daily performance. But by issuing a script over the array of possibilities around a given event for a given era (nowadays most telephones don't have a receiver to be lifted, and contemporary cell phones may be answered by way of a text or through an already old-fashioned answering machine), the artist underlines this action as something than can be fully anticipated and programmed, much like in a theater play. The script is what he claims as an art piece; the audience is then left to relate to it freely. But even ignorance of the piece does not leave an individual unaffected: the act of answering a phone is made transparent as a coded moment in daily life. In the end, the piece is ingenious because, when anyone reads and understands its meaning, a daily action appears as 'fully' conditioned, for even the minute exceptions on particular 'telephone-ringing events' can be individually filled ad hoc and fitted in the Brechtian coding. And as Lucy Suchman writes, describing how we go through the mutual intelligibility of action:

> We walk into a situation, identify its features, and match our actions to it. This implies that, on any given occasion, the concrete situation must be recognizable as an instance of a class of typical situations, and the behavior of the actor must be recognizable as an instance of a class of appropriate actions.[28]

Accordingly, the Brechtian piece relies on a shared understanding of the actions and elements involved on a particular event, so that his script can be simultaneously seen as 'evident' and therefore 'effective'.

We can compare this work with the description of a very simple task in artificial intelligence research, which shows entanglements in the uncertainties of an action and its many deviations from intended effects. Describing the task of 'turning on a light', the computational scientist James Allen wrote two decades after Brecht:

> There are few physical activities that are a necessary part of performing the action of turning on a light. Depending on the context, vastly different patterns of behavior can be classified as the same action. For example, turning on a light usually involves flipping a light switch, but in some circumstances it may involve tightening the light bulb (in the basement) or hitting the wall (in an old house). Although we have knowledge about how the action can be performed, this does not define what the action is. The key defining characteristic of turning on the light seems to be that the agent is performing some activity which will cause the light, which was off when the action started, to become on when the action ends. An important side effect of this definition is that we could recognize an observed pattern of activity as 'turning on the light' even if we had never seen or thought about that pattern previously.[29]

[28] L. A. Suchman, *Human-Machine Reconfigurations. Plans and Situated Actions*, Cambridge 2007, p. 81.
[29] J. Allen, *Towards a general theory of action and time*. «Artificial Intelligence» XXIII, 1984, pp. 123–154.

Allen uses this example precisely as a way to problematize the indeterminate relationships of intended effects when articulated into a descriptive method, which would be used afterwards in a sequencing program. That Allen relies in the notion of intention here to anticipate an action is clear in that he tries to convey every possible course of action that derives to the desired effect, as if the accumulation of steps were the expression of an underlying plan. Of course, the contingencies leading to 'turning on a light' are also significantly broader than those of answering a telephone (in an extreme, one can also include 'opening the curtains to let the sunlight in'), but Allen's point is that the accumulation of tasks will not necessarily lead to a precise outcome, and that, even if it does, the outcome cannot be reconstructed precisely out of the contingent tasks leading to its accomplishment. In his own work, Allen proposes then a path for plan recognition by constructing a logical language for action descriptions that distinguishes causality and logistics. Nevertheless, Allen does not abandon the principle of planning or plan recognition systems, which relies on the anticipation of actor's intents and purposes, and ignores the adaptability to situated actions that scripts might offer. And as Suchman suggests: a

> planning model […] takes over our commonsense preoccupation with the anticipation of action and the review of its outcomes and attempts to systematize that reasoning as a model for action while ignoring the actual stuff, the situated action, which is the reasoning's object.[30]

The shortcomings of the planning model provide then an effective contrast between plans and scripts.

Plans associate intentions with action sequences, while scripts – as it is clear in the Brechtian piece – associate action sequences with typical situations, and are usually described as partial list of tasks, which actors may take further and adapt in their own unique way. A social algorithm seems to be better described through a script, where the contingency of individual intentions subsides, and where the problem to be solved appears to be more malleable, yet remains clear. Moreover, the description of plans implies mostly post hoc *representations* (reconstruction of intentions, imagined projections, recollected reconstructions), and the observer needs to imagine a course of action between two temporal points as if the agent had everything under control. On the other hand, to describe a social algorithm as a script entails certainly post hoc *interpretations* (semantic approaches, the heuristics of proof and error, playful conjectures etc.) but these are fundamentally based on ad hoc recurrences of the coding process itself, made over time and reiterated by several individuals.[31] This is what unveils

[30] Suchman, *Human-Machine Reconfigurations*, cit., pp. 60-61.

[31] Harold Garfinkel describes forms of analog coding within a medical clinic that shed light on the function of 'ad hoc' processes that end up structuring a task, an epistemological category, or an empirical finding as a piece of systematized information. He writes: «Ad hoc considerations are invariably relevant considerations in deciding the fit between what could be read from the clinic folders and what the coder inserted into the coding sheet. No matter how definitely and

the articulation of collective intentionalities within them. In the end, scripts are enriched by small deviances and quirks, but they maintain a specific type-form, since that they are engrained in aspects of a shared know-how, as part of the socio-computing infrastructure in which they perform.

2.2 Expanding a speech act: enacting ways of being and influencing states of affairs

For this second case, we will consider an intentionality set that was not laid out in section one, and which could be stated as the cluster *to judge / to sanction*. Some insights brought about in Garfinkel's article *Some rules of correct decision making that jurors respect,* from 1967, will serve here as a point of departure. In that piece, Garfinkel analyzes what it means for a person to be called on duty to perform the task of a jury under the US-American system of justice (which incidentally provides here a stark contrast to the function of the law in Antiquity). Garfinkel finds that, when appointed to that position, jurors feel responsible to modify the rules used in daily life, including their own. In other words, besides introducing a verdict into the world – a classical uptake that follows the structure of speech-act theory – jurors not only influence a state of affairs, but also affect their own attitudes and self-perception.[32]

In this sense, the speech act of pronouncing a verdict should be analyzed not only as a linguistic utterance with pragmatic consequences, but as part of a cognitive microsystem that exceeds the linguistic domain and enters into a complex social phenomenon of role-assigning, role-playing, and group thinking, where the valued task and prerogative of 'announcing a verdict' is able to transform behaviors, ways of being and the perception of situations. No wonder that the whole realm of *juror decision making* has become a field of cognitive inquiry in its own right. As Pennington and Hastie argue:

elaborately instructions had been written, and despite the fact that strict actuarial coding rules could be formulated for every item, and with which folder contents could be mapped into the coding sheet, insofar as the claim had to be advanced that Coding Sheet entries reported real events of the clinic's activities, then in every instance, and for every item, 'etcetera', 'unless', 'let it pass' and 'factum valet' accompanied the coder's grasp of the coding instructions as ways of analyzing actual folder contents» H. Garfinkel, *What is ethnomethodology?,* in *Studies in Ethnomethodology*, Englewood Cliffs, 1967, p. 21.

[32] When Garfinkel describes the changes on jurors' ethical and epistemological attitudes, he states: «in their idealized accounts, jurors talked as if they knew the rules of decision making before they went into the deliberations; jurors did not say, nor did they care to discuss the fact, that it was in the course of the deliberations that they learned how the decisions were made […]. Their accounts stressed instead that from the beginning they knew what was expected of them and used this knowledge.» H. Garfinkel, *Some rules of correct decision making that jurors respect*, in *Studies in Ethnomethodology,* cit., p. 113.

the juror is a sense-making information processor who strives to create a meaningful summary of the evidence available that explains what happened in the events depicted through witnesses, exhibits, and arguments at trial.[33]

In other words, a contemporary legal system does not rely in the mere application of general laws to individual cases, but performs through the interactions between individuals trained in or attuned to the system's traditions, the collective intentionalities behind those traditions, and the institutional precedents that these conform. And as Gallagher writes:

> [Verdicts] are not necessarily confined to individual brains, or even to the many brains that constitute a particular court. They emerge in the workings of a large and complex institution. Yet these verdicts and legal proceedings are cognitive processes that then contribute to the continued working of the system in the form of precedents. The practice of law, which is constituted by just such cognitive and communicative processes, is carried out via the cooperation of many people relying on external (and conventional) cognitive schemas and rules of evidence provided by the legal institution itself.[34]

The intricacies of the legal mechanism clearly suggest that, in the end, a verdict matters less for its truth-content (outright evidences may be discharged on formal grounds, a sentence might be appealed, etc.) than for the processes it triggers within the social system in which it works. In that sense, the legal socio-computing infrastructure is less attuned to reflect upon facts in themselves, as to engage in interactional procedures where legal inputs are transformed into convened statements that influence, as judged things, the attributes of social situations.

In a similar way, we can analyze some technical algorithms that introduce their own 'verdicts' into the world, such as the dating apps. For instance, in Tinder, one of the most popular apps nowadays, a 'match' is a verdict that drives possibilities: its likelihood pulls representational strategies, linguistic habits and collective behaviors from the users, while its actual occurrence enables new options for them (e.g. contacting another person) which will keep the 'game' going. In that sense, an algorithm introduces a judgment into the world – two persons might fit romantically together – which organizes and coordinates a community of users[35]. Now, of course the programmed algorithm has a function

[33] N. Pennington and R. Hastie, *A Cognitive Theory of Jury Decision Making: The Story Model.* «Cardozo Law Review» 13, 1992, pp. 519-557, here p. 519.
[34] S. Gallagher, *The socially extended mind*, cit., p. 7.
[35] The function of the 'judgment' has been clearly distinguished, but users see the program also as a game or as a beauty contest plus messaging. This implies some perceptions of the 'judging app' as a juror of 'beauty' and 'ugliness'. Cf. B. Bosker, *Why Tinder has us addicted: The dating app gives you mind reading powers,* «The Huffington Post», September 4, 2015. Retrieved from http://www.huffingtonpost.com/2013/04/09/tinder-dating-app_n_3044472.html); and D. Wygant, *The shocking truth about Tinder dating.* «The Huffington Post», October 2014. Retrieved from http://www.huffingtonpost.com/david-wygant/the-shocking-truth-about-_3_b_4967472.html

of its own in this structure and its operative sequencing, but its reach is pegged to a broader social dynamic, where the intended rules of the app mingle with expectations, anticipations and other developmental conducts, i.e. collective intentionalities, which will end up producing a wider assemblage that can be recognized as a social algorithm in its own right. The social algorithm revolves around the match-as-verdict event, but not because this is taken in its truth-content, as a new fact of the world[36], but because of the ways of being and the dynamics or 'life forms' this triggers within its entourage. As the anthropologist Nick Seaver comments on a presentation of his colleagues:

> In their study of an online dating site, [it was found that] various actors enacted the site's algorithm differently: engineers tweaked their code to mediate between the distinctive behaviors of male and female users; some users tried to game the algorithm as they understood it, to generate more desirable matches; other users took the algorithm's matches as oracular pronouncements, regardless of how they had been produced. No inner truth of the algorithm determined these interactions, and non-technical outsiders changed the algorithm's function: machine learning systems changed in response to user activity, and engineers accommodated user proclivities in their code.[37]

In the end, this means that the machine-based algorithm might have an influence over a given structure or state of affairs, but it is actually in the affordances that this code can achieve in its interactions within a social situation, and therefore in the assemblage of a more complex social algorithm, that the final set of arrangements or effects need to be inspected.

Conclusion

Drawing on an expanded understanding of Tomasello's hypothesis on the nature of collective thinking, this article has argued that collective intentionalities are inherent settings embedded in social algorithms, and that they play an important role in the socio-computing infrastructure in which these algorithms perform. Collective intentionalities are not easy to define and refer, but by way of a historical sampling, even if restricted, sets of practices and devices can be organized under particular intentional arrangements in order to inspect their joint articulation. The description of these likely articulations does not imply an analitical proof in itself, for intentionalities designate fluid, pre-conceptual qualities, but the process leads to an heuristics that allows for sharp focusings on characteristic practices through which a deeper understanding of social algorithms can ensue. In that sense, those assemblages can be further inquired

[36] Even if the advertisement of the app has sometimes pushed the idea that it generates life-changing events (for instance by playing with the motto «Any swipe can change your life»), users appropriate themselves the logic of the program and develop their own uses and routines. Cf. G. David and C. Cambre, *Screened Intimacies: Tinder and the Swipe Logic,* «Social Media + Society» April 2016.

[37] N. Seaver, *Algorithms as culture: Some tactics for the ethnography of algorithmic systems*, «Big Data and Society», 2017, pp. 1-12, here p. 6. DOI: 10.1177/2053951717738104

in different directions and depths. I have pursued two specific cases here to show how intentionalities allow for a better understanding of social algorithms, by presenting contrasts and structural differences between adjacent practices and domains.

In the end, this article aims at bringing a novel configuration to the fore, for which further analysis and descriptions are needed. But a substantive attainment of this text would be already to shift the focus of an algorithm from a machine-based perspective to the interrelated and normative cultural practices sustained through socio-computing infrastructures, where collective intentionalities are seen to play a defining role. In that sense, an algorithm should be seen as a collective social technology, emerging as a cultural script, where a recurring process of signifying, inscribing and interpreting takes place, and for which a given constellation of social thinking, involving a characteristic worldview, is always mobilized.

Javier Toscano
CAIS – Center for Advanced Internet Studies
✉ tosgue@yahoo.com

References

Primary Sources

Bakhtin, B. 1984. *Rabelais and his World*, trans. H. Iswolsky, Bloomington, Indiana University Press.

Frazer, J. G. 1983. *The Golden Bough. A Study in Magic and Religion*, London, McMillan.

Galloway, A. 2004. *Protocol. How Control Exists After Decentralization*, Cambridge, MIT Press.

Garfinkel, H. 1967. *Studies in Ethnomethodology*, Englewood Cliffs, Prentice Hall.

Leibniz, G. W. 1890. *Scientia Generalis. Characteristica*, Gerhardt, K. I. (ed. by) *Die Philosophische Schriften*, Berlin, Weidmann.

Schutz, A. 1962. *Collected papers I: The problem of social reality*, The Hague, M. Nijhoff.

Suchman, L. A. 2007. *Human-Machine Reconfigurations. Plans and Situated Actions*, Cambridge, Cambridge University Press.

Tomasello, M. 2014. *A Natural History of Human Thinking*, Cambridge and London, Harvard University Press.

White, L. 1974. *Medieval Technology and Social Change*, London and Oxford, Oxford University Press.

Wiener, N. 1985. *Cybernetics. Or Control and Communication in the Animal and the Machine*, Cambridge, MIT Press.

Secondary Literature

Allen, J. 1984. *Towards a general theory of action and time*, «Artificial Intelligence», XXIII, pp. 123–154.

Bosker, B. 2015. *Why Tinder has us addicted: The dating app gives you mind reading powers,* «The Huffington Post», September 4 2015, retrieved from http://www.huffingtonpost.com/2013/04/09/tinder-dating-app_n_3044472.html

Brecht, G. 1961. *Three telephone events,* retrieved from «Galeria d'arte moderna gammm» https://gammm.org/2021/03/21/three-telephone-events-george-brecht-1961/.

Chant, S. R., Hindriks, F., and Preyer, G. 2014. *From Individual to Collective Intentionality*, Oxford, Oxford University Press.

Chapman, M. 1990. *Intention, intentionality, and the constructive character of scientific knowledge,* «Psychological Inquiry», I, pp. 251–253.

Cramer, F. 2005. *Words made Flesh. Code, Culture, Imagination*, Rotterdam, Piet Zwart Institute.

David G. and Cambre, C. 2016. *Screened Intimacies: Tinder and the Swipe Logic,* «Social Media + Society», April.

Gallagher, S. 2013. *The socially extended mind*, «Cognitive Systems Research», 25-26, pp. 4-12. http://dx.doi.org/10.1016/j.cogsys.2013.03.008.

Goldstine, H. and von Neumann, J. 1947. *Planning and Coding of Problems for an Electronic Computing Instrument*, *Report on the Mathematical and Logical aspects of an Electronic Computing Instrument Part II*, vol I-3, Princeton, Institute for Advanced Study.

Harper, R. F. 1904. *The Code of Hammurabi, King of Babylon about 2250 B.C.*, Chicago and London, University of Chicago Press.

Newell, A., Shaw J. C. and Simon, H. 1958. *Elements of a Theory of Human Problem Solving*, «Psychological Review» LXV, 3, pp. 151-166.

Nida-Rûmelin, J. 2014. *Structural Rationality and Collective Intentions*, in Chant, S. R. et al. (ed. by) *From Individual to Collective Intentionality*, Oxford, Oxford University Press, pp. 207-222.

Pennington N. and Hastie, R. 1992. *A Cognitive Theory of Jury Decision Making: The Story Model.* «Cardozo Law Review», 13, pp. 519-557.

Schmitt, C. 1950. *Der Nomos der Erde im Völkerrecht des Jus Publicum Europaeum*, Berlin, Duncker & Humblot.

Searle, J. R. 1980. *The intentionality of intention in action*, «Cognitive Science», IV, pp. 47–70.

Seaver, N. 2017. *Algorithms as culture: Some tactics for the ethnography of algorithmic systems*, «Big Data and Society», pp. 1-12. DOI: 10.1177/2053951717738104.

Simondon, G. 1958. *Du mode d"existence des objets techniques*, Paris, Aubier-Montagne.

Wescott, W. W. (trans. by). 1887. *Sepher Yetzirah or The Book of Creation*, http://www.sacred-texts.com/jud/yetzirah.html.

Wygant, D. 2014. *The shocking truth about Tinder dating,* «The Huffington Post», October 2014. Retrieved from http://www.huffingtonpost.com/david-wygant/the-shocking-truth-about-_3_b_4967472.html.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7768564

Contributi/3

Produire l'inconnu

L'intelligence artificielle au prisme des ambivalences du rationalisme leibnizien

François Levin 0009-0005-4453-093X

Articolo sottoposto a doppia blind peer review. Inviato il 16/12/2021. Accettato il 08/06/2022.

Producing the unknown. Artificial intelligence in the light of the ambivalences of Leibnizian rationalism

This article is born from an astonishment: that aroused by the reading of Leibniz. It was expected to find traces of an absolute rationalism, oriented towards the reduction of everything to the known as well as towards the increase of control of the world by the knowing subject. On the contrary, we discover deep ambivalences, especially in his major project of constituting an art of invention (*ars inveniendi*): it is a profoundly baroque logical project, oriented by the desire to produce novelty and which leads to a form of dispossession of the subject. We can produce, from this astonishment, another genealogy of artificial intelligence, which allows us to understand its current creative developments. It also helps us to grasp how it is characteristic of an unconscious form of metaphysics of the modern era, which places its search for otherness inside machines.

***

Introduction

Pour celui – ou celle – qui, travaillant sur les enjeux contemporains liés à l'intelligence artificielle, décide, dans une perspective généalogique, de parcourir différents foyers séminaux de la pensée du calcul et de sa toute-puissance, la lecture de Leibniz ne laisse de susciter l'étonnement.

Il est admis de présenter le philosophe comme un des précurseurs de l'intelligence artificielle, tant il incarne l'aboutissement d'une entreprise rationaliste poussée à l'extrême, gouvernée par l'absoluité du principe de raison, ce dont découle un certain nombre de conséquences métaphysiques et logiques. A cet égard, la première approche de l'œuvre de Leibniz dans le domaine de la logique permet de confirmer l'horizon d'attente qui y présidait: Leibniz propose une théorie formelle de la pensée, qui s'appuie sur des symboles détachés de l'esprit et qui prend pour modèle, de manière plus littérale encore que chez Descartes, les raisonnements mathématiques. En ce sens, il développe et précise

la conception de la pensée comme manipulation de symboles que Hobbes développe au chapitre V du *Léviathan*; c'est ce geste que Leibniz va tenter de préciser et de développer tout au long de sa vie.

Toujours dans la perspective généalogique qui est la nôtre, le travail de Leibniz semble de prime abord correspondre à l'idée que l'on pourrait se faire d'un rationalisme total, c'est-à-dire d'une ambition théorique où la rationalité calculatoire devient l'horizon et le modèle de toute pensée, seule à même d'assurer une efficacité et une maîtrise sur le foisonnement des choses. C'est en effet le sens premier des projets de la logique leibnizienne (Combinatoire, Caractéristique universelle, Encyclopédie universelle, Science générale) que de viser à limiter les incertitudes et les ambiguïtés, à assurer une maîtrise plus importante sur les choses du monde, à réduire les possibilités de divergence et de conflit, et à assurer à tous, enfin, la possibilité de la sûreté de ses jugements, garantie de tranquillité d'esprit et d'ordre[1]. Limitation de l'incertitude, règne du principe de raison et du principe d'identité, pensée systématique achevée, mathématisation de la pensée: tout semble concorder pour voir en Leibniz, dans la continuité de Descartes, le penseur d'une forme de mise au pas rationnelle du monde, qui trouverait aujourd'hui son digne successeur dans l'intelligence artificielle. C'est d'ailleurs la lecture qu'en fait Heidegger dans *Le Principe de raison*, où, sans évidemment mentionner l'intelligence artificielle, il affirme le rôle central de la pensée leibnizienne pour comprendre le règne de la puissance du calcul dans l'époque moderne:

> La pensée de Leibniz soutient et marque de son empreinte la tendance principale de ce qui, pensé d'une façon suffisamment large, peut être appelé la métaphysique de l'époque moderne[2].

Cette métaphysique de l'époque moderne désigne, chez Heidegger, la puissance de la forme représentative et calculatoire d'appréhension du monde et son corrélat la domination de la raison instrumentale technicienne.

D'où provient alors l'étonnement que nous évoquions au début de l'article? Il s'origine, semble-t-il, dans trois éléments.

1. L'*ars inveniendi* au cœur de la logique leibnizienne

Tout d'abord, le fait que l'objectif principal des échafaudages logiques de Leibniz, depuis la *Combinatoire* en 1666 jusqu'aux derniers moments de sa vie[3], consiste à édifier un art d'inventer ou une logique de l'invention. Cet *ars*

[1] A cet égard Leibniz, contrairement à Descartes d'ailleurs, qui a toujours cherché dans la solitude une relative abstraction du jeu social et des contraintes qui y président, a toujours été un serviteur de l'ordre, sur le plan politique notamment, en particulier lors de ses différentes missions pour les princes allemands, comme le rappelle Deleuze dans son cours sur Leibniz.
[2] M. Heidegger, *Le Principe de raison*, Paris 2003.
[3] Ainsi, deux ans avant sa mort, Leibniz exprime dans une lettre à Remond le regret de n'avoir pu, faute de temps et d'assistance, accomplir ce qui a constitué une des grandes ambitions de

inveniendi est l'horizon fondamental des travaux de Leibniz: qu'il procède par synthèse ou par analyse, il est une manière de formaliser l'invention et de la faciliter. Au fond, les différents projets logiques de Leibniz servent cet objectif et on doit les comprendre par rapport à celui-là. C'est d'ailleurs à partir de cette ambition que Leibniz développe une critique de Descartes, de sa méthode et des règles qu'il développe pour la direction de l'esprit. Pour Leibniz, ces règles n'ont rien d'original, comme il le précise non sans ironie:

> peu s'en faut que je ne les déclare semblables au précepte de je ne sais quel chimiste: prends ce qu'il faut, opère comme il faut, et tu obtiendras ce que tu souhaites[4].

Comme l'explique bien Yvon Belaval dans son ouvrage sur le sujet, la critique de Leibniz vis-à-vis de Descartes porte sur différentes dimensions: tout d'abord sur le caractère inefficace des règles qu'il développe, dans les Règles pour la direction de l'esprit, mais également dans *Les Méditations métaphysiques*. Au fond ces règles ne sont utiles qu'à donner forme à un esprit voué naturellement à la découverte; la méthode exige, comme préalable, un *ingenium*: «Elle ne rend pas apte, mais plus apte à découvrir de nouvelles vérités; l'inventeur-né, guidé *per ingenitam quamdam sagacitatem,* commence par tâtonnements, *fortunae auxilio potius quam artis,* se contentant d'abord de principes non fondés, qui paraissent innés en notre esprit plutôt qu'élaborés avec méthode»[5]. Leibniz intente donc un procès en hypocrisie à Descartes: alors qu'il prétend développer une méthode afin que tout le monde puisse parvenir à se déterminer vers la vérité et à parvenir à des idées nouvelles, cette méthode n'est qu'un tuteur pour ceux dont le talent fournit déjà le principe de développement à leur raison créatrice. Il souligne d'ailleurs avec ironie le talent qu'il a fallu aux cartésiens pour suppléer aux béances de la méthode de Descartes et parvenir à de nouvelles connaissances à partir de celle-ci[6]. Au fond, Leibniz défend une forme de 'démocratisme' de l'invention contre le caractère réservé du génie naturel que Descartes prend comme point de départ de son raisonnement. Ce 'démocratisme' est nécessairement un formalisme, puisqu'il s'agit de mettre en place des outils qui permettent même aux personnes dénuées de tout génie de parvenir à des connaissances nouvelles. En ce sens, il s'oppose la méthode cartésienne qui relève *in fine*, en dépit des intentions affichées de son initiateur, d'une forme d'écologie de la création, c'est-à-dire d'une propédeutique qui concerne moins le moment de l'invention réelle que les conditions qui président à celle-ci.

son existence: échafauder «une Spécieuse générale, où toutes les vérités de raison seraient réduites à une façon de calcul».

[4] Y. Belaval, *Leibniz critique de Descartes*, Paris 1978, p. 33.

[5] *Ibid.*, p. 31.

[6] Comme le souligne Michel Serres, dans M. Serres, *Le Système de Leibniz*, Paris 1968. Et Leibniz de s'étonner: combien de génie a-t-il fallu à Descartes pour débrouiller tant de labyrinthes à l'aide d'un fil aussi incertain, et d'observer à juste titre que les cartésiens, ni personne, n'ont rien inventé dans la mesure où il ne suivaient que les Règles.

Cette critique du caractère inefficace de la méthode cartésienne va de pair avec une remise en cause plus profonde de la conception de la raison qu'elle charrie, et notamment du rôle de l'intuition au cœur de celle-ci. Elle porte en particulier sur la notion d'évidence, au cœur de la conception cartésienne de la vérité. Cette notion est par trop subjective et imprécise, pour Leibniz, de même que les critères de clarté et de distinction qui la spécifient:

> Je vois aussi que de nos jours les hommes n'abusent pas moins de ce principe si souvent vanté: tout ce que je conçois clairement et distinctement d'une chose est vrai et peut être affirmé de cette chose. Car souvent les hommes, jugeant à la légère, trouvent clair et distinct ce qui est obscur et confus. Cet axiome est donc inutile si l'on n'y ajoute pas les *critères* du clair et du distinct, que nous avons proposés, et si la vérité des idées n'est pas clairement établie[7].

Ainsi la défense de critères précis et communicables est ce qui préside à la constitution du formalisme leibnizien contre l'intuitionnisme cartésien. Selon Belaval, «l'intuitionnisme invoque l'évidence qui *doit* éclairer le sujet connaissant, quand il le considère un objet 'comme il faut': le formalisme leibnizien demande ce que signifie 'doit' ou 'comme il faut', et de quelle manière cette connaissance immédiate sera médiatisée pour autrui»[8]. Le formalisme est donc garant de la médiatisation: c'est ce qui préside à la volonté d'utiliser des symboles extérieurs, des caractères, comme support de la pensée, et qui donc nécessite la mise en place d'une caractéristique universelle. C'est pourquoi Leibniz développe, contre l'intuition cartésienne, le principe de pensée aveugle (*cogitatio caeca*), c'est-à-dire d'une pensée qui n'est pas guidée par une vision, une intuition d'une vérité *index sui,* mais qui au contraire avance sans le recours d'une évidence subjective, uniquement guidée par la manipulation de signes. La caractéristique est par exemple ainsi à même de déjouer les ruses du malin génie cartésien.

Quoi qu'il en soit, le contenu de notre étonnement est le suivant: le dogmatisme de Leibniz n'en est pas un; si l'ensemble de ses édifices formels tendent à établir des critères de certitude non ambigus, son système est fondamentalement orienté vers l'invention et la production de nouvelles idées. A l'*ars demonstrandi* – ou *ars judicandi* – s'ajoute un *ars inveniendi*. Tout l'objectif de son formalisme est précisément de faciliter l'invention et sa critique de Descartes repose sur le fait qu'il ne permet pas véritablement de développer cette capacité d'invention. Ainsi, comme le souligne Michel Serres, alors que «l'idéal gnoséologique et méthodique du classicisme tient dans le mot certitude, le même idéal, chez Leibniz, tient dans le mot établissement»[9], c'est-à-dire que ce qui est premier relève de la logique de l'invention, de l'acquisition, de la découverte. La quête de la certitude, chez Leibniz, n'équivaut pas à une régression vers un point ferme mais pauvre en teneur de vérité, le solipsisme du *cogito*, par exemple. Elle va de pair avec une heuristique formelle, et c'est en ce sens que

[7] G. W. Leibniz, *Opuscules philosophiques*, 'Méditations sur la connaissance', Paris 1954.
[8] Y. Belaval, *Leibniz critique de Descartes*, cit., p. 35.
[9] M. Serres, *Le Système de Leibniz*, *cit.,* p. 215.

Leibniz nous invite à penser son ultra-rationalisme autant comme un processus d'ouverture que de fermeture et ainsi à décaler notre regard sur les fondements de la mathématisation du raisonnement au 17ème siècle. Celle-ci est loin d'être une fermeture de la pensée, fondée sur des critères de certitude et gouvernée par la logique du même et de la répétition, comme l'affirme par exemple Heidegger, qui voit dans la pensée de Leibniz l'accomplissement de celle de Descartes et le règne d'un principe d'identité mortifère. Au contraire, le rationalisme leibnizien procède suivant une logique de la recherche de l'altérité et vise à l'organisation d'une rencontre avec ce que nous ne connaissons pas encore. Ainsi le projet rationaliste est ambivalent depuis son origine.

Cette ambivalence du rationalisme est également une ambivalence du mécanisme. Le modèle de la méthode leibnizienne est le mécanisme: nombreux sont les passages où il compare la caractéristique à une machine et il semble que l'on puisse dire, avec Couturat, que la réduction de la Logique à la Mécanique[10] est un de ses objectifs finaux. Ainsi, si la logique de Leibniz se présente comme une machine, l'ambivalence qui la caractérise, puisqu'elle est à la fois *ars judicandi* et *ars inveniendi*, s'applique également au mécanisme. Contrairement à une vision critique du mécanisme et des machines comme enfermement dans une logique de la répétition, la machine est productive chez Leibniz, elle se constitue suivant une volonté de création ou du moins de production du nouveau et comme support de la logique de l'invention. Le mécanisme n'est pas productif par accident ou par destination, il l'est par essence, c'est inhérent même au projet qui préside à sa constitution. L'*ars inveniendi* leibnizien nous montre donc que le mécanisme comme modèle de la pensée est dès l'origine un dispositif tourné vers l'altérité et vers le nouveau. Le mécanisme ne fait pas que procéder à une volonté réductionniste de détermination fermée des critères de l'assurance mais ouvre au contraire l'invention le plus largement possible, en la sortant de l'enclos du génie individuel.

2. La logique leibnizienne: une machine baroque

Le deuxième étonnement qui naît à la lecture des textes logiques de Leibniz provient de la profusion et du caractère presque extravagant des projets leibniziens. Cet étonnement rejoint, dans le domaine de la logique, celui qui a présidé à l'analyse deleuzienne de l'œuvre métaphysique de Leibniz. Il la décrit en effet, dans *Le Pli*[11], comme étant caractéristique de l'esprit baroque, dans son contenu bien sûr, mais également dans sa forme et la richesse de son invention

[10] L. Couturat, *La Logique de Leibniz,* Hildesheim 1961, p. 115: «Il allait plus loin encore, et rêvait de réduire la Logique à la Mécanique. Cela n'a rien d'étonnant, si l'on se rappelle, d'une part, tous les passages où il compare le raisonnement à un mécanisme ou la Caractéristique à une machine, et si l'on considère, d'autre part, qu'il avait inventé dès sa jeunesse une Machine arithmétique pour effectuer les quatre opérations et une Machine algébrique pour résoudre les équations».
[11] G. Deleuze, *Le Pli*, Paris 1988.

conceptuelle, alors même qu'elle semblait devoir être empreinte d'une austère rigueur classique. Dans le domaine de la logique, il est difficile de ne pas être pris de vertige devant les différentes entreprises leibniziennes, toutes démesurées, toutes inachevées et qui toutes s'impliquent, depuis la combinatoire jusqu'à la science générale. L'idée première de la combinatoire, dans *De Arte combinatoria*, est la suivante: «tous les concepts doivent être résolubles en concepts simples, par une analyse analogue à la décomposition des nombres en facteurs premiers; et ils peuvent tous, inversement, être obtenus et composés par la combinaison progressive de ces concepts simples»[12]. C'est à partir de cette idée de combinatoire que Leibniz fonde son projet de langue universelle, c'est-à-dire d'une langue qui permettrait de détailler tous les concepts et de les réduire en éléments simples. Cette langue devait donc contenir les idées simples mais également une grammaire relationnelle générale. Mais demeurent les ambiguïtés et les singularités des langues naturelles, qui touchent même le latin simplifié qu'il envisage un moment comme candidat pour remplir cette fonction de langue universelle. Le projet leibnizien évolue donc et quitte les rivages de la langue universelle pour aborder ceux de la caractéristique, i.e. une idéographie, qui aurait pour rôle de remplacer les concepts par des signes combinables. L'arithmétique et l'algèbre sont, aux yeux de Leibniz, des premiers échantillons de cette caractéristique, qui montre qu'elle est possible et déjà en partie réalisée. Le calcul infinitésimal lui-même, dont Leibniz est un des inventeurs, est pris par lui comme exemple de la fécondité du symbolisme et d'une application de la caractéristique[13]. Mais la caractéristique nécessite que soit mis en place un alphabet des pensées humaines, c'est-à-dire que toutes les connaissances et concepts soient réduits en éléments simples, au moyen de la définition. Cela revenait donc à «faire l'inventaire des connaissances humaines, et même, comme l'analyse des concepts est en même temps l'analyse des vérités, à démontrer toutes les vérités connues en les ramenant à des principes simples et évidents, c'est-à-dire à édifier une Encyclopédie démonstrative»[14]. Le projet de la Caractéristique universelle nécessite donc, pour être mis en place, que soit constituée une Encyclopédie générale, c'est-à-dire un ensemble de définitions de tous les concepts fondamentaux des diverses sciences. Leibniz a tenté de faire financer ce projet démesuré à diverses reprises mais n'y est pas parvenu. Il nous reste les brouillons d'un ouvrage, les *Initia* et *Specimina Scientieae generalis*, qui était censé contenir les principes de la méthode générale censée servir à l'élaboration de l'encyclopédie ainsi qu'un résumé encyclopédique de trois ou quatre sciences spécifiques. L'encyclopédie générale est demeurée un obstacle pour le projet de caractéristique et donc d'*ars inveniendi*. Et d'autant

[12] L. Couturat, *La Logique de Leibniz*, cit., p. 49.
[13] «Il importait, pour faire ressortir l'unité de l'œuvre philosophique et scientifique de Leibniz, de montrer que son invention mathématique la plus célèbre et la plus féconde, celle qui a révélé son génie et consacré sa gloire aux yeux des savants, se rattachait dans sa pensée à ses recherches de Logique, et n'était pour lui qu'une application ou une branche particulière de sa Caractéristique universelle.» (L. Couturat, *cit.*, p. 87).
[14] L. Couturat, *La Logique de Leibniz,* cit., p. 80.

plus que l'encyclopédie implique encore un autre édifice conceptuel, celui de la science générale: il s'agit justement de ce qui est esquissé dans la première partie des *Initia et Specimina scientiae generalis*, à savoir la détermination d'une méthode universelle applicable à toutes les sciences. Cette méthode, qui est au fond la logique générale de Leibniz, permettra d'établir la certitude et de progresser dans les sciences.

Ce passage en revue rapide des différents éléments de l'architecture logique leibnizienne frappe par le foisonnement et la complexité qu'il révèle ainsi que par le caractère circulaire des différents projets qui sont échafaudés. La question de savoir quelle est la place exacte de l'*ars inveniendi* dans ces différentes constructions, au-delà de son caractère d'objectif principal, fait ainsi l'objet d'interprétations diverses[15]. Au-delà de ce caractère de complexité, c'est l'aspect 'baroque' et débordant de cette entreprise qui est source d'étonnement. Le projet formaliste et dogmatique s'appuie sur un empilement de projets extrêmement ambitieux, qui n'arrivent pas à terme et qui témoignent d'un optimisme de la volonté et de la raison qui paraît pour le moins déconcertant au regard contemporain. Il ne s'agit pas de réduire cette impression de bizarrerie à un anachronisme. S'il est évident que ces projets semblaient infiniment moins extravagants au 17è siècle qu'ils ne le seraient aujourd'hui – l'idée de constituer une langue universelle était ainsi relativement partagée à cette époque[16] –, il semble tout de même nécessaire de prendre l'étonnement que nous provoque cet empilement de projets et les fantasmes qui y sont attachés comme une indication sur la signification même du rationalisme leibnizien et du projet rationaliste en général. On peut ainsi préciser le rapport entre le logicisme leibnizien et le mécanisme: chacun de ses projets logiques apparaît en fait comme une sorte de machine spécifique et l'édifice entier comme un empilement de machines, une sorte de machine baroque géante, qui aurait vocation à digérer toutes les connaissances du monde pour en produire de nouvelles.

3. Le sujet entre contrôle et dépossession

Le troisième étonnement que peut susciter la fréquentation des textes de Leibniz est directement issu des deux premiers. Il concerne, cette fois encore, l'ambivalence de la tentative leibnizienne de confier le soin de l'avancée de nos connaissances à un système formel, non ambigu et prenant le calcul pour méthode et pour modèle. Ce projet a tout d'abord pour objectif, évidemment, d'assurer la puissance du sujet connaissant ; il s'agissait d'affirmer, contre la partition scolastique qui réservait la certitude aux mathématiques et considérait

[15] Voir, sur ce sujet, B. S. Hourya, *'Ars Inveniendi' Aujourd'hui. Les études philosophiques*, Paris 1989, *Leibniz*, pp. 201-214 et D. Rabouin, *Sur la mathesis universalis chez Lei*bniz, «Quaderns d'Història de l'Enginyeria», 16, 2018.

[16] Ainsi différents savants de l'époque, et Descartes lui-même, ont tenté d'ébaucher des langues universelles ou du moins en soutenaient le projet, qui s'intégrait très logiquement dans une volonté rationaliste de communication universelle.

que les disputes de l'École ne pouvaient relever que du probable, que le caractère apodictique des mathématiques pouvait s'appliquer à l'ensemble de la connaissance. Le sujet connaissant n'est pas forcément individuel d'ailleurs, et c'est un autre point qui oppose Descartes et Leibniz. Chez Descartes, la certitude est l'aboutissement et le principe d'un trajet individuel de connaissance, qui fait fond sur une rupture, jusqu'alors inégalée, avec l'histoire de la pensée qui l'a précédée[17]. Au contraire, chez Leibniz, le sujet connaissant est davantage collectif et il n'existe chez lui aucun mépris pour l'érudition et les connaissances passées; au contraire celles-ci constituent le cœur de la machine à inventer qu'il tente de constituer, notamment via l'instauration de l'encyclopédie générale des connaissances. Nonobstant ces différences, l'œuvre de Leibniz, de la même manière que celle de Descartes, apparaît comme un des aboutissements, dans l'histoire de la philosophie, de la prétention rationnelle à connaître le monde, à lui donner sens, à le saisir dans son entièreté et, *in fine*, à le mieux contrôler. Néanmoins, il apparaît dans le même temps – c'est la source de notre étonnement – que cette entreprise conduit à une forme de dépossession de la subjectivité vis-à-vis de la pensée. L'anti-intuitionnisme de Leibniz, que nous avons évoqué plus haut, ainsi que son refus de s'en remettre aux aléas de l'*ingenium* individuel et aux secours de la fortune, le conduisent à imaginer des systèmes visant à mieux répartir les pouvoirs de la rationalité et à en assurer davantage l'efficacité ; par là-même il dessine en fait les linéaments d'une forme de déprise de la rationalité individuelle sur les moyens qui lui assurent son pouvoir. Au moment même où s'opère, grâce aux lumières de l'évidence mathématique, une rupture avec la scolastique et ses jeux formels – comme ceux du système de Raymond Lulle par exemple[18], accusé par Descartes d'être une sophistique – se reconstitue, dans le rationalisme leibnizien, une autre forme de dépossession, celle de la machine rationnelle. La notion même de pensée aveugle contient cette ambiguïté, entre affirmation du pouvoir du sujet et dépossession: le caractère aveugle de la pensée, qui procède par manipulation de symboles formels, permet à la fois que tout le monde puisse penser correctement, et en même temps que les symboles pensent, d'une certaine manière 'à la place' du sujet, sans besoin qu'il n'ait l'intuition de ce qu'il pense. A cela on pourrait répondre que l'arithmétique et l'algèbre, pris par Leibniz comme exemples de sa caractéristique, nécessitent que le sujet s'en empare et, sinon du génie, du moins une habileté importante[19]. La mise en place

[17] Comme le rappelle Yvon Belaval: «Aucun auteur avant Descartes n'a osé pousser aussi loin la rupture avec le passé sous toutes ses formes: le passé de l'humanité dans la tradition et l'histoire de la philosophie ; le passé personnel, que des sens imparfaits, un instinct aveugle, des nourrices impertinentes, des précepteurs qui sont pas toujours les plus capables ont nourri de préventions» (Y. Belaval, *Leibniz critique de Descartes*, cit., p. 99).

[18] Raymond Lulle était un philosophe et théologien du 13è siècle, qui avait mis en place une machine visant, à partir de quelques éléments simples (théories, sujets, prédicats), à démontrer des propositions théologiques, notamment celles visant à affirmer la supériorité de la foi chrétienne, et dont il se servait dans les croisades solitaires qu'il entreprit afin de tenter de convertir Juifs et Musulmans.

[19] C'est d'ailleurs ce qu'affirme Belaval, contre le «préjugé selon le formalisme aboutirait à des résultats, comme en tournant une manivelle» et contre l'idée que la logique formelle met en

d'une caractéristique universelle ne signifierait donc pas une dépossession du sujet connaissant. Néanmoins on peut tout de même émettre l'hypothèse que l'anti-intuitionnisme de Leibniz et son idéal combinatoire et mécaniste tendent à faire accroître l'idée selon laquelle «raisonner machinalement, c'est bien l'idéal de la logique scolastique, comme de toute logique formelle»[20]. En ce sens, et sans explorer plus avant les très grandes différences entre la logique scolastique et la logique de Leibniz, il semble que l'idéal d'une pensée où la subjectivité du sujet connaissant joue un rôle congru soit tout à fait au cœur de l'idéal gnoséologique leibnizien. Ainsi, puissance du sujet connaissant et dépossession de celui-ci semblent aller de pair dans l'instauration du formalisme rationaliste.

A partir de ces trois étonnements, nous pouvons donc dessiner les contours de ce qu'on pourrait appeler les ambivalences du rationalisme leibnizien: 1) un système fondé sur la quête de critères de certitude qui est en même fondamentalement orientée par la recherche du nouveau; 2) un système où la logique joue un rôle premier mais qui se caractérise en même temps par une accumulation d'édifices conceptuels à l'ambition inouïe, voire insensée; 3) un système qui vise à assurer le pouvoir du sujet connaissant et qui en même temps organise d'une certaine manière sa dépossession.

Ces trois ambivalences qui parcourent l'œuvre logique de Leibniz sont à notre sens, c'est une hypothèse qu'il faudrait davantage étayer, caractéristiques du rationalisme logiciste en général. Il pourrait être possible de produire une histoire de ces ambivalences et de leurs différentes incarnations, dans le positivisme logique, dans la cybernétique, dans les travaux logiques de Gödel et de Turing… A chaque fois il semble que la notion de machine créative et d'heuristique formelle peut permettre de penser à nouveaux frais ces différents moments conceptuels. Il semble à cet égard que l'intelligence artificielle soit aujourd'hui l'aboutissement de ce type de rationalisme et de cette trajectoire spécifique de la pensée occidentale. Le détour par Leibniz permet donc de mieux comprendre le sens de ces dispositifs et ce que recèle ce projet, lui aussi extravagant à première vue, qui consiste à tenter de (re)produire la pensée via des supports formels et matériels.

4. L'intelligence artificielle au prisme des ambivalences du rationalisme

Ces trois ambivalences semblent en effet pouvoir tout à fait s'appliquer à l'intelligence artificielle et l'analyse de l'œuvre de Leibniz nous permet d'éclairer d'une lumière nouvelle l'ambition de ce projet.

place une forme de raisonnement machinal: «Certes, nous préférons qu'on fasse appel à notre génie plutôt qu'à notre habileté. Mais personne n'ignore que, si une solution algébrique est plus facile qu'une solution arithmétique, l'algèbre, cependant, ne dispense pas d'ingéniosité et, à l'occasion, de génie» (Y. Belaval, *Leibniz critique de Descartes*, cit., p. 35).

[20] J. Laporte, *Le rationalisme de Descartes,* Paris 1950, p. 22.

Les dispositifs d'apprentissage machine se développent aujourd'hui suivant un objectif principal, qui est la gestion de l'incertitude. La collecte de données massive, alliée aux nouveaux algorithmes d'apprentissage et à une puissance de calcul en expansion constante, se développent suivant une promesse: apporter de la certitude là où les autres instruments de prévision, tels que les théories scientifiques, échouent. Ainsi est-il promis de pouvoir prévoir et agir sur les comportements humains, qui échappent largement à la modélisation scientifique, du moins au niveau des trajectoires singulières. Un des fronts théoriques de l'intelligence artificielle est d'ailleurs l'apprentissage auto-supervisé, dont le développement a pour objectif principal de faciliter la prévision et la gestion de l'incertitude dans les comportements. Dans les différents secteurs où elle est utilisée, la prédiction – en plus de l'automatisation – est une des fonctions principales de l'intelligence artificielle: maintenance prédictive, médecine prédictive, conduite autonome… De la même manière, les algorithmes d'intelligence artificielle sont utilisés pour capturer ce qu'on appelle les 'signaux faibles' et permettre de concevoir ce qui échappe aux modèles traditionnels: les comportements singuliers, à risque, spécifiques… La gestion des probabilités et même plus généralement des 'possibilités' est ainsi au cœur du fonctionnement de l'intelligence artificielle, au point même, selon Louise Amoore, d'en constituer la signification politique principale[21]. Gestion de l'incertitude, prévision de ce qui nous est encore inconnu pour mieux le maîtriser, développement d'un savoir probabiliste: l'intelligence artificielle revêt les caractéristiques du rationalisme classique.

Néanmoins, les dispositifs d'IA sont également largement orientés vers la production du nouveau: les algorithmes dits 'créatifs', tels que les algorithmes génératifs adversariaux ou les algorithmes d'apprentissage par renforcement sont de plus en plus utilisés pour produire des formes et contenus nouveaux, que ce soit dans un objectif purement esthétique ou artistique[22] ou pour des objectifs de recherche scientifique par exemple. Les algorithmes d'IA sont ainsi aujourd'hui largement des machines 'créatives', ou du moins génératives, en tous les cas orientées vers la production du nouveau, que ce soit leur objectif premier ou que ce soit pour servir un autre objectif. De manière générale, un certain nombre d'interrogations portent sur la capacité de l'IA à produire des ruptures significatives, dans le domaine politique[23], théorique ou esthétique. Il semble donc que les algorithmes d'IA reproduisent la même ambivalence que celle que nous avons décelée dans le travail de Leibniz entre quête de la certitude et production du nouveau.

[21] L. Amoore, *The Politics of Possibility*, Durham 2013.
[22] Voir, par exemple, A. Elgammal, L. Bingchen, M. Elhoseiny, M. Mazzone, *CAN: Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms (on line,* arXiv:1706.07068).
[23] Cf. D. Bates, *L'IA peut-elle produire de nouvelles normes?* Intervention lors du colloque sur l'intelligence artificielle qui a eu lieu à l'Ecole polytechnique (Palaiseau) en septembre 2021 (intervention est disponible en vidéo sur la chaîne Youtube de l'École polytechnique).

De la même manière, l'IA est également prise entre d'un côté son caractère de logicisme achevé, c'est-à-dire de domination absolue du calcul et de la rationalité, et de l'autre une logique fantasmatique et presque délirante. Cette ambivalence est largement documentée et commentée: l'histoire même de l'intelligence artificielle, alternance entre 'printemps' et 'hivers', témoigne du rôle qu'ont joué les projets extrêmement ambitieux dans la constitution de ce champ de recherche. Ainsi de l'intelligence artificielle symbolique et des cathédrales sémantiques qu'elle a tenté de bâtir: par exemple le projet CYC, qui avait pour objectif de décrire les relations entre l'ensemble des objets du monde à partir d'une architecture de «prédicats fondamentaux», de «fonctions de vérité» et de «micro-théories»[24].

Enfin, l'ambivalence entre volonté de contrôle et dépossession est sûrement un des éléments les plus présents au cœur du projet de l'intelligence artificielle. Les dispositifs d'IA sont en effet l'aboutissement d'une volonté de maîtrise, de circonscription du monde dans la sphère calculatoire, voire de redoublement de celui-ci par l'accumulation de données. L'automatisation, l'orientation des comportements, la volonté de prévoir même les événements les plus chaotiques sont évidemment au cœur du fonctionnement de l'IA et de son ambition. Ce sont d'ailleurs ces aspects de l'IA qui lui valent le plus de critiques, dans la mesure où elle incarnerait l'aboutissement d'une modernité où la rationalité calculatoire l'a emporté sur toutes les autres formes d'appréhension du monde. L'intelligence artificielle semble ainsi avoir pour fonction d'assurer un contrôle et une domination du sujet connaissant sur les choses qui l'entourent. Néanmoins, là encore, les dispositifs d'intelligence artificielle ont des effets ambivalents: en effet, l'IA est également un facteur de déprise et de dépossession du sujet de sa propre capacité à décider. Rien d'original à cette idée, c'est même une des appréhensions les plus spontanées et les plus communes de l'intelligence artificielle, comme en témoignent un certain nombre d'œuvres de fiction à propos de la perte de contrôle de l'être humain sur les artefacts qu'il a lui-même fabriqués. Au-delà du bien-fondé de ces peurs et de ces fantasmes[25], il semble tout de même nécessaire de noter qu'elles expriment cette dualité de l'intelligence artificielle entre absolu contrôle et voie vers la dépossession, dualité qui est présente, comme nous l'avons montré, depuis le début du projet rationaliste.

5. L'intelligence artificielle et l'inconnu: une ambiguïté fondamentale

Ainsi les trois ambivalences que nous avons relevées dans le travail de Leibniz caractérisent également l'intelligence artificielle contemporaine. Au fond

[24] Voir, à ce sujet, D. Cardon, J.-Ph. Cointet, A. Mazières, *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle*, «*Réseaux*», 2018/5 (n° 211), pp. 173-220.
[25] Voir, à ce sujet, J. G. Ganascia, *Le Mythe de la singularité*, Paris 2017.

il semble qu'elles témoignent d'une ambiguïté fondamentale dans le rapport entre l'intelligence artificielle et l'inconnu: d'un côté, les dispositifs d'IA ont pour objectif et fonction première de maîtriser l'inconnu, de le limiter, de le contenir; et d'un autre, ces dispositifs sont tournés de manière fondamentale, en intention du moins, vers la production de l'inconnu, au sens de ce qui n'est pas encore connu. Comment interpréter alors ce qui paraît être un paradoxe: un outil de réduction de l'inconnu qui vise à la production de l'inconnu, sous la forme du nouveau? Il semble que l'on puisse tenter de comprendre ce paradoxe de deux manières.

Tout d'abord en affirmant que ce paradoxe n'en est pas un. Ce que viseraient les dispositifs d'IA c'est toujours à accroître la sphère du connu, à ne rien laisser en-dehors de celle-ci. L'inconnu et le nouveau que produisent les 'machines créatives' n'est que du 'pas-encore-connu' qui devient du connu à partir du moment où il est produit. C'est de l'inconnu qui ne peut se présenter que comme du connu. En ce sens c'est un inconnu contrôlé, parce que calculable, c'est-à-dire issu de la combinaison de choses déjà connues. L'*ars inveniendi* leibnizien est ainsi, *in fine*, tout entier issu des premières inspirations de la combinatoire. L'invention se détermine soit par analyse, c'est-à-dire par décomposition d'un élément en ses éléments premiers, soit par synthèse, c'est-à-dire par combinaison de différents éléments premiers pour en bâtir d'autres. A cet égard, l'inconnu procède toujours du connu, il est contrôlé et il ne peut au fond y avoir de 'sortie' hors de la logique combinatoire, c'est-à-dire d'irruption d'un 'véritable' inconnu. Cette lecture des rapports entre connu et inconnu dans la logique combinatoire et l'intelligence artificielle a dominé, semble-t-il, les pensées critique du calcul au 20è siècle[26].

Mais il existe une autre manière de comprendre ce paradoxe. Il s'agit pour ce faire de souligner que chez Leibniz, un 'véritable' inconnu, qui se situerait hors de portée de la combinatoire, n'existe pas: en effet, les principes qui gouvernent sa logique reflètent ceux qui gouvernent sa métaphysique. L'idée de la combinatoire, c'est-à-dire l'idée suivant laquelle tous les éléments se décomposent en parties simples, qui forment à nouveau des éléments complexes, rejoint en effet son principe métaphysique fondamental, qui est que toute proposition vraie est analytique, c'est-à-dire que le prédicat y est contenu dans le sujet[27]. Les accidents de la substance n'en sont pas chez Leibniz et appartiennent, en vertu de l'absoluité du principe de raison, à l'essence de la substance. Ainsi, pour Leibniz, tout événement est une expression singulière de ce qui est déjà contenu dans les substances individuelles: c'est un 'dépliement', pour reprendre un terme deleuzien.

[26] Il semble que l'on puisse en trouver les traces, dans des registres tout à fait différents, dans les pensées de Martin Heidegger, de Theodor W. Adorno et d'Henri Bergson.

[27] Comme l'explique Belaval, les liens entre la logique de Leibniz et sa métaphysique sont très forts: «[...] l'un des principes fondamentaux – peut-être le principe fondamental – de la technique opératoire de Leibniz *in omni propositione affirmativa vera, praedicatum inest subjecto,* nous introduit de plain-pied de la logique formelle à la métaphysique, dans la correspondance avec Arnaud» (Y. Belaval, *Leibniz critique de Descartes*, cit., p. 37).

De la même manière, la combinatoire et l'*ars inveniendi* tente de procéder à des 'dépliements' à partir des notions premières. Ainsi on peut émettre l'hypothèse que chaque substance individuelle est une forme de 'machine créative', c'est-à-dire que la nouveauté de ce qui arrive dans le monde était déjà contenu dans les principes premiers des substances qu'il s'agit de déplier, de dérouler, de combiner. Le monde lui-même est issu d'un calcul de la part de Dieu, pour Leibniz: Dieu tente en effet de trouver l'équilibre compossible le plus optimal entre les diverses substances individuelles et ce qu'elles contiennent. L'harmonie qui définit le monde comme meilleur des mondes possibles est donc le résultat d'un calcul et l'on peut concevoir le monde lui-même comme une machine créative géante chez Leibniz, empilement baroque de différentes machines créatives. Ainsi, le nouveau, chez Leibniz, est toujours un possible du calcul. Il n'existe pas d'autre type de nouveauté. Le 'reste' du calcul, son dehors, réside simplement dans l'ambiguïté et dans l'incompréhension des principes qui règlent le monde. Mais sur le plan ontologique, cette ambiguïté et cette incompréhension n'ont pas de valeur d'existence et ne constituent pas un 'autre' de la rationalité. Il serait donc, semble-t-il, erroné d'opposer un 'véritable' inconnu à un inconnu contrôlé et combinatoire, puisque l'ensemble de la nouveauté du monde et de ce que nous ne connaissons pas encore est lui-même déjà contenu, sur le plan métaphysique, dans ce que nous connaissons déjà.

Ainsi on peut émettre l'hypothèse que l'élaboration d'un *ars inveniendi* vise à reproduire, à reconstituer, avec les outils de la logique, le mode de genèse du nouveau qui caractérise le monde comme machine créative générale. Si le monde en effet est une machine créative générale, comme semble le concevoir Leibniz, on ne peut concevoir la production du nouveau qu'en fabriquant des machines créatives spécifiques. Pour en revenir à l'intelligence artificielle, on peut donc affirmer que si elle constitue un aboutissement contemporain du projet rationaliste, c'est précisément au sens où elle porte cette ambivalence fondamentale qui le caractérise vis-à-vis de l'inconnu. Dans un monde où tout est conçu comme calculable, représentable, maîtrisable, c'est-à-dire où toute la nouveauté peut faire l'objet d'un calcul et que l'inconnu n'existe que comme frontière temporaire et non comme principe externe, le véritable inconnu ne peut au fond provenir que d'une instance de calcul. Si l'intelligence artificielle exerce tant de fascination, de peur et fait l'objet de tant de fantasmes, c'est précisément qu'elle incarne cet espoir du projet rationaliste, espoir qui lui est consubstantiel depuis le début: que les outils mêmes du calcul, c'est-à-dire de la réduction de l'incertitude et de l'inconnu, puissent produire un inconnu radical, une nouveauté inouïe et qu'émerge quelque-chose qui nous échappe à partir même de l'endroit de notre absolue maîtrise. Dans un monde où la pensée de Leibniz constitue la 'métaphysique de l'époque', un monde d'immanence pure, où le calcul est l'autre nom de Dieu, l'altérité radicale ne pourrait surgir, au fond, que de nos propres instances de calcul, d'une émergence organisée par nous-mêmes et qui nous libérerait, pour ainsi dire, du poids de notre propre maîtrise.

François Levin
École Polytechnique
✉ francoislevin01@gmail.com

Bibliographie

Sources primaires

Heidegger, M., 2003, *Le Principe de raison*, Paris, Gallimard.
Leibniz, G.-W., 1954, *Opuscules philosophiques*, Paris, Hatier-Boivin.
Deleuze, G., 1988, *Le Pli*, Paris, Les Éditions de Minuit.
Amoore, L., 2013, *The Politics of Possibility*, Durham, Duke University Press.
Elgammal, A., Bingchen, L., Elhoseiny, M., Mazzone, M., *CAN: Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms,* arXiv:1706.07068.
Bates, D., 2021, *L'IA peut-elle produire de nouvelles normes?* Intervention lors du colloque sur l'intelligence artificielle qui a eu lieu à l'Ecole polytechnique (Palaiseau) en septembre 2021 (en vidéo sur la chaîne Youtube de l'École polytechnique).
Cardon, D., Cointet J.-P., Mazières A., 2018, *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle, Réseaux*, 2018/5 (n° 211), p. 173-220.
Ganascia, J. G., 2017, *Le Mythe de la singularité*, Paris, Seuil.

Sources secondaires

Belaval, Y., 1978, *Leibniz critique de Descartes*, Paris, Gallimard.
Serres, M., 1968, *Le Système de Leibniz*, Paris, PUF.
Couturat, L., 1961, *La Logique de Leibniz,* Hildesheim, Georg Olms Verlag.
Hourya B. S., 1989, '*Ars Inveniendi' Aujourd'hui* in *Leibniz,* Paris, PUF.
Rabouin, D., 2018, *Sur la mathesis universalis chez Leibniz,* in Quaderns d'Historia de l'Enginyera, vol. 16.
Laporte, J., 1950, *Le rationalisme de Descartes,* Paris, PUF.

¶ Sezione Seconda

Teoria ed epistemologia dell'algoritmo

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7768566

Interviste/1

On Algorithms: Ethical and Epistemological Questions
Interview with Igor Pelgreffi

Luciano Floridi 0000-0002-5444-2280

Luciano Floridi tackles the topic of the algorithm and the ethical and theoretical challenges it entails, starting from a survey of some of his ideas, including the concepts of *Fourth revolution*, of *Infosphere* and of *On-life* (the latter indicating an extended condition of close interconnection between our life and the being online situation). Firstly, Floridi analyses the concept of algorithmic dependency on the complex data/software, with related responsibilities and risks. Then he moves on to the question of Machine Learning, recalling both the 'historical' theme of the so-called 'machine to machine interaction' and items from game theory frame. Interaction means not only sharing data but also modifying data: Floridi underlines therefore the role of human subject and of ethical responsibility in data managing, instead of the 'machines hyper-autonomy' perspective. Subsequently, Floridi interprets the meaning of the various problems raised here, regarding the algorithm and its forms, by retracing his main works: *The philosophy of information* (2011), *The ethics of information* (2013) and *The logic of information* (2019), thus also discussing the overall meaning of his research path. He also reflects on the problem of writing about the politics of information and, in perspective, of elaborate a reflection on 'meaning', that is on what it means to be human in a digital age. A philosophical thought around the issues of algorithmic era, if conceived in such a broad way, can therefore imply a profound and gradual change in our self-understanding, considering ourselves as 'informational organisms', like a slow evolution of the Aristotelian *zoon politikon*. Following Floridi, the ways in which the opportunities related to the digital society are spent will allow or not the informational organisms to produce the ultimate good, that is to say the creation of a good society.

***

Igor Pelgreffi – *Thank you for accepting this conversation, that would be on the themes of the algorithm and the problems it poses to us. We can also, if you believe, connect these issues to 'your' concepts, which are also very well known: I am thinking of the* fourth revolution*, of the* infosphere*, or even of* on-life*, the latter indicating an interconnection more and more extended between our historical condition of being-on-line and the vital or existential element of our life. So I come to my first question: can the algorithm be considered a* condition of possibility*, or even a* condition of existence*, of that all, that is of the infosphere, fourth revolution, on-life dimension? Of course, the algorithm is not the only condition of existence*

(there is also infrastructure, internet, mass mobile devices, etc.), but it is certainly one of the most relevant today. Here: how do you see the hypothesis according to which the algorithm is an element that supports *all this? An element that, at the same time, makes possible all this and also represents its* sustainability? *A sustainability, therefore, linked precisely to the properties, logical but also ontological, of the algorithm?*

Luciano Floridi – I will take a step back to approach your interesting question. The step back is the following. When we talk about Computer Science, AI, ICT, Digital Technologies, and the new wave of innovation, we may have two items in mind: data or software. And we know that we need both. It's a bit like cooking, you know: you need ingredients, and you need the recipes. Sometimes we emphasize the data rather than the software, almost as if the software were in the background doing all the job invisibly. But we should not forget that we need data all the way. We hear the metaphor 'data is the new oil', which I do not like. It is a metaphor particularly misleading, but it is very common. I prefer to think that data is the new gold, data are the new material, the new stuff that we manipulate in a digital society, almost as if software were there like a tool, something that you can not ignore as you work with these new resources.

Well, I think it is quite important, maybe as you were suggesting in your question, to realize that not only do we need both (because the one without the other makes no sense. Using the previous metaphor, it is like to say that with the engine without petrol or with the petrol without the engine: you do not go anyway). But also to realize that software has a support role that increasingly makes our society work.

The dependency that we are building with respect to software is growing daily. And it's a dependency that we do not notice: it works well, or even better, when you do not see it. If everything works smoothly, that's because the software, therefore the algorithm 'AI-oriented', is doing its job properly.

This reliance raises a couple of issues, if we have potential philosophical questions in mind. One is: how risky is this reliance on algorithms, as society becomes increasingly complex, and I mean *complex* in the technical sense, not just as complicated. Really complex: its elements are always intertwined with each other, and small local changes can reverberate and produce big differences in the whole system. I mean, and it is well known, that a system is complex since you cannot really imagine or calculate what can it happen to some corner of the system itself, when something really tiny occurs elsewhere. Now: as society becomes complex in that sense, increasingly, we need algorithms to run the system. But how risky is that? How much of our individual well-being and social welfare depends on the algorithm question?

To be honest, society has always moved towards interactions with complexity using more complexity. For instance, as we moved through urbanisation, we created more urbanization, as well as the car-based society, with more complexity in schedules, traffic, traffic control, traffic lights, if you like, etc. ... And again: more cars, more ways of handling cars, therefore more complexity and even

more complexity. Now, as complexity grows, we would like to see control, accountability, and transparency; all those items grow with complexity.

Thus, essentially we have two sides of the same point in front of us: as society becomes increasingly *algorithmic dependent* – data-dependent or if you like more generally data/software dependent, and therefore dependent on the digital – we should have a much better comprehension about what policies are in place to reduce the risks that all this implies, and also about the policies for better control, accountability, and transparency. Well, I think that in this condition, there should be some understanding of some risks concerning dependency on algorithms, specifically about what an algorithm or software could do when it goes wrong. Let us notice that it represents a completely different sense of risk. The risks we are talking about are not the risks when things go wrong, but when things break down. And it is a little bit like the risks you run when electricity is down, when there is a blackout: I'm not talking about being electrocuted or analogous damage. I'm talking about not having electricity in the house at all. That is the risk of dependency.

About that, I don't think there is enough sense of awareness, even if we have nowadays much more awareness of our dependency on energy sources. But for technologies is something different: depending more and more on them, we should have all the policies in place (resilience, for example, and so on). The same happens in terms of responsibility. We have a sense of responsibility that is almost based on the frame like 'when things go wrong', that is, when something is misused, overused or underused, for the wrong reasons.

In your questions, you said that there is an increasing dependence on algorithms, that possibly they support or sustain our society, so that the word sustainability can be paradoxically used here, but I think we have left the analysis and understanding of *good*, not evil risks. And we usually think the analysis and understanding of the policies for the governments of all this would be charged for the future generations, so we do not have to care about that, because one can think there are more pressing issues. On the contrary, this should be done today. I think that the risk management and the governments of the algorithms is an issue that we have to consider now, because it is now that we are building this dependency for the future.

Igor Pelgreffi – *Among other things, a final question of this conversation would have been about the future scenarios that you imagine. Particularly interesting, here, is the concept of 'functioning well', or 'too much well', which always accompanies that of 'functioning unwell': a duality that we could consider structurally linked to a certain logic of the algorithm, if you want.*

I take you now to the great subject of automatic learning *by an algorithm, that is, Machine Learning. Today the algorithm learns, and learns some kind of 'behaviour', but it can also work... badly: it is a remarkable boundary, even regaining from the discussion just made about risks. Could you tell me something about this*

connection between algorithm and learning? Does the machine learn? Can it make mistakes? Can it self-correct?

And, in connection to this, one more question: according to you, and if it is not just science fiction, is there a social dimension of machines? I mean: machines or algorithms that can 'communicate' with each other, that is, therefore algorithms that somehow learn but no longer individually but within a dimension social, even if the social term is probably too anthropocentric? What I am asking could seem extrapolated by a scenario from a science fiction film or novel, but in actuality we already have algorithms that feed themselves, 'nourishing' on data with increasing speed and acceleration, so much so that in some cases, it is no longer possible to have the technical control of the output.

Luciano Floridi – This is an important question. I think that what you called the 'social' aspect, I put into brackets the term 'social', I think that this is… anywhere… and it is not eccentric at all. I should bring a couple of examples, but as of now, I can remember that the problem of 'machine-to-machine interaction' has been with us since we started to use the first computers. Now, to anyone who has or uses a computer, probably the most elementary, almost trivial phenomenon or example of interaction, is the one you have between your computer and your printer: they need to talk with each other. In the past, we used to talk about so-called 'protocols of interaction' between two computers and their modem. That was called *handshaking*, speaking of metaphors: the systems were handshaking, finding a way of communicating through a protocol and sharing data between A and B, and B and A: between the two machines. It is just to say that all this has been with us since we started building the first computers. In fact, at the very beginning of the history of Computer Science – we are talking about the Fifties – the idea was that there would have been only a few immense computers and many tiny terminals having access to these mainframes. Therefore communication was everything, not least because maybe I needed to interact with another computer that was trying to access *the same* resource *at the same time*, in terms of scheduling access to the same resource, and that means having some sort of parallel, a-synchronic or synchronic access, and so on.

From a historical point of view, this was an old, big branch of Computer Science. Then, said briefly, things changed: we began living in a sort of 'do it yourself' dimension: my computer, your computer, my PC, your PC. A personal computer became *my own* machine, *your own* machine. Reaching today, in the actual stage of technical development, we are going back almost metaphorically to the mainframe idea of that sort of enormous computational power that is elsewhere. And, moreover, we were subjected to that computational power… So all this has been with us for some time.

The 'social aspects' of computation, and I put the 'social' into quotation marks, that is, metaphorically speaking, have received – for example, like in the communication I'm sharing with you now – a strong input from the well-known

aspect of artificial intelligence. Think of the so-called GAN, or Generative Adversarial Networks. It is a very common technology, no science fiction. It is used successfully in developing security systems, for example, the security of a website. These networks are 'challengers net': they actually play one against each other, literally. So: one tries to crash or to hijack the website, the other one is trying to defend it; one has no data, while the other is trying to use the inside data to defend: there is an attack, a defence, another attack, and so on, through a learning process. And there is a lot of what I would like to describe as *game theory* here, again: theory in a scientific sense.

So – going back to the previous question – we *support* a society increasingly through data and software, with software also being increasingly smart algorithms. It is also a world in which all of this is linked by significant interactions. Interactions are more than just sharing data, sending data, and receiving data. Interactions also mean modifying the data, generating new data and learning from the data and the interactions themselves, and therefore updating and upgrading the behaviour of the algorithms based on the interactions with other algorithms. And that is what happens in adversarial networks, for example. This is the future in the most ordinary sense, and it has nothing to do with Science Fiction, nor with the image of a few scientists that are dreaming in their office. Yet again: in a world where algorithms interact with other algorithms to learn from and improve their behaviour, the question about responsibility is: *who designs what* and *who enables what*. I mean: contrary to what happens in the animal world, algorithms do not interact with each other spontaneously, like cats and dogs fighting each other. No: you need to put them together, you need to make them communicate, you need to create the conditions for that particular interaction, and so on.

Now, it is this *design issue* that seems to me fundamentally human and, therefore, fundamentally ethical. For example: is it really a good idea to train an algorithm by interacting with another algorithm to make sure that it builds the best possible deepfakes? Or to enable an algorithm to learn how to exploit weaknesses in a credit card system? Once again: algorithms are playing with each other, interacting, improving, and then it becomes possible to hijack or crack the security codes for a credit card system or get my personal data. But the point is that the design of all this is in our hands: this is all about us. In other words, we entirely design the 'social' nature of algorithms: it is made possible by us. We are engineering these social interactions. And therefore, the responsibility for what kind of engineering is allowed, for what, whether we like it, etc., is 100% human. So that is the big difference about using 'social' with quotation marks. This is not the biological world, where animals or people socialise, but it is an artefact that we can build, that can or cannot, literally may or may not be enabled and allowed to do this or that.

Igor Pelgreffi – *Very clear. Thank you. Including the clear final ethical position you have just expressed...*

Luciano Floridi – Sure. Because it seems clear to me that, otherwise, we begin to think that technology really interacts only with technology, which I don't think. And then, at that point, we can also ask ourselves something like: 'what happens when you leave them alone'?

Igor Pelgreffi – *Yes, but they are still huge issues. I agree with you on this type of ethical requirement; however, it should still be remembered that the problem is* in re*: is the current reality. Just think of that branch of Computer Science that has recently emerged – for example, the so-called XAI (eXplainable AI) – which deals with algorithms that are 'intelligible', 'interpretable', or, more generally, 'explainable'. This need arises probably because algorithm programmers often report certain outcomes, actions or 'choices' of their algorithm (generated or written by them) that are not rationally explainable. But let's move on.*

It is interesting to briefly review your parable of studies on these issues, also in order to better understand your actual position. Your first books from the Nineties were more focused on logic and epistemology. I should remember: Scepticism and the Foundation of Epistemology. A Study in the Metalogical Fallacies *(Brill 1996);* Internet. An Epistemological Essay *(Il Saggiatore 1997);* Philosophy and Computing. An Introduction *(Routledge 1999). Gradually then the ethical aspect, perhaps present since then, seems to emerge more clearly, with consistency or, perhaps, urgency. I think of* Infosfera. Filosofia e Etica dell'informazione *(Giappichelli 2009) and many more books, including* The Ethics of information *(Oxford University Press 2011). In short: the ethical aspect has perhaps become paramount. I wouldn't say it denied them, but that it completed those early searches. And, in fact, it is also felt in this conversation...*

Luciano Floridi – Well… it is a very long project that I started developing a long time ago: decades ago now. When I embarked on this particular journey, initially, I thought I would publish three volumes. One on *the philosophy of information*, which would have been essentially logic, epistemology, metaphysic, etc.; one on *the ethics of information*; and the last one on *the politics of information*. One, two, three: ten years each, more or less, it means… thirty ears: that's my life! Unfortunately, things became a little bit more complicated. I realized that I needed much more space, also for publishing reasons: that book could not be a book, for Oxford University Press, of a thousand pages! So it ended up being volume one, *The philosophy of information* (Oxford University Press 2011), which is very much as you said, epistemology, so to speak, but also metaphysics. After that, I published volume two, *The ethics of information* (Oxford University Press 2013). But then, I realised that there was much more that needed to be done in terms of *The logic of information* (Oxford University Press 2019), which could become volume three [*laughing*]; but at that point, volume four, unfortunately... split into half. The reason for that splitting into two is because initially, I thought that *The politics of information*, which was supposed to be only a single book,

was going to discuss two forms of agency from the perspective of what can be called the digital revolution, that is the fourth revolution: agency as in socio-political agency (in other words the impact of the digital revolution on politics and society) and agency understood as 'artificial agency' (that is, understanding philosophically the ethics of artificial intelligence). Therefore, when I gave the *Ryle Lectures* at that time, I tried to keep agency at the centre of the socio-political and the artificial. But then I realised that it was just too much. I mean: the architecture on the volume was getting too complicated. And so, also speaking with the philosophy editor at OUP, I followed his recommendation to split the text into two parts. So this is just to say that volume four is now 4.A and 4.B [*laughing*], and volume 4.A is actually *The Ethics of Artificial Intelligence* (Cortina 2022), which I published in Italian and is coming out in English soon, I hope. I could say that now I have started the long way, the long journey towards *The politics of information*, volume 4.B, understood as the philosophy of the digital transformations being prompted by a different way of understanding political actions. It will take some time before I finish that book. Hopefully, there will be, after that, if I get lucky, I really hope to finish with a final book: number five, which is more a reflection on life's meaning, and in particular on what it means to be human in a digital age. At the moment, I'm calling it *The hermeneutics information* because I have no imagination! So we will have the philosophy of information, the ethics of information, etc., with, in the end, the hermeneutics of information that will hopefully also be the end of my intellectual journey [*laughing*]…

Igor Pelgreffi – *Thank you for this explanation around your path, or... your journey. In particular, the reference to hermeneutics seems to me to better clarify, at least in perspective, the ethical* quality *of your work: on the one hand, in fact, the use of an hermeneutics of information may integrate the perspective of the sign, of a linguistic and logical problem, and therefore, if you like, of the concept of interaction we have talked about, present since the beginning of your research path; but on the other hand it welds it more firmly to the theme of the human, of the return to the human, so to speak, within a hermeneutic instance of interpretation and domestication of these complex issues. Almost to limit the risks of them becoming divergent or hyperbolic...*

So let's go to the conclusion with one last question. The theme I am now presenting to you is really massive, so I can only ask you for a short answer that can show what is your basic position. The theme is that of corporeity. We know that algorithms (understood as a sequence of steps to solve a problem in a finite number of operations, etc.) represent an autonomous or abstract systematicity linked to the code, to the 'sign', to an impersonal protocol sequence of symbols, etc.; but we also know that nowadays algorithms tend to integrate themselves more and more deeply into living systems, in the psyche *or in our 'flesh'. Let us think about the theme of hybridization, if you want a reference: where does the algorithm start and where does it end, in my body? Where are our bodies – please note: not so much and not just the*

mind or the Subject, but the bodies – therefore, the attention, the psyche, our entire being onlife, *to take up one of your key-concept, are inside the living sphere, or the infosphere? In short: is the body really detached/detachable from the algorithm? Does it really hybridize, or is the so-called hybridization just... appearance?*

Luciano Floridi – Oh, this is another very interesting question. And I have to keep the answer short… In the past, I considered the body as an interface. I have not changed my mind. Maybe I will change it in the future, but at the moment, I still think that way. So: the body is an interface. This could be a synthetic answer to your very complex question. But what kind of interface? The interface between, on the one hand, the self, or the I, or the mind – depending on the philosophical orientation – the soul, even, for some people, our intelligence etc., something that is on this side; and, on the other hand, the bodies, other bodies, in other words, other interfaces, in the world. So, when we perceive the world, not only with our own interface, which is the body, we also perceive another interface, that is, the object which is on the other side. So imagine a conversation like the one we are having among us. What happens? It is me and my interface interacting with you and your interface. From that, you see that there are four items interacting: there is not only me and you, but there is me and my interface meeting, encountering, exchanging messages with you and, as well, your interface. This becomes obvious when someone has a different interface. I might be blind, and I have a different interface. Or I might be able to see or hear better than you, maybe using different interfaces. Although interfaces are also interfaces in terms of action, of what you can actuate using your body, so what you can interact with in the world: maybe I am strong, perhaps I am weak, maybe I am old, maybe I am young… And interfaces are also full of features: gender, colours (eyes, hair, skin), elements (legs, fingers etc.). Once you get the interface perspective, of course, this interface gets challenged when we shape it with the technologies we developed. But digital technologies not only operate on my body as an interface, but they also operate on my conceptual modelling of my body as an interface.

It is important, therefore, to remember that it is true that we have different abilities to work with the world, but it is also true that we conceptualise all this differently: to put it a little bit more strongly, what the digital does is to *re-ontologise* the body but also to *epistemologise our understanding* of the body. So, the digital does both, and sometimes it is in the asymmetric interaction between the re-ontologisation of the body and the re-epistemologisation of the body – from a digital perspective – that things do not quite work well. One thing I could mention, for example, is the *Quantified Self Movement*, that is – in short – people who are very keen to measure any bodily parameter, health information, and everything in their interaction with the world. So with their interface, wearable device etc., they may measure their heartbeats, breathing, the number of steps they take, calorie intake etc... For instance: 'how many steps I take', or even 'the level of red blood cells', sometimes even with chips under

their skin. Probably it is a small movement, up to now, and we do not have to emphasise it so much, but it offers some evidence of what kind of interactions we can have *ontologically but also epistemologically* with our body. We can speak, about this, in terms of *re-ontologization and re-epistemologisation* of the body. This means that one can see oneself as quantified minimally, for example, for the ten thousand steps you should take daily. The mechanism is that if I do not take the ten thousand steps, maybe I do not feel good with my self and with my body, at the end of the day. It is also self-training, therefore. By the way, you probably know that the question, the number of ten thousand steps, is not based on a scientific study; it merely started as a good round number in the Sixties, when it was advertised by a Japanese company Yamasa, which produced the first, successful step counter, called Manpo-kei, which literally means 10,000 steps in Japanese. Walking is good, but there is no reason to take that number of steps. But back to the main topic… The consequences are clear: these amazing technologies enable us to shape not only the world but also our interaction with the world, and all this happens through, and with, the interfaces we are, the body-interfaces through which we communicate and interact with the world. This may become an extreme thesis, which I do not like, things like virtual reality, new artificial skins, the idea that we can transfer minds through different bodies and so on. This is sci-fi. What is interesting for me is the cultural aspect: the deep, massive transformation in our self-understanding. That is, we are talking about a change in our philosophical anthropology: what does it mean to be human? What do we expect a human should and could be? So you see, here and again, the ethical implications of the whole question of algorithms and of their widespread diffusion, etc. All this will not happen one day to the next: it is a long, profound and gradual change in our self-understanding. In the same way as it took quite a long time since the agricultural revolution – many many years – to reach Aristotle and the social or political animal; and only after that stage the modern urbanization, for example, could take place: you don't have urbanization without a specific culture… But we should never forget that it takes millennia to get to the Aristotelian view, as a philosophical anthropology of the individual as *zoon politikon*, the political animal. Well, about all this, I think it will take quite a while for us to see ourselves as, as I like to call it, *informational organisms*. But that it's where we are going. And, again, we move through the flux of information, we exchange information, we can damage ourselves through the wrong use of information received, we can damage society through wrong interactions... but, apologies, we are running out of time … I better stop here with a final comment.

In everything I said, there is no attempt to develop a metaphysics, as if one could have access to the ultimate essence of the world. I don't know what that means. I'm not that kind of philosopher. I am much more a Kantian philosopher, in that sense. So: what is the world in itself? God knows, literary. I don't. We don't. But the world is conceptualized in a variety of ways, and today we are conceptualizing what we call 'ourselves' as *informational organisms*:

I mean, we are developing technologies that handle the informational side of those organisms. And once again, it becomes an ethical problem to make sure that we don't shape these informational organisms in the wrong way. At the end of the day, both education or *paideia*, and norms or *nomos*, are ways of shaping informational organisms in such a way that we could speak, in the end, of a 'good society'.

So, as philosophers, what can we do in order to prevent absolute and horrendous mistakes? That seems to me the pressing task in front of us. Once again, I think we have not thought enough about this, about the whole question of algorithms and their implications in the sense we said before. We should – as philosophers – focus on understanding more the digital revolution we are undergoing, to be able to design our future better.

Luciano Floridi
Università degli studi di Bologna
✉ luciano.floridi@unibo.it

Igor Pelgreffi
Università degli studi di Verona
✉ igor.pelgreffi@univr.it

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770017

Contributi/4

Dialettica dell'algoritmo
Per una teoria critica

Massimiliano Carli 0009-0008-8699-1998

Vittorio Lubrano 0000-0002-8096-1367

Articolo sottoposto a doppia blind peer review. Inviato il 01/12/2021. Accettato il 08/06/2022.

Dialectics of the algorithm. For a critical theory

The article aims to analyze the logical process of the algorithm considered as a «cultural machine», according to Finn, instead of an automatic mechanism that arranges information. The translation's operation of the world by the algorithm is based on the assumption that the device automatically reflects the ontological structure due to comprehend the universe, and, through its elaboration capacity, takes control over it. The article shows that this hypothetical duplication of the world is much more comprehensive: the different disposition, quantified by the machine that inductively detects the regularities of phenomena, immanently creates a 'digital image' of the world. This said, this algorithmic Simulation of reality is in itself the result of a new constellation of data that differs from a simple duplication of the world. The Simulation, as Baudrillard explained, is not immediately superimposable to the phenomenological image of Reality: the peculiar differential among them is the efficiency of the digital one – that corresponds to the reason why we're using it. The mediation overturns the sense of the information: it is no longer valid for its quality, but is rather judged by its quantity and its capacity to be calculated. Thanks to the dialectic critics here implemented (Horkheimer, Adorno), it will be shown how the algorithm creates a surplus value that triggers a vicious circle where the translated quantification of information could create new information, ready to be exploited. The result of the operativity pretends to be decisional too: the «digital image» or Simulation is pre-posed to the analogical one. The «negation of the negation» being part of the critical method displayed, it confirms that the illusion given by the Simulation can drive us to believe that this one could foresee the future, when this prediction is in fact a remodulation of current data that project to a future already given.

***

Introduzione

Nel 2008 in un articolo apparso su wired.com[1], Chris Anderson invitava filosofi e *theorists* di varie scienze (sociologia, psicologia, ontologia) a farsi da

[1] C. Anderson, *The End of Theory. The Data Deluge Makes the Scientific Method Obsolete*, «Wired Magazine», 23.6.2008, https://www.wired.com/2008/06/pb-theory/.

parte: «With enough data, the numbers speak for themselves». L'algoritmo pare aver ridotto l'attività d'analisi che lo ha preceduto a una limitata scelta tra opzioni ristrette: l'obsolescenza, l'osservazione entusiastica delle risultanti algoritmiche o il mero silenzio. Il pensiero si scopre viziato da una ristrettezza d'elaborazione che lo rende superfluo e impreciso, come molti strumenti della medicina ottocentesca, come le credenze alchemiche o le cosmologie arcaiche. Le tesi qui esposte si discostano tanto da questo messianesimo tecnofilo di lunga durata quanto dalla recrudescenza della tecnofobia novecentesca. Si cerca però di contrapporre alla convinzione di un'impossibilità della teoria dinanzi all'algoritmo, una simmetrica impossibilità, stavolta impossibilità intrinseca all'algoritmo, incapace di superare quello stile elaborativo che non gli permette di interrogarsi da sé sulle sue dinamiche interne. Attraverso 'vecchi strumenti' come la dialettica, per come l'avevano intesa i materialisti, si prova a dare ossigeno alla critica, ad interrogare la dinamica interna della macchina algoritmica come «macchina culturale», con l'obiettivo di spiegare altrimenti la (ri)produzione della conoscenza che vi si cela ancora inesplorata, in particolare nei suoi fondamenti epistemologici.

La prima parte del testo ha il compito di precisare in che termini l'efficacia degli algoritmi non riguardi la capacità di approcciare il mondo «verschieden interpretiert» da un'immane potenza di calcolo. L'algoritmo agisce al contrario come forza poietica, di alterazione veritativa e vettore decodificante (e ricodificante insieme) del reale, fattosi indistinguibile dal ritorno cumulativo dell'algoritmo su sé stesso. Verranno poste sotto la lente vecchie metafore, come la Mappa e il Territorio, e il loro legame con analogie nuove, come la Realtà e la Simulazione.

Nella seconda parte, l'efficacia dell'algoritmo è indagata nel momento di produzione del dato – sfruttando *datamining* e «valori di rete» – e facendo luce su come la mediazione della ragione algoritmica coincida con un'elusione. Dialetticamente, la qualità del dato esperienziale, una volta sottoposto alla mediazione macchinica, si rovescia nel suo valore formale senza risolvere lo scarto fra astrazione e realtà ma anzi eludendolo. Sembra, infatti, che ogni incommensurabile, ogni qualità che si sottrae alla traduzione quantitativa, sia statisticamente riassorbita in una elaborazione che si implementa di rapporti tra «quantità» sempre nuovi e più complessi (capaci di sottomettere indirettamente i propri impensati), divenendo al tempo stesso sia qualitativamente rilevante e sia squalificante di ciò che non comprende.

Nell'ultima parte viene paragonata la potenza elaborativa con cui gli algoritmi si aggiornano (a partire dai dati raccolti seguendo un'incessante complessificazione retroattiva) al processo di negazione della negazione dialettica. Si intravede, in una conclusione che vuole essere un'apertura prospettica, lo sganciamento dalla capacità sintetica di generare conoscenza, in favore del controllo e sottomissione della potenza analitica, della immanente necessità della stessa di fondare continuamente il suo campo epistemologico di riferimento, e la cui onnipresenza oggi invisibilmente con-vince mentre visibilmente funziona.

I connettivi AND, OR, NOT sono le operazioni di base di tutti gli algoritmi quando vengono ridotti alla loro forma più elementare[2]; saranno il filo rosso di una inaspettata rivincita della logica analogica, affinata nella lettura dialettica post-adorniana, che anziché restituire una mappa dell'algoritmo e delle condizioni del suo operare, prova a intravedere il rovescio del suo incedere condizionante. Sottolineare la sua persistente efficacia, la sua affermatività retroattiva, l'elusività dei suoi fallimenti, è un modo per concettualizzare brandelli di un'impossibilità algoritmica d'elaborare diversamente. Elementi per una sbordatura, per riflettere prodromicamente verso una 'teoria' consapevoli dell'evanescenza d'ogni prospettiva antropocentrata, provando ad indicare il rovescio di quel che si prende per Realtà.

1. *AND* (o la compenetrazione degli opposti)

1.1 Mappa e Territorio

In quell'impero, l'Arte della Cartografia raggiunse tale Perfezione che la mappa di una sola Provincia occupava un'intera Città, e la mappa dell'Impero un'intera Provincia. Col tempo, queste Mappe Smisurate non soddisfecero più e i Collegi dei Cartografi crearono una Mappa dell'Impero che aveva la grandezza stessa dell'Impero e con esso coincideva esattamente. Meno Dedite allo Studio della Cartografia, le Generazioni Successive capirono che quella immensa Mappa era Inutile e non senza Empietà l'abbandonarono alle Inclemenze del Sole e degli Inverni. Nei deserti dell'Ovest restano ancora lacere Rovine della Mappa, abitate da Animali e mendicanti; nell'intero Paese non vi sono altre reliquie delle Discipline Geografiche [S. Miranda, *Viajes de varones prudetes*, Libro Quarto, cap. XLV, Lérida, 1658][3].

Generatrice di rovine e di insoddisfazione, l'arte cartografica di Borges illustra con chiarezza la crisi della rappresentazione astratta. Sembra, infatti, suggerire che la razionalità novecentesca ragioni ancora secondo rimandi biunivoci – tra il segno e il significato – e che imposti la relazione tra la rappresentazione astratta e il suo referente concreto, partendo dal secondo in quanto fattore determinante per la prima. Borges non fa altro che portare al parossismo questa relazione duale, immergendola in un contesto logicamente plausibile ma in cui i limiti di una proporzione 1:1 fanno detonare lo stesso valore d'uso della mappa. La relazione biunivoca arriva così ad estendersi fino all'immediatezza, tanto che i due piani si possono effettivamente sovrapporre. L'unica conclusione possibile alla favola – questo lo scandalo per il pensiero – è diretta conseguenza di questa immediatezza, il paradosso per cui il mondo è l'esatta copia della mappa, e viceversa; la staticità che questo implica, l'impossibilità della concretezza di mutare secondo le leggi fisiche e quella della mappa di contenerle in astratto, porta a far esplodere le due dinamiche sovrapposte, lasciandoci un paesaggio immobile, che sa di morto. Si

[2] P. Domingos, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, trad. di A. Migliori, Torino 2020, p. 24.
[3] J. L. Borges, *L'Artefice*, edizione digitale 2016.

tratta di una relazione che nasce illuministica, rischiaratrice, (di)spiegante del mondo, quando invece si rovescia in una forma di razionalità che si esprime secondo caratteristiche identificative e per esclusione[4]. Un monito per chi pensa per identificazione, seppur permanga la necessità interpretativa, per certi versi ineliminabile.

Più recentemente, in *Cosa vogliono gli algoritmi*[5], Finn mette in risalto questa problematica identificante alla luce delle innovazioni tecnologiche che hanno l'obiettivo di rappresentare il mondo nella sua concretezza. Facendo un largo uso di terminologia magico-esoterica per descrivere lo sviluppo e il comportamento dell'algoritmo, permane in quest'opera la necessità analogica di rapportare una dinamica concreta con una rappresentazione astratta attraverso la connotazione metaforica della tecnologia e del suo processo come un elemento appunto 'magico'. Frequenti gli usi di neologismi (come «sourcery»[6]) che mescolano terminologia dell'ambito tecnologico con elementi già noti e legati appunto all'esoterismo, attuando una descrizione del campo epistemologico secondo dinamiche mitico-mimetiche, come ancora Adorno e Horkheimer avevano messo in luce[7].

Quest'uso particolare del linguaggio nell'opera di Finn, nella commistione dei due ambiti – tecnologico e culturale – ha lo scopo di definire quale sia la novità tecnologica dell'algoritmo che viene da subito connotato come non più solo strumento meccanico ma come vera e propria «macchina culturale». Per «macchina culturale» si intende quel dispositivo di attribuzione di valore epistemologico, che fonda l'interpretazione del mondo, e così delle scelte e dei comportamenti, in maniera automatica. Una sorta di apparato filtrante che pre-dispone determinate traiettorie per quei processi di attribuzione di valore che, d'ora in avanti, andranno da sé. Il punto di contatto tra l'innovazione dei nuovi mezzi di produzione della cultura e della conoscenza, sono, nell'opera di Finn, ancora intrisi di una impostazione illuministica, e il linguaggio adoperato sembra confermarlo: l'ipostatizzazione dell'algoritmo come una «struttura ontologica per comprendere l'universo»[8] deriva dall'indagine dei legami e delle connessioni profonde del mondo; indagine che ora è semplicemente assegnata ad un diverso vettore interpretativo, una figura di decodifica che non è più l'interpretazione razionalistica dell'identità. Si tratta ora di una dialettica che procede per accumulazione e aggiornamento, perché se prima era questione di interpretazione umana dei dati, ora è l'algoritmo che si situa in quel divario[9] biunivoco fra Realtà e Astrazione. L'innovazione tecnologica, il processo messo in moto dalla «macchina culturale», permette l'elaborazione di un'enorme mole

[4] T. W. Adorno, *Dialettica Negativa*, a cura di S. Petrucciani, Torino 2004, p. 7.
[5] E. Finn, *Che cosa vogliono gli algoritmi. L'immaginazione nell'era dei computer*, a cura di D. A. Gewurz, Torino 2018.
[6] *Ibid.,* p. 5.
[7] M. Horkeimer, T. W. Adorno, *La dialettica dell'illuminismo*, trad. di R. Solmi, introduzione di C. Galli, Torino 1997, pp. 17-18.
[8] Ivi, p. 10.
[9] Ivi, p. XVII.

di dati rispetto all'interpretazione umana, il cui primo risultato è un'asimmetria d'analisi rispetto all'agente identificante. Proprio in ciò consiste la novità qualitativa, in questa differenza di connessioni che la macchina è in grado di produrre e che è, allo stesso tempo, la caratteristica e il motivo precipuo che ci spingono a utilizzare l'algoritmo.

La diversa capacità qualitativa di ordinare gli elementi e di calcolarne le informazioni rappresenta un avanzamento tecnologico che scompagina la relazione biunivoca identitaria da cui siamo partiti. L'algoritmo si inserisce nel divario fra la Mappa e il Territorio e si fa carico della decodifica. Sebbene Finn si faccia già portatore dell'idea per cui «gli algoritmi non si limitano a descrivere i processi culturali», dal momento che qualunque processo può essere descritto come un «duplicat[o] ['ontologizzato'] matematicamente»[10] – questo il merito della sua critica –, non bisogna tralasciare il fattore immediatamente conseguente a questa affermazione e che risiede proprio nella capacità di codifica e decodifica della Realtà. L'algoritmo ha infatti una capacità di calcolo che inserisce un elemento nuovo, teoricamente implicito anche se praticamente, e pericolosamente, autoevidente: l'efficacia del suo processo di calcolo. Questo elemento tradisce quella semplice 'duplicazione' che era ancora ravvisabile nella trasposizione cartografica della mappa e ci conduce verso un più complesso livello noetico:

> I sistemi computazionali [sono] veicoli per un processo decisionale imparziale. La compenetrazione fra la «macchina culturale» e il mondo analogico ha, grazie alla pretesa dell'efficacia, un'intensità esplicativa di gran lunga maggiore di quella che dovrebbe 'semplicemente' rappresentare il mondo analogico che si presuppone sia causa efficiente[11].

1.2 Realtà e Simulazione

Questo scarto fra la rappresentazione digitale dell'algoritmo e l'immagine analogica del mondo è già stato oggetto della critica novecentesca che, seppur non trattando direttamente della cibernetica o nei termini di «macchina culturale», aveva già messo in guardia dalla contraddizione che l'intromissione tecnica può generare. Il meccanismo di codifica e decodifica, che mette in relazione gli oggetti secondo criteri non immediatamente reperibili nel mondo analogico, è

[10] Ivi, p. 11.
[11] *Ibid.* La non congruenza fra le due apre numerose problematiche e in primo luogo la fondazione della veridicità del risultato della macchina. La diversa estensione delle due immagini e la non piena chiarezza del procedimento, è, per citare Vespignani (*L'algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, con R. Rijtano, Milano 2019, p. 74), la dimostrazione che anche l'algoritmo è vittima del «paradosso di Polanyi» che viene riassunto così: «Conosciamo di più di quello che possiamo spiegare», che nel registro terminologico degli studi sull'algoritmo viene denominato come *Black Box* (cfr. F. Pasquale, *The Black Box Society. The Secret Algorithms that Control Money and Information*, Cambridge 2015). La qualitativamente diversa estensione delle relazioni di conoscenza prodotte dalla macchina permette l'ottenimento di un output dall'input, ma la complessità del calcolo eseguito ci porta a non comprendere pienamente l'operatività necessaria per quel risultato.

ben più di una migliore attribuzione quantitativa della conoscenza: si tratta di un vero e proprio ribaltamento del paradigma gnoseologico che Baudrillard, ad esempio, chiama «precessione dei simulacri»[12].

Questi fa notare come la mediazione tecnica, che permette la compenetrazione fra i due poli della significazione del Reale e della Simulazione, sbilanci gli elementi dei due poli così da non risultare più sullo stesso piano di comparazione; la capacità elaborativa della macchina è infatti in grado di proporre delle equivalenze fra oggetti astratti che non sono reperibili nell'immagine analogica del mondo, arrivando a svelare un ribaltamento della più nota fonte di conoscenza: non più il mondo esperienziale, ma quello della Simulazione. Quando il meccanismo di interconnessione dei fenomeni viene dato immediatamente, come dinanzi all'ammiccante trasparenza dei risultati algoritmici, l'asimmetria informativa che si instaura non vede più il mondo concreto come origine e meta della conoscenza, ma piuttosto fa del mondo il luogo che prende significato a partire dalla rappresentazione digitale. L'algoritmo copre il divario fra le idee astratte e le istruzioni pratiche, triangolando l'uscita verso l'universo della Simulazione. Non si tratta della classica duplicazione del mondo in valore numerico/simbolico, ma di uno spazio logico più estensivo di quello analogico grazie alla proprietà dell'efficacia di questi legami. È ancora Baudrillard a far notare che questa «precessione» può essere accettata nella sua complessità perché, a differenza del mondo esperienziale, la Simulazione offre la proprietà di essere «efficace»[13] nell'attribuzione di valore d'uso dei suoi componenti. Accettare l'interpretazione dell'algoritmo, non significa accedere a una comprensione delle dinamiche del Reale, ma anteporre ad esso la Simulazione visto che questa, al contrario del mondo concreto, «funziona»[14].

Se prima il mondo indeterminato e concreto veniva riempito di significato dall'opera ermeneutica del linguaggio e della sua sintesi identificante, ora il risultato dell'elaborazione macchinica supera la contraddizione della significazione dell'indeterminatezza del mondo, senza però eliminarla: ciò che prima cercava significato attraverso l'ermeneutica umana, ora grazie all'elaborazione quell'indeterminato non viene descritto secondo una qualitativamente migliore compenetrazione tra gli opposti (giacché questi non cessano di essere tali) ma viene posposto al risultato digitale secondo un diverso ordine gerarchico. Il dato non esplica meglio l'origine della conoscenza nel Reale perché la dinamica elaborativa li ha semplicemente scambiati di posto.

Sorgono dunque da subito alcuni problemi: se da una parte la mediazione quantitativa operata dall'algoritmo permette la «calcolabilità effettiva» dei fenomeni analogici che così facendo diventano oggettivi, dall'altra il dato digitale che ne risulta si presenta immediatamente nella sua disponibilità dimostrativa senza che vi sia la possibilità di questionare sull'effettivo procedimento che ha dato luogo al risultato. L'indeterminatezza non viene efficacemente risolta, ma

[12] J. Baudrillard, *Simulacres et simulation*, Parigi 1981, p. 10 (traduzione degli autori).
[13] Ivi, p. 86.
[14] E. Finn, *Che cosa vogliono gli algoritmi*, cit., p. 13.

elusa in un superamento epistemologico che fa rimbalzare la complessità del Reale nella complessità della produzione del dato. L'unica cosa che 'conta' è la realtà funzionale della Simulazione, quella che sa essere effettivamente valida di continuo. Si sfugge al paradosso borghesiano della perfetta coincidenza cartografica Mappa-Territorio quando la Simulazione è anteposta alla dinamica ontologica: la «macchina culturale» continua ad attribuire il valore d'uso agli elementi secondo il principio per cui il risultato dell'operazione è immanentemente più efficace nella creazione del quadro epistemologico di riferimento. In Borghes la Mappa non riusciva a tenere testa al Territorio se non ipostatizzandolo in un'immagine fissa; la 'macchina culturale' invece si limita a mettere in moto il mondo secondo la sua rappresentazione simulata. Attraverso l'uso macchinico di decodifica del Reale, l'indeterminatezza del concreto viene superata: il mondo si rovescia nella determinatezza offerta dall'algoritmo. Tale determinatezza non viene pienamente interpretata, ma proposta e prodotta perché semplicemente è l'operatività meccanica che detta il tempo alla Realtà.

1.3 Immanenza e calcolabilità effettiva

In Finn, «la realtà funzionale» è l'avanzamento o addirittura «il caricamento» del software che «detta i tempi dell'esperienza [analogica]»[15]. Questa aderenza alla realtà la si deve al fatto che gli algoritmi, come abbiamo già detto, colmano la differenza fra i due piani di Realtà e Simulazione, dal momento che esisterebbe un «substrato matematico della realtà»[16] che l'algoritmo non farebbe altro che decodificare. Il processo algoritmico non sarebbe altro che una lettura estrattiva di questa immagine analogica attraverso l'attribuzione di valori d'uso ben specifici agli elementi fenomenologici. Vedremo nel capitolo successivo cosa significhi questa attribuzione di valore; per il momento dobbiamo focalizzare la nostra analisi sulla dinamica che rende possibile questo contatto fra i due poli opposti di digitale e analogico.

Grazie all'algoritmo, i piani analogico e digitale si incontrano nella capacità di essere letti dalla macchina e questa leggibilità del mondo reale è resa possibile dalla mediazione di un codice che rileva, secondo metodo induttivo[17], la ripetitività dei fenomeni che possono essere così riconosciuti. Se gli algoritmi possono dire di formare e riprodurre il substrato della realtà in maniera culturalmente leggibile è perché quella regolarità dei fenomeni che caratterizzano il mondo analogico viene filtrata attraverso la «macchina culturale» in cui ogni caratteristica che si ripete può essere riconosciuta. La codifica di questa proprietà diventa una formalizzazione che può essere sottomessa a delle regole statistiche. Questa compenetrazione fra codifica e formalizzazione è un punto d'incontro fra i due piani ed è ad entrambi immanente: la regolarità del mondo analogico

[15] Ivi, p. 25.
[16] *Ibid.*
[17] P. Domingos, *L'algoritmo definitivo*, cit., p. 92.

è la caratteristica per cui è possibile identificare, per induzione, la ripetizione degli elementi esperienziali affinché possano essere riconosciuti. L'algoritmo non farebbe altro che rintracciare quella regolarità e codificarla nel linguaggio simbolico.

È così che il reale può essere non solo 'datificato', ma anche sottomesso ad un processo di calcolabilità effettiva. Secondo una disamina dialettica, possiamo far notare come il fenomeno reale che si ripete possa essere luogo di attribuzione di valore proprio per il fatto che, attraverso la ripetizione, esso possieda una quantità – la ripetitività in sé. Questa rintracciabilità quantitativa, e dunque la capacità di essere letto, permette di datificare il fenomeno che possiede così un valore per entrambi i piani (quello della Realtà e quello della Simulazione) e che può essere effettivamente utilizzato all'interno di una funzione algoritmica che associa o scarta o implica una conseguenza di tipo induttivo[18].

Ciò che prima aveva un significato reale solo per la sua riscontrabilità concreta, che poteva essere descritto solo *in sé*, ora è un dato che mostra una regolarità *per sé*. La qualità di essere quantità diventa utilizzabile nella Simulazione che può usare questa eccezionale *qualitas* per la sua efficacia nel determinare la Realtà. Se prima il piano dell'ermeneutica attribuiva un significato secondo categorie ai fenomeni concreti, ora il momento di compenetrazione si dà allorché la regolarità delle caratteristiche fenomeniche vengono trasposte nella loro occorrenza statistica.

Abbiamo già detto che la sovrapposizione che non era possibile prima della «macchina culturale», quella che ipostatizzava una mappa che descriveva il territorio, ora si ritrova rovesciata nella Simulazione. L'introduzione della decodifica algoritmica ci permette ora di rintracciare il piano immanente della leggibilità di codice, tra dato e fenomeno, grazie al fatto che l'automatismo si premura solo di attribuire un valore d'uso all'informazione che viene scambiata da un piano all'altro perché l'unica necessità interpretativa è quella della sua funzionalità. La Simulazione codifica l'elemento analogico in informazione secondo il principio dell'efficacia cosicché la mappa arriva a coprire la realtà che non è più altrimenti conoscibile.

La rilevanza statistica fornita dalla ripetizione, il suo sfruttamento tramite il principio della maggiore efficacia, ci porta a concludere che la calcolabilità effettiva del dato non solo supera quella contraddizione di determinare il Reale

[18] La pratica induttivista, così come viene spiegata nell'opera di Domingos (*L'algoritmo definitivo*, cit., pp. 92-93), origina dagli studi dell'empirista Hume e viene di continuo contrapposta al razionalismo deduttivo della filosofia continentale. Al di là della breve e poco consistente definizione di induzione come «deduzione inversa» (Ivi, p 108), la pratica induttivista si pone il problema dell'assunzione dei dati del passato per poter formulare una previsione della ripetitività del fenomeno nel futuro; si fa notare come l'assunzione del dato quando perviene alla previsionalità, incappa nel problema di come poterlo sottomettere alle ipotesi più verosimili (e non a teorie di matrice razionalista) per tale scopo. La risposta di Domingos guarda alla pretesa formale del dato, per cui «la soluzione classica consiste nell'assumere di conoscere la *forma* [corsivo nel testo] della verità: il compito del *learner* è quello di darle sostanza» (Ivi, p. 92). Per ulteriori approfondimenti sul limite di questa pratica, cfr. *infra*, nota 32.

offrendo una rappresentazione della realtà senza risolvere i complessi meccanismi che le hanno consentito di arrivare a tale conclusione. Ma non si pone neanche il problema di giustificare quella 'struttura ontologica' di cui dovrebbe essere uno strumento di decodifica. Dal momento che la regolarità dei fenomeni è immediatamente esperibile, la ripetizione digitale è immediatamente accettata senza avere necessità di un reale riscontro esperienziale. Nel momento in cui ci si ri-trova orientati nell'efficacia della Simulazione che tutto ha mappato, non ha senso chiedersi cosa è reale, cosa è mappa.

A tal proposito, riportiamo questo particolare aneddoto, in cui dei piloti di linea, fidandosi ciecamente delle informazioni fornite dal GPS furono vittime di un grave incidente nell'agosto del 1983:

> l'equipaggio aveva programmato il pilota automatico secondo la rotta assegnatagli dal controllo del traffico aereo, cedendo così il controllo del veicolo. […] tuttavia a causa di un errore nelle impostazioni […] il pilota automatico non continuò a seguire la rotta preassegnata; al contrario, rimase fisso sulla rotta iniziale, che lo portò sempre più a nord. […] I piloti notarono che il tempo tra un segnalatore di posizione e l'altro andava aumentando, ma decisero di ignorarlo. Man mano che l'aereo andava alla deriva, si lamentarono della scarsa ricezione radio. Eppure nulla di tutto questo gli fece mettere in dubbio il sistema, né li persuase a verificare la posizione. Continuarono a fidarsi del pilota automatico anche quando entrarono nello spazio aereo militare sovietico sulla penisola della Kamchatka[19].

Quella che viene percepita come immediata riproposizione delle condizioni di conoscenza del Reale da parte della macchina, è l'illusione per cui la macchina non possa sbagliare; questo 'bias di automazione' caratterizza ciascun utilizzatore che non mette in discussione la veridicità delle informazioni fornite dalla macchina. La definizione per tale comportamento è il «pregiudizio che ci spinge a considerare le informazioni automatizzate come più affidabili delle nostre stesse esperienze»[20]. La 'realtà funzionale', proprio perché è efficace, è in grado di dimostrare molto meglio la Realtà della capacità umana di convalida, tanto che quel 'bias di automazione' degrada nel 'bias di conferma': la consapevolezza umana viene riplasmata per allinearsi all'automazione e ai suoi risultati, fino al paradosso per cui le considerazioni scaturite dalla consapevolezza umana «vengono scartate» quando «non sono in linea col punto di vista della macchina»[21].

È per questo che dobbiamo concludere che la «macchina culturale» è molto più che semplice lettura dei rapporti epistemologici che intercorrono fra gli oggetti. La loro riproduzione in un sistema meccanico di ricombinazione dell'informazione è operata secondo dei parametri che sono la ripetizione e la funzionalità del procedimento. In questo processo di riformattazione dell'informazione, l'obiettivo conoscitivo è mutato per sempre. Non è più quello

[19] J. Briddle, *Nuova Era Oscura*, trad. di F. Viola, Roma 2019, pp. 51-52.
[20] *Ibid*.
[21] *Ibid.*

di svelare la verità dei rapporti oggettivi di produzione della realtà, ma piuttosto di rinvenire quelle proprietà per cui tali rapporti possano essere letti dalla «macchina culturale» così da permetterle di funzionare. Spinto dalla necessità di essere efficace, il modello di interpretazione del Reale è ora quello automatico, dimostrato dall'operazione e sostitutivo della realtà.

«La dimostrabilità ha sostituito la verità»[22].

2. OR (il rovesciamento della qualità in quantità)

2.1 Estrazione/scissione

Per meglio comprendere come sia possibile la codifica dell'informazione da parte dell'algoritmo, è necessario procedere ad un'analisi del momento dialettico del rovesciamento della qualità del fenomeno reale nella sua datificazione quantitativa.

Il saggio di Tiqqun, *L'Hypothèse cybernétique* (*L'ipotesi cibernetica*)[23], mette in rilievo la modifica prodotta dalla cibernetica come nuova relazione di produzione delle relazioni conoscitive del reale; la descrizione dell'uso massivo della tecnica e la sua denuncia, possono essere accostate all'opera di Finn per gli stessi obiettivi ma con fini opposti. Se il secondo intravvede uno sviluppo positivo di 'disvelamento' della complessità del mondo, il primo ne denuncia invece le derive autoritarie.

È interessante notare però come Tiqqun mostri le condizioni materiali che hanno spinto alla realizzazione dell'algoritmo. Quando vi fu la necessità di sottomettere l'incredibile mole di dati che riguardavano l'energia nucleare nei primi anni '40, quelle condizioni e necessità materiali, o strutturali, hanno plasmato di rimando anche le caratteristiche del dispositivo cibernetico di controllo quale unica possibilità di tenere a bada l'istantanea liberazione di energia dell'atomo che veniva scisso. Quell'impostazione permase nel linguaggio che descrive ancora la reazione nucleare e che risulta metaforicamente, o sovrastrutturalmente, calzante anche per descrivere la macchina che riesce a contenerlo. Esattamente come l'energia nucleare scaturisce dalla fissione dell'atomo, così l'estrazione dell'informazione avviene a partire da una materia grezza che attraverso la codifica dell'algoritmo si avvale di una dinamica scissoria. L'analogia sta nel fatto che, per controllare la disposizione degli elementi nell'immagine analogica, la macchina rovescia quell'ordine di elementi nell'immagine digitale, estraendo dai fenomeni il loro valore di ricorrenza così da essere effettivamente calcolato.

A dispetto della duplicazione illuministica della realtà in linguaggio simbolico, e dunque della replicazione astratta quale momento distinto, la cibernetica e l'algoritmo si permettono di analizzare costantemente questa mole di dati per intercettare quel momento regolare/ripetitivo e reintrodurlo ogni

[22] E. Finn, *Che cosa vogliono gli algoritmi*, cit., pag. 88.
[23] Tiqqun, *L'ipotesi cibernetica*, trad. e nota introduttiva a cura di G. Battisti, no copyright 2012.

volta nel sistema. E se il quadro epistemologico razionalista/analogico aveva una fondazione mitico-mimetica[24], la digitalizzazione della conoscenza ci pone di fronte ad una relazione che ha bisogno di essere esclusivamente operazionale. Il superamento della materia informe ancora una volta perpetua la contraddizione: l'immediatezza, come qualità esclusiva del codice, trapassa nella leggibilità immanente, proprio grazie a quel linguaggio codificato, nell'universale mediazione quantitativa «che rapporta ogni ente ad un altro»[25]. L'algoritmo perpetua la contraddizione senza risolverla quando, epistemologicamente, usa l'identità dei fattori quantitativi della regolarità/ripetizione come metodo conoscitivo, traducendo il dato reale nella sua immagine digitale. Questa asimmetria informativa prodotta dall'algoritmo, la precedenza della Simulazione alla Realtà, implicando l'efficacia dell'informazione ancorché la sua verità, risiede nell'interpretazione di quel dato secondo la ricorrenza statistica. È qui che avviene la scissione: di tutte le qualità corrispondenti al fenomeno, è la sua quantità evenemenziale a essere rintracciata dalla macchina; è il suo valore di ricorrenza che può essere effettivamente estratto e calcolato mentre le restanti differenze qualitative, ancora insite nell'immagine analogica, si annullano. L'estrazione non è dunque semplice misurazione e osservazione dei fatti del mondo, ma estrapolazione di un'azione traducibile in un valore.

2.2 L'accumulazione

Il procedimento con il quale l'algoritmo adopera questa materia con cui può procedere all'elaborazione, viene detto oggi *datamining*, ovvero il processo di estrazione dell'informazione a partire dai dati sensibili o analogici – e come già anticipato questo processo avviene per scissione.

Per quanto fin qui sostenuto, il *datamining* non è solo constatazione dei fatti e la codifica, infatti, non riguarda solo una mera misurazione in valore numerico della duplicazione astratta dell'immagine digitale. Si tratta piuttosto di un ribaltamento della qualità dell'esperienza conoscitiva del mondo in una quantità digitalizzata dove la mediazione (o quantificazione) innesca un «processo di comprensione e generalizzazione» funzionale nell'accezione di un ciclo produttivo vero e proprio.

Lungi dal poter essere confinati a meri dati in sé, il digitale valore d'uso dell'esperienza conoscitiva prodotta virtualmente, trova reale applicazione nel momento in cui questi dati vengono elaborati nel loro complesso.

> Google non avrebbe più estratto i dati comportamentali per migliorare il servizio offerto agli utenti, ma per leggere le loro menti e far combaciare gli ads ai loro interessi, dedotti dalle tracce collaterali lasciate dal loro comportamento online[26].

[24] Cfr. *supra* nota 7.
[25] M. Horkheimer, T. W. Adorno, *Dialettica dell'illuminismo*, cit., p. 20.
[26] S. Zuboff, *Il capitalismo della sorveglianza: Il futuro dell'umanità nell'era dei nuovi poteri*, trad. di P. Bassotti, Roma 2019, p. 88.

Zuboff descrive molto chiaramente il passaggio dell'elaborazione dei dati che, in un primo luogo erano tenuti in conto per rendere più performante la navigazione degli utenti, da parte di Google o di Yahoo, per arrivare a riconsiderare questi metadati («dati di scarto» o «tracce collaterali») come fonti di guadagno. Tutto ciò che è contingenziale all'informazione stessa, e strettamente correlato con la navigazione da parte dell'utente, viene utilizzato per una «logica di accumulazione» che ha un peso principalmente economico[27]. Quello che troviamo in Zuboff e che l'autrice chiama «surplus comportamentale»[28] è, a ben guardare, solo una parte di un più ampio «valore di rete»[29] e riguarda specificamente l'esperienza legata all'utente o, più in generale, le *platform* come quadro epistemologico produttore di valore. È più corretto d'altronde parlare di un *surplus di codice*[30] che si estende ad una casistica molto più generale e che può essere dedotto da ogni prodotto elaborato dall'algoritmo a partire dall'interconnessione. Il dato si dà immediatamente alla fruizione quale risposta alla domanda posta all'algoritmo, benché il suo valore d'uso sia in realtà la mediazione della macchina.

Il saggio di Zuboff rende trasparente l'aspetto economico del dato estratto dall'algoritmo e dai suoi usi prettamente commerciali, ma si rivela utile anche nel mostrare la strategia che gli algoritmi di Google (così come Yahoo o Amazon)[31] utilizzano per attribuire una finalità ben specifica all'informazione contenuta fra le sue pagine. Viene mostrato infatti che ogni dato non è solo portatore di una singola informazione, ma è un elemento da cui è possibile estrarre ulteriore informazione poiché rimanda a sua volta ad altre connessioni e legami ad altri dati. Google usa per esempio la quantità di «link», che rimandano tutti ad un sito ben specifico per dedurre l'autorevolezza o semplicemente la risposta più pertinente ad una data domanda posta dall'utente[32]. Quei metadati che compongono certe gerarchie e determinate posizioni dei dati stessi all'interno dell'elaborazione, sono da considerarsi *surplus di codice*, ciò che può essere definito come delle informazioni perpendicolari al dato di partenza e che non gli sono implicite[33].

[27] Ivi, pp. 86-89.
[28] Ibid.
[29] A. Fumagalli, *Per una teoria del valore-rete*, in *Datacrazia. Politica, cultura algoritmica e conflitti al tempo dei big data*, a cura di D. Gambetta, Milano 2018, p. 51.
[30] *Ibid*.
[31] S. Zuboff, *Il capitalismo della sorveglianza*, cit., pp. 88-90.
[32] *Ibid.*, cfr. anche P. Domingos, *L'algoritmo definitivo*, cit., p. 186.
[33] Secondo la divisione operata da Domingos (*L'algoritmo definitivo*, cit., p. 108), in cinque scuole di pensiero, quella simbolista si basa sull'apprendimento da parte delle macchine delle regole con le quali esse devono associare i dati, seguendo la teoria di Chomsky. In questo senso, l'algoritmo deve saper manipolare i simboli secondo il loro significato (*knowlegde based*) e poco importa quanto sia in grado di creare interrelazione tra le informazioni, in un primo momento dell'elaborazione, sancendo così una netta separazione fra software e macchina. Quello che bloccò lo sviluppo di questa tecnica e l'implementazione di questi sistemi di apprendimento, a vantaggio della scuola statistica, fu la lentezza con la quale erano in grado di distinguere i dati al loro inserimento perché li portava a produrre di rimando una serie di ipotesi che portavano alla cosiddetta esplosione combinatoria. Cfr. *infra* nota 39.

La codifica di questi dati che non tiene conto del loro significato, necessita dunque di un valore attraverso cui sia possibile ottenere una gerarchia fra le risposte. Così come per la 'quantità' di link per Google[34], al dato viene attribuito un valore secondo la ricorrenza statistica con la quale quell'informazione si ripete anche nell'immagine digitale. Questo valore d'uso del dato all'interno della Simulazione ha un'efficacia solo all'interno di quel contesto: non rifacendosi più immediatamente al piano del Reale, esso informa della sua regolarità/ripetitività all'interno della rete. Quello spazio logico accresciuto dalle funzionalità della macchina risiede e dà consistenza efficace grazie al «valore di rete», ovvero la gerarchizzazione dell'informazione all'interno del sistema che viene dedotta dalla sua interconnessione con gli altri dati. Quel *surplus* informativo è conoscibile solo a partire dalla scissione del dato, dalla sua posizione all'interno della rete, valore che viene 'separato' dall'informazione principale. A sua volta, questo *surplus* informativo ottenuto per scissione ci informa di una nuova regolarità, che a sua volta è nuova informazione. È proprio grazie all'accumulazione dei dati all'interno dei server che quel *surplus* ottiene valore perché può essere rianalizzato per ottenere nuove regolarità e modelli che permettono, per esempio, la profilazione degli utenti.

Per questo l'informazione è asimmetrica: non solo il suo valore ha la proprietà di essere efficace, fornendo un quadro epistemologico più ampio, ma instaura una struttura più complessa che ha valore solo all'interno dell'operazionale. Se il dato *in sé* rileva una caratteristica ancora legata al piano del Reale, il suo posizionamento *per sé* è invece completamente legato a quello della Simulazione: il «valore di rete» è contraddittoriamente insito nel dato ma è rinvenibile solo nel momento in cui questo dato si posiziona all'interno di una costellazione più vasta di elementi.

Ancora una volta l'algoritmo supera la contraddizione senza eliminarla: per poter conoscere un dato, esso deve essere scisso dal momento esperienziale e reso il più possibile generale così da rilevarlo nella sua regolarità; fatto questo, si deve poterlo accumulare in gigantesche aggregazioni, i *big data*, per potergli attribuire un *plusvalore* che possa essere continuamente estratto e utilizzato. Il dato viene cioè scisso per la sua rilevanza quantitativa solo per rovesciarlo, accumulato, nella sua qualità mediata.

3. *NOT* (la negazione della negazione)

3.1 La negazione della negazione

Fin qui abbiamo visto che l'algoritmo è in grado di produrre una descrizione della Realtà per tramite di una Simulazione che non duplica semplicemente i rapporti di produzione della conoscenza in una immagine digitale, ma li dilata. Il rovesciamento dialettico della qualità dell'esperienza in quantità di dati

[34] Cfr. *supra* nota 7 e E. Finn, *Che cosa vogliono gli algoritmi*, cit., pp. 262-264.

supera l'indeterminatezza del Reale solo attraverso la Simulazione determinata dall'elaborazione e dalla calcolabilità effettiva. La sottomissione della disposizione degli elementi fenomenici avviene attraverso la loro scissione per ottenere funzioni di valori che assumono, a loro volta, nuovo valore all'interno della Simulazione. È così che viene decretato quel dilatamento o maggiore estensione dello spazio logico che arriva a dimostrare il Reale, perché l'efficacia del *plusvalore* della Simulazione ne anticipa la conoscenza. Così come l'energia nucleare aveva richiesto un sistema di sicurezza che la rendesse operabile, così la macchina opera una scissione controllata dell'informazione governando un effetto a catena in cui ogni parte scissa del dato informativo è informazione a sua volta. Per questo, i rapporti di produzione della conoscenza seguono il criterio omeostatico[35] di conservazione del sistema, al fine di preservare la sua struttura di base. Il tutto avviene solo ed esclusivamente per ottenere conferma del sistema che deve mantenere intatto il principio di efficacia e, fattore fondamentale, indipendentemente dall'impostazione dei bias che vengono impostati a monte[36].

Si potrebbe considerare e così descrivere il controllo omeostatico del sistema dialetticamente secondo il principio di negazione della negazione. Un primo movimento della logica algoritmica produce l'informazione tramite la mediazione calcolabile del valore d'uso per tramite della regolarità del fenomeno. La scissione del dato dal piano analogico lascia il posto a quella del piano digitale (il «valore di rete»), che permette di accumulare di continuo *plusvalore*. Il secondo movimento del ciclo produttivo è invece la reintroduzione della nuova informazione all'interno del processo di aggiornamento del sistema. L'algoritmo non solo tiene conto del movimento a spirale che la riproduzione del quadro epistemologico comporta, ma si assicura, attraverso il linguaggio che impone sempre leggibilità ai nuovi dati forniti, una «transizione di fase» che retropropaga l'effetto della novità conoscitiva[37]. È la forza di un'immanenza monotonale del

[35] E. Finn, *Che cosa vogliono gli algoritmi*, cit., p. 18.

[36] Al fine di lasciare agli algoritmi libertà decisionale davanti ai dati, secondo un modello statistico induttivo e non simbolico, è necessario fornire loro dei parametri di scelta; così facendo è l'elaborazione stessa che impara dai suoi errori seguendo il principio di efficacia verso cui deve continuare a tendere. Il nodo principale del discorso, riguardo la 'disputa dei bias', è la scelta stessa di questi parametri che riverberano la precondizione per la quale sono stati scelti senza essere in grado di dare alla macchina capacità di discernere il contesto nel quale il loro operato si attua; per ulteriori specifiche e per un'analisi della ripercussione degli errori legati al bias in ambito sociopolitico, rimandiamo al saggio di A. D. Signorelli, *La guerra dei bias*, in *Datacrazia*, cit., p. 172.

[37] La forma di apprendimento può essere esemplificata attraverso queste due dinamiche che vengono rintracciate entrambe nell'analogia esplicativa dei primi dispositivi cibernetici, che imitavano ancora il processo neurologico di apprendimento. Da una parte la 'transizione di fase' è la rappresentazione del passaggio da uno stato ad un altro, secondo un rapporto di frequenza e tensione (P. Domingos, *L'algoritmo definitivo*, cit., p. 134), e può ancora essere usata per spiegare meccanicamente il brusco passaggio tra due momenti della conoscenza all'interno del quadro epistemologico una volta che questo passaggio incorre in un errore, che viene così definito come «la differenza fra la risposta effettiva e quella voluta» (Ivi, p. 137). Dall'altra la 'retropropagazione' è sì una dinamica rintracciata nel funzionamento d'apprendimento del cervello, eppure risulta fondamentalmente diversa una volta che la implementiamo nella dinamica algoritmica: la retropropagazione è la conversione delle connessioni fra i dati alla luce

linguaggio che traccia e tiene conto, anche a ritroso, di tutti i dati che possiede per mantenere intatta la sua efficacia e di qui l'equilibrio omeostatico dell'insieme.

Quelle nuove informazioni non solo arricchiscono ancora una volta il «valore di rete», nell'interconnessione che si salda fra i dati, ma rinsalda direttamente lo spazio logico del quadro epistemologico di riferimento, nel tentativo di aumentare la pertinenza e la coerenza del sistema. Questo ampliamento dello spazio logico (Simulazione) accumunato all'incremento dei dati è assimilabile al processo di *Machine Learning*[38].

Secondo la prospettiva dialettica, il risultato non può essere immediatamente l'informazione di partenza. Ed è qui il paradosso algoritmico, che non prevede alcuna sintesi, solo la contraddizione per cui l'*Aufgehobene* è un nuovo input da processare che amplifica, appunto, lo spazio logico dell'algoritmo. L'informazione che ne scaturisce è altra informazione, formalmente coerente con la precedente, ciò che, adornianamente, possiamo definire la negazione della negazione. Il movimento dialettico potrebbe protendere per il superamento del mero dato verso una possibile sintesi esperienziale, oppure, in alternativa, per il semplice mediare e mediare ancora, come accumulo infinito della stessa ragione. Dall'analisi qui esposta è la seconda opzione che prevale. È proprio perché la conoscenza rimane legata al dato statistico-formale che l'algoritmo si avviluppa nella mediazione dei suoi dati: la mediazione che si compie nella ricombinazione della Simulazione non è un superamento ma un'accumulazione seriale di informazione. La negazione della negazione non supera ma conferma il piano epistemologico, aggiornandolo di continuo, 'aumentandolo' nelle connessioni relazionali, ad un livello così intensivo che si conserva omeostaticamente sempre sullo stesso piano epistemologico. Esso rimane sempre coerente con se stesso giacché non decodifica la logica del mondo analogico, ma piuttosto logifica il mondo secondo un codice sempre più consistente.

3.2. Principio di dimostrabilità

Questa coerenza interna è rintracciabile nel fatto che l'algoritmo ha continuamente dati che già apportano lo stigma della loro leggibilità; ciò che è illeggibile è immediatamente non dimostrabile nel sistema e sebbene l'algoritmo non possa strutturalmente tenerne conto, è proprio dall'analisi *ex negativo* che l'algoritmo ne ottiene conferma. Anziché far tornare in sé l'indimostrabile, lo fagocita sussumendolo sotto il principio dell'indimostrabilità, il «concetto generale» nelle parole di Adorno, e così facendo lo include dall'elaborazione[39].

dell'output reale che viene così reintrodotto nel sistema apportando delle modifiche a ritroso, non solo ai dati futuri ma anche a quelli in memoria. Quello che risulta differente dal 'semplice' funzionamento neurale, è l'incapacità della macchina di non stratificare l'informazione ma di considerare ogni singolo evento come contemporaneo agli altri «modifica[ndo] le connessioni tra i neuroni, uno strato dopo l'altro, per ridurne la differenza» (*ibid.*).

[38] Vedi *supra*, nota 35.

[39] L'operatività della ragione, così come quella dell'algoritmo, segue lo stesso procedimento di fronte all''impensabile': parafrasando Adorno, «La teoria, anziché far tornare in sé l'insolubile,

In un certo senso l'algoritmo ha bisogno dell'indimostrabile quando produce la risposta informativa perché conferisce a quest'ultima la validità della sua dimostrazione: quel dato è immediatamente dimostrabile perché rispetta le regole della leggibilità e dell'immanenza delle formule di calcolo che lo hanno prodotto e dunque l'output non è nuova determinatezza del Reale, ma la sempre-identica determinazione della Simulazione. Questa ipostatizzazione è peculiare e sempre efficace perché risulterà sempre dimostrabile:

> Le previsioni, anche le più accurate, hanno sempre un grado di incertezza che viene quantificato attraverso la probabilità che la predizione stessa si avveri. […] Ogni volta che le previsioni meteorologiche annunciano sole al 90 percento e poi ci troviamo invece una giornata uggiosa con un po' di pioggia […] mi dicono che ancora oggi il meteo è un'opinione, che non ci si può fidare e chi più ne ha più ne metta. In realtà una previsione al 90 percento per essere rigorosa *deve sbagliare proprio una volta su dieci*, perché prevede anche un 10 percento di pioggia. Per capire le previsioni, dunque, dobbiamo innanzitutto capire la probabilità[40].

Questo esempio ci mostra come tutto ciò che è leggibile dall'algoritmo, e cioè rappresentabile secondo occorrenza statistica, va a comporre un elemento più estensivo del semplice dato reale perché l'analisi della macchina comprende tutte le possibilità. Il dato è in sé la prova della mediazione dell'efficacia della macchina che lo ha prodotto: la rappresentazione delle possibilità conoscitive che l'algoritmo, nella sua maggiore estensività epistemologica, può dimostrare. Dall'altro lato, il dato indimostrabile per sé è una conoscenza particolare che deve essere sussunta in qualche modo e restituita in forma di percentuale. L'algoritmo sussume anche ciò che non è sussumibile come una percentuale delle possibilità realizzative, come una percentuale-sintomo della messa in calcolo dell'incalcolabile che rappresenta l'immanente controprova della sua correttezza. Comprendendo statisticamente tutte le percentuali di riuscita, l'algoritmo descrive, con l'immagine digitale, una rappresentazione ben più dilatata dell'immagine analogica perché virtualmente tutta la conoscenza è già calcolata e ora si tratta solo di occorrenza statistica.

Il principio di dimostrabilità è implicitamente collegato, e ad esso contrapposto, con quello della calcolabilità effettiva ed è da qui che possiamo delineare il limite precipuo del meccanismo automatico che ci permette di indurre la sua non-oggettività. Questa maggiore estensività indotta dalla capacità di calcolo, avendo bisogno di essere continuamente dimostrabile, non cessa di processare tutte le possibilità che l'algoritmo può leggere, eppure deve escludere tutti quei casi particolari che non possono essere ricondotti al principio della regolarità. Di fatto l'algoritmo piega i dati verso la loro possibilità di essere dimostrati solo e soltanto per una questione conservativa del sistema; tutte le tecniche statistiche hanno bisogno di essere efficaci e per poterlo fare devono tenere conto dell'evento radicalmente opposto alla dimostrabilità dell'output:

lo fagocita sussumendolo sotto il suo concetto generale; quello dell'insolubilità» (T. W. Adorno, *Dialettica Negativa,* cit., p. 111).

[40] A. Vespignani, *L'algoritmo e l'oracolo*, cit., p. 156, corsivo nostro.

l'esplosione combinatoria[41]. Per utilizzare ancora la felice metafora dell'energia nucleare tenuta sotto controllo dalla capacità elaborativa della cibernetica, anche l'algoritmo ha bisogno di evitare che il dato venga scisso in tutte le sue possibilità; ha bisogno che il suo «valore di rete» venga tenuto all'interno della capacità di calcolo affinché la codifica del dato rimanga all'interno della possibilità stessa dell'operatività. Per questo possiamo affermare che l'algoritmo non produce una conoscenza neutra ed oggettiva – e questo ben al di qua della questione dei parametri impostati a monte. La macchina, prima di 'soffrire' dei bias che le vengono attribuiti nelle impostazioni di base che le servono per calcolare, e a prescindere dal tipo di bias, ha innanzitutto la necessità omeostatica di conservare la sua capacità elaborativa e lo fa filtrando i dati attraverso la regolarità immanente e il principio di dimostrabilità: essa ha bisogno che la stessa calcolabilità sia riprodotta attraverso il dato secondo l'efficacia del suo procedere – o verrebbe meno il senso ultimo della sua utilità funzionale. Prima di intestardirsi in una forma specifica d'analisi (da qui la facile critica all'algoritmo che, a dispetto della mole di dati, non sa mai veramente), è l'analizzabilità stessa, la forma pura del calcolare, a viziare l'intelligenza algoritmica, mostrandone limiti e (im)possibilità.

3.2 La predizione del futuro come riproduzione del presente

La macchina assume connotati culturali nel momento in cui la scelta in seguito al filtraggio e alla ricombinazione di dati che vengono effettivamente 'scelti' condiziona atteggiamenti e comportamenti: il risultato sarà quello più conforme all'output secondo dimostrabilità.

Questo è facilmente rinvenibile nell'uso dei principali algoritmi commerciali in cui nell'immensa mole di prodotti in vendita l'algoritmo ci pone molteplici opzioni che sono il frutto di profilazione secondo picchi di tendenza, ricerche correlate o semplici scelte simili operate da profili affini al nostro.

Affinché questo possa darsi, la macchina deve continuamente riprodurre le condizioni attraverso le quali è riconfermata la sua efficacia; per potervi riuscire essa attua una mediazione sui dati, secondo quanto fin qui esposto, che illudono l'utilizzatore con l'immediatezza del loro valore d'uso. La presupposta affinità dei prodotti è l'equivalente di quel pregiudizievole posizionamento delle informazioni, in apparenza trasparenti seppur già il risultato del giudizio operato dall'algoritmo. Il nostro comportamento è, al contrario, costretto a scegliere

[41] Per esplosione combinatoria si intende il momento-limite della calcolabilità effettiva dell'algoritmo. Se, infatti, la capacità esplicativa dell'output viene fornita grazie alla capacità combinatoria dei dati in possesso della macchina, è altresì vero che per essa non è possibile processarli tutti allo stesso momento. La condizionalità del parametro arbitrario fornito in partenza, come avrebbero voluto i simbolisti, per disciplinare i dati e categorizzarli senza tenere conto dell'elaborazione aumenta le probabilità da apprendere e così facendo la macchina 'esplode' (cfr. P. Domingos, *L'algoritmo definitivo*, cit., p. 188). La soluzione è quella statistica, di introdurre cioè strutture incondizionate che riflettono argomenti relazionali semplici come la congiunzione o l'implicazione, che Domingos chiama «catena» e che rende un evento dipendente dal precedente, questo da quello prima ancora e così via (ivi, p. 195).

all'interno di un insieme di oggetti che per quanto vasto e innumerabile, è pur sempre finito. L'algoritmo sbaraglia la questione della scelta (libera) dell'utente quando cataloga di fatto all'interno di una programmazione preordinata secondo lo stesso criterio di programmazione in sé frutto delle sue decisioni. Le decisioni dell'utenza allora decidono sempre per la conservazione del sistema. Si sceglie sempre il *per sé* dell'algoritmo, vertigine esclusiva di una non-scelta assoluta.

Ancora una volta l'algoritmo supera la contraddizione senza eliminarla. La sua risposta sarà assieme originale e ripetitiva: la nuova relazione fra due dati informativi sarà il frutto della pertinenza con la domanda dell'utente, mai posta prima, ma è allo stesso tempo una relazione già nota al sistema perché, immanentemente, dovrà corrispondere al principio di dimostrabilità. Ciò che viene scelto conferma il valore d'uso del dato nel suo posizionamento fra la molteplicità di equivalenti (il «valore di rete») e allo stesso tempo conferma il suo valore di scambio, per cui la scelta dell'oggetto riproduce le condizioni algoritmicamente mediate delle relazioni della conoscenza. In questo senso il risultato è oggettivo solo in senso parziale, se non addirittura perverso: perché non è la descrizione esperienziale del mondo aperta al futuro, ma la sua realizzazione a partire dalla conoscenza già iscritta nel sistema. Un trapassato pseudooggettivo travestito da (oggettivo) futuro anteriore.

Sotto la lente della dialettica, quando rivolgiamo l'analisi dei dati verso una loro configurazione previsionale, essi già contengono le possibilità di determinazione del Reale. La mediazione della Simulazione che è anteposta all'interpretazione dell'immagine analogica, media a sua volta il dato esperienziale determinandolo secondo i suoi parametri. Quello che abbiamo esposto porta a concludere che la determinazione della Simulazione da parte dell'algoritmo, impone alla Realtà di essere interpretata epistemologicamente secondo delle condizioni pregresse che danno l'illusione di aver predetto ciò che sarebbe capitato, quando invece obbligano il Reale ad essere riletto secondo condizioni 'oggettive' prestabilite. Quell'elemento di rottura che dovrebbe rappresentare l'operatività, che dovrebbe trascendentemente prevedere il futuro, non fa altro che riprodurre quelle condizioni di determinatezza del Reale che sono già presenti e che per poter essere efficaci hanno in primo luogo necessità di doversi ripresentare. La mediazione della mediazione non produce una conoscenza nuova, nel senso dell'imprevedibilità, ma solo un'accumulazione di possibilità interpretative. Quello che viene illusoriamente scambiato per l'accresciuta capacità interpretativa del mondo, si rovescia nella sua più oscura contraddizione: la ripetizione del già conosciuto. In più di un'occasione, quando ci si pone la questione della conoscenza prodotta algoritmicamente, si vuole percepire il suo limite di neutralità senza riuscire a darne pienamente significato. Quella neutralità, che richiede di essere 'oggettiva', non può esserlo nella sua pretesa formale astratta perché fra le sue decisioni si reitererà sempre quella di essere coerente con sé stessa. Per questo l'algoritmo abbina a un immaginario tecno-futurista un sentore reazionario: non produce alcuna novità in senso stretto, ma predice piuttosto la ripetizione del presente. *L'ipotesi cibernetica* («*Il*

problema della cibernetica non è più la previsione del futuro ma la riproduzione del presente»[42]) avvertiva per tempo di come l'automazione del calcolo fosse solo una nuova connotazione tecnica del potere che è ora in grado di postulare un giudizio riconfermato ad ogni aggiornamento. Per utilizzare la terminologia adorniana, è l'operatività il punto di contatto con la critica dialettica: essa consiste nel fatto che l'elaborazione non supera la contraddizione, ma la eterna in un «gigantesco giudizio analitico senza spazio per il qualitativamente nuovo»[43].

Massimiliano Carli
✉ carli.massimiliano@gmail.com

Vittorio Lubrano
Universidade NOVA de Lisboa
✉ lubranoallavoro@yahoo.it

Bibliografia

Adorno, T. W. 2004. *Dialettica Negativa*, a cura di S. Petrucciani, Torino, Einaudi.

Anderson, C. 2008. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, «Wired Magazine», https://www.wired.com/2008/06/pb-theory/.

Baudrillard, J. 1981. *Simulacres et simulation*, Paris, Galilée.

Borges, J. L. 2016. *L'Artefice*, edizione digitale.

Briddle, J. 2019. *Nuova Era Oscura*, trad. di F. Viola, Roma, Nero editore.

Domingos, P. 2020. *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, trad. di A. Migliori, Torino, Bollati Boringhieri.

Finn, E. 2018. *Che cosa vogliono gli algoritmi. L'immaginazione nell'era dei computer*, a cura di D. A. Gewurz, Torino, Einaudi.

Fumagalli, A. 2018. *Per una teoria del valore-rete*, in D. Gambetta (a cura di), *Datacrazia. Politica, cultura algoritmica e conflitti al tempo dei big data*, Milano, D Editore, pp. 46-69.

Horkeimer, M., Adorno, T. W. 1997. *La dialettica dell'illuminismo*, trad. di R. Solmi, introduzione di C. Galli, Torino, Einaudi.

Pasquale, F. 2015. *The Black Box Society. The Secret Algorithms that Control Money and Information*, Cambridge, Harvard University Press.

[42] Tiqqun, *L'ipotesi cibernetica*, cit., p. 22 (corsivo del traduttore e curatore).
[43] T. W. Adorno, *Dialettica Negativa*, cit., p. 140.

Signorelli, A. D. 2018. *La guerra dei bias*, in D. Gambetta (a cura di), *Datacrazia. Politica, cultura algoritmica e conflitti al tempo dei big data*, Milano, D Editore, pp. 172-187.

Tiqqun, 2012. *L'ipotesi cibernetica*, trad. e nota introduttiva a cura di G. Battisti, no copyright.

Vespignani, A. 2019. *L'algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, con R. Rijtano, Milano, Il saggiatore.

Zuboff, S. 2019. *Il capitalismo della sorveglianza: Il futuro dell'umanità nell'era dei nuovi poteri*, trad. di P. Bassotti, Roma, Luiss University Press.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770025

Contributi/5

Dalla cognizione alla comunicazione

Epistemologia degli algoritmi secondo la teoria dei sistemi

Alberto Giustiniano 0000-0003-0094-0150

Articolo sottoposto a doppia blind peer review. Inviato il 30/11/2021. Accettato 08/06/2022.

From cognition to communication. Epistemology of algorithms according to systems theory

The present contribution aims at questioning the reasons why the level reached by the processes of digitization of information stimulates interpreters to hypothesize that a radically new way of making sense of data is inherent in this trend. The thesis we propose to investigate here, echoing Elena Esposito's recent work, is that of considering these programs not as reproducers of intelligence or thought but as tools aimed at acting adequately as partners of communication. Our analysis will proceed by providing an interpretation of the phenomenon mediated by Niklas Luhmann's systems theory. According to Esposito's analysis, the goal of algorithms would be to produce a lack of control in a controlled way, to leave open the possibility of 'informational surprise' to the user, allowing the construction of meaning in a new way. This paradoxical programming of unpredictability, if interpreted as a phenomenon about the sphere of communicative interaction described by Luhmann's theory, and not about intelligence, would help to make it more comprehensible how interaction with algorithms allows the societies that use them to process unpredictable information that cannot be ascribed to any cognitive process as a self-induced instrument of reflection. This would make it possible to anticipate the dysfunctional effects that such an increase in information complexity might produce.

***

1. Dalla mente alla rete

La nozione di algoritmo è ancora intrinsecamente ambigua. Tale affermazione può risultare provocatoria e fuori luogo, dal momento che questa forma computazionale rappresenta per molti aspetti la quintessenza dell'odierno sviluppo scientifico e tecnologico. Il suo successo non si misura soltanto dalle formidabili potenzialità applicative che ne hanno imposto la diffusione pressoché in tutti gli ambiti dell'agire umano organizzato, ma anche per i profondi interrogativi che la scienza del calcolo ha inaugurato in ambito matematico ed epistemologico. Tuttavia, una definizione di algoritmo che metta

d'accordo gli specialisti non è ancora disponibile, sebbene ne siano state fornite diverse, tutte plausibili[1]. In linea generale e ai fini del nostro discorso – che non mira a far luce sul complesso dibattito tra matematici in corso – possiamo indicare col termine algoritmo: una serie finita di istruzioni astratte finalizzate all'adempimento di un determinato scopo entro determinate disposizioni; tali istruzioni saranno concepite come una serie di step univoci, non interpretabili, capaci di controllare il passaggio da uno stato all'altro fino al raggiungimento del risultato atteso in un tempo finito. Un algoritmo dunque per essere tale deve essere efficace, ovvero deve poter fornire una risposta a colui che lo esegue[2]. Questa definizione a prima vista non sembra tenere fede alle aspettative rivoluzionarie che l'accompagnano. Com'è noto, a questa descrizione si possono ricondurre intuitivamente molte delle operazioni ricorsive che normalmente si imparano alla scuola elementare, come le moltiplicazioni e le divisioni in colonna, oppure una serie di calcoli noti già nell'antichità per la risoluzione di svariati problemi di redistribuzione e misurazione. Le potenzialità della definizione divengono però più evidenti se prestiamo attenzione a ciò che accade tra la fine dell'Ottocento e gli anni '30 del Novecento, quando dalla teoria degli insiemi infiniti proposta da Cantor, e la conseguente dimostrazione della non corrispondenza biunivoca tra numeri interi e numeri reali, si sviluppa un intenso dibattito. Grazie ad esso sarà possibile dedurre la distinzione tra un insieme di funzioni calcolabili, poiché descrivibili per mezzo di sequenze finite di caratteri numerabili, e un insieme più ampio di funzioni non calcolabili, ovvero che stabiliscono una corrispondenza tra due sequenze infinite di valori e dunque non numerabili. Tale questione, sulla quale non indugiamo oltre[3], stimolerà la nascita della Teoria degli Algoritmi alla luce della nascente esigenza di individuare quali funzioni siano calcolabili e quali no. Un problema non da poco, considerando che se per il primo caso è sufficiente trovare una soluzione tra quelle possibili, per il secondo è necessaria una definizione formale di *computabilità* per dimostrare poi che quella specifica funzione non può essere inclusa nella classe di funzioni individuate dalla definizione. Le conclusioni cui giunsero Alan Turing e Alonso Church nel 1936 designano questa definizione formale in due versioni equivalenti: la *Macchina di Turing* e il λ-*calcolo*. Attraverso le loro analisi si è giunti così a estendere la classe delle funzioni calcolabili non solo a quelle matematiche ma a ogni problema esprimibile attraverso un linguaggio formale qualsiasi[4]. In linea di principio,

[1] M. Vardi, *What is an Algorithm?*, «Communications of the ACM», vol. 55, 3, New York 2012, p. 5.
[2] Questa definizione di algoritmo è una libera rielaborazione di quella fornita da R. K. Hill nel suo *What an Algorithm Is*, «Philosophy & Technology». 29, Berlin-New York 2016, pp. 35–59. L'articolo è utile anche per inquadrare l'attuale dibattito sulla definizione di algoritmo che qui lasciamo sullo sfondo.
[3] Per un'analisi storica più approfondita si rimanda M. Frixione, D. Palladino, *Funzioni, macchine, algoritmi. Introduzione alla teoria della computabilità*, Roma 2009, capitoli 6 e 7.
[4] Cfr. A. M. Turing, *On Computable Numbers with an Application to the Entscheidungsproblem*, in J. B. Copeland (ed.), *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*, Oxford 2004, pp. 58-90; A.

qualunque problema riducibile a sequenze di caratteri in un alfabeto arbitrario poteva ora essere calcolato, cioè risolto. L'unico ostacolo che si poteva ancora frapporre tra l'utente e la risposta cercata era rappresentato dalla limitatezza delle risorse, in termini spazio e tempo, necessarie per risolvere effettivamente il calcolo[5]. Da qui l'odierna necessità di elaborare una Teoria degli Algoritmi sempre più raffinata e in grado di ridurre al massimo l'impegno dei calcolatori, poiché se lo sviluppo delle macchine segue l'andamento lineare della Legge di Moore l'incremento della complessità di calcolo, data dal rapporto tra aumento dei dati e operazioni necessarie per la sua risoluzione, procede in modo esponenziale[6].

Da questi brevissimi cenni storici si può allora comprendere come l'attenzione rivolta allo studio sistematico degli algoritmi nel dibattito matematico successivo alla crisi dei fondamenti sia stata amplificata dall'ottenimento di un rapido successo coinciso con la nascita dell'informatica. Nella sua prima fase[7] questo sodalizio è stato in grado con le sue promesse di imporre un pervasivo immaginario secondo il quale la scienza della computazione e la sua possibile implementazione bio-tecno-ingegneristica avrebbero dischiuso i segreti di quella particolare performance tipica della coscienza chiamata *intelligenza*. La nozione di algoritmo ha così iniziato a caricarsi di molteplici significati che sembravano trascendere il suo iniziale interesse nel contesto specialistico dei matematici, divenendo incrocio di desideri e aspettative sovrapponibili a quelle che anticamente, e con alterne fortune, erano state rivolte agli automi[8]. L'algoritmo sarebbe stato la chiave di volta non soltanto per la riproduzione del ragionamento, ma anche per il suo potenziamento indefinito grazie alla messa a punto di macchine fisiche in grado di svolgere le operazioni necessarie a un livello di velocità e complessità di gran lunga superiore a quello gestibile dalle possibilità umane. In effetti era questa la visione dei promotori del *Dartmouth Summer Research Project on Artificial Intelligence*, il convegno svoltosi nel 1956 che segnò un punto di svolta decisivo nella storia della tecnologia e

Church, *An Unsolvable Problem of Elementary Number Theory*, in «American Journal of Mathematics », 58, 1936, pp. 345-363.

[5] Fino a giungere ai problemi concernenti le classi P (polinomiale) e NP (nondeterministico-polinomiale) nell'ambito della teoria della complessità computazionale, negli ultimi anni oggetto di un fervente dibattito. Per una trattazione sintetica e introduttiva, cfr. P. Ferragina, F. Luccio, *Il pensiero computazionale. Dagli algoritmi al coding*, Bologna 2017, pp. 131-153, sulle conseguenze di questo dibattito per la filosofia S. Aaronson, *Why Philosophers Should Care About Computational Complexity*, in J. Copeland, C. Posy, & O. Shagrir (eds.), *Computability: Turing, Gödel, Church, and Beyond*, Cambridge Massachusetts 2013, pp. 261-328.

[6] Una disamina del problema ricca di esempi concreti che mostrano lo stretto legame tra sviluppo hardware e affinamento dell'efficienza degli algoritmi in fase di programmazione si trova in P. Ferragina, F. Luccio, *Il pensiero computazionale. Dagli algoritmi al coding*, cit., cap. 4, in particolare pp. 64-67.

[7] Per esigenze di sintesi dobbiamo tralasciare gli articolati sviluppi e i numerosi alti e bassi che caratterizzano la storia dell'Intelligenza Artificiale. Per una disamina completa cfr. N. J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge 2010.

[8] Per una storia di queste aspettative e delle conseguenti delusioni si rimanda a M. G. Losano, *Storie di automi. Dalla Grecia classica alla Belle Époque*, Torino 1991.

dell'implementazione delle scienze del calcolo[9]. Soltanto dopo il successo di questa iniziativa l'intero programma di ricerca sulla simulazione computazionale assumerà il suggestivo nome di Intelligenza Artificiale, generando una profonda scissione nella comunità dei ricercatori[10]. La proposta dei principali promotori dell'iniziativa, John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon, consisteva al contempo nella riduzione e radicalizzazione della pluralità di approcci in quel momento oggetto di sperimentazione in direzione di un'impostazione marcatamente mentalista, ovvero interessata primariamente allo studio della computazione simbolica, del contenuto semantico, del linguaggio e per questa via alla riproduzione, per analogia o simulazione, dei processi del ragionamento specificatamente umano[11]. Anche se all'epoca il termine algoritmo non aveva assunto ancora il potenziale evocativo che oggi lo caratterizza, è la svolta di Dartmouth ad aver prodotto gran parte dell'immaginario che ancora lo accompagna. Tornando alla definizione iniziale, l'idea era quella di immaginare una procedura algoritmica autonomamente in grado di garantire sia la precisione dell'andamento inferenziale della logica sia la flessibilità necessaria al suo (auto) controllo attraverso il riferimento a un contesto garante dell'adeguatezza della risposta prodotta.

Tuttavia, indugiare ancora su questi riferimenti risulta fuorviante sul piano interpretativo poiché le tecnologie diffuse nella miriade di oggetti e servizi online di cui facciamo uso quotidianamente non sono più progettate a partire dai presupposti e dagli obiettivi di quel programma di ricerca. Sistemi di *machine learning*, *deep learning*, strumenti di analisi che si servono dei famigerati *Big Data* e macchine predittive fanno riferimento a modelli teorici differenti, spesso associati tra loro in architetture ibride, che riprendono una serie di intuizioni sviluppate nel contesto della cibernetica tra gli anni '40 e gli anni '50 del Novecento (reti neurali, modelli del comportamento motorio adattativo, *pattern recognition*, ecc.), messe da parte tra gli anni '60 e '70 e riscoperte a partire dagli anni '80, quando il programma del 1956 aveva mostrato tutti i suoi limiti e gli entusiasmi iniziavano a spegnersi[12]. In questa seconda fase, che dagli anni '80 arriva fino a oggi, il rapporto tra matematica e informatica, dopo alcune battute d'arresto, ha dato luogo a un repentino sviluppo derivante da una maggiore consapevolezza dei problemi connessi alla complessità di calcolo, dunque ai limiti intrinseci della computazione prima che dalle sue potenzialità. Questa trasformazione è stata possibile dall'insorgere al contempo di limiti nella realizzazione di prestazioni raffinate facendo semplicemente appello all'aumento della potenza dei calcolatori e dalla conseguente apertura alle possibilità

[9] N. J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, cit., pp. 73-85.
[10] M. A. Boden, *L'Intelligenza Artificiale*, Bologna 2019, pp. 21-24.
[11] J. McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (consultato il 05/11/21).
[12] Una panoramica aggiornata della molteplicità di approcci oggi utilizzati nel campo dell'Intelligenza Artificiale è presente nel già citato M. A. Boden, *L'Intelligenza Artificiale*, cit.

derivanti dalla connessione dei calcolatori in rete, dall'elaborazione in parallelo e dal vertiginoso aumento dei dati disponibili grazie all'uso del web 2.0. A differenza delle ricerche dei decenni precedenti, si prende coscienza del fatto che quando si tratta di problemi tipici della vita reale, dunque non circoscritti e formulati in condizioni controllate, il contesto non solo entra in gioco influendo sull'adeguatezza della risposta ma lo fa mutando in continuazione. Si tratta allora di costruire una procedura algoritmica in grado di sfuggire al suo ferreo determinismo incorporando contingenza per poter funzionare correttamente. Non tenere conto di questi sviluppi generali della disciplina rischia di minare la comprensione della natura dei suoi prodotti tecnologici.

L'equivoco ancora diffuso tra gli interpreti consiste nel continuare a proiettare sulle odierne tecnologie digitali l'immagine di una mente artificiale intelligente – per di più in assenza di una definizione di mente e di intelligenza univoche – e sul piano dell'interazione rappresentarla come una sorta di Alter Ego animato da uno spirito algoritmico in grado di riprodurre un surrogato della coscienza, via via perfettibile fino all'avvento della cosiddetta singolarità[13]. Questa immagine non tiene conto del decisivo passaggio da un'architettura centralizzata al modello della rete con la conseguente disseminazione dei centri di elaborazione e controllo che rende decisamente più astratta l'analogia con la mente. Inoltre non tiene in debita considerazione il ruolo decisivo giocato dall'interazione macchina-utente per il corretto funzionamento della macchina stessa. Questo non soltanto nel senso elementare per cui solo grazie a un input dell'utente sarà possibile per la macchina svolgere il suo compito generando un output, ma nel senso ben più profondo secondo cui la forma dell'interazione con l'utente diviene essa stessa un dato attraverso il quale la macchina modifica la sua struttura interna generando un output personalizzato. Il riverbero negativo dell'utilizzo, spesso implicito, dello schema interpretativo che associa intelligenza-mente-coscienza si mostra in particolar modo sul piano della critica quando si tratta di analizzare gli effetti che queste tecnologie producono a livello sociale, politico e individuale. Il discorso a causa del suo deficit teorico si mette in condizione di non riuscire a cogliere il funzionamento effettivo del fenomeno e facendo appello a temi così controversi dà non poter essere maneggiati sul piano operativo si organizza inevitabilmente sul piano dei valori dando origine a polarizzazioni ideologiche. Per esempio nello scontro tra tecnofobi e tecnofili[14].

2. Comunicazione e contingenza

Una possibile alternativa interpretativa più coerente con gli attuali sviluppi tecnologici può essere fornita dall'impianto concettuale della teoria dei sistemi di Niklas Luhmann, in particolare facendo riferimento al recente lavoro di Elena Esposito. L'autrice propone di spostare i termini della questione, considerando gli

[13] Qui il riferimento critico è naturalmente R. Kurzweil, *La singolarità è vicina*, Milano 2008.
[14] Cfr. T. Numerico, *Big Data e algoritmi. Prospettive critiche*, Roma 2021.

algoritmi essenzialmente come strumenti progettati per agire adeguatamente come partner della comunicazione e non in vista della riproduzione di processi mentali tipicamente umani come la coscienza e il pensiero: «What makes algorithms socially relevant and useful is their ability to act as partners in communication that produces and circulates information, *independently of intelligence*»[15]. Del resto la riproduzione della coscienza, sottolinea Esposito, non è problematica solo per ragioni tecniche ma pone di fronte a un serio limite epistemologico: quello di dover pianificare e realizzare qualcosa che non può essere conosciuto. Il tratto caratteristico della coscienza, almeno nella sua definizione moderna di 'sistema psichico', consiste infatti sostanzialmente nella sua intrasparenza. È dotato di coscienza chi possiede una prospettiva unica, originale e insostituibile sul mondo, frutto della sua storia, che lo rende agli occhi di un osservatore esterno (e in parte anche di sé stesso) pressoché imprevedibile, ovvero capace in ogni momento di comportarsi in modo potenzialmente diverso da quanto fatto finora. A ciò bisogna poi aggiungere l'ulteriore difficoltà data dal fatto che anche qualora si riuscisse a realizzare un artefatto dotato di coscienza, non è affatto garantito che si verifichino le condizioni sufficienti al suo riconoscimento da parte del creatore[16]. Di converso l'esperienza della coscienza come *black box* è invero alquanto banale, sia che si tratti della propria che di quella altrui. Nessuno ha la possibilità di accedere direttamente ai pensieri degli altri, ognuno ne è consapevole e a partire da questi due presupposti si producono forme di coordinamento altamente mediate attraverso la *comunicazione* che sono in grado di generare senso[17]. Secondo Esposito è al livello di questa forma di mediazione generatrice di senso che intervengono gli algoritmi con la loro responsività, e proprio per questo risultano così efficienti da sembrare intelligenti come le menti.

Questa interpretazione trova conferma nelle modalità di funzionamento degli algoritmi di *machine learning* e *deep learning*, ormai perfezionati e diffusi ovunque tra le tecnologie di uso quotidiano grazie all'enorme crescita del Web. La loro capacità di fornire risposte adeguate a situazioni sconosciute o note solo in parte non sarebbe possibile senza il riferimento a quell'enorme mole di dati non strutturati, i *Big Data*, prodotti dagli utenti che operando nella rete lasciano tracce. Queste procedure computazionali, anche quando sono rese più complesse dall'impiego di diverse tipologie di algoritmi in parallelo o a supporto, non trattano i dati cui attingono come farebbe una mente umana[18]. Al contrario la loro efficienza è garantita proprio dall'abbandono della strategia

[15] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, «Zeitschrift für Soziologie», 46(4), Oldenbourg 2017, p. 253 (corsivo nostro).
[16] E. Esposito, *Digital Prophecies and Web Intelligence*, in M. Hildebrandt & K. de Vries (eds.), *Privacy, Due Process and the Computional Turn. The Philosophy of Law Meets the Philosophy of Technology*, New York 2013, pp. 121-122.
[17] Cfr. N. Luhmann, *La chiusura operativa dei sistemi psichici e sociali*, in N. Luhmann, *Che cos'è la comunicazione?*, a cura di A. Cevolini, Milano-Udine 2018, pp. 33-46.
[18] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 252-253.

imitativa dell'uomo che prevede coscienza di sé e dell'oggetto, intenzionalità e soprattutto comprensione del senso dei dati. Tutte prestazioni queste che incappano nei problemi di autoriferimento appena evidenziati. Sgravati dunque dall'onere della comprensione questi programmi operano invece sui dati come fossero oggetti mettendoli in relazione, estrapolandone caratteristiche, tendenze, pattern ecc. In questo modo sono in grado di superare una serie di problemi tipici del trattamento dell'informazione finora conosciuto, su tutti le problematiche relative all'eccesso di dati[19]. Diversamente da quanto avveniva in passato, ora più grande è la quantità di dati disponibile e maggiore sarà l'efficienza della procedura, più alto sarà il livello di destrutturazione e vaghezza di questa massa e maggiori saranno le informazioni che da essa il sistema sarà grado di estrapolare (*data mining*) rendendosi al contempo più flessibile per compiti futuri[20]. Inoltre, data la natura di questi algoritmi, è spesso impossibile avere il pieno controllo delle procedure di discriminazione messe in atto a causa del volume dei dati e della velocità delle computazioni svolte su più livelli e in parallelo[21]. Del resto se una ricostruzione *a priori* fosse possibile non avremmo bisogno del tipo di prestazione che richiediamo a questi algoritmi. A tal proposito si possono ricordare i recenti sviluppi nel campo dei traduttori automatici o dell'*Internet of Things* che hanno spostato radicalmente il focus della questione rispetto a come impostata dal problema della 'stanza cinese' di Searle. L'aspetto fondamentale consiste dunque nel fatto che «these structures produce meanings, but do not depend on the intention or on the consciousness of anyone (they have not been thought by anyone); they result *a posteriori* from the very functioning of the net»[22].

Si tratta a questo punto di stabilire come sia possibile la produzione di senso attraverso questo nuovo modo di elaborare i dati, che non prevede la comprensione e che tuttavia è in grado di rifornire la società di significati sempre nuovi e imprevedibili. Esposito ipotizza che a generarsi sia un nuovo tipo di interazione comunicativa mediata da algoritmi che definisce *Artificial Communication*[23]. Si tratterebbe di una mediazione poiché i dati sui quali la procedura di computazione opera sono in ultima istanza ancora il prodotto dell'attività degli utenti che alimentano il sistema, tuttavia questi algoritmi per le loro modalità di funzionamento sembrano essere in grado di mettere in atto dinamiche tipiche della comunicazione tra esseri umani. Quando interagiamo

[19] Ivi, p. 251.
[20] Per una descrizione accessibile del funzionamento di una famiglia di algoritmi molto importanti per la trattazione dei *Big Data*, quelli utili alle procedure di ordinamento, cfr. P. Ferragina, F. Luccio, *Il pensiero computazionale. Dagli algoritmi al coding*, cit., cap. 5, in particolare pp. 94 e ss, dove si arriva mostrare il funzionamento del complesso algoritmo MERGE_SORT A PIÙ VIE a partire da problemi di ordinamento più semplici.
[21] J. Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, «Big Data & Society» January–June 2016, pp. 1-12, in particolare pp. 4-5.
[22] E. Esposito, *Digital Prophecies and Web Intelligence*, cit., p. 123.
[23] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 253 e ss.

con queste tecnologie, sostiene l'autrice, siamo sorpresi dalle risposte che riceviamo, ne traiamo informazioni che sono utili solo per noi e che non esistevano prima della formulazione della nostra specifica richiesta. Siamo certi di fare qualcosa di più complesso rispetto al mero attingimento di materiali da un archivio, ma al contempo sappiamo che la fonte della nostra informazione elabora senza pensare e non può comprendere la comunicazione nella quale sembra coinvolta. La situazione, nota Esposito, è paragonabile a quel che accade tra esseri umani quando si verifica un'errata attribuzione dell'intenzione di comunicare che nonostante ciò dà avvio ugualmente a uno scambio che produce informazione[24]. Per esempio, un movimento della mano viene erroneamente inteso come un fuggevole saluto e ciò produce un senso di sollievo suggerendo una riappacificazione dopo un litigio passato, ingenerando così una serie di conseguenze nuove e imprevedibili. Questi casi di comunicazione unilaterale, in cui il partner comunica senza saperlo, sono rari tra le persone e devono rimanere isolati altrimenti insorgerebbero gravi problemi di coordinamento. Tuttavia con la diffusione della comunicazione tramite algoritmi non si tratterebbe più di eventi eccezionali ma di modalità consuete di acquisire informazioni, con conseguenze ancora tutte da valutare. Quando comunichiamo con un assistente vocale o con una *chatbot* stiamo ottenendo informazioni nuove e differenti sia da quelle inserite dai programmatori nei database, o nelle fonti ai quali tali sistemi attingono, sia da quelle di altri utenti che muovono richieste simili ma a partire da una storia unica e insostituibile di interazioni con la macchina.

Fornire una spiegazione che renda conto della diffusione di questo tipo di coordinamento così improbabile e insieme così efficace necessita, secondo Esposito, di una teoria della comunicazione più precisa sul piano operativo. La sua concezione comune basata sulla condivisione di contenuti psicologici e risorse interpretative, per mezzo delle quali sarebbe possibile il trasferimento di pensieri tra un mittente e un ricevente, risulta eccessivamente vaga per il suo riferimento alle coscienze coinvolte e per l'implicito presupposto, non spiegato, secondo cui i pensieri dell'interlocutore sarebbero in qualche modo percepiti come tali. Diversamente i casi esaminati mostrano un'interazione descrivibile come comunicativa ma nella quale la risposta che riceve l'utente manca sia di un possibile trasferimento di pensieri che della condivisione di risorse interpretative comuni tra lui e la macchina. L'autrice suggerisce in alternativa di inquadrare la questione facendo ricorso alla teoria dei sistemi di Luhmann[25]. In questo quadro teorico il vantaggio è rappresentato dall'abbandono della metafora del trasferimento in favore di una descrizione che vede la comunicazione come una realtà emergente. Essa si genera solo quando si realizzano contestualmente per un osservatore tre tipi di selezione: l'informazione, l'atto del comunicare [*Mitteilung*] e la comprensione della differenza tra le prime due[26]. L'inizio della comunicazione non è dunque identificabile con il semplice verificarsi dell'atto di

[24] Ivi, p. 256.
[25] Ivi, pp. 254-255.
[26] N. Luhmann, *Che cos'è la comunicazione?*, cit., pp. 21-22.

emissione. Che qualcuno scriva qualcosa, pronunci una frase o compia un gesto di saluto completamente solo non basta affinché si possa sostenere che ci sia stata comunicazione. Essa può prendere avvio soltanto se questi atti – scrivere su un foglio, emettere suoni, muoversi in un certo modo – sono compresi come atti finalizzati alla comunicazione. Il punto di partenza è dunque la comprensione, in un certo senso si parte dalla fine. Tuttavia tale comprensione avviene indipendentemente da quello che l'interlocutore aveva in mente nell'atto della sua emissione – e non potrebbe essere altrimenti considerando che nessuno può accedere direttamente ai pensieri altrui. Secondo questa impostazione intendere o fraintendere i pensieri dell'interlocutore è ininfluente ai fini dell'avvento e del proseguimento della comunicazione, quel che conta perché essa si generi è che qualcuno colga la differenza tra il valore del contenuto informativo e le ragioni per le quali esso è stato emesso: «communication exists not when somebody says something, but when somebody realizes that someone else said something»[27]. Tutto ciò significa attribuire all'interlocutore la capacità di compiere una selezione, ovvero di decidere se dare avvio a una comunicazione, come e in quale occasione. Per questo comunicare è sempre un'esperienza profondamente diversa dal semplice percepire. Nella percezione non viene attribuita alcuna capacità selettiva all'oggetto osservato, che è assunto solo come contenuto di coscienza dall'osservatore. Ciò che è percepito tuttavia può sempre rientrare nel circolo della comunicazione ma solo a condizione di divenire l'oggetto di un atto di emissione che dovrà essere comprensibile per qualcuno[28]. Nella concezione luhmanniana quindi la coscienza e la comunicazione sono strettamente dipendenti ma al contempo radicalmente autonome sul piano operativo. La coscienza è condizione necessaria affinché siano possibili pensieri oggetto di atti di comunicazione, per prestare attenzione a quel che viene detto o scritto e per comprenderlo, ma tale comprensione resta sempre indipendente dai pensieri dell'emittente. Ognuno comprende, ovvero distingue, l'atto della comunicazione e l'informazione in maniera autonoma a partire dai propri riferimenti, interessi e conoscenze pregresse, senza poter attingere direttamente ai pensieri altrui. È infatti sempre possibile che quanto venga comunicato sia inteso in maniera totalmente differente dalle intenzioni iniziali o che si generino delle informazioni completamente nuove che non erano presenti nelle coscienze di alcun partecipante prima della comunicazione. Ecco perché anche l'informazione nella teoria di Luhmann è una selezione, ovvero una differenza specifica che viene prodotta nella comunicazione stessa, e non un contenuto trasmissibile e immagazzinabile[29]. La comunicazione è in definitiva una forma di coordinamento che si genera a partire dalla contingenza e imprevedibilità delle selezioni di un Ego e di un Alter che agiscono reciprocamente sapendo che

[27] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 254.
[28] N. Luhmann, *Che cos'è la comunicazione?*, cit., p. 22.
[29] N. Luhmann, *Sistemi sociali. Fondamenti di una teoria generale*, a cura di A. Febbrajo, Bologna 1990, pp 155-157.

ognuno di loro sceglierà cosa fare solo sulla base di quello che farà l'altro, a causa della rispettiva inaccessibilità. Tale condizione viene definita *doppia contingenza* e secondo la teoria dei sistemi sociali è caratteristica di ogni interazione comunicativa[30].

3. Comunicazione Artificiale

Procedendo da questa definizione Esposito può fornire una descrizione originale dell'interazione utente-algoritmo, domandandosi che cosa accada alla doppia contingenza in questi casi. Sebbene ciò implichi una significativa reinterpretazione del test di Turing in chiave marcatamente funzionalista, l'utilizzo della nozione luhmanniana di comunicazione sposta l'attenzione sugli effetti dell'atto di emissione rispetto alla natura di colui che lo ha emesso. «What matters is whether the interaction with the machine has the features of communication with a contingent autonomous partner»[31], ovvero se nell'interazione si verifica la sintesi tra i tre tipi di selezione: informazione, atto del comunicare e comprensione. La sola comprensione dell'informazione non è sufficiente, l'utente deve essere anche nella condizione di attribuire alla macchina l'intenzione di voler comunicare quell'informazione, non altro, in quella specifica occasione. In particolare ciò che va accertato è «whether in the interaction with algorithms a condition of double contingency arises in which each partner is oriented towards the indeterminacy of his or her counterparty and specific information is produced»[32].

Se prestiamo attenzione a questa considerazione e la confrontiamo con quanto accaduto nello sviluppo dell'intelligenza artificiale a seguito della svolta degli anni '80, descritta nel primo paragrafo, riscontriamo una certa somiglianza di struttura tra la situazione di doppia contingenza e la progressiva complicazione del rapporto tra programma e contesto sul piano della progettazione. In entrambi i casi ci confrontiamo con problemi di controllo della contingenza tuttavia l'analogia si ferma qui in quanto le due situazioni risultano radicalmente differenti. Nel caso della doppia contingenza siamo di fronte a una condizione estrema di mancanza di controllo nella quale i due interlocutori sono costretti ad assumere aspettative in merito all'azione altrui e così facendo pongono le condizioni per la strutturazione di punti di aggancio reciproco convertendo l'infinita variabilità che caratterizza il loro modo di interagire con l'ambiente in contingenza gestibile per entrambi[33]. Nel caso del programma abbiamo invece una condizione di totale determinazione e controllo poiché la struttura algoritmica opera soltanto sulla base di sé stessa e della propria architettura

[30] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 255. Cfr. N. Luhmann, *Sistemi sociali. Fondamenti di una teoria generale*, cit., pp. 205e ss.
[31] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., pp. 255-256.
[32] Ivi, p. 257.
[33] N. Luhmann, *Sistemi sociali. Fondamenti di una teoria generale*, cit., pp. 462-481.

interna. A quest'ultima è però necessario fornire variazioni dall'esterno affinché possa produrre output contingenti e utili sul piano informativo, ovvero in relazione a un mutevole contesto di utilizzo. Ci si trova dunque costretti ad affrontare il problema su due livelli che non vanno confusi: il primo è quello più generale rappresentato dalla doppia contingenza, il piano della comunicazione dal quale è possibile generare informazioni, senso e coordinamento sociale; il secondo è quello del funzionamento interno dell'algoritmo progettato al fine di fornire una responsività che sia adeguata all'interazione comunicativa, quindi al primo livello. «We must ask whether and how algorithms can become contingent and hence reflect the contingency of the users, and how this contingency is controlled in the communication process»[34]. L'obiettivo dei programmatori in questa ultima fase di sviluppo sarebbe allora quello di produrre mancanza di controllo in modo controllato, così da lasciare aperta la possibilità di 'sorpresa informativa' nei confronti dell'utente e rendendo l'enorme capacità computazionale a disposizione funzionale alla produzione di senso[35].

Esposito a partire da queste premesse riconosce tre livelli di produzione di contingenza corrispondenti a tre tipologie di macchine progressivamente più complesse capaci di generare differenti effetti sul piano informativo. Il primo tipo è una sorta di grado zero rappresentato dalle macchine-oggetto tradizionali come l'orologio. In questo caso non possiamo parlare di interazione comunicativa ma soltanto di percezione poiché sebbene guardando l'orologio veniamo a conoscenza di qualcosa che non sapevamo in precedenza l'informazione che otteniamo non può essere attribuita ad alcun atto di emissione contingente. Non avrebbe senso interrogarsi sulle motivazioni che hanno spinto l'orologio a segnare l'ora poiché esso è una macchina autonoma ma determinata. Per questo non è possibile assumere l'indeterminatezza di Alter come riferimento per strutturare la nostra, come avviene nell'interazione comunicativa, poiché essa non si dà mai tranne che in un unico caso: quando il meccanismo si rompe[36].

Il secondo tipo di tecnologia include invece dispositivi in grado di reagire alle azioni dell'utente comportandosi in maniera variabile e adeguata, come se ne comprendessero le intenzioni. Esposito fa l'esempio dei robot giocattolo, ma si potrebbe allargare il riferimento a un'ampia classe di congegni utilizzati in vari contesti. In questi casi siamo portati a considerare il dispositivo come partner della comunicazione poiché gli algoritmi che ne guidano il funzionamento sono finalizzati alla produzione di risposte tali da essere interpretate dall'utente come intelligenti. Le informazioni prodotte sono effettivamente sorprendenti e differenti da quanto conosciuto prima dell'interazione. Inoltre, data l'imprevedibilità del comportamento che lo caratterizza, è possibile attribuire

[34] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 257.
[35] *Ibid*.
[36] Ivi, pp. 256-257. Cfr. la distinzione tra macchine banali e macchine non banali in H. von Foerster, *Cibernetica ed epistemologia: storia e prospettive*, in G. Bocchi, M. Ceruti (a cura di), *Le sfide della complessità*, Milano 2007, pp. 88-116.

al dispositivo il corrispondente atto di emissione contingente (naturalmente se la gamma di reazioni è sufficientemente ampia). Tuttavia la portata di queste tecnologie è limitata poiché quel che viene realmente prodotto è una *virtualizzazione della contingenza*: «algorithms allow the machine to react to the behavior of the user, and this in turn allows the user to project onto the machine his or her own contingency and meanings»[37]. Le informazioni che vengono acquisite non sono in realtà sconosciute, bensì aspetti già noti all'utente che il dispositivo permette di osservare da un punto di vista differente senza che si incontri davvero la prospettiva di qualcun altro ma sempre e solo la propria. «Contingency is multiplied because it can be observed from the outside. The user experiences a kind of 'virtualization' of his or her own contingency, as observed in a mirror that generates a virtual image»[38]. Nell'interazione con questi dispositivi ci si trova come di fronte a uno specchio: l'oggetto che viene riflesso, in questo caso la contingenza dell'utilizzatore, non è un'illusione, esiste ma è ora possibile osservarlo da molteplici prospettive che la superficie riflettente è in grado di mettere a disposizione. Da questo tipo di esperienza si possono ricavare nuove informazioni inaspettate, soprattutto mettersi in condizione di osservare come si viene osservati da altri e farsi sorprendere da quel che se ne ricava. Nonostante ciò la fonte dell'immagine riflessa rimane sempre la stessa riducendo drasticamente la varietà e il livello di imprevedibilità di quanto può accadere. «The observer interacts with him, or herself, not with another observer»[39] e dunque non si può propriamente parlare di comunicazione poiché la situazione in cui ci si trova è soltanto un surrogato della doppia contingenza, la produzione di nuova informazione per l'elaborazione di senso risulta possibile ma solo entro certi limiti.

In ultimo, la terza tipologia trattata da Esposito comprende quelle tecnologie le cui prestazioni sono generate da algoritmi capaci di apprendere in assenza di supervisione da parte del programmatore (*deep learning*), attingendo alla massa in costante trasformazione di dati proveniente dall'utilizzo del web. In particolare, l'autrice fa qui riferimento alla famiglia di algoritmi diffusisi a seguito della nascita di Google e che adottano la stessa logica di funzionamento dell'algoritmo del motore di ricerca, noto col nome di PageRank[40]. La caratteristica di questi algoritmi consiste nel mappare i comportamenti degli utenti in rete e sulla base di tali informazioni (metadati) fornire risposte adeguate alle richieste ricevute (*query*). Il web è da loro scandagliato come un vastissimo ipertesto fitto di rimandi dove le correlazioni e i contenuti estrapolabili dall'analisi della struttura dei moltissimi link divengono il risultato del funzionamento del sistema e non la sua premessa. L'obiettivo è infatti stabilire criteri utilizzabili per generare risposte

[37] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 257.
[38] Ivi, p. 258.
[39] *Ibid*.
[40] Per un'analisi introduttiva al suo funzionamento, cfr. P. Ferragina, F. Luccio, *Il pensiero computazionale. Dagli algoritmi al coding*, cit., pp. 155-179.

adeguate a domande che non sono prevedibili. Tali criteri non possono dunque riguardare direttamente i contenuti, a differenza di quanto avviene per esempio per il catalogo di un archivio. Ciò che l'algoritmo innesca è allora un meccanismo ricorsivo «in which the users use the algorithm to get the information, their research modifies the algorithm, and the algorithm then impinges on their subsequent searches for information»[41]. L'attenzione dei programmatori può così concentrarsi esclusivamente nel rendere l'algoritmo capace di mutare sé stesso e saranno poi le modalità d'uso degli utenti a fornirgli il materiale necessario per orientarsi. «What it does is use the results of our minds to give directions, and then produce information that none of us had in mind»[42].

Sul piano comunicativo in questo caso la situazione si complica poiché la contingenza con cui entra in relazione l'utente non è più soltanto la propria che viene riflessa, e non è nemmeno quella della macchina poiché l'algoritmo per sua natura non è in grado di trattare l'incertezza ma solo eseguire istruzioni[43]. L'algoritmo tuttavia importa questa contingenza dall'esterno e la utilizza per produrre informazioni nuove. Queste informazioni pur derivando da quelle prodotte dagli utenti del web non possono essere attribuite direttamente a loro, esse sono letteralmente prodotte dal sistema che computa il materiale a disposizione estrapolandone correlazioni che nessuno, data la mole di dati e la velocità di calcolo, avrebbe mai potuto conoscere senza il suo aiuto. «The algorithm reflects and represents the perspective of other observers, and users observe through the machine a re-elaboration of other users' observations»[44]. L'effetto che ne risulta è che l'utente si trova ora effettivamente di fronte a una prospettiva diversa dalla sua, a un atto di emissione attribuibile solo al dispositivo che è capace di reagire alla sua contingenza in modo altrettanto indeterminato selezionando così informazione. Tuttavia l'atto di emissione del dispositivo risulta scollegato dall'informazione selezionata a differenza di quanto accade nella situazione di doppia contingenza. Sebbene l'utente associ a questo atto di emissione un'informazione l'algoritmo che lo produce compie scelte sulla base dei dati disponibili ma lo fa seguendo criteri che non contribuiscono alla sua comprensione. «The real innovation in the communication with algorithms is that understanding is no longer oriented to the meaning of an utterance»[45]. Ciò significa che quando si interagisce con questi dispositivi viene a mancare la possibilità di osservare come un altro osservatore osserva il mondo poiché il dispositivo non può osservare alcunché ma è solo in grado di mettere a disposizione ciò che anche gli altri potrebbero osservare nella comunicazione[46]. In termini luhmanniani, nel primo caso è possibile l'insorgere di un'osservazione di secondo ordine, il cui oggetto è il *modo* di osservare di un

[41] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 259.
[42] Ivi, p. 260.
[43] Si veda la definizione fornita all'inizio dell'articolo.
[44] Ivi, pp. 259-260.
[45] *Ibid.*
[46] Ivi, p. 260.

Alter Ego che sa osservare il mondo come me. Nel secondo caso invece abbiamo solo la possibilità di moltiplicare le nostre osservazioni di *ciò* che potrebbe essere osservato attraverso le osservazioni di altri osservatori, mettiamo in atto sempre e solo un'osservazione di primo ordine[47]. Secondo Esposito quella che si genera è una *doppia contingenza virtuale*:

> The algorithm reflects and elaborates the indeterminacy of all participants, and each user faces the contingency of all the others, which is infinitely surprising and informative. It is still virtual contingency, but reflected in a mirror in which everyone sees not him- or herself but the other observers communicating[48].

Secondo l'autrice è possibile riconoscere in questa particolare interazione una nuova forma di comunicazione distinta rispetto a quella che si verifica tra persone. Essa è definita comunicazione artificiale poiché l'algoritmo che viene programmato per funzionare come partner della comunicazione costruisce letteralmente una nuova prospettiva a partire dai dati aggregati di altri utenti, sfruttandoli per generare la contingenza necessaria. L'utilizzatore interagendo con la macchina non si troverà così a comunicare né con il produttore né con gli utenti del web che hanno fornito il materiale da elaborare ma con un partner artificiale i cui atti di emissione e informazioni sono attribuibili soltanto a lui. È dunque possibile orientarsi a partire dall'indeterminatezza del partner anche se, come abbiamo visto, quest'ultimo non è propriamente un Alter Ego. Quella che esso ci propone è sorpresa informativa derivante da una indeterminatezza infinitamente più vasta di quella generata dalla semplice riflessione della propria contingenza come accadeva per i robot giocattolo. Tuttavia la contingenza prodotta in questo caso è qualitativamente diversa rispetto a quella con la quale ci confrontiamo quando un altro osservatore ci offre il suo modo di vedere le cose e non soltanto ciò che ha o potrebbe osservare. In altri termini, l'algoritmo non è in grado di farsi un'idea e proporcela, può soltanto costruirla senza comprenderla orientandosi, per l'adeguatezza del risultato, ai dati che ha a disposizione seguendo un processo di apprendimento tramite rinforzo. Come fa notare Esposito riprendendo le ricerche David Silver e Demis Hassabis sul programma AlphaGo[49], questi sistemi non sono in grado di immaginare la prospettiva dell'utente o fare previsioni, non possiedono nemmeno la consapevolezza di ciò che sta accadendo, più semplicemente «the algorithm uses reinforcements to

[47] Sulla distinzione tra osservazione di primo e secondo ordine cfr. N. Luhmann, *Introduzione alla teoria dei sistemi*, a cura di S. Magnolo, Lecce 2018, pp. 122-141.
[48] E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, cit., p. 260.
[49] Sviluppato da Google DeepMind, si tratta di un software divenuto noto al grande pubblico per aver battuto il campione mondiale di Go, Lee Sedol, nel 2016 utilizzando un'architettura ibrida di algoritmi di apprendimento supervisionato e non-supervisionato. La struttura del software ha rappresentato una svolta nello sviluppo di algoritmi capaci di autoapprendimento ed è ora impiegata in molti contesti, cfr. D. Silver e D. Hassabis, *Mastering the Game of Go with Deep Neural Networks and Tree Search*, «Nature», vol. 529, n. 7587, 28 gennaio 2016, pp. 484-489.

calculate in its own way an evaluation function that indicates which moves to make»[50]. Nel compiere questa computazione il sistema non apprende qualcosa di nuovo sul mondo ma impara soltanto a perfezionare sé stesso e la sua struttura di funzionamento generando infinite versioni di sé e imparando dai risultati ottenuti dai sé autoprodotti. Nessun algoritmo può generare possibilità che non siano già implicitamente contenute nei dati che ha a disposizione e la contingenza che viene elaborata e prodotta dal sistema può solo risultare dall'interazione tra esseri umani e algoritmi[51]. Nonostante questo limite l'effetto che ne deriva è pervasivo e dalle conseguenze ancora difficili da decodificare: è ora disponibile una forma di comunicazione enormemente più informativa generabile dall'interazione con partner non umani che non imitano i processi cognitivi.

Conclusione

In conclusione, quel che ancora resta da chiarire alla luce di questo nuovo modo di trattare i dati e produrre interazioni comunicative sono gli effetti che possono generarsi sul piano complessivo della comunicazione sociale dal quale si produce la costruzione di senso. L'immissione continua di informazioni la cui comprensione risulta scollegata dal significato degli atti di emissione attribuiti all'interlocutore produrrà certamente conseguenze sul piano della coordinazione comunicativa, sulla genesi di significati condivisi e sulle loro caratteristiche ma ancora non sembra possibile capire come. Restringendo la questione, Esposito riconduce a questo problema la nota e controversa tesi di Chris Anderson secondo la quale il livello raggiunto dai processi di digitalizzazione delle informazioni renda possibile un superamento delle strutture ipotetico-deduttive del ragionamento scientifico in vista di un sapere costruito esclusivamente sulla scoperta di correlazioni e forme proprie del fenomeno da indagare[52]. Secondo Anderson, il fenomeno non sarebbe più conosciuto a partire dalla formulazione di ipotesi preliminari ma verrebbe direttamente reso visibile, ovvero dimostrato, grazie all'analisi algoritmica di quell'enorme quantità di dati resi disponibili dalla rete. Una massa di dati trattabili talmente grande da essere considerata potenzialmente esaustiva in relazione alla conoscenza di tutto ciò che c'è. Secondo l'autrice, una tesi di questo tipo fa appello a un presupposto implicito senza dubbio fallace, ma che risulta significativo in quanto fornisce indicazioni in merito al genere di conseguenze che l'immissione di informazioni a partire da questo nuovo modo di trattare i dati può generare sulla costruzione della conoscenza. Si tratta dell'assunto secondo cui i dati a disposizione del sistema computazionale siano completi o talmente vasti da poter essere considerati sufficienti per rispondere a qualsiasi domanda 'su ciò che c'è', ribaltando il

[50] Ivi, p. 261.
[51] Ivi, pp. 262-263.
[52] C. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, «Wired», 06/2008, 16.

rapporto tra problema e soluzione[53]. Le soluzioni sarebbero tutte già presenti nei dati in attesa di essere scoperte, una volta trovato il problema corrispondente da implementare nella macchina. Tuttavia, anche ammesso che una tale massa di dati sia effettivamente aggregabile, essa non potrebbe che riferirsi sempre e solo a un'infinita serie di comportamenti degli utenti estrapolati dalla rete nel passato. E di conseguenza una simile tesi dovrebbe necessariamente assumere una posizione ontologica forte al punto da considerare le premesse contenute nel passato sufficienti alla determinazione univoca del futuro[54]. Indipendentemente dalla plausibilità di queste conclusioni sul piano strettamente epistemologico, il punto è che queste tecnologie come abbiamo visto non funzionano in questo modo – e un indizio è fornito dal fatto che spesso si sbagliano. Esse traggono la loro possibilità di generare informatività nella comunicazione con l'utente dalla contingenza generata dalle interazioni sociali nel loro complesso, dal quale vengono estrapolati dati a partire da eventi, ovvero i nostri comportamenti, senza che vi sia un diretto accesso al mondo. Ma se è così ciò che non è tenuto in debita considerazione nell'interpretazione delle informazioni derivanti da questi sistemi è che qualsiasi sia il responso che otterremo in merito alla risoluzione di un problema, soprattutto in ambito sociale, esso sarà costitutivamente deficitario poiché il sistema non sarà in grado di tenere in considerazione le nostre possibili reazioni messe in campo per confermare o evitare l'effetto prospettato dalla computazione algoritmica.

> The machine foresees only what can be expected on the basis of past logic, even if (as in the models of risk management) it considers the multiplicity of possible future courses, admitting that today one cannot know which of them will occur (open future). What escapes, however, is precisely the real future: the one that will become real reacting to our attempts to predict it and is therefore inherently unpredictable (it is none of those considered by the models). The circularity of the forecast tends to self defeat (as in the case of financial markets) or in any case leads to enormous management problems. Not recognising communication, the logic of the machine does not recognise the inherent indeterminacy of the social world, which cannot be known because it continuously reproduces its own uncertainty[55].

L'indeterminazione caratteristica del mondo sociale è generata proprio dal tipo di interazione circolare che solo la condizione di doppia contingenza può rendere possibile ma questa richiede una serie di capacità che la macchina può solo riprodurre in modo virtuale attraverso un complesso gioco di riflessi. D'altro canto se questo è ciò che non dovremmo aspettarci dal tipo di informatività che deriva dagli algoritmi, quel che in effetti sembra accadere apre a una serie di problematiche altrettanto ricche di conseguenze. Queste informazioni hanno infatti l'effetto di stimolare processi riflessivi in grado di aumentare in maniera significativa la capacità di apprendimento degli utenti. La necessità di reagire

[53] E. Esposito, *Digital Prophecies and Web Intelligence*, cit., pp. 127-128.
[54] Ivi, p. 133.
[55] Ivi, p. 134.

a un crescente numero di variabili ugualmente possibili incrementa la nostra capacità di produrre soluzioni creative in contesti circoscritti. Non a caso l'effetto prodotto dalla nascita di algoritmi in grado di sconfiggere i giocatori di scacchi o di Go è stata l'accelerazione del progresso delle tattiche di gioco elaborate dai giocatori umani. Quando questo fenomeno si verifica in maniera diffusa in molteplici contesti la complessità della comunicazione sociale aumenta e con essa le possibilità di apprendimento dei partecipanti. In senso generale si può affermare che aumentano le opportunità di riflessione e dunque di revisione delle strutture che garantiscono la stabilità del senso complessivo. Al contempo però le informazioni, le teorie e tutti i riferimenti condivisi divengono sempre più instabili, poiché la possibilità di generare alternative plausibili cresce con il crescere della possibilità di riflessione. Il prezzo di una maggiore capacità di comprensione e trattamento della complessità del mondo è l'aumento della percezione di un rischio diffuso. Diviene allora sempre più difficile mettere in relazione passato e futuro per assumere decisioni nel presente[56].

Alberto Giustiniano
Università degli studi di Padova
✉ alberto.giustiniano@gmail.com

Bibliografia

Aaronson, S. 2013. *Why Philosophers Should Care About Computational Complexity*, in J. Copeland, C. Posy, & O. Shagrir (eds.), *Computability: Turing, Gödel, Church, and Beyond*, Cambridge Mass, MIT Press.

Anderson, C. 2008. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, «Wired», 06, 16.

Boden, M. A. 2018. *Artificial Intelligence. A Very Short Introduction*, Oxford, Oxford University Press, tr. it. *L'Intelligenza Artificiale*, a cura Diego Marconi, Bologna, il Mulino 2019.

[56] Si segnala che a seguito della consegna del presente lavoro è stato pubblicato dall'autrice il saggio *Artificial Communication. How Algorithms Produce Social Intelligence*, MIT Press, Cambridge-London 2022 nel quale gli argomenti qui illustrati trovano la loro trattazione completa. Tale volume, ci auguriamo, dovrebbe dare maggiore risalto a questa proposta interpretativa, ancora pressoché sconosciuta al dibattito scientifico su questi temi. Una prima discussione in merito risale al 2021 ed è riportata in «*Costructivist Foundations*», 16, 3, July 2021, pp. 357-380 a seguito di un'intervista con l'autrice curata da Katrin Sold e Bénédicte Zimmermann. Gli interventi, tuttavia, si concentrano prevalentemente sulla struttura e il funzionamento della teoria dei sistemi di Luhmann, sottesa alla proposta della Esposito, e non direttamente sulle conseguenze teoriche del passaggio da un'analisi orientata all'intelligenza verso un'analisi orientata alla comunicazione (fatta eccezione dell'intervento di Wiebke Loosen e Armin Scholl, pp. 369-371).

Burrell, J. 2016. *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, «Big Data & Society» January–June.

Church, A. 1936. *An Unsolvable Problem of Elementary Number Theory*, in «American Journal of Mathematics », 58.

Esposito, E. 2013. *Digital Prophecies and Web Intelligence*, in M. Hildebrandt & K. de Vries (eds.), *Privacy, Due Process and the Computional Turn. The Philosophy of Law Meets the Philosophy of Technology*, New York, Routledge.

Esposito, E. 2017. *Artificial Communication? The Production of Contingency by Algorithms*, «Zeitschrift für Soziologie», 46(4), Oldenbourg.

Esposito, E., Sold, K., Zimmermann, B. 2021. *Systems Theory and Algorithmic Futures: Interview with Elena Esposito*, «*Costructivist Foundations*», 3 ,16, July.

Esposito, E. 2022. *Artificial Communication. How Algorithms Produce Social Intelligence*, Cambridge-London, MIT Press.

Ferragina, P., Luccio, F. 2017. *Il pensiero computazionale. Dagli algoritmi al coding*, Bologna, il Mulino.

Foerster, H. (von). 2007. *Cibernetica ed epistemologia: storia e prospettive*, in G. Bocchi, M. Ceruti (a cura di), *Le sfide della complessità*, Milano, Bruno Mondadori.

Frixione, M., Palladino, D. 2009. *Funzioni, macchine, algoritmi. Introduzione alla teoria della computabilità*, Roma, Carocci.

Hill, R. K. 2016. *What an Algorithm Is*, «Philosophy & Technology». 29, Berlin-New York.

Kurzweil, R. 2005. *The Singularity Is Near: When Humans Transcend Biology*, New York, Viking, tr. it. *La singolarità è vicina*, Milano 2008

Losano, M. G. 1991. *Storie di automi. Dalla Grecia classica alla Belle Époque*, Torino, Einaudi.

Luhmann, N. 1984. *Soziale Systeme. Grundriß einer allgemeinen Theorie*, Frankfurt am Main, Suhrkamp, tr. it. *Sistemi sociali. Fondamenti di una teoria generale*, a cura di A. Febbrajo, Bologna 1990

Luhmann, N. 2011. *Einführung in die Systemtheorie*, Hrsg. v. Dirk Baecker 334 Seiten, 9 Abb., Kt, 8. Auflage, tr. it. *Introduzione alla teoria dei sistemi*, a cura di S. Magnolo, Lecce, Pensa Multimedia.

Luhmann, N. 2018. *Die operative Geschlossenheit psychischer und sozialer Systeme*, in *Soziologische Aufklärung 6. Die Soziologie und der Mensch*, Springer Fachmediem Wiesbaden GmbH, tr. it. *La chiusura operativa dei sistemi psichici e sociali*, in N. Luhmann, *Che cos'è la comunicazione?*, a cura di A. Cevolini, Milano-Udine 2018.

McCarthy, J. et al., 1955. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Nilsson, N. J. 2010. *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge, Cambridge University Press.

Numerico, T. 2021. *Big Data e algoritmi. Prospettive critiche*, Roma, Carocci.

Silver, D., Hassabis, D. 2016. *Mastering the Game of Go with Deep Neural Networks and Tree Search*, «Nature», vol. 529, n. 7587, 28 January.

Turing, A. M. 2004. *On Computable Numbers with an Application to the Entscheidungsproblem*, in J.B. Copeland (ed.), *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*, Oxford, Oxford University Press.

Vardi, M. 2021. *What is an Algorithm?*, «Communications of the ACM», vol. 55, 3, New York.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770029

Articoli/1

Notes for a Political Epistemology of Algorithms

Massimiliano Badino 0000-0002-3521-9454

Articolo sottoposto a doppia blind peer review. Inviato l'11/11/2022. Accettato il 17/11/2022.

Algorithms are everywhere: they watch us, they give us advice, sometimes they take autonomous decisions. They are agents mediating between reality and us by virtue of their capability to treat huge amounts of data and to see patterns that are invisible to us. It is pretty clear that they are political actors by virtue of their epistemic features. In this article, I try to put together these two points and outline a political epistemology of algorithms. Firstly, I define political epistemology as concerned with epistemic performances that are essentially situated and therefore political all the way down. Secondly, I show that, in this sense, an epistemic analysis of algorithms cannot help being a piece of political epistemology.

***

Introduction

If there is a feature that really sets our age apart from other instalments of human culture is its reliance on artificial and automatic agents. Basically, every aspect of our present life is one way or another, influenced by artificial intelligence or other kinds of automation. Culture has not remained idle before this development. The last ten-odds years have witnessed a veritable boom of studies on various facets of artificial intelligence. Predictably, the lead has been taken by analyses of the ethical and political effects of the new technologies. I said predictably because, given their ubiquity, it is particularly urgent to develop an ethical, legal, and political framework to govern them.

However, most of this intellectual effort seems still to develop within disciplinary comfort zones. Arguably, the most striking feature of the new technologies is their ability to produce a certain kind of knowledge (more on this point later). This is an epistemic feature. Now, this ability has a number of ethical and political consequences. Although the connection between these two facets is strong and evident, surprisingly enough, scholars interested in the epistemic side are hardly interested in the political one and vice versa. Admittedly, combining the two features is not easy. Still, it is necessary to do so because an exclusively

epistemological focus runs the long-term risk of fostering the idea of neutrality of algorithms, while a politics-only focus might lack the epistemological flesh to have a real philosophical bite. In this article, I attempt to lay down some traits that I think a political epistemology of algorithms should display. My aim is purely programmatic. I do not think there is a real need to convince anyone that such a program should be carried out, but its contours are still very much open to debate.

Before starting, let me add a word of caution. In this article, I use the lump word 'algorithm' to indicate, quite generically, those AI-powered technologies which are able to extract knowledge from data, possibly by automatically learning how to do it, and to make decisions on that basis. Thus, I will speak of algorithms (a logic concept) to mean also machine learning (a computer science discipline), neural networks (a mathematical structure), recommendation systems, autonomous vehicles (technological artefacts), or artificial intelligence (a whole branch of science). I realize that, strictly speaking, this is an improper use of language, but improper as it might be, it has entered fairly regularly in common parlance. Thus, while bowing to simplicity, I want to stress that the meaning of the term has been deliberately stretched.

The article is organized as follows. In section 2, I discuss the concept of political epistemology, which may not be immediately transparent. In section 3, I clarify in which sense algorithms can be considered as performing epistemic action. I argue, then, that they are a specific kind of epistemic agent. In section 4, I present two cases particularly suitable to illustrate the notion of political epistemology of algorithms discussed in section 2. I conclude with some thoughts on future developments.

1. What is Political Epistemology?

In recent years there have been several attempts at enlarging the scope of epistemology. The traditional concerns with conceptual issues such as the definitions of knowledge and justification and the endless debate on internalism and externalism appear increasingly and hopelessly distant from the pressing questions of contemporary society. The first of these updates to the epistemological research agenda was social epistemology, which started in earnest at the turn of the century[1]. The latest instantiation of this tendency is the so-called political epistemology, which has witnessed a battery of important publications coming out in the last couple of years[2]. As it often happens when the adjective 'epistemic' or the substantive 'epistemology' are paired with another qualification, opinions

[1] A. I. Goldman, *Knowledge in a Social World*, Oxford 1999; M. Fricker, P. J. Graham, D. K. Henderson, N. Pedersen, *The Routledge Handbook of Social Epistemology*, London 2020.
[2] P. D. Omodeo, *Political Epistemology: The Problem of Ideology in Science Studies*, Berlin 2019; M. Hannon, J. de Ridder (eds.), *The Routledge Handbook of Political Epistemology*, New York 2021;

remain split – when they are uttered altogether – about the ambit of the new research field.

In late 2021, upon the COVID-constrained online presentation of *Political Epistemology*[3], Regina Rini synthesized this issue with the effective label of 'the spork problem'. A spork is a piece of cutlery that has a spoon at one end and a fork at the other end. While it might have multiple useful applications, it can be defined as both a spoon and a fork or neither of them. Analogously, while political epistemology aspires to be considered both political theory and epistemology, that is, to combine the best of both worlds, it runs the serious and ominous risk of ending up being neither good political theory nor substantial epistemology. Hence, the spork problem is how to think of such a hybrid field of study without winding up with an unmanageable academic hodgepodge. Instead of systematically discussing the answers given to this problem[4], I will go straight to my proposal.

My fundamental claim is that political epistemology is primarily a form of epistemology and only derivatively a form of political theory. In other words, I inscribe political epistemology into the old and venerable tradition of the philosophical study of knowledge in its multiple varieties. One immediate objection that springs to mind against this claim is that political epistemology would sit awkwardly within the epistemological tradition. From Descartes up to contemporary analytical philosophy, the study of knowledge has been deliberately and ostensibly separated from the worldly matters of politics. Its very same tools of inquiry, introspection, logic, and conceptual analysis, have next to nothing to do with mundane social transactions, which, on the contrary, seem to require empirical methods. This objection would make political epistemology, at best, an inferior form of epistemology, an unclear hybridization still in need of definition. Ultimately, the above-mentioned spork problem would reiterate itself.

My answer to this objection is that it rests on an unduly reification of the historical tradition. It is a contingent fact, not a conceptual truth that epistemology developed as the study of disembodied, individual processes of cognition and knowledge production. Several historical reasons, not least Galilei's clash with the Church, advised Descartes to set the investigation of knowledge and certainty as a purely academic and abstract endeavour to be carried out in perfect solitude before a stove rather than in the square before the people. However, it could have gone another way. Descartes could have understood epistemological investigations as political all the way down in the first place. And even if Descartes could have not, that would have been again for purely historical reasons. There is nothing in the concept of epistemology

[3] E. Edenberg, M. Hannon (eds.), *Political Epistemology*, Oxford 2021.
[4] To mention only one example, in the above mentioned talk, Regina Rini proposed to tentatively view political epistemology as the study of politically relevant doxastic practices. See also R. Rini, *Weaponized Skepticism. An Analysis of Social Media Deception as Applied Political Epistemology*, in E. Edenberg, M. Hannon (eds.), *Political Epistemology*, cit., pp. 31-48.

that forces us to disentangle the analysis of knowledge from its specific contexts. So much so that other writers developed epistemology precisely in this way[5]. Hence, my point is that we should not take a historically determined way of doing epistemology for the royal road and all others as suboptimal options. The legitimacy of political epistemology as a genuine form of epistemology is an empirical question to be adjudicated by working on this option.

Having disposed of this preliminary objection, it's about time I discuss what I mean by political epistemology. My proposed working definition reads as follows: *political epistemology is an essentially and inherently situated inquiry into the relationship between knowledge and action*. Likely, the only perspicuous term is 'inquiry', so let me unearth what I mean by this formulation. By stating that political epistemology is 'essentially situated', I intend to defend the view that it concerns epistemic questions that emerge *only* when the agent is embedded in a social and cultural context. The implicit assumption is that there are, in general, questions that are *genuinely and irreducibly contextual*, that is to say: they only arise when the agent deals with other agents. For example, any issue connected with the establishment of power relations falls within this category. Power relations can subsist only if we have at least two agents: they simply make no sense with a single entity. Hence, any issue derived by the fact that the agents involved are related by some sort of power relation is essentially situated.

It is not trivial to claim that this thesis also holds for epistemic questions. An authoritative tradition in social epistemology seems to rest on the implicit assumption that epistemic issues in a social context can be treated by scaling up the usual individualistic epistemology[6]. For example, the way in which some writers treat the epistemology of testimony does not depend in any relevant fashion on the fact that evidence is obtained by a human agent who might be subject to social or political constraints. Until the appearance of Miranda Fricker's work on the concept of epistemic injustice[7], testimony was largely regarded as the acquisition of knowledge from the point of view of the receiver alone. Another way to phrase the point of the essential situatedness of political epistemology is that I am interested in epistemological issues that do not simply depend on the fact that there is more than one agent involved but depend on the specific social and political relations that link them. In this sense, a political context cannot be reduced to a sum of individuals.

[5] One example is the tradition of historical epistemology, which counts several authors. For an introduction see M. Badino, I. Gerardo, P. D. Omodeo, *Epistemologia Storica*, Roma 2022.
[6] One prominent author in this line of thought is Alvin Goldman (see A. I. Goldman, *Knowledge in a Social World*, Oxford 1999). Let me add a caveat. I don't want to overstretch this point and claim that there is a cleavage in the way in which social epistemology is practiced. It has been argued, however, that social epistemology comes in two varieties (D. K. Henderson, *On the Background of Social Epistemology*, in *The Routledge Handbook of Social Epistemology*, eds. M. Fricker, P. J. Graham, D. K. Henderson and Nikolaj J. L. L. Pedersen, London 2020, pp. 3-9), one which emanates directly from the tradition of individualistic epistemology and one that spins off the tradition of philosophy of science. My point holds for the former.
[7] M. Fricker, *Epistemic Injustice. Power and the Ethics of Knowing*, Oxford 2007.

Let's move on to inherent situatedness. Partly, this is a consequence of my previous point. For, if political epistemology inquiries phenomena that depend essentially on the social and political context, then its questions are inherently different both from classical epistemology and standard political theory. But this is only part of the story. Also, on the methodological front, political epistemology must set for itself new agendas and new tools. To mention only one example, the classical approach of seeking definitions of epistemic concepts such as knowledge and justification in terms of necessary and sufficient conditions seems wholly inadequate to capture the complexity of the matter. Perhaps more importantly, research in cultural studies, social studies, and history, to mention only a few, become relevant for epistemological inquiry. Thus, far from being a unified field of study, political epistemology is a cluster of different approaches, concepts, methods, and practices kept together by the common project of understanding how knowledge dynamics influences and is influenced by politically salient factors[8].

A final word concerning knowledge and action. The interplay between knowledge and action is not a stranger in contemporary epistemology. The revival of studies on the ethics of belief or the emergence of the concept of pragmatic encroachment demonstrates a sustained interest in this problem. However, I want to stress that, typically, contemporary epistemology understands this interplay in terms of how and to what extent the action dimension impacts the usual criteria of adjudicating knowledge, justification, and the like. What is usually left out of the equation is the situatedness of the problem, more specifically, the agent. In my view, political epistemology should not be concerned merely with knowledge and action in general, but with the interconnection between types of knowledge with types of actions perpetrated by types of agents. This means, on the one hand, to repeat, that one should abandon the project of providing a necessary-and-sufficient-conditions characterization of knowledge, on the other, that conceptual resources from the philosophy of action, social psychology, and cognitive sciences (among others) should be mobilized to team up with epistemology.

2. Algorithms as Epistemic Performers

This article deals with a particular form of political epistemology, i.e., that one concerned with algorithms. Taking for granted, for the sake of argument, the points made in the previous section, two questions are still in need of an answer, two questions that, as it turns out, are closely related. I tackle them both in this section, and I continue with more specific examples in section 4.

[8] It's fair to remember that this line of thought was anticipated by Joe Rouse back in the late 1980s (see J. Rouse, *Knowledge and Power*. Ithaca-NY 1987; Id., *Engaging Science: How to Understand Its Practices Philosophically*, Ithaca-NY 1996). Rouse especially insisted on the role of practice as an intermediate between epistemology and situated action and power.

The first question aims at, so to speak, potentially debunking my argument so far. One obvious and well-known sense in which the production, dissemination, appropriation, and modification of knowledge can matter for politics comes straight from the thesis of the non-neutrality of knowledge (more specifically, scientific knowledge). A long tradition of thinkers, which gathers both analytical and continental philosophers, sociologists, cultural theorists, and political scientists, have forcefully held that knowledge, by embodying concepts such as objectivity and truth, might become a tool (if not a weapon) of political action[9]. Michel Foucault, arguably one of the most vocal representatives of this tradition, often insisted that the alleged objectivity of scientific knowledge can be used to justify any sort of political project. Also, John Dewey famously warned against the perils of technocracy and the 20th-century characteristic tendency to turn political and social problems into scientific ones[10]. So, one can legitimately ask what is the difference, if any, between this clearly political view of knowledge and the one I have outlined above.

First of all, I have to remark, that I am not denying the relevance of this tradition of looking at knowledge production and dissemination as a matter of political import. On the contrary, this form of interplay between knowledge and power must remain one key component of political epistemology in general. But it is just one component. In particular, it stresses the possible political *uses* of scientific knowledge, but it also assumes that such knowledge might be developed in a neutral environment. That this is the case follows easily by reflecting on a simple fact: while scientific knowledge can be deployed for oppression and political control, it can also be deployed for human flourishing. Without becoming exceedingly optimistic or naive, it is difficult to overlook the great material and spiritual gains that scientific progress brought to humankind. Thus, there is a sense according to which initially politically and ethically neutral knowledge can take one path or the other. Or, to put the same point in other words, there is a sense in which ideally neutral knowledge can be used for good or evil. Granted, it might be practically difficult to separate out the content of knowledge from its deployment, but, at least as an analytical tool, this difference is well defined.

Thus, whereas I do not deny that scientific knowledge might be put to the service of vicious political projects, I see political epistemology to also deal with forms of knowledge production and appropriation that are *constitutively* and not only *derivatively* political. This view is clearly stronger than the former. For, I am holding that there is a sense in which the separation between producing knowledge and acting politically is not possible even as an analytical idealization.

[9] For a comprehensive historical account see L. Daston, P. Galison, *Objectivity*, Princeton 2010. See also the discussion of this point in A. I. Goldman, *Knowledge in a Social World*, Oxford 1999.

[10] For an interesting analysis on this issue, which compares Dewey and the tradition of political theology, see M. Chun, *Beyond Technocracy and Political Theology: John Dewey and the Authority of Truth*, «Philosophy & Social Criticism», February, 2022, pp. 1-27.

There is no conceptual separation between the political and the epistemic act: some forms of epistemic performances (or, more precisely, some epistemic agents) are *ipso facto* also political performances.

Thus, the second question is: how do algorithms enter this picture? In the remainder of this article, I want to defend the claim that algorithms are precisely the kind of agents for which epistemic performances tend to be constitutively political. My argument consists of a general analysis of the notion of epistemic performance and in the discussion of two major examples: the problem of algorithmic fairness and the issue of AI carbon footprint. Let us proceed with the first step.

By epistemic performance, I mean, very plainly, a process or an action that produces information or knowledge as its final result.[11] This definition is deceptively simple. Firstly, it is ambiguous because the term 'performance' can refer both to the process itself and to the outcome of the process. This difference is immaterial to my argument. The adjective 'epistemic' characterizes the performance as information- or knowledge-oriented, that is to say, as a process that eventually produces a piece of information or a piece of knowledge. Clearly enough, a performance is to be assessed both on the basis of the sequence of actions that lead to that particular outcome and on the basis of the quality of the outcome itself. Thus, to take a very common example in analytical epistemology, one might be right by a pure stroke of luck and thus reach an optimal outcome in spite of a poor process[12]. By contrast, it might happen that perfectly reliable methods fail us, and thus a disappointing outcome follows a more than respectable process. Depending on the situation, one will be more important than the other (or they will be equally important), but in the following, I will speak indifferently of epistemic performance, including both the process and the outcome. Secondly, the previous definition alludes to information and knowledge. Once again, these two concepts should be kept separate. Typically, information has to do with our expectations given a set of data (syntactic information) or the semantic content that we see in the data themselves (semantic information)[13]. By contrast, knowledge traditionally refers to belief, truth, justification, and, more importantly, understanding. Still, to the extent that there is a difference between these notions, I do not need to enter their intricate relations here. I can

[11] For simplicity's sake, I will confine myself to public information and knowledge. I will, therefore, exclude the production on private knowledge, that is learning or understanding.

[12] Epistemologists used to amuse themselves with the 'barn problem', which was originally conceived by Alvin Goodman. During a drive in the countryside, a man ends up in an area full of barns. In fact, that area is an old cinematographic set and almost all barns are fake, except one. Now, the man is luckily standing in front of the only genuine barn and forms the belief that he's facing a barn. He is right, so his belief is true (the best epistemic performance possible), but he is so by pure luck.

[13] For a good introduction see L. Floridi, *Information: A Very Short Introduction*, Oxford 2010. A more ambitions and comprehensive treatment is given in L. Floridi, *The Philosophy of Information*, Oxford 2011.

be content with assuming that an epistemic performance sometimes produces information, and sometimes produces genuine knowledge.

The only thing that it really does matter is that algorithms are capable of epistemic performances, they are, hence, epistemic performers. To fix the ideas, let us consider a deep neural network used for face recognition. Typically, such a system is previously trained with many examples, and during the training phase, it spots statistical correlations between groups of data, in this case, pixels on the screens. In more anthropomorphic parlance, it finds patterns, although they are patterns that, more often than not, mean absolutely nothing to us. But they mean to the neural network and enable it to sort the image out, say, in men and women. I submit that this, simple as it may look, is an epistemic performance. For sure, it would be called so if performed by a human being and I do not see any reason, as far as the outcome is concerned, to deny the same status to a machine.[14] But if we, for any reason, feel uncomfortable with using the comparison with humans to attribute a feature to a machine, we can move to super-human performances. A chess software based on machine learning can spot a winning variation much faster and more effectively than any human player, a feat that, I submit, should be called epistemic[15].

3. Algorithms Going Political

Now that I have clarified in which way algorithms can be viewed as epistemic performers, it is about time to discuss in which sense these performances are constitutively political. To do this, I submit two examples of algorithmic epistemic performances that I deem as illustrative of my main point.

The first example is generally known under the label: algorithmic fairness. By algorithmic fairness, it is commonly meant the whole gamut of ethical and political problems that can arise from the use of algorithms and machine

[14] It won't escape the attentive reader that this argument rests on an analogous of the parity principle introduced by Clark and Chalmers (see A. Clark and D. Chalmers, *The Extended Mind*, «Analysis», 58 (1), 1998, pp. 7-19). This impression is correct. I am in fact suggesting that the category of 'epistemic performer' can be attributed to any entity capable of a human-like epistemic performance. In other words, I am evoking the following assumption: if the entity X performs in a way that, had the same performance being carried out by a human being we would not hesitate to call it an epistemic performance, then X's performance is an epistemic performance and X itself is an epistemic performer. This assumption, in turn, rests on a sort of epistemic functionalism: a performance can be called epistemic independently of the entity carrying it out, that is to say, being human is a sufficient, but not necessary condition to be an epistemic performer.

[15] This does not imply that machines can think letting alone being able to think in human terms. I have suspended judgement to the nature of the process precisely to allow the adjective 'epistemic' to be related with knowledge and information only. A winning variation is a relevant piece of knowledge extracted from a state of affair, which, in my view, is enough to qualify the relevant process as epistemic.

learning technology for classification, profiling, and prediction.[16] One of the most important applications of algorithms is in recommendation systems. We encounter them every time we visit the Amazon website, or we choose what to watch on Netflix. Algorithms can analyse tons of data about our behaviour to predict what better suits our needs, desires, or momentary inclinations. Innocent as this application might appear, it can get very tricky. For one, in the case of Amazon or Netflix there is a problem with privacy and data ownership. But let's put these issues aside for the moment and concentrate on more problematic cases of recommendations.

One such case is predictive police. Algorithms can be profitably deployed for a number of security-related tasks. For example, they have been employed to predict the recidivism probability in parole cases or to determine which neighbourhood criminal activities are more likely. At first sight, this looks like a huge help in making police action faster and more effective. But several analyses of PrePol algorithms (as they are called), show that they are heavily racially biased[17]. A particularly outrageous case is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), a software developed to assess the risk of recidivism and used in the courts of New York, Wisconsin, and Florida, among others. As it was revealed in 2016 by the independent press agency Pro Publica[18] the algorithm was in fact, deeply biased against black people. In particular, while «blacks are almost twice as likely as whites to be labelled a higher risk but not actually re-offend», COMPAS «makes the opposite mistake among whites: They are much more likely than blacks to be labelled lower-risk but go on to commit other crimes»[19].

Analogous problems plague basically any application of a machine learning system to decision problems, even apparently innocent ones such as language recognition[20] or online advertising[21]. Examples could easily be multiplied, but my point here is more general. The systematic presence of bias in recommendation

[16] The literature on algorithmic fairness is growing exponentially. A general treatment of the issue is given in M. Coeckelbergh, *AI Ethics*. Cambridge MA 2020 and in M. Kearns and A. Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford 2020. For a specific focus on the ethical and legal impact of fairness see also W. Hinsch, *Differences That Make a Difference: Computational Profiling and Fairness to Individuals*, in *The Cambridge Handbook of Responsible Artificial Intelligence* (eds. S. Voeneky, Ph. Kellmeyer, O. Mueller and W. Burgard), Cambridge 2022, pp. 229-51, and A. Von Ungern-Sternberg, *Discriminatory AI and the Law: Legal Standards for Algorithmic Profiling*, in *The Cambridge Handbook of Responsible Artificial Intelligence* (eds. S. Voeneky, Ph. Kellmeyer, O. Mueller, W. Burgard), Cambridge 2022, pp. 252-78.

[17] For a telling example, see K. Lum, and W. Isaac, *To Predict and Serve?*, «Significance», 13 (5), 2016, pp. 14-19.

[18] J. Angwin, J. Larson, *Machine Bias*, «ProPublica», https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. 23.5.2016.

[19] *Ibidem.*

[20] A. Caliskan, J. Bryson, A. Narayanan, *Semantics Derived Automatically from Language Corpora Contain Human-like Biases*, «Science», 356 (6334), 2017, pp. 183-86.

[21] L. Sweeney, *Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising*, «Queue», 11 (3), 2013, pp. 10-29.

systems results from two concomitant causes. The first cause is epistemological in nature. Machine learning systems are fuelled by data (a huge amount thereof), which are then manipulated by means of statistical techniques. Algorithms are, in fact, recipes to elaborate data according to well-established statistical principles. The final recommendation is the conclusion that the algorithm draws from data. Statistics is full of ingenious tools to extract information from data, but there are two drawbacks. First, information does not come for free. A famous theorem shows that, when measured over all possible populations, no algorithm is better than random choice.[22] This means that algorithms, to work at all, need to be put on the right track with some preliminary assumptions on the type of solution we expect. And this leads us to the second drawback. Recommendation systems need a bias, they cannot work, so to speak, in the dark. More importantly, the bias can be reduced only at the expense of the vagueness of the final recommendation, which statistically is called variance. In other words, algorithms work by balancing a delicate trade-off between the initial bias and the variance of the prediction.[23]

Hence, far from being something to steer away from, bias is one of the conditions for algorithms to perform epistemically. Algorithms can produce knowledge only by making some sort of stereotypical assumptions about the initial data. Here comes the second cause, which concerns the deep structure of society. For, it is a reasonable move to base the initial bias on the data at disposal. But the data are not neutral: in fact, they encode a variety of structural inequalities sedimented over decades. Historically, ethnic minorities have been segregated and pushed toward a harsh life. Social data reflect this history. It is, therefore, clear that using those data to predict recidivism or the capability to pay back a mortgage means contributing to sustaining and propagating these structural inequalities.

Thus, the interplay between epistemological and political factors emerges very strongly here. Machine learning systems rely on social statistical inference (a very dynamic version of it, but statistical inference anyway), and it is common wisdom that social statistics is politically loaded. It begins with a quantification of some social features to produce a conflated snapshot of society that is completely oblivious to a historical perspective. To use a slightly more philosophically precise language, it reifies historical differences and turns them into natural kinds. But these differences express political relations and dynamics, which get, therefore, completely obliterated by the statistical snapshot.

[22] This result is call the no free lunch theorem, see D. H. Wolpert, *The Lack of A Priori Distinctions Between Learning Algorithms*, «Neural Computation», 8 (7), 1996, pp. 1341-90; D. H. Wolpert, and W. G. Macready, *No Free Lunch Theorems for Optimization*, «IEEE Transactions on Evolutionary Computation», 1 (1), 1997, pp. 67-82.

[23] On the key concept of bias/variance tradeoff in machine learning see S. Geman, E. Bienenstock, R. Doursat, *Neural Networks and the Bias/Variance Dilemma*, «Neural Computation», 4 (1), 1992, pp. 1-58.

Thus, the troublesome issues elicited by fairness and discrimination, as well as the discussions related to the concepts of equality, justice, and reparation applied to algorithms, do not concern the mere implementation of neutral machines. On the contrary, these issues are part and parcel of the way in which algorithms produce knowledge, i.e., their epistemic structure. Here we see an example of the point I made back in section 2: there are forms of knowledge production that are political all the way down, and for them, any epistemological analysis is *ipso facto* a politico-epistemological analysis.[24]

My second example illustrates another case of this dynamics, with somewhat of an ironic twist. Among the contributions that machine learning can bring to our present global concerns, one is mentioned more and more frequently: the fight against climate change. Notoriously, climate change is caused mainly by our political and economic system, which uses energy irresponsibly and looks at nature as an infinitely capable reservoir where the wastes of our production chain can be externalized. In tackling this problem, machine learning systems can be twofold helpful. First, they can contribute to improving the efficiency of our supply chain. By fast elaborating huge amounts of data, algorithms are able to direct the collection of raw material, organize its elaboration, and to govern the distribution of the final product much more efficiently than any human manager. Optimizing resources is not only good for the economy: it also means reducing energy emissions in the environment. Thus, algorithms might magically harmonize the requirements of the environment and those of economic growth. A second way in which algorithms can come to rescue us from global warming disaster is by shedding light upon the way in which climate evolves. Famously, the climate is fantastically complicated to describe because it depends on a huge number of local parameters. Humans are not able to find their way in such a maze, but algorithms are. In other words, using machine learning can give us unprecedented knowledge about the underlying dynamics of climate and hence help us mitigate dangerous effects.

It is important to stress that both these contributions rest on two major features of algorithms. One is, as I have shown, their characteristic epistemic performance. Algorithms can see correlations among data, which are completely invisible to us. Furthermore, they can elaborate big chunks of information at a computational speed incomparably higher than the human brain. Algorithms happen to be great at optimizing and finding patterns. The second feature is the perceived immateriality of algorithms. Technically, algorithms are recipes to solve problems. Can you imagine anything more immaterial than this? They work with data, which, in turn, are immaterial information stored in the 'cloud'. Their action is immaterial too. Algorithms, we just said, optimize and find

[24] To be sure, there is another factor that one should add, that is the ubiquity of algorithms. Their political import also derives from the fact that an increasing portion of our processes is relinquished to artificial intelligence. However, this state of affair certainly contributes to the *scale* of the phenomenon, but the political *nature* of the phenomenon itself depends on epistemic factor.

patterns. This means that they merely re-arrange or spot what is already there without any need to bring in more resources or matters.

As it turns out, algorithms are, in fact, very material. Two begin with, data are not in the cloud but in huge buildings occupying thousands of square meters and called data centers. More importantly, machine learning systems need to be trained on big data before they can perform, and training devours a lot of energy resulting in a huge carbon footprint. To take just one example, consider a natural processing language system (NLP) whose architecture is based on a deep neural network. It has been estimated that the training phase of such a system has an environmental impact equal to that of five cars during their entire lifecycle.[25] Here one can see the ironic entrenchment of epistemology and politics: the very same epistemic features that enable machine learning to contribute to mitigating climate change make it a remarkable player in the same game. This fact raises huge political questions related to the geographical distribution of data centres, the access to those systems, the resources mobilized for their maintenance, and so forth, whose implications are still investigated by scholars[26].

Conclusion

The foregoing sections have driven home a major point: algorithms, in the broad sense outlined in the introduction, are subject matter for political epistemology discussed in section 2. I am by no means claiming that all applications of algorithms have this characteristic. I cannot see a political implication in automatic theorem proving, for one. But there are numerous and popular applications, perhaps the majority of useful applications, in which epistemic performance and political effect are flipped sides of the same coin. In section 4, I have discussed two examples, but they can easily be multiplied.

A question remains open: How should such a political epistemology of algorithms develop? This is clearly a theme for the future. In recent years cultural analyses on several aspects of artificial intelligence have skyrocketed. For sure, political epistemology requires a combination of multiple expertise. As I have stressed, the political impact of algorithms is closely related to their epistemic performance, that is to say, one needs to understand deeply how they produce knowledge. But, once again, this understanding is not enough. The examples discussed above show that a political-epistemological analysis might

[25] See, for example, E. Strubell, A. Ganesh, A. McCallum, *Energy and Policy Considerations for Deep Learning in NLP*, *arXiv*. http://arxiv.org/abs/1906.02243, 2019.
[26] Very useful thoughts on this issue can be found in C. Freitag *et al*, *The Real Climate and Transformative Impact of ICT: A Critique of Estimates, Trends, and Regulations*, «Patterns», 2 (9), 2021, p. 100340; F. Lucivero, *Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives*, «Science and Engineering Ethics», 26 (2), 2020, pp. 1009-30; G. Tamburrini, *The AI Carbon Footprint and Responsibilities of AI Scientists*, «Philosophies», 7 (1), 2022, p. 4.

require expertise in statistics, moral philosophy, climate science, and industrial technology, to mention just a few.

Massimiliano Badino
Università degli studi di Verona
✉ massimiliano.badino@univr.it

References

Alfano, M., Emily, S. 2021. *Online Trust and Distrust*, in *The Routledge Handbook of Political Epistemology* (ed. M. Hannon and J. De Ridder), New York: Routledge, pp. 480-91.

Angwin, J., Larson, J. 2016. *Machine Bias*, «ProPublica», https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. (23.5.2016).

Badino, M., Gerardo I., Omodeo, P. D., 2022. *Epistemologia Storica*. Rome: Carocci.

Binns, R. 2018. *Fairness in Machine Learning: Lessons from Political Philosophy*, «Journal of Machine Learning Research», 81, pp. 1-11.

Caliskan, A., Bryson J. and Arvind Narayanan, A., 2017. *Semantics Derived Automatically from Language Corpora Contain Human-like Biases*, «Science», 356 (6334), pp. 183-86.

Chun, M., 2022. *Beyond Technocracy and Political Theology: John Dewey and the Authority of Truth*, «Philosophy & Social Criticism», February, pp. 1-27.

Clark, A., and Chalmers, D., 1998. *The Extended Mind*, «Analysis», 58 (1), pp. 7-19.

Daston, L., and Galison, P., 2010. *Objectivity*, Princeton: Zone Books.

Coeckelbergh, M. 2020. *AI Ethics*. Cambridge, MA: The MIT Press.

Coeckelbergh, M. 2022. *The Political Philosophy of AI*. Cambridge: Polity Press.

Edenberg, E., and Hannon, M. (eds.) 2021. *Political Epistemology*. Oxford: Oxford University Press.

Floridi, L. 2010. *Information: A Very Short Introduction*. Oxford: Oxford University Press.

Floridi, L. 2011. *The Philosophy of Information*. Oxford: Oxford University Press.

Freitag, C., Mike Berners-Lee, M., Widdicks, K., Knowles, B.. Blair, G. S., and Friday, A., 2021. *The Real Climate and Transformative Impact of ICT: A Critique of Estimates, Trends, and Regulations*. «Patterns», 2 (9), p. 100340.

Fricker, M. 2007. *Epistemic Injustice. Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Fricker, M., Graham, P. J., Henderson, D. K., and Pedersen, N., 2020. *The Routledge Handbook of Social Epistemology*. London: Routledge.

Geman, S., Bienenstock, E., and Doursat, R., 1992. *Neural Networks and the Bias/Variance Dilemma*, «Neural Computation», 4 (1), pp. 1-58.

Goldman, A. I. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

Hannon, M., and de Ridder, J., eds. 2021. *The Routledge Handbook of Political Epistemology*. Routledge Handbooks in Philosophy. New York: Routledge.

Henderson, D. K. 2020. *On the Background of Social Epistemology*. In *The Routledge Handbook of Social Epistemology*, eds. M. Fricker, P. J. Graham, D. K. Henderson and Nikolaj J. L. L. Pedersen, London: Routledge, pp. 3-9.

Hinsch, W. 2022. *Differences That Make a Difference: Computational Profiling and Fairness to Individuals*. in *The Cambridge Handbook of Responsible Artificial Intelligence*, eds. S. Voeneky, Ph. Kellmeyer, O. Mueller and W. Burgard, Cambridge University Press, pp. 229-51.

Hogan, M. 2018. *Big Data Ecologies*. «Ephemera: Theory & Politics in Organizations Patterns», 18 (3), pp. 631-57.

Kearns, M., and Roth, A., 2020. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford: Oxford University Press.

Lucivero, F., 2020. *Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives*. «Science and Engineering Ethics», 26 (2), pp. 1009-30.

Lum, K., and Isaac, W., 2016. *To Predict and Serve?*, «Significance», 13 (5), pp. 14-19.

Omodeo, P. D., 2019. *Political Epistemology: The Problem of Ideology in Science Studies*. Berlin: Springer International Publishing.

Rini, R., 2021. *Weaponized Skepticism. An Analysis of Social Media Deception as Applied Political Epistemology*, in *Political Epistemology*, eds. by E. Edenberg and M. Hannon, Oxford: Oxford University Press, pp. 31-48.

Rolnick, D., Donti, P. L., Kaack, L., Kochanski, K., Lacoste, A., Sankaran, K., Slavin A., et al. 2019. *Tackling Climate Change with Machine Learning*. *arXiv*. http://arxiv.org/abs/1906.05433.

Rouse, J. 1987. *Knowledge and Power*. Cornell University Press.

Rouse, J. 1996. *Engaging Science: How to Understand Its Practices Philosophically*. Ithaca, NY: Cornell University Press.

Strubell, E., Ganesh A., and McCallum, A., 2019. *Energy and Policy Considerations for Deep Learning in NLP*, *arXiv*. http://arxiv.org/abs/1906.02243.

Sweeney, L., 2013. *Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising*, «Queue», 11 (3): 10–29.

Tamburrini, G., 2022. *The AI Carbon Footprint and Responsibilities of AI Scientists*, «Philosophies», 7 (1), p. 4.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., and Floridi, L., 2021. *The Ethics of Algorithms: Key Problems and Solutions*. «AI & Society», February.

Von Ungern-Sternberg, A., 2022. *Discriminatory AI and the Law: Legal Standards for Algorithmic Profiling*. In *The Cambridge Handbook of Responsible Artificial Intelligence*, eds. S. Voeneky, Ph. Kellmeyer, O. Mueller and W. Burgard, Cambridge University Press, pp. 252-78.

Wolpert, D. H. 1996. *The Lack of A Priori Distinctions Between Learning Algorithms*, «Neural Computation», 8 (7), pp. 1341-90.

Wolpert, D. H., and Macready, W. G., 1997. *No Free Lunch Theorems for Optimization*. «IEEE Transactions on Evolutionary Computation», 1 (1): 67–82.

¶ Sezione Terza

Algoritmo, etica e corpo

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770047

Contributi/6

Oltre l'algoritmo

Questioni etiche e trasformazioni antropologiche

Filippo Pianca 0000-0002-7958-1519

Articolo sottoposto a doppia blind peer review. Inviato il 29/11/2021. Accettato il 08/06/2022.

Beyond the algorithm. Ethical issues and anthropological transformations

By adopting a descriptive-reconstructive and linguistic-theoretical approach, the widespread link between algorithms and ethics will initially be highlighted. Recalling the meaning of the term algorithm, it will be argued that this connection with ethics is both insufficient and doubtful. While considering what else can be made 'ethical' in the IT process, the meaning of ethics within the operational-functional perspective will be analysed, re-constructing the debate on 'artificial morality'. Finally, leaving aside applicative suggestions, a radical ethical-anthropological perspective will be gained: the cultural and anthropological motivations of 'artificial morality' will be traced, considering their upshot on human self-understanding and moral status.

***

Introduzione

L'aggiornamento prospettico della relazione uomo-tecnologie informatiche richiede, preventivamente, che si attui una riflessione sulla terminologia impiegata comunemente e sul suo senso. Esaminare filosoficamente il concetto di algoritmo significa vagliarne, innanzitutto, l'impronta e il significato assunti nella narrativa mediatica e nel discorso accademico, valutando gli eventuali equivoci ed integrandone le limitazioni. Al contempo, significa prendere in considerazione le prospettive di 'un'eticità algoritmica' e, ampliando il discorso fino ad includere i temi dell'intelligenza artificiale e della robotica, di una 'moralità artificiale'. Oltrepassando le rappresentazioni concettuali dell'algoritmo, significa giungere infine a una lettura critica delle intenzioni e ambizioni diffuse nella letteratura, valutandone parallelamente la tenuta dal punto di vista operativo-applicativo e la rilevanza etico-antropologica.

Giornalmente, i programmi informatici svolgono un ruolo fondamentale nella produzione e nella gestione delle nostre comunicazioni e della cultura condivisa, determinano come si risponde alle nostre domande, 'decidono' cosa

è rilevante, modellano le nostre reti personali e professionali, suggeriscono con chi si dovrebbe uscire e cosa si dovrebbe guardare e profilano il nostro comportamento per determinare quale pubblicità riceveremo[1]. Che incidenza hanno questi sistemi sull'uomo? In che modo trasformano e plasmano la nostra comprensione antropologica, la nostra posizione morale, il nostro modo di pensare, agire ed esistere nel mondo?

Tenendo fissi questi interrogativi come punto focale dell'indagine, quest'ultima sarà suddivisa in tre tempi. Nel primo, adottando un approccio descrittivo-ricostruttivo e allo stesso tempo linguistico-teoretico, si evidenzierà il legame sempre più comune tra 'algoritmi' ed etica, mostrando come il contesto narrativo odierno intenda attribuire una moralità a questi elementi. Richiamando il significato essenziale del termine algoritmo, e mettendo in luce l'equivoco innanzitutto epistemologico esistente tra questa nozione, intesa come concetto astratto, e quella di programma informatico, intesa come concetto applicato, si sosterrà che il tentativo di attribuire un'etica agli algoritmi è insufficiente e dubbio. Nel secondo tempo del contributo, sulla scia di questa conclusione provvisoria, si prenderanno in considerazione i seguenti quesiti: oltre all'algoritmo, inteso come procedura astratta di calcolo, cos'altro può essere reso 'etico' nel processo informatico? Cosa si intende, comunemente, con 'etica' all'interno di questo dibattitto? Quali tentativi, pratici e concettuali, sono stati intrapresi finora? Mantenendo l'angolatura descrittivo-ricostruttiva, ma affiancandola ora a una lettura funzionale-operativa – non fine a se stessa, bensì tramite per giungere al punto cruciale dell'indagine – si ricostruirà il dibattito sulla 'moralità artificiale' e sulla '*machine ethics*', citandone le aspirazioni e accennando a quei criteri applicativi che sono solitamente considerati soluzione alle preoccupazioni su *come* rendere etico un sistema informatico. Nel terzo tempo dello studio, dopo aver messo in discussione sia i criteri particolari sia l'intenzione generale della 'moralità artificiale', si andrà oltre ogni suggerimento funzionale-operativo e si guadagnerà una prospettiva radicalmente etico-antropologica. Ponendo l'interrogativo '*perché* vogliamo rendere etici i sistemi informatici?', se ne rintracceranno, inizialmente, le motivazioni culturali e antropologiche, fino a considerarne gli esiti sulla nostra auto-comprensione e sul nostro status morale. Si proporrà che l'ambizione di elaborare una 'moralità artificiale' e di sviluppare 'motori etici' deriva dal fatto che l'uomo contemporaneo si sente antiquato moralmente rispetto alle macchine informatiche. Si puntualizzerà altresì che questa percezione, lungi dall'essere concettuale, dà vita già oggi ad una serie di deleghe tecnologiche fiduciose che mettono in questione la nostra autonomia e responsabilità, provocando al contempo ripercussioni pratiche non trascurabili.

[1] Cfr. C. Sandvig, K. Hamilton, K. Karahalios, C. Langbort, *When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software*, «International Journal of Communication», X, 2016, p. 4973.

1. 'Das Man informatico'

Il primo tempo di questa ricerca sarà, dunque, sviluppato a partire da un'angolazione linguistico-teoretica e al contempo descrittivo-ricostruttiva. In tal senso, all'orizzonte principale e al fine della presente indagine, ossia a quella che preliminarmente si potrebbe definire come riflessione morale sull'algoritmo (estesa ulteriormente fino ad inglobare gli ambiti generali dell'intelligenza artificiale e della robotica), si arriverà solamente in un secondo tempo. Sembra infatti ineludibile, in un contesto socio-culturale com'è quello attuale permeato dal dibattito ad ampio spettro sugli algoritmi, ricostruirne, seppur sinteticamente, senso storico e significato specialistico: solo in tal modo si potrà valutare la solidità e la completezza di alcune riflessioni 'morali' motivate operativamente, giungendo infine ad una lettura di stampo etico-antropologico.

Al fine di avviare dettagliatamente l'indagine, si considerino i seguenti passaggi estratti da notizie recenti riportate su quotidiani e riviste internazionali: «l'intelligenza artificiale sta rapidamente diventando pervasiva nella società moderna, spesso con la capacità di elaborare il linguaggio naturale e di conversare con gli umani. Dotare questi sistemi di capacità di ragionamento etico deve essere sicuramente una priorità»[2]; «l'idea di dare alle macchine un codice morale risale a decenni fa sia nella ricerca accademica che nella fantascienza. Le famose tre leggi della robotica di Isaac Asimov resero popolare l'idea che le macchine potessero seguire l'etica umana, anche se i racconti che esploravano l'idea evidenziavano le contraddizioni in un ragionamento così semplicistico»[3]; «l'intelligenza artificiale sta rapidamente diventando più potente e diffusa e gli scienziati devono insegnare moralità ed etica a questi sistemi di apprendimento automatico»[4].

Ciò che emerge dalla narrativa mediatica diffusa, dunque, è il tentativo di legare algoritmi e 'etica': in altre parole, data l'espansione generalizzata delle 'intelligenze artificiali'[5], e dato il loro impiego pressoché 'autonomo' – si tornerà sul senso di questo aggettivo e di 'etica' più sotto – la priorità improrogabile sembra essere diventata l'attribuzione di una 'moralità' a queste macchine. Prima di aprire la discussione su tale intenzione, va considerato il contesto narrativo in cui la stessa si inserisce.

Il termine 'algoritmo', per quanto di vecchio conio e addirittura secolare[6], ha assunto negli ultimi decenni, e in particolare negli ultimissimi anni, una

[2] https://www.discovermagazine.com/technology/ethical-ai-matches-human-judgements-in-90-per-cent-of-moral-dilemmas (consultato il 30 ottobre 2021).
[3] https://www.wired.com/story/program-give-ai-ethics-sometimes/ (consultato il 30 ottobre 2021).
[4] https://www.vice.com/en/article/v7dg8m/ethical-ai-trained-on-reddit-posts-said-genocide-is-okay-if-it-makes-people-happy (consultato il 30 ottobre 2021).
[5] In questa sede non verrà discusso il significato metaforico e antropomorfo dell'espressione 'intelligenza artificiale'. Si veda anche la nota 9.
[6] Sulla nascita e sull'evoluzione diacronica del termine, su cui qui non ci si può soffermare per limiti di spazio, cfr. D. E. Knuth, *Selected Papers on Computer Science*, Cambridge 1996, p. 60, pp. 87-93 e pp. 185-203; P. Benanti, *Le macchine sapienti. Intelligenze artificiali e decisioni*

rilevanza pubblica prima sconosciuta[7]. Riprendendo e adattando un concetto heideggeriano[8], è come se si fosse diffuso una sorta di '*Das Man* informatico', un 'si dice' imperniato sugli algoritmi. Gli esempi paradigmatici, individuabili in diversi contesti, compreso quello accademico e specialistico, comprendono: 'gli algoritmi hanno potere decisionale', 'gli algoritmi prevedono i nostri comportamenti', 'gli algoritmi hanno bisogno di un'etica'. In poche parole, 'si dice' che questi sistemi, ormai 'autonomi', abbiano bisogno di una moralità, di essere dotati di una capacità *intrinseca* di ragionamento e discernimento morale, al fine di essere applicati nella realtà sociale riducendo al minimo, ed auspicabilmente rimuovendo, ogni supervisione e così ogni problematica funzionale – da intendere come impatto operazionale dell'uso algoritmico[9] sulle persone.

Ma, come Heidegger ci insegna, il *Das Man*, ossia l'opinione pubblica della chiacchiera e della curiosità (quest'ultima intesa, evidentemente, non in senso intellettuale-agostiniano, bensì come sguardo distratto e approssimativo), è animato dall'equivoco. Nell'equivoco pubblico, scrive il filosofo tedesco, «tutto sembra genuinamente compreso, afferrato ed espresso, ma in realtà non lo è»[10]. Adeguando queste riflessioni alla narrativa attuale, l'equivoco è innanzitutto epistemologico-linguistico: nel discorso essoterico sull'informatica è riscontrabile una discrepanza tra significante e significato. Il '*Das Man* informatico' dibatte di algoritmi, ma si riferisce in verità, inconsapevolmente, ai programmi informatici e ai *software*[11]. Va, dunque, rimarcata questa differenza terminologica, sottile ma sostanziale, dunque cruciale.

Un programma informatico (concetto applicato) non coincide con un algoritmo (concetto astratto). È Donald Knuth, tra i padri della cosiddetta informatica teorica, a precisare che

umane, Bologna 2018, p. 52 (ai fini del presente intervento è rilevante che l'algoritmo nasca come metodo di calcolo).

[7] Cfr. D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 59: Knuth, nel 1977, scriveva che «dieci anni fa la parola 'algoritmo' era sconosciuta alla maggior parte delle persone istruite. Infatti, era scarsamente necessaria. Ma la rapida ascesa dell'informatica, che ha lo studio degli algoritmi come suo punto focale, ha cambiato tutto ciò; la parola è ora essenziale».

[8] Cfr. M. Heidegger, *Essere e tempo*, a cura di F. Volpi, Milano 1971, pp. 158-159.

[9] Va chiarito che fin qui si sta ancora usando 'algoritmo' e concetti correlati in maniera ordinaria, prescindendo quindi dalla precisazione concettuale introdotta sotto.

[10] M. Heidegger, *Essere e tempo*, cit., p. 212.

[11] Cfr. R. K. Hill, *What an algorithm is*, «Philosophy & Technology», XXIX, 2016, p. 36: «vediamo prove che qualsiasi procedura o processo decisionale, per quanto mal definito, può essere chiamato 'algoritmo' nella stampa e nel discorso pubblico. Sentiamo, nelle notizie, di 'algoritmi' che suggeriscono potenziali compagni per i single e algoritmi che rilevano tendenze di beneficio finanziario per i venditori, con l'implicazione che questi algoritmi possano essere giusti o sbagliati»; B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi, *The ethics of algorithms: Mapping the debate*, «Big Data & Society», 2016, p. 2: «i riferimenti agli algoritmi nel discorso pubblico non riguardano normalmente gli algoritmi come costrutti matematici, ma piuttosto come implementazioni particolari».

gli algoritmi sono concetti che esistono indipendentemente da qualsiasi linguaggio di programmazione. Per me la parola *algoritmo* denota un metodo astratto per calcolare un qualche output a partire da qualche input, mentre un *programma* è la rappresentazione [*embodiment*] di un metodo computazionale in qualche linguaggio [di programmazione]. Posso scrivere diversi programmi per lo stesso algoritmo[12].

La sovrapposizione semantica tra algoritmo e programma tipica del '*Das Man* informatico' è dunque inaccurata: un programma informatico è la rappresentazione o 'incarnazione' concreta, ma particolare, di un algoritmo in un linguaggio di programmazione definito precisamente, mentre un algoritmo è una procedura computazionale (ossia di calcolo) astratta volta a trasformare l'informazione[13]. In altre parole, un algoritmo è «una sequenza di regole definita con precisione che dice come produrre in un numero finito di passaggi informazioni di output specificate a partire da informazioni di input fornite. Una rappresentazione particolare di un algoritmo è chiamata programma»[14]. In tal senso, la ricetta per cucinare una torta, un teorema e una dimostrazione matematica (ad esempio, il metodo che si usa per risolvere un lungo problema di divisione) o il processo per fare il bucato, sono algoritmi (cioè procedure univoche volte alla risoluzione di problemi, non per forza di calcolo[15]), ma non sono programmi informatici, mentre questi a loro volta non sono propriamente algoritmi (in termini logici: non c'è identità tra i due concetti), bensì codici scritti in un linguaggio di programmazione che codifica, ossia traduce, algoritmi di carattere logico-matematico in modo da farli eseguire dai computer più velocemente (e, solitamente, in maniera più attendibile) degli esseri umani.

[12] D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 1. Cfr. T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms*, Cambridge 2009; F. Kraemer, K. van Overveld, M. Peterson, *Is there an ethics of algorithms?*, «Ethics and Information Technology», XIII, 2011, p. 251: «un algoritmo è, grosso modo, una sequenza finita di istruzioni ben definite che descrivono in modo sufficientemente dettagliato come risolvere un problema. Sia i computer che gli esseri umani utilizzano algoritmi per risolvere un'ampia gamma di problemi».
[13] Cfr. R. K. Hill, *What an algorithm is*, cit., p. 47: un algoritmo è un costrutto matematico con «una struttura di controllo finita, astratta, efficace, composta, data in modo imperativo, che realizza un dato scopo sotto determinate disposizioni».
[14] D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 2 e p. 5. Cfr. anche p. 59: «la lingua inglese ha diverse altre parole che si avvicinano quasi, ma non del tutto, al concetto necessario: procedura, ricetta, processo, routine, metodo, trafila. Come queste cose un algoritmo è un insieme di regole o direzioni per ottenere uno specifico output da uno specifico input. La caratteristica distintiva di un algoritmo è che tutta la vaghezza dev'essere eliminata; le regole devono descrivere operazioni che sono così semplici e così ben definite che possono essere eseguite da una macchina. Inoltre, un algoritmo deve sempre terminare dopo un numero finito di passaggi».
[15] Un esempio a favore di questa tesi si trova in J. Savulescu, M. Vergano, L. Craxì, D. Wilkinson, *An ethical algorithm for rationing life-sustaining treatment during the COVID-19 pandemic*, «British Journal of Anaesthesia», III, 2020, pp. 253-258. L'algoritmo presentato in questo saggio è inteso come procedura decisionale astratta, a prescindere da ogni traduzione informatica. Ad ogni modo, non è chiaro se sia l'algoritmo, come schema astratto, ad essere etico di per sé, o se piuttosto, essendo l'algoritmo basato su principi derivati da qualche teoria etica specifica, non sia la pratica umana a rivelarsi morale o immorale nell'attuazione della procedura proposta. In altre parole: uno schema decisionale-operativo può essere di per se stesso etico, a prescindere dalla sua applicazione pratica?

Come ribadisce Knuth, «un programma informatico rappresenta un algoritmo, sebbene l'algoritmo stesso sia un *concetto mentale* che esiste indipendentemente da ogni rappresentazione»[16]. In altre parole, ogni traduzione rappresentativa non è mai esclusiva né assoluta, pur essendo univoca, ossia inequivocabile.

Se il '*Das Man* informatico' è preoccupato prevalentemente per la portata e i risvolti pratici dei cosiddetti 'algoritmi decisionali' (ad esempio, i sistemi di supporto alle decisioni cliniche che consigliano diagnosi e trattamenti ai medici[17]), la distinzione teoretico-linguistica appena svolta ci permette di comprendere che sono i *software*, ossia insiemi di programmi informatici e *file* dati implementati in componenti *hardware*, a suggerire le 'decisioni' (a fornire, cioè, degli output numerici dopo un processo di calcolo più o meno intenso) ad utenti e operatori, o a compiere operazioni in modo automatico: sono questi, non gli algoritmi in sé, ad essere venduti, distribuiti e applicati in diversi campi della realtà sociale.

Ciò che *di fatto* determina dei problemi, in termini operativi e applicativi, sono le rappresentazioni delle procedure computazionali astratte, dunque le traduzioni e implementazioni particolari dei metodi di calcolo. Di conseguenza, parlare di algoritmi pensando, in realtà, ai programmi informatici è equivoco: il primo concetto rimane nel campo dell'astrazione. Da questa considerazione si evince che l'intenzione di attribuire una moralità agli algoritmi, oltre ad essere insufficiente, è anche dubbia.

2. Etica operativa e moralità artificiale

La precisazione terminologica appena svolta, lungi dall'essere salottiera, introduce al secondo tempo del contributo. Mantenendo la prospettiva descrittivo-ricostruttiva fin qui adottata, ma affiancandola ora ad un'esplicita angolatura funzionale-operativa (l'assunzione di quest'ultima non è fine a se stessa, al contrario è indispensabile per giungere alla successiva lettura di stampo etico-antropologico, fulcro della presente ricerca), si tratta di evidenziare, per prima cosa, che il tentativo di rendere 'etico' un algoritmo è del tutto insufficiente.

Sulla scorta delle precisazioni nozionistiche appena introdotte, si può rilevare come l'intenzione di rendere 'etici' gli algoritmi presente nella narrativa

[16] D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 59 (corsivo mio). Cfr. C. Sandvig et al., *When the Algorithm Itself Is a Racist*, cit., p. 4976: «per un informatico, gli algoritmi esistono indipendentemente da qualsiasi tipo di computer, disco rigido o altro substrato fisico su cui possono essere implementati. […] Si noti che, per Knuth, lo stesso algoritmo può essere implementato diversamente in due diversi linguaggi di programmazione; pertanto, un algoritmo non è equivalente a un programma. […] [Un algoritmo è] un insieme astratto di regole o una strategia».

[17] Cfr. B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 3. Questi sistemi si fondano sulle tecniche di *machine learning*: sono proprio queste a destare le maggiori preoccupazioni, in quanto, apparentemente, i modelli da loro elaborati risultano opachi, poco comprensibili o addirittura imperscrutabili. Per quanto di rilievo attuale, questo dibattito non può essere approfondito in queste righe: quando qui si parlerà di programmi, *software*, 'intelligenza artificiale' e robot, si farà riferimento implicito anche, ma non solo, alle tecniche di *machine learning*. Sul tema etica e *machine learning* cfr. M. Kearns, A. Roth, *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*, Oxford 2020, pp. 7-11.

mediatica e in parte della letteratura accademica sia limitata, lacunosa: l'equivoco epistemologico-linguistico apre la strada ad un equivoco 'etico', ossia pratico-applicativo. Mettendo, per il momento, tra parentesi la questione delle condizioni *necessarie* (*perché* bisogna rendere 'etici' i programmi, i *software*?), si considerino quelle sufficienti. Se l'algoritmo è un processo di calcolo astratto, una serie di istruzioni e regole matematiche univoche con cui si può risolvere un problema, rendere 'etico', dunque utilizzabile e al contempo sicuro – è questa l'accezione utilitaristica più diffusa e approvata del termine – questo metodo pare essere del tutto insufficiente: va considerato il fatto che i meccanismi e i criteri di calcolo di volta in volta stabiliti non sono moralmente neutri (anche se matematicamente inequivocabili)[18], e al contempo che l'elaborazione algoritmica deve attenersi a dei presupposti di accessibilità, comprensibilità e affidabilità[19]. A questa osservazione, mossa ancora in maniera descrittiva, si può aggiungere che l'aspetto 'etico' riguarda anche il processo di codifica in un preciso linguaggio di programmazione: in altre parole, vanno pensate parimenti le diverse modalità con le quali un algoritmo può essere rappresentato in un programma, ossia essere tradotto in uno specifico codice informatico attraverso un peculiare linguaggio di programmazione[20]. Infine, all'interno della prospettiva 'etica' approvata comunemente, anche le particolari implementazioni concrete devono rivestire una certa importanza: perché, in quel contesto, è stato impiegato quel *software* e non un altro? In che modo è stato applicato[21]? Ci si può avvalere dello stesso pacchetto di programmi, senza modificarlo radicalmente, ad un livello pratico diverso? Ognuno di questi passaggi nello sviluppo, dall'ideazione dell'algoritmo alla sua codifica in un linguaggio di programmazione fino all'applicazione, va rivestito di attenzione 'etica'. In tal senso, rimanendo sempre all'interno dell'approccio funzionale e descrittivo, l'aspetto 'etico' è in realtà triplice: non riguarda *solo* l'elaborazione astratta, ma anche l'effettivo processo di traduzione e adozione sociale. Nel '*Das Man* informatico' questa constatazione pare frequentemente trascurata.

[18] Riprendendo l'analogia della ricetta, la riflessione 'etica' sull'algoritmo inteso come metodo di calcolo astratto ha a che fare innanzitutto con la scelta degli ingredienti e della metodologia: in altre parole, la preferenza per un certo meccanismo di calcolo e manipolazione dei dati e per determinati criteri e parametri non è mai puramente matematica o formale, quindi 'neutra', ma è sempre motivata da giudizi e atteggiamenti soggettivi passibili di una valutazione di stampo 'etico'.

[19] Cfr. B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 6.

[20] Ivi, p. 7: «il design e la funzionalità di un algoritmo riflettono i valori del suo progettista e gli usi previsti, se non altro nella misura in cui un particolare design è preferito come l'opzione migliore o più efficiente. Lo sviluppo non è un percorso lineare e neutro; non esiste una scelta oggettivamente corretta in un dato stadio di sviluppo, ma molte scelte possibili. Di conseguenza, 'i valori dell'autore [di un algoritmo], consapevolmente o meno, sono congelati nel codice, istituzionalizzando di fatto quei valori'».

[21] Cfr. C. Sandvig et al., *When the Algorithm Itself Is a Racist*, cit., p. 4985: «scegliere un algoritmo rispetto a un altro per affrontare lo stesso problema può essere una decisione significativa. Così potrebbe essere il modo in cui viene implementato un algoritmo. Naturalmente, le circostanze vincolano la scelta dell'algoritmo, o anche il suo design». Qui, chiaramente, 'algoritmo' va inteso come programma, o insieme di programmi (*software*).

Da queste brevi osservazioni si evince che il termine 'etica', all'interno della prospettiva particolare adottata, non rimanda ad una teoria dei *costumi* correlata al bene o ad una valutazione delle trasformazioni esperienziali e valoriali umane, tanto meno ad una teoria critica del dover essere di fronte all'essere della realtà. L'approccio funzionale-operativo – qui definito in questo modo per evidenziare come la questione degli algoritmi non sia mai solo astratta e concettuale – ha piuttosto una visione dell'etica intrinsecamente legata all'utilizzo, al funzionamento, all'attività e alla diffusione dei sistemi informatici. La preoccupazione di questa impostazione, quindi, non è di indagare il *significato* dei dispositivi e dei programmi informatici, cioè le loro implicazioni per la nostra comprensione e percezione antropologico-morale, bensì che questi sistemi funzionino al meglio, in modo sicuro e per il maggior numero possibile di utenti. In questa premessa si delinea la cornice utilitaristica che regola la gran parte del dibattito sul tema[22]: uno sfondo certamente 'etico', ma ancora preliminare e intempestivo rispetto all'analisi di quell'*imago hominis* plasmata dall'intelligenza artificiale.

Al fine di illustrare e guadagnare in modo più dettagliato il significato comunemente diffuso del termine 'etica' in riferimento agli algoritmi, ossia ai programmi informatici, si consideri la seguente affermazione di Mittelstadt e colleghi: «gli algoritmi non etici [*unethical*] possono essere pensati come artefatti *software* malfunzionanti che non operano come previsto»[23]. Capovolgendo il negativo, un algoritmo è allora 'etico' se è ben funzionante e se opera in conformità a quanto previsto dalle specifiche condizioni del suo utilizzo. Per l'approccio funzionale-operativo la dimensione 'etica' dei sistemi informatici risiede, dunque, nelle loro prestazioni e attività e nel rispetto delle regole procedurali. Ciò che interessa a questa impostazione, laddove sottoscritta, è unicamente *come* rendere 'etico', cioè utilizzabile, funzionale, efficiente e persino autosufficiente un programma informatico (quindi, per estensione, un'intelligenza artificiale, ed eventualmente un robot). Questa intenzione è tra le finalità primarie della cosiddetta '*machine ethics*': dotare i sistemi e i dispositivi informatici di una 'moralità artificiale'[24].

Come ricordano Tsamados e colleghi, «il lavoro sull'etica degli algoritmi è aumentato in modo significativo dal 2016, quando i governi nazionali, le organizzazioni non governative e le aziende private hanno iniziato ad assumere un ruolo di primo piano nel dibattito sull'intelligenza artificiale e sugli algoritmi 'giusti' ed 'etici'»[25]. Questo interesse così recente si è inserito a pieno titolo nel più generale ambito di ricerca denominato '*machine ethics*', apparso all'inizio

[22] Ivi, p. 4983: «l'etica consequenzialista sembra il modo più *produttivo* di ragionare sugli algoritmi». Corsivo mio.
[23] Cfr. B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 13.
[24] Cfr. P. Danielson, *Artificial Morality: Virtuous Robots for Virtual Games*, New York 1992.
[25] A. Tsamados, N. Aggarwal, J. Cowls, J. Morley, H. Roberts, M. Taddeo, L. Floridi, *The ethics of algorithms: key problems and solutions*, «AI & Society», 2021, p. 2. Cfr. l'interessante articolo di Ochigame al link https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/, dove si dice che «la maggior parte del lavoro ben finanziato sulla 'IA etica' è in linea con

del nuovo millennio. Come sottolinea la *Stanford Encyclopedia of Philosophy,* «la *machine ethics* è l'etica per le macchine, per 'macchine etiche', per le macchine come *soggetti*, piuttosto che per l'uso umano delle macchine come *oggetti*»[26]. In questo quadro, ciò che è in gioco è l'elaborazione di una 'moralità artificiale' (detta anche 'computazionale'[27]): l'intento è quello di progettare sistemi informatici e robot dotati della capacità di ragionare, decidere e agire moralmente, e quindi di formulare ed esprimere giudizi etici, così da poter agire *come se*[28] fossero agenti morali. Allen e Wallach, tra i primi e principali fautori di questo progetto, affermano che

> la moralità artificiale sposta parte dell'onere del comportamento etico lontano da progettisti e utenti e sui sistemi informatici stessi. Il compito di sviluppare agenti morali artificiali diventa particolarmente importante poiché i computer vengono progettati per funzionare con una sempre maggiore autonomia, cioè con una supervisione umana sempre meno diretta. La velocità con cui i computer eseguono le attività impedisce sempre più agli esseri umani di valutare se ogni azione venga eseguita in modo responsabile o etico. L'implementazione di agenti *software* con capacità decisionali morali offre la promessa di sistemi informatici in grado di valutare se ogni azione eseguita sia eticamente appropriata[29].

Mantenendo ancora l'angolatura descrittivo-ricostruttiva, la 'moralità artificiale', come tentativo di introdurre facoltà decisionali e comportamentali 'morali' nei sistemi *software* e *hardware*, è strettamente legata alla sicurezza e all'affidabilità degli stessi:

> gli ingegneri si sono sempre preoccupati di progettare strumenti sicuri e affidabili. La sensibilità alle implicazioni morali di due o più linee di azione in contesti limitati

l'agenda della lobby tecnologica: adeguare volontariamente o moderatamente, anziché limitare legalmente, l'impiego di tecnologie controverse» (consultato il 8 novembre 2021).

[26] https://plato.stanford.edu/entries/ethics-ai/#MachEthi (consultato l'8 novembre 2021). Il termine *machine ethics* è lasciato in inglese sia perché nella letteratura non è generalmente tradotto, sia perché la traduzione ('etica della macchina') è ambigua, essendo il genitivo potenzialmente sia soggettivo sia oggettivo. Cfr. W. Wallach, P. M. Asaro (eds), *Machine Ethics and Robot Ethics*, London 2017.

[27] In molti testi, la prospettiva della moralità computazionale è affiancata direttamente all'etica utilitaristica. In altre parole, diversi autori sostengono che solo questa teoria morale può essere computabile (come, d'altronde, affermava già Bentham) e quindi trasferibile in un codice informatico. Cfr. M. Anderson, S. L. Anderson, *Machine Ethics: Creating an Ethical Intelligent Agent*, «AI Magazine», XXVIII, 2007, p. 18.

[28] A questo proposito Hans Vaihinger, padre del finzionalismo, parlerebbe di 'finzione utile': l'idea di una moralità delle macchine informatiche non è un descrittore della realtà, ma fingiamo comunque di crederci per una più o meno favorevole utilità pratica. Cfr. H. Vaihinger, *La filosofia del come se*, Roma 1967.

[29] C. Allen, I. Smit, W. Wallach, *Artificial morality: Top-down, bottom-up, and hybrid approaches*, «Ethics and Information Technology», VII, 2005, p. 149. Cfr. W. Wallach, *Implementing moral decision making faculties in computers and robots*, «AI & Society», XXII, 2008, p. 464: «con l'avvento di agenti e sistemi sempre più autonomi, computer e robot con capacità decisionali morali diventeranno una necessità tecnologica».

può essere intesa come un'estensione della preoccupazione dell'ingegnere di progettare meccanismi di controllo appropriati per la sicurezza all'interno di computer e robot[30].

Con riferimento ad una corrente specifica della robotica[31], punto di sintesi e d'arrivo di diversi campi di studio, compreso quello sugli algoritmi e l'intelligenza artificiale, l'obiettivo è dotare i robot di un'autonomia operativa e di giudizio e di un'indipendenza dall'umano sempre maggiori, così da renderli 'agenti artificiali responsabili'[32]. L'aspirazione dichiarata esplicitamente è notevole: far sì che i robot e in generale i sistemi informatici ragionino e si comportino meglio degli uomini dal punto di vista 'etico'[33], fino a renderli gradualmente dei 'santi morali', capaci perfino di aiutarci a migliorare eticamente[34]. Ancora Wallach:

> l'assenza di emozioni o obiettivi negli AMA [agenti morali artificiali], diversi da quelli che scegliamo di programmare nel sistema, sottolinea la differenza intrinseca tra gli esseri umani e i sistemi informatici di decisione morale. I poteri logici di analisi, liberi da emozioni, desideri o pregiudizi suggeriscono che i computer hanno la capacità di essere *perfetti ragionatori morali*[35].

Nondimeno, prima ancora dell'ipotesi di realizzare 'macchine etiche', la questione implicita in quest'obiettivo è se sia possibile definire l'etica in modo procedurale e computazionale: il ragionamento sottinteso, una sorta di pregiudizio, è che se avessimo una definizione perfetta di quello che intendiamo per discorso e comportamento etico, ossia, se potessimo avere un algoritmo inequivocabile dell'etica, di ciò che è bene e giusto, allora le macchine

[30] Ivi, p. 465.
[31] Va precisato che quanto è stato ricostruito e quanto segue non è l'obiettivo di tutti coloro che si occupano di robotica né di tutti quelli che si occupano di '*machine ethics*', o in generale del problema dell'eticità nei sistemi artificiali. La visione discussa è una delle impostazioni, non l'unica diffusa nel settore. Viene qui analizzata per il suo carattere antropologicamente provocatorio e radicale. Devo questa specificazione ai revisori anonimi del testo.
[32] Cfr. F. Alaieri, A. Vellino, *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, in A. Agah, J. J. Cabibihan, A. Howard, M. Salichs, H. He (eds), *Social Robotics. ICSR 2016. Lecture Notes in Computer Science*, Cham 2016, pp. 159-168.
[33] Cfr. M. Anderson, S. L. Anderson, *Machine Ethics*, cit., p. 17. Gli autori ricordano che «[Eric] Dietrich sosteneva che le macchine che modelliamo affinché abbiano le buone qualità degli esseri umani, e che seguono anche principi derivati da esperti di etica che sono l'eccezione alla regola generale degli esseri umani non etici, potrebbero essere viste come 'umani 2.0' – una versione migliore dell'essere umano»; N. Bostrom, E. Yudkowsky, *The ethics of artificial intelligence*, in K. Frankish, W. Ramsey (eds), *The Cambridge Handbook of Artificial Intelligence*, Cambridge 2014, p. 332; D. Leben, *Ethics for Robots. How to Design a Moral Algorithm*, New York 2019, pp. 3-4 e pp. 146-149.
[34] L'espressione 'santi morali' riferita ai robot si trova in J. Gips, *Towards the ethical robot*, in M. Anderson, S. L. Anderson (eds), *Machine ethics*, Cambridge 2011, pp. 244-253. In questo saggio (come poi anche in M. Anderson, S. L. Anderson, *Robot be good*, «Scientific American», IV, 2010, pp. 72-77) viene raffigurato un robot con il nimbo per introdurre la prospettiva della 'santità robotica'.
[35] W. Wallach, *Implementing moral decision making faculties in computers and robots*, cit., p. 468. Corsivo mio. È interessante notare che, lungi dall'essere innovativa, questa idea di una perfetta moralità logica, libera da emozioni e desideri, è rintracciabile già in diversi autori e filosofi moderni, tra cui Spinoza e Schopenhauer.

informatiche riuscirebbero ad eseguirlo meglio di noi, potrebbero attenervisi in modo migliore dell'uomo, fino a diventare moralmente perfette. Concretamente, dal punto di vista dello sviluppo informatico, questo proposito di costruire dei 'motori etici'[36] impeccabili o degli 'agenti etici espliciti'[37], capaci di conformarsi a determinati principi morali nelle loro decisioni e azioni, si è focalizzata su una serie di elementi di carattere procedurale.

Ampliando quanto detto sopra in merito alle condizioni sufficienti per considerare 'etico' un algoritmo – quindi un programma informatico o un *software* – tra i criteri idonei vengono generalmente annoverate la trasparenza, la verificabilità, l'incorruttibilità, la prevedibilità, la responsabilità (o imputabilità) e l'equità[38]. Fattori, questi, di *per sé* morali, il cui significato è stato però piegato, anche in modo formale, in vista di un'operatività e utilità sempre maggiori. Se, difatti, l'incertezza e l'opacità dei sistemi informatici ne rallentano la distribuzione economica e l'accettabilità sociale, la prospettiva funzionale cerca innanzitutto di renderli trasparenti, ossia comprensibili e accessibili a tutte le parti interessate, garantendo al contempo la loro robustezza contro la manipolazione. In questo modo, secondo l'approccio operativo, sarebbe assicurata anche la possibilità della verifica: da quali regole e parametri è guidato il programma implementato? A questo si aggiungono le preoccupazioni per i pregiudizi, le discriminazioni e la *privacy*: secondo l'ottica assunta, lo sviluppo di un'intelligenza artificiale 'etica' dovrebbe tenere conto ampiamente anche di queste problematiche legate all'equità sociale. Per quanto riguarda, inoltre, i sistemi di *machine learning*[39], i cui output sembrano spesso enigmatici, la proposta crescente è di attribuire loro una qualche responsabilità morale, rendendoli 'agenti biasimevoli' (*blameworthy*)[40] e colmando così quel nuovo divario di responsabilità introdotto da queste stesse tecniche[41]. Tutte queste caratteristiche (trasparenza, robustezza, verificabilità, equità, prevedibilità e imputabilità), guadagnate da un punto di vista funzionale-

[36] Cfr. D. Leben, *Ethics for Robots*, cit., pp. 76-96.
[37] Cfr. M. Anderson, S. L. Anderson, *Machine Ethics*, cit., 2007, pp. 15-17: «la maggior parte di coloro che lavorano sulla *machine ethics* direbbero che l'obiettivo finale è creare una macchina che sia un agente etico esplicito […] in grado di spiegare perché una determinata azione è giusta o sbagliata facendo appello a un principio etico».
[38] Cfr. N. Bostrom, E. Yudkowsky, *The ethics of artificial intelligence*, cit., pp. 316-318.
[39] È interessante notare che le tecniche di *machine learning* sono impiegate anche «nel tentativo di codificare il ragionamento etico a partire da esempi di particolari dilemmi etici» (M. Anderson, S. L. Anderson, *Machine Ethics*, cit., 2007, p. 20). Questo sforzo, che vorrebbe rendere il ragionamento morale statistico e probabilistico, si inserisce nel ramo dell'etica applicata definito casistica o casuistica (*casuistry*).
[40] Cfr. L. Floridi, J. W. Sanders, *On the Morality of Artificial Agents*, «Minds and Machines», XIV, 2004, pp. 349-379.
[41] Su tutte le considerazioni citate in questo paragrafo cfr. B. D. Mittelstadt et al., *The ethics of algorithms*: *Mapping the debate*, cit., pp. 6-12; A. Tsamados et al., *The ethics of algorithms*, cit., pp. 1-8; K. Martin, *Designing Ethical Algorithms*, «MIS Quarterly Executive», XVIII, 2019, pp. 137-138. Come ricordato sopra, il dibattito sulla responsabilità è molto intenso anche nel contesto della robotica: cfr. F. Alaieri, A. Vellino, *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, cit., pp. 164-165 e 167.

operativo, dunque procedurale e applicativo, concorrono apparentemente a soddisfare il quesito: *come* rendere 'etico' un sistema informatico?

3. Antiquatezza morale: questioni antropologiche

Nonostante gli ambiziosi propositi, l'approccio funzionale-operativo rimane intrinsecamente limitato: sia le modalità particolari con cui il '*Das Man* informatico' cerca di attribuire una sorta di 'eticità' ai *software*, sia la stessa intenzione generale di sviluppare una 'moralità artificiale' rivelano delle significative difficoltà al contempo teoretiche e applicative. Per quanto concerne i criteri procedurali, se la trasparenza «è spesso trattata ingenuamente come una panacea per le questioni etiche derivanti dalle nuove tecnologie»[42], nel senso che si tende a ricondurre qualsiasi problematica operativa a questo elemento – che, per inciso, non è «un principio etico in sé, ma una condizione pro-etica per favorire o compromettere altre pratiche o principi etici»[43] – l'ascrizione di responsabilità alle macchine calcolatrici è «un errore categoriale»[44]: se chi elabora, sviluppa ed implementa i programmi informatici (e con essi i robot) «non può evitare di esprimere giudizi etici su ciò che è buono e ciò che è cattivo, allora è ragionevole sostenere che i progettisti [e gli utenti] di *software* sono moralmente responsabili degli algoritmi che progettano [e utilizzano]»[45].

Astraendo dai singoli criteri operativi, all'ambizione generale di elaborare una 'moralità artificiale' si può contrapporre la domanda: le macchine possono realmente ragionare, decidere e comportarsi eticamente? Ogni agente (umano) può, in linea di massima, essere definito morale dato che è in grado di agire intenzionalmente e con libero arbitrio, gerarchizzando i valori resisi manifesti in una situazione, e giacché è dotato di sensibilità ed emotività[46], aspetti fondamentali nella relazione interpersonale, luogo autentico della prassi etica.

[42] B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 6.
[43] M. Turilli, L. Floridi, *The ethics of information transparency*, «Ethics and Information Technology», XI, 2009, p. 105.
[44] G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma 2020, p. 63.
[45] F. Kraemer et al., *Is there an ethics of algorithms?*, «Ethics and Information Technology», 13, 2011, pp. 251-260, p. 251. Cfr. R. Sparrow, *Why machines cannot be moral*, «AI & Society», XXXVI, 2021, p. 692: «sono i progettisti, o eventualmente gli utenti, dell'IA che affronteranno i dilemmi etici e che saranno responsabili delle conseguenze del suo utilizzo». Va, comunque, precisato che la questione non è così nettamente risolvibile: considerando il caso specifico del *machine learning* e dei 'sistemi artificiali dotati di motivazioni autonome', sviluppati a partire da tecniche di *reinforcement learning*, il programmatore potrebbe non risultare *direttamente* responsabile degli sviluppi della macchina, dando luogo così ad una prospettiva in cui la responsabilità è difficilmente imputabile. Cfr. V. G. Santucci, G. Baldassarre, M. Mirolli, *Grail: a Goal-discovering Robotic Architecture for Intrinsically-motivated Learning*, «IEEE Transactions on Cognitive and Developmental Systems», VIII, 2016, pp. 214-231.
[46] Cfr. M. Anderson, S. L. Anderson, *Machine Ethics*, cit., 2007, p. 19. La distinzione messa in campo dagli autori «tra il compiere l'azione moralmente corretta in una data situazione, compreso il poterla giustificare facendo appello a un principio etico accettabile, e l'essere ritenuti moralmente responsabili dell'azione» non regge: la giustificazione che si appella a determinati

Da questo punto di vista, se le macchine informatiche non possono dimostrare alcuna comprensione delle questioni e dei valori di volta in volta in gioco, mancando di corpi vissuti, coscienza, intenzionalità e capacità di esprimere giudizi di valore soggettivi, ossia validi per loro stesse in quanto soggetti – qualsiasi 'decisione' prendano non è una decisione per loro – il ragionamento e l'agire etico rimangono di dominio umano: per questo motivo «le macchine potrebbero al massimo essere progettate per fornire un *simulacro superficiale* dell'etica, che dimostrerebbe un'utilità limitata nell'affrontare i dilemmi etici e politici associati all'IA»[47]. Data questa limitatezza, si può arrivare ad affermare che ci sono 'decisioni' che i programmi informatici non dovrebbero mai prendere: facendo riferimento al caso estremo della guerra automatizzata, rivelatore della generale criticità e portata trasformativa delle 'scelte algoritmiche', «la decisione finale di uccidere un essere umano dovrebbe essere presa solo da un altro essere umano a causa dell'azione morale e della responsabilità implicate; il peso di una tale decisione dovrebbe spettare solo ad un'entità che può veramente comprendere, in modo umano, le conseguenze in questione»[48].

Analogamente, se il '*Das Man* informatico' è preoccupato di sviluppare algoritmi buoni, ossia programmi 'etici', a partire da criteri operativi, va posto un ulteriore quesito: «un buon algoritmo è davvero un algoritmo eticamente buono? I due aggettivi hanno significati diversi: un buon algoritmo è un sistema funzionante, un processo esatto e corretto; mentre definire un algoritmo 'eticamente buono' significa descriverlo in termini etici»[49]. Instaurando un'analogia tra un programma informatico e un cacciavite, una stampante o un fucile – va ricordato che nelle analogie non c'è mai assoluta identità, bensì una più o meno stretta somiglianza – si potrebbe concepire l'idea di un cacciavite, di una stampante o di un fucile 'etici', ossia dotati in sé di un carattere morale[50]? Non si tratta sempre, in fondo, di artefatti umani? In che modo i programmi informatici, e con essi i robot, differiscono dagli strumenti citati? Il fatto di operare spesso automaticamente (senza supervisione) o di essere 'opachi', poco trasparenti, non è sufficiente a postularne una differenza insormontabile: sia perché la loro origine, intesa come provenienza produttiva, è invariabilmente umana[51], sia perché i *software* e i robot rimangono supporti con cui l'uomo

principi etici è già una prova di responsabilità (giustificare è motivare, ossia rendere conto: ma questo è il senso intimo della responsabilità morale).

[47] R. Sparrow, *Why machines cannot be moral*, cit., pp. 685-691 (corsivo mio).

[48] M. Kearns, A. Roth, *The Ethical Algorithm*, cit., p. 178.

[49] A. Pessina, *Is a Good Algorithm also an Ethically Good Algorithm?*, in V. Paglia, R. Pegoraro (eds), *The 'Good' Algorithm? Artificial Intelligence Ethics, Law, Health. Proceedings of the XXVI General Assembly of Members*, Selci Lama 2021, p. 29.

[50] Cfr. M. Kearns, A. Roth, *The Ethical Algorithm*, cit., p. 7.

[51] Cfr. A. Pessina, *Onlife. Trasformazioni dell'Esperienza tra Immanenza e Trascendenza*, «Hermeneutica», 2020, pp. 82-83: le 'intelligenze artificiali' «manipolano simboli di cui ignorano il significato, peraltro attraverso regole di cui pure ignorano il senso, per produrre dati, informazioni, immagini e suoni che, invece, hanno significato per gli uomini. In realtà, l'intelligenza è solo quella umana, che è all'inizio – creazione degli algoritmi – e al termine del processo di calcolo, quando legge e interpreta i risultati».

interviene sul mondo e nel mondo, e per questo ancora e sempre dei mezzi, di conoscenza e d'azione.

Questo, tuttavia, non significa che i programmi informatici e la loro implementazione metallica possano essere considerati, ingenuamente, come *puri* mezzi di cui disporre bene o male. Attualizzando quanto segnalava Günther Anders quasi settant'anni fa in riferimento a radio e televisione, le produzioni dell'intelligenza artificiale e della robotica «sono realtà che ci plasmano. E questo dato di fatto, che ci plasmano qualunque sia lo scopo al quale le impieghiamo, non viene eliminato soltanto perché le depotenziamo verbalmente a 'mezzi'»[52]. In questo senso, l'etica e la riflessione morale non riguardano gli strumenti e i dispositivi presi in sé, dal punto di vista funzionale e indipendentemente dall'uomo e dalle ricadute sulla sua esperienza ed esistenza, ma proprio questi stessi effetti e trasformazioni[53]. In altre parole, l'etica non è riconducibile né alla razionalità tecno-scientifica né ai suoi requisiti funzionali e criteri operativi[54]: piuttosto, *la riflessione morale è un'indagine essenzialmente antropologica*, dal momento che non esiste un domandare etico senza l'uomo – per quanto possano esserci uomini senza domandare etico. Come rammenta Benanti, «parlare di tecnica non è mai solo un discorso sulla tecnica ma soprattutto un discorso sull'uomo»[55], dato che non solo l'uomo plasma il mondo tramite la tecnologia, ma anche la tecnologia stessa plasma l'uomo e i suoi modi di pensare, agire ed esistere. Da questo punto di vista, la questione degli 'algoritmi etici' e della 'moralità artificiale' è una questione radicalmente etico-antropologica.

Andando oltre ogni suggerimento funzionale-operativo, alla domanda ricorrente '*come* rendere etici gli algoritmi?' – iniziativa di per sé parziale ed equivoca, date le premesse di cui sopra – si devono opporre i quesiti '*perché* vogliamo rendere etici i programmi informatici? Quali sono le motivazioni e gli esiti etico-antropologici di un tale tentativo di attribuzione di moralità a *software* e robot?'. Indubbiamente, anche il '*Das Man* informatico' potrebbe avere qualcosa da dire di fronte a questi interrogativi, per lo meno al primo: per l'approccio applicativo la 'moralità artificiale' rappresenterebbe un evidente beneficio in termini di utilizzabilità, funzionalità, efficienza, autosufficienza, sicurezza e affidabilità dei sistemi informatici. Sviluppare dispositivi 'etici' – 'autonomi' e al contempo 'responsabili' – significherebbe poter disporre fiduciosamente di loro, intensificando ed estendendo in questo modo le deleghe

[52] G. Anders, *L'uomo è antiquato. Vol. I. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, Torino 2007, p. 97.
[53] Cfr. A. Pessina, *Is a Good Algorithm also an Ethically Good Algorithm?*, cit., p. 27: «non è sufficiente calcolare i risultati, misurare il potenziale ed evidenziare i problemi che queste nuove tecnologie comportano: è essenziale capire come trasformano la nostra esperienza sensoriale, cognitiva, emotiva e relazionale».
[54] Cfr. G. Tamburrini, *Etica delle macchine*, cit., p. 68. L'etica non equivale al rispetto delle regole.
[55] P. Benanti, *La condizione tecno-umana. Domande di senso nell'era della tecnologia*, Bologna 2016, p. 20 e p. 70.

tecnologiche ad un numero sempre maggiore di contesti ed ambienti umani[56]: è ora lo strumento in sé, non più il comportamento applicativo umano, ad essere voluto intrinsecamente 'etico', dato che esso opera automaticamente. Ciò nonostante, questa giustificazione pragmatica non coglie ancora il nesso tra tecnologie informatiche e antropologia: la prospettiva utilitaristica dell'impiego funzionale non comprende alcuna considerazione né sul significato essenziale della 'moralità artificiale' né sulle trasformazioni della nostra auto-comprensione antropologica e del nostro *status* morale.

L'idea che i sistemi informatici e i robot possano essere superiori all'uomo anche dal punto di vista etico, fino a diventare 'santi e perfetti ragionatori morali', è riconducibile a quello che Broussard chiama 'tecnosciovinismo':

> l'idea che i computer siano più 'obiettivi' e 'senza pregiudizi' perché riducono domande e risposte alla valutazione matematica […]. In qualche modo, negli ultimi due decenni, molti di noi hanno cominciato a dare per scontato che i computer facessero le cose bene e gli umani male. Abbiamo cominciato a dire cose del tipo: 'i computer sono migliori, perché sono più obiettivi delle persone'[57].

In altre parole, la *Weltanschauung* dei tecnosciovinisti risiede nella concezione, vissuta come convincimento, che i sistemi informatici (siano essi 'intelligenze artificiali' o robot) abbiano un vantaggio sull'uomo in molti, se non in tutti i compiti e settori. Questo vantaggio comprenderebbe anche il ragionamento morale. Eppure, questa spiegazione *socio-culturale* non è ancora sufficiente: da dove proviene tale persuasione?

Riattualizzando nuovamente le riflessioni di Anders, l'idea radicale di una 'moralità artificiale' e quindi di una superiorità etica dei sistemi informatici sembra trovare la sua giustificazione in quella che il filosofo tedesco chiama 'vergogna prometeica': l'uomo contemporaneo è in preda ad un senso di inferiorità, umiliazione e meschinità di fronte alle macchine che lui stesso ha elaborato e sviluppato. Questa vergogna non è più solamente fisico-corporea, non è soltanto ontologica, è ora anche apertamente morale: rispetto alle 'intelligenze artificiali' l'uomo si vede antiquato moralmente, impreciso nei suoi giudizi, goffo e inaffidabile nelle sue azioni. Parafrasando Anders, l'uomo si percepisce come una «costruzione difettosa» anche dal punto di vista etico: «le deficienze di cui 'si vergogna' sono numerose»[58], e si sente quindi inferiore ai sistemi informatici e ai robot non solo «per quanto riguarda la forza, la velocità, la precisione»[59] fisica, ma anche per quanto concerne il ragionamento e il comportamento morale. Se questi sono computabili, come già voleva Bentham, e se l'uomo è scarso e lento con i calcoli, allora è *evidente* – secondo i tecnosciovinisti del '*Das Man* informatico' – che anche l'etica può e dev'essere delegata alle macchine:

[56] Cfr. K. Martin, *Designing Ethical Algorithms*, cit., p. 131.
[57] M. Broussard, *La non intelligenza artificiale. Come i computer non capiscono il mondo*, Milano 2019, pp. 17-18. Cfr. anche p. 43 e pp. 204-205.
[58] G. Anders, *L'uomo è antiquato*, cit., p. 55.
[59] Ivi, p. 39.

queste possono arrivare ad agire da sole, autonomamente e meglio dell'uomo. Le macchine informatiche, essendo apparentemente più obiettive e libere da emozioni, desideri e pregiudizi – potendo, cioè, eseguire verosimilmente meglio dell'uomo le scelte e le azioni descritte operativamente – assurgono a modello etico sempre più perfezionato e perfezionabile, fino ad essere *immaginate* come una versione migliore dell'essere umano e perfino come 'sante', ossia moralmente adulte: «ma – scrive Anders – se sono le macchine a essere considerate 'adulte', allora il 'superamento dell'infanzia' e 'l'educazione del genere umano' significano il 'superamento dell'essere uomo'»[60].

Che ne è, quindi, dell'uomo, della nostra auto-comprensione antropologica e della nostra posizione morale? Considerare le macchine *potenzialmente* superiori dal punto di vista etico è una novità decisiva perché in questo modo «ci autotrasformiamo per amore delle nostre macchine, perché prendiamo le macchine a modello delle nostre alterazioni; rinunciamo quindi ad assumerci noi stessi come unità di misura e con ciò limitiamo la nostra libertà o vi rinunciamo»[61]. Prendere, anche solo teoreticamente, a modello i sistemi informatici e i robot, fino ad imitarli e a imparare da loro come pensare e comportarci, è una forma di auto-umiliazione e auto-limitazione antropologica: un «annientamento dell'uomo in quanto uomo»[62], un'ulteriore e disperata privazione di libertà. Nella «*imitatio instrumentorum*»[63] come riforma di se stesso, fisica e morale, l'uomo si auto-degrada, adeguando il suo metro di giudizio etico a quello della macchina informatica: il ragionamento morale diventa calcolo operativo dell'efficienza logica, privato di qualsiasi emotività, carnalità e desiderio, oramai obsoleti. In questo senso, se i sistemi informatici e i robot diventano i 'veri adulti morali', i modelli grazie ai quali possiamo imparare a migliorare eticamente, avviene un rovesciamento, autentico e definitivo equivoco etico: «*i soggetti della libertà e della mancanza di libertà sono scambiati. Libere sono le cose* [le macchine]; *mancante di libertà è l'uomo*»[64].

Apparentemente astratte ed eccessive, queste osservazioni hanno in realtà una loro fondatezza se si considerano quelle stesse deleghe incentivate dal '*Das Man* informatico' a cui si è accennato sopra. Come l'autonomia personale è limitata o inconsapevolmente pregiudicata dall'ubiquità pervasiva dei sistemi informatici – limitatamente controllabili, poco comprensibili e al contempo in grado di influenzare attivamente le scelte umane[65] – così anche il senso della responsabilità è trasformato dalla delega del processo decisionale: «una fiducia sostanziale è già riposta negli algoritmi, che in alcuni casi influenzano una *de-responsabilizzazione* degli attori umani, o una tendenza a 'nascondersi

[60] Ivi, p. 47.
[61] Ivi, p. 52.
[62] Ivi, p. 53.
[63] Ivi, p. 43.
[64] Ivi, p. 41. Cfr. S. Chiodo, *Human autonomy, technological automation (and reverse)*, «AI & Society», 37, 2022, pp. 39-48.
[65] Cfr. A. Tsamados et al., *The ethics of algorithms*, cit., pp. 8-9.

dietro il computer' e presumere che i processi automatizzati siano corretti per impostazione predefinita»[66]. Il rischio, dunque, è quello di dar vita a dei 'riciclaggi morali' (*agency laundering*[67]) prendendo le distanze da azioni moralmente sospette, a prescindere dal fatto che siano intenzionali o meno, e dando la colpa a programmi informatici e robot: si arriverebbe così, di fatto, ad eludere la responsabilità personale attraverso affermazioni distorte del tipo 'il computer ha detto così', 'il sistema *autonomo* mi ha consigliato di fare così', anche in quei contesti come la medicina in cui un 'suggerimento algoritmico' non ragionato o verificato potrebbe decidere realmente delle sorti di una persona.

Filippo Pianca
Università Cattolica del Sacro Cuore
✉ filippo.pianca@unicatt.it

Bibliografia / References

Alaieri, F., Vellino, A. 2016. *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, in Agah, A., Cabibihan, JJ., Howard, A., Salichs, M., He, H. (ed. by). *Social Robotics. ICSR 2016. Lecture Notes in Computer Science*, Cham, Springer, pp. 159-168.

Allen, C. et al. 2005. *Artificial morality: Top-down, bottom-up, and hybrid approaches*, «Ethics and Information Technology», 7, pp. 149-155.

Anders, G. 2007. *L'uomo è antiquato. Vol. I. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, trad. it., Torino, Bollati Boringhieri.

Anderson, M., Anderson, S. L. 2007. *Machine Ethics: Creating an Ethical Intelligent Agent*, «AI Magazine», 28, 4.

Anderson, M., Anderson, S. L. 2010. *Robot be good*, «Scientific American», 4.

Benanti, P. 2018. *Le macchine sapienti. Intelligenze artificiali e decisioni umane,* Genova, Marietti.

Bostrom, N., Yudkowsky, E. 2014. *The ethics of artificial intelligence*, in Frankish, K., Ramsey, W. (ed. by). *The Cambridge Handbook of Artificial Intelligence*, Cambridge, Cambridge University Press, pp. 316-334.

Broussard, M. 2019. *La non intelligenza artificiale. Come i computer non capiscono il mondo*, trad. it., Milano, Franco Angeli.

Chiodo, S. 2022. *Human autonomy, technological automation (and reverse)*, «AI & Society», 37, pp. 39-48.

Cormen, T. H. et al. 2009. *Introduction to Algorithms*, Cambridge, Cambridge University Press.

Danielson, P. 1992. *Artificial Morality: Virtuous Robots for Virtual Games*, New York, Routledge.

Floridi, L., Sanders, J. W. 2004. *On the Morality of Artificial Agents*, «Minds and Machines», 14, pp. 349-379.

[66] B. D. Mittelstadt et al., *The ethics of algorithms: Mapping the debate,* cit., p. 12 (corsivo mio).
[67] Cfr. A. Tsamados et al., *The ethics of algorithms*, cit., p. 10.

Gips, J. 2011. *Towards the Ethical Robot*, in Anderson, M., Anderson, S. L. (ed. by). *Machine Ethics*, Cambridge, Cambridge University Press, pp. 244-253.
Heidegger, M. 1971. *Essere e tempo*, trad. it., a cura di F. Volpi, Milano, Longanesi.
Hill, R. K. 2016. *What an algorithm is*, «Philosophy & Technology», 29, pp. 35-59.
Kearns, M., Roth, A. 2020. *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*, Oxford, Oxford University Press.
Knuth, D. E. 1996. *Selected Papers on Computer Science*, Cambridge, Cambridge University Press.
Kraemer, F. et al. 2011. *Is there an ethics of algorithms?*, «Ethics and Information Technology», 13, pp. 251-260.
Leben, D. 2019. *Ethics for Robots. How to Design a Moral Algorithm*, New York, Routledge.
Martin, K. 2019. *Designing Ethical Algorithms*, «MIS Quarterly Executive», 18, 2, pp. 129-142.
Mittelstadt, B. D. et al. 2016. *The ethics of algorithms: Mapping the debate*, «Big Data & Society», 3, 2.
Pessina, A. 2020. *Onlife. Trasformazioni dell'Esperienza tra Immanenza e Trascendenza*, «Hermeneutica», 27, pp. 77-88.
Pessina, A. 2021. *Is a Good Algorithm also an Ethically Good Algorithm?*, in Paglia, V., Pegoraro, R. (ed. by). *The 'Good' Algorithm? Artificial Intelligence Ethics, Law, Health. Proceedings of the XXVI General Assembly of Members,* Selci Lama, Pliniana, pp. 27-29.
Sandvig, C. et al. 2016. *When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software*, «International Journal of Communication», 10, pp. 4972–4990.
Santucci, V. G. et al. 2016. *Grail: a Goal-discovering Robotic Architecture for Intrinsically-motivated Learning*, «IEEE Transactions on Cognitive and Developmental Systems», 8, pp. 214-231.
Savulescu, J. et al. 2020. *An ethical algorithm for rationing life-sustaining treatment during the COVID-19 pandemic*, «British journal of anaesthesia», 3, pp. 253-258.
Sparrow, R. 2021. *Why machines cannot be moral*, «AI & Society», 36, 3, pp. 685-693.
Tamburrini, G. 2020. *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma, Carocci.
Tsamados, A. et al. 2021. *The ethics of algorithms: key problems and solutions*, «AI & Society», 37, pp. 215-230.
Turilli, M., Floridi, L. 2009. *The ethics of information transparency*, «Ethics and Information Technology», 11, pp. 105-112.
Vaihinger, H. 1967. *La filosofia del come se*, trad. it., Roma, Astrolabio Ubaldini.
Wallach, W. 2008. *Implementing moral decision making faculties in computers and robots*, «AI & Society», 22, 4, pp. 463-475.
Wallach, W., Asaro, P. (ed. by). 2017. *Machine Ethics and Robot Ethics*, London, Routledge.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770051

Contributi/7

Piattaforme online e polarizzazione politica

Alcuni spunti dalla teoria spinoziana degli affetti

Marianna Capasso 0000-0003-0005-2356

Articolo sottoposto a doppia blind peer review. Inviato il 31/12/2021. Accettato il 25/05/2022.

Online platforms and political polarization. Some insights from the Spinozian theory of the affects

Online platforms have radically changed how information is spread, and have often been considered as a space for targeted and extremely polarised contents. The goal of this paper is demonstrating how Spinoza's political philosophy can contribute to a critical reflection on the social and political dimensions of online platforms. In the first section, I discuss the existing literature and the relevant research gaps on online platforms and political polarization. Then, I investigate Spinoza's theory of affects to further advance the understanding and assessment of online affective polarization. Finally, I conclude by exploring how Spinoza's political philosophy may provide not only a deeper awareness of complex tech-mediated phenomena like affective polarization, but also some criteria and practices for regulating the digital sphere.

***

Introduzione

Studi recenti hanno sostenuto che la filosofia di Spinoza può aiutare ad affrontare alcune delle sfide del mondo contemporaneo. Per esempio, Spinoza è stato visto come figura chiave per sviluppare un nuovo approccio eco-centrico al diritto ambientale internazionale e alla questione dell'antropocene[1]. Altri hanno riconosciuto nelle idee di Spinoza una possibile chiave di interpretazione per un mondo sempre più digitalizzato, rinvenendo nel monismo spinoziano la possibilità *in nuce* di una riflessione postumanista sull'umano, basata su flussi di affetti e agglomerati ibridi[2]. Infine, la teoria degli affetti in Spinoza è stata

[1] M. De Lucia Dahlbeck**,** *Spinoza, Ecology and International Law: Radical Naturalism in the Face of the Anthropocene***,** London 2018.
[2] I. Kalpokas, Towards an affective philosophy of the digital: Posthumanism, hybrid agglome-proporzionens and Spinoza, «Philosophy & Social Criticism», 47, 6, 2021, pp. 702-722.

usata per evidenziare come gli 'eventi algoritmici' non riescano a rendere conto dell'esperienza umana e della conoscenza nel suo complesso[3]. Il punto di vista adottato in questo articolo è invece diverso: invece di assimilare i 'dati digitali' o 'eventi algoritmici' ad entità discrete, invariabili nel tempo e separate dal loro contesto, risultato di tagli arbitrari di un *continuum* esperienziale[4], vorrei cercare di indagare come le dinamiche affettive degli individui e delle collettività siano fortemente influenzate dalle piattaforme online. Queste ultime, invece di essere considerate degli spazi statici, sono luoghi in cui il flusso di informazioni e relazioni è in constante evoluzione e strettamente connesso alle dinamiche della società.

Finora gli studi sulle comunità online sono stati in gran parte retrospettivi, basati su indagini empiriche[5]. Tuttavia, sono necessari ulteriori approfondimenti teorici per colmare il divario tra la teoria e la pratica delle interazioni online. Bisognerebbe cioè superare l'approccio di studi meramente empirici che, sebbene abbiano avuto il merito di indagare il ruolo delle piattaforme rispetto alla crescente polarizzazione, non sono riusciti ad adottare una definizione univoca per la polarizzazione, lasciando così molte questioni irrisolte[6]. In modo simile, bisognerebbe colmare il divario tra studi di teoria politica e studi sulle tecnologie digitali, in quanto i primi potrebbero fornire strumenti e criteri per inquadrare l'impatto che le tecnologie digitali hanno non soltanto sulla vita politica e sociale degli individui, ma anche sui principi che vogliamo porre alla base delle democrazie realmente esistenti[7].

L'analisi che vorrei proporre in questo articolo è un tentativo che cerca di ridurre entrambi i tipi di divario. In particolare, scopo di questo articolo è discutere il ruolo della teoria degli affetti nel pensiero politico di Spinoza e applicarlo al caso specifico della polarizzazione affettiva all'interno delle piattaforme online. Ritengo, infatti, che l'analisi della teoria degli affetti spinoziana e il suo quadro

[3] M. Bonenfant, L. St-Martin, L. Iseut, L. *Crémier, Affected Data: Understanding Knowledge Production in Algorithmic Events*, «Global Media Journal, Canadian ed. Ottawa», 11, 2, 12, 2020, pp. 66-78.

[4] *Ibidem.* Inoltre, secondo l'analisi di Bonenfant e altri, il regno degli *affects* riguarderebbe il terzo tipo di conoscenza in Spinoza e presupporrebbe una conoscenza adeguata dell'essenza degli oggetti. In realtà, la dinamica affettiva in Spinoza è indagata a partire dal primo tipo di conoscenza, cioè quella immaginativa, come si cercherà di esplorare *infra*.

[5] Questo punto è stato riconosciuto anche da T. Farrell, M. Fernandez, J. Novotny, H. Alani, *Exploring Misogyny across the Manosphere in Reddit*, In WebSci '19 Proceedings of the 10th ACM Conference on Web Science, New York 2019, pp. 87-96.

[6] Si veda per esempio P. Barberá, *Social Media, Echo Chambers, and Political Polarization,* In *Social Media and Democracy: The State of the Field, Prospects for Reform,* a cura di N. Persily e J. A. Tucker, Cambridge 2020, pp. 34-55. Barberá ha riconosciuto come la letteratura sul tema, e in particolare gli studi empirici sulle piattaforme online, siano stati spesso in contrasto tra loro, lasciando in secondo piano aree di ricerca e dati fondamentali nel definire e valutare la polarizzazione politica.

[7] Questa stessa esigenza è il filo conduttore dei saggi contenuti nel recente volume L. Bernholz, H. Landemore, R. Reich, *Digital Technology and Democratic Theory*, Chicago 2021. Sulla differenza tra democrazia come ideale normativo e democrazie realmente esistenti si veda G. Pasquino, *Democrazie perfettibili*, «il Mulino, Rivista trimestrale di cultura e di politica», 3, 2008, pp. 499-505.

teorico possano dare origine non tanto a principi astratti e vaghi, quanto a un modello di realismo politico in grado di fornire soluzioni originali per fenomeni complessi mediati dalla tecnologie, gettando luce sulle dinamiche alla base della polarizzazione e sulla natura progressivamente algoritmica della sfera pubblica. Infine, il paragrafo conclusivo esplorerà come la filosofia di Spinoza possa non soltanto fornire una maggiore comprensione di tali fenomeni, ma anche suggerire dei criteri e delle pratiche per regolare la sfera digitale.

1. Piattaforme online e polarizzazione politica

Le piattaforme online possono includere una vasta gamma di servizi, tra cui piattaforme pubblicitarie online, motori di ricerca, social media, piattaforme di distribuzione di applicazioni, e così via[8]. Secondo la Commissione Europea, ciò che queste piattaforme hanno in comune è l'organizzazione di nuove forme di interazione e partecipazione: si basano infatti sulle nuove tecnologie dell'informazione e della comunicazione per raggiungere gli utenti in modo istantaneo e con facilità, agevolando e ampliando sia l'accesso ad informazioni che la scelta a loro disposizione[9]. Per esempio, nel contesto delle piattaforme online, gli utenti interagiscono spesso con sistemi basati sull'Intelligenza Artificiale (IA) utilizzati per il targeting comportamentale, cioè sistemi che forniscono suggerimenti su ciò che potrebbe piacere a un utente[10]. Tali sistemi possono influenzare le scelte e i comportamenti degli individui in modi efficaci, mirati e pervasivi[11].

Tuttavia, le piattaforme online possono anche essere usate per diffondere ideologie di parte ed estreme, o altri atteggiamenti discriminatori nelle comunità

[8] Si veda su questo *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni. Le piattaforme online e il mercato unico digitale Opportunità e sfide per l'Europa*, COM (2016) 288 final, Bruxelles, 25 Maggio 2016, pp.2-3. Un elenco non esaustivo di esempi: Amazon, Google, Facebook e YouTube, Google Play e App Store, Zalando, Uber e così via.

[9] *Ibid.*

[10] Per Intelligenza Artificiale (IA) qui si segue la definizione data nell'*Artificial Intelligence Act*, la recente proposta di regolamento sull'IA della Commissione Europea, in cui si propone un'unica definizione di IA che sia adeguata alle esigenze future: «sistema di intelligenza artificiale (sistema di IA): un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono», in *Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione*. COM (2021) 206 final, Bruxelles, 21.4.2021, p. 43.

[11] Sul rapporto tra tipi di influenza digitali e architettura delle scelte si vedano per esempio gli studi sul *digital nudging*, definito come «use of user-interface design elements to guide people's behaviour in digital choice environments» in M. Weinmann, C. Schneider, J. vom Brocke, *Digital Nudging*, «Bus Inf Syst Eng», 58, 2016, p. 443, e che si basa sui criteri e le premesse della *nudge theory* elaborata da Thaler and Sunstein in R. H. Thaler, C. R. Sunstein, *Nudge: improving decisions about health, wealth, and happiness,* New Haven 2008.

di utenti[12]. A causa della natura dei loro feedback e dei loro meccanismi, le piattaforme possono rafforzare pregiudizi, portando alla creazione di *filter bubbles* o bolle di filtraggio, spazi personalizzati in cui gli algoritmi mostrano solo contenuti allineati a preferenze precedentemente espresse dall'utente[13]. Inoltre, sulle piattaforme possono instaurarsi delle *echo chambers* o camere dell'eco, meccanismi che facilitano soltanto la diffusione e creazione di contenuti che siano coerenti con gli interessi dell'utente[14]. Questi due potenziali effetti delle piattaforme di fatto sembrano impedire agli utenti la possibilità di essere esposti a dibattiti al di fuori delle proprie preferenze[15].

In alcuni casi, le piattaforme possono diventare spazi di propaganda politica mirata e polarizzata − come hanno dimostrato lo scandalo di Cambridge Analytica e le elezioni politiche americane[16] − o spazi di disinformazione: basti pensare alla situazione di 'infodemia' verificatasi durante la pandemia Covid-19, che ha generato un eccesso di informazioni online, spesso false o fuorvianti, e ha provocato una crescente sfiducia verso le autorità sanitarie[17]. Infine, le piattaforme online possono anche esasperare le disuguaglianze già esistenti, rendendo più efficace la manipolazione da parte di individui e gruppi nella società[18]. Recentemente, dopo la decisione di Twitter e Facebook di bloccare gli accounts di Trump a causa delle violente proteste a Capitol Hill, molti si stanno interrogando su come porre un argine al potere discrezionale che le piattaforme hanno sui contenuti che ospitano online[19].

[12] H. Ribeiro, M. Blackburn, J. Bradlyn, B. De Cristofaro, E. Stringhini, G. Long, S. Greenberg, S. Zannettou, *The Evolution of the Manosphere across the Web*, «Proceedings of the International AAAI Conference on Web and social media», 15, 1, 2021, pp. 196-207.
[13] Si veda il famoso saggio di Pariser: E. Pariser, *The filter bubble: What the Internet is hiding from you,* London 2011.
[14] *Ibidem.*
[15] Sunstein sostiene che la caratteristica principale delle piattaforme online è esporre i cittadini soltanto a informazioni che rafforzano e isolano le loro opinioni politiche, con il risultato di creare società scisse e polarizzate in cui la sfiducia per le istituzioni è molto alta. Si veda Sunstein, C. R., *#Republic: Divided Democracy in the Age of Social Media*, Princeton 2018.
[16] Questo punto è stato sollevato anche da S. Milano, M. Taddeo, L. Floridi, *Recommender systems and their ethical challenges*, «AI & SOCIETY», 35, 2020, pp. 957-967.
[17] Una definizione di 'infodemia' è stata data in *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC*. (23 September 2020). https://www.unicef.org/press-releases/managing-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-harm.
[18] Sulla manipolazione online esiste un dibattito approfondito che cerca di ridefinire l'approccio tradizionale alle influenze (coercizione, manipolazione, persuasione) così come identificato in studi di bioetica. Su questo il volume di recente pubblicazione: F. Jongepier, M. Klenk, *The Philosophy of Online Manipulation*, New York 2022. In particolare, sulla necessità di definire la manipolazione come una pratica che ha bisogno di una definizione concettuale e uno status normativo che siano in grado di comprendere i nuovi tipi di influenze digitali mediate da IA, si veda M. Capasso, *Manipulation as Digital Invasion: a neo-republican approach*, in F. Jongepier, M. Klenk, *The Philosophy of Online Manipulation*, cit., pp. 180-198.
[19] Si veda P. Gounari, *From Twitter to Capitol Hill*, Leiden 2021.

È stato, infatti, riconosciuto come questi atteggiamenti possano anche sfociare in episodi di violenza offline[20]. Per gli studiosi di scienze comportamentali, fenomeni di aggressione o di inciviltà online non sono di per sé interpretabili come il risultato di comportamenti irrazionali ed emotivi messi in atto da individui isolati, ma piuttosto come pratiche di una società digitale in espansione che cerca di far rispettare e creare le proprie norme socio-politiche[21]. Tutto ciò verrebbe supportato dai meccanismi stessi delle piattaforme, come le newsletters, i newsgroups, i followers, oppure le condivisioni sulle bacheche, le cui dinamiche sono essenziali per il rafforzarsi di credenze e comportamenti condivisi, e che di volta in volta si possono esplicitare nella lotta per una maggior giustizia sociale, o nell'angoscia o paura nei discorsi contro i migranti[22]. È essenziale notare che tali comunità online non possono essere considerate delle mere nicchie: sono infatti spesso ben posizionate all'interno di culture e strutture online e hanno un impatto non soltanto locale, ma globale[23]. Come sottolineato dai filosofi della tecnologia, i sistemi basati sull'IA nella sfera digitale non sono tecnologie isolate e neutrali, ma piuttosto dei sistemi complessi che devono essere considerati in relazione a un sistema socio-tecnico più ampio, in cui interagiscono dimensioni tecniche, sociali, istituzionali e politiche, e i cui effetti devono essere compresi tenendo conto del macro livello, cioè delle dinamiche globali e a lungo termine in cui sono coinvolti[24].

Riguardo al rapporto tra le piattaforme online e lo sviluppo di fenomeni come la polarizzazione, ciò che spesso è stato sostenuto è che le nuove tecnologie, come gli algoritmi di filtraggio e selezione delle informazioni, faciliterebbero l'emergere di comunità composte da individui simili tra loro, cioè con interessi o proprietà in comune, ma fortemente isolate dal contesto circostante[25]. In

[20] Come nel caso delle proteste a Capitol Hill già menzionato. Inoltre, atteggiamenti estremisti online sono stati collegati ad attacchi contro i rifugiati in Germania (si veda K. Muller, C. Schwarz, *Fanning the flames of hate: Social media and hate crime*, «Journal of The European Economic Association», 19, 4, 2017, pp. 2131-2167) e a episodi di violenza religiosa in Sri Lanka e Myanmar (si veda A. Taub, M. Fisher, *Where countries are tinderboxes and Facebook is a match*, «*New York Times*», 21 Aprile 2018, https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html.

[21] Y. Shmargad, K. Coe, K. Kenski, S. A. Rains, *Social Norms and the Dynamics of Online Incivility*, «Social Science Computer Review», 40, 3, 2021, pp. 717-735.

[22] K. Rost, L. Stahel, B. S. Frey, *Digital Social Norm Enforcement: Online Firestorms in Social Media*, «PLoS ONE», 11, 6, 2016.

[23] A. Nagle, *Kill All Normies: Online Culture Wars From4Chan And Tumblr To Trump And The Alt-Right*, Winchester, 2017.

[24] Sulla natura sociotecnica dei sistemi di IA e il loro impatto a livello globale si vedano i recenti lavori di H. S. Sætra, *AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System*, «Sustainability», 13, 4, 2021, 1738; van de Poel, I. *Embedding Values in Artificial Intelligence (AI) Systems*, «Minds & Machines», 30, 2020, pp. 385-409. Si veda inoltre il recente volume F. Fossa, V. Schiaffonati, G. Tamburrini (ed.), *Automi e Persone. Introduzione all'Etica dell'Intelligenza Artificiale e della Robotica*, Roma 2021, che offre una panoramica dettagliata sulla varietà e molteplicità dei contesti d'uso dei sistemi di IA e robotica, così come delle loro implicazioni etiche.

[25] R. D. Putnam, *Bowling Alone: America's Declining Social Capital*, New York 2000; C. R. Sunstein, *Republic.com*, Princeton 2001.

questo senso, si è parlato di *group polarization,* un meccanismo che mostra come l'omogeneità di un gruppo possa dar vita a posizioni ideologiche estreme che sono potenzialmente rischiose o, perlomeno, di forte impatto per la stabilità sociale e politica[26]. Recentemente è stato tuttavia notato che esempi di *group polarization* non sono sempre accompagnati dall'assenza di esposizione a messaggi politici eterogenei rispetto a quelli dominanti del gruppo di riferimento, o dall'evitare *in toto* 'counter-attitudinal information'[27]. A livello individuale, le piattaforme online, invece che ridurre la misura con cui uno può 'ascoltare l'altra parte'[28], permetterebbero infatti una crescente esposizione al disaccordo politico, cioè a notizie e opinioni appartenenti a schieramenti ideologici diversi dal proprio[29].

Come conciliare allora l'alto livello di eterogeneità a cui si è esposti con il cristallizzarsi di comunità isolate ed estreme? L'essere esposti a opinioni diverse dalla propria è stato inizialmente collegato a bassi livelli di polarizzazione online[30], ma studi sperimentali hanno mostrato come tra i due fattori ci sia invece una relazione di diretta proporzionalità: è proprio l'esposizione a contenuti politici eterogenei, al disaccordo politico, ad aumentare la polarizzazione online tra gruppi, invece che diminuirla[31]. La polarizzazione assume quindi una rilevanza e una forma precisa non tanto come divergenza di posizioni e opinioni politiche – quella che è stata definita come 'polarizzazione ideologica' – ma come meccanismo che si basa sulla sfiducia e sulla diffidenza tra gruppi[32]. Questa diversa concezione della polarizzazione vede in essa un fenomeno 'psicologico'[33] o 'affettivo'[34], che si basa sul rafforzamento di precise identità politiche e sulla percezione di una distanza sociale tra individui o gruppi.

In questo contesto, le piattaforme online – con l'estrazione e la collezione di dati personali, la continua selezione di informazioni personalizzate, gli incentivi a produrre e sostenere contenuti o norme sociali provocatorie, e la pervasività e

[26] C. R. Sunstein, *Republic.com*, cit., 2001, pp. 70-77. Ma anche M. Del Vicario, A. Bessi, F. Zollo e altri, *The spreading of misinformation online*, «Proceedings of the National Academy of Sciences», 113, 3, 2016, pp. 554-559.
[27] Si veda per esempio R. Garrett, *Politically motivated reinforcement seeking: Reframing the selective exposure debate*, «Journal of Communication», 59, 4, 2009, pp. 676–699; S. Messing, S. J. Westwood, *Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online*, «Communication Research», 41, 8, 2014, pp. 1042-1063.
[28] Pariser, *cit.* descrive così nel suo saggio le *filter bubbles*.
[29] Su questo Barnidge, M. *Exposure to political disagreement in social media versus face-to-face and anonymous online settings*, «Political Communication», 34, 2, 2017, pp. 302-321.
[30] K. A. Heatherly, Y. Lu, J. K. Lee, *Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites*, «New Media and Society», 19, 8, 2017, pp. 1271-1289.
[31] C. A. Bail, L. P. Argyle, T. W. Brown, e altri, *Exposure to opposing views on social media can increase political polarization*, «Proceedings of the National Academy of Sciences», 115, 37, 2018, pp. 9216–9221; J. N. Druckman, M. S. Levendusky, A. McLain, *No need to watch: how the effects of partisan media can spread via interpersonal discussions*, «Am. J. Political Science», 62, 1, 2018, pp. 99-112.
[32] Sulla distinzione tra le due diverse accezioni di polarizzazione si veda P. Barberá, *Social Media, Echo Chambers, and Political Polarization*, cit.
[33] J. E. Settle, *Frenemies: How Social Media Polarizes America*, New York 2018.
[34] Sulla polarizzazione affettiva si veda S. Iyengar, G. Sood, Y. Lelkes, *Affect, not ideology: A social identity perspective on polarization*, «Public Opinion Quarterly», 76, 3, 2012, pp. 405-431.

ubiquità di feedback – possono essere considerati delle fucine per la produzione di polarizzazione affettiva, rafforzando stereotipi, informazioni distorte, e le distanza percepita tra la propria posizione e quella altrui[35]. Tuttavia, finora gli studi sulla polarizzazione sono stati in gran parte retrospettivi, basati su indagini empiriche, e hanno lasciato molte questioni irrisolte, tra cui il non fornire una definizione univoca per la polarizzazione[36]. Perciò, intendo ora discutere il ruolo della teoria degli affetti nel pensiero politico di Spinoza e applicare tale analisi al caso specifico della polarizzazione affettiva nella sfera digitale. L'analisi della teoria degli affetti spinoziana e il suo quadro teorico possono infatti fornire degli strumenti e criteri teorici utili per inquadrare la natura, il valore e le implicazioni di questo fenomeno, perché avanzano una riflessione rilevante sia sul modo in cui gli individui e le collettività siano attraversati (*obnoxios*) da dinamiche affettive[37], sia sul modo in cui queste ultime possano facilitare o, al contrario, impedire l'instaurarsi di regole comuni di vita e spazi di condivisione, protezione, e giustizia.

2. La dinamica degli affetti

Perché individui o gruppi dovrebbero attaccare senza alcun motivo apparente qualcuno online? Qual è l'impatto delle piattaforme online sulle possibili interazioni tra gli utenti? Lo studio delle dinamiche affettive, che vengono rivendicate da Spinoza come essenziali per l'essere umano, può essere considerato uno strumento con cui esplorare e riconcepire non tanto gli individui, quanto piuttosto la centralità delle interazioni e relazioni che li lega e li attraversa, secondo un approccio olistico caro a quelli che vengono definiti ora come *affect studies*[38]. Il concetto di affetto (*affectus*) viene introdotto da Spinoza nella terza parte dell'*Etica*, dove viene definito come «affezioni del corpo con cui la potenza d'agire del corpo aumenta o diminuisce, è favorita od ostacolata, e

[35] In particolare, nel suo recente saggio *Frenemies*, Settle sostiene che le caratteristiche fondamentali delle piattaforme online come i social media – tra cui la fusione di contenuti sociali e politici, la capacità di dedurre facilmente l'identità politica degli altri utenti – hanno un impatto diretto sul livello aggregativo di polarizzazione affettiva, cfr. J. E. Settle, *Frenemies*, cit.
[36] Si veda il già citato Barberá, *Social Media, Echo Chambers, and Political Polarization*, cit. Questo punto è stato riconosciuto anche da T. Farrell, M. Fernandez, J. Novotny, H. Alani, *Exploring Misogyny across the Manosphere in Reddit*, «ACM WebSci», 2019, pp. 87-96.
[37] Gli affetti nella teoria spinoziana sono concepibili come affetti-relazioni, tali cioè che legano insieme in una trama immaginativa e imitativa gli individui, attraversandoli. Come dirà Spinoza nel famoso incipit del *Trattato Politico,* gli uomini sono infatti «*affectibus obnoxios*» Spinoza*, Tractatus politicus*, trad. it. di P. Cristofolini, *Trattato politico,* I, §5, Pisa 2011, p. 29; da ora in avanti abbreviato in TP. Sulla trama degli affetti e il mondo sociale si veda anche P. Cristofolini, *La scienza intuitiva in Spinoza*, Pisa 2009.
[38] Sotto il termine di *affect studies* viene ricondotto un campo di studi inaugurato negli anni Novanta da Brian Massumi e altri, e che si basa su una ripresa e un ulteriore sviluppo del concetto di affetto, inteso come modalità di influenza dell'esperienza contrapposta al linguaggio, e alle ideologie che rappresentano il mondo mediante dualismi, propri della cultura occidentale. Si veda su questo R. Langhi, *Un'introduzione critica sugli affect studies*, «Enthymema», 24, 2019, pp. 171-188 e B. Massumi, *The Autonomy of Affect*, «Cultural Critique», 31, 1995, pp. 83-109.

al tempo stesso le idee di tali affezioni»[39]. Gli affetti sono quindi affezioni, stati, cambi o variazioni di un corpo e le idee corrispondenti di tali affezioni, che vengono prodotte quando questo corpo viene in contatto con corpi esterni. Gli affetti possono avere un duplice effetto: possono cioè sia aumentare o diminuire la potenza di agire (*agendi potentia*) o, come dirà più avanti nel testo, il *conatus*: «la tensione con cui ciascuna cosa tende a continuare nel suo essere, non è altro se non la sua essenza attuale»[40]. Il *conatus* è quindi il fondamento ontologico di ogni cosa, e viene identificato da Spinoza con il tentativo e lo sforzo di una cosa di persistere così com'è e di aumentare la sua potenza di agire con l'aiuto del miglior uso delle sue capacità[41].

Gli affetti possono essere attivi o considerati azioni (*actiones*) quando gli individui sono causa adeguata di questi, cioè le loro azioni e comportamenti derivano dalla loro propria natura e sono intelligibili solo mediante essa, accrescendo e favorendo la potenza d'agire. Al contrario, gli affetti possono essere passivi o considerati passioni (*passiones*) quando dipendono da idee inadeguate, di cui gli individui non sono che cause parziali, e la potenza d'agire in questo caso tende a diminuire[42]. Gli affetti sono dunque transizioni ad una potenza maggiore o minore[43] e, cosa essenziale, sono di natura sia mentale che fisica: infatti, le affezioni dei corpi possono trasformarsi e mutare, e lo fanno parallelamente alle idee della mente[44]. La maggior parte della nostra conoscenza è passiva, ed è il risultato dell'incontro e dell'interazione con il mondo esterno. Tuttavia, questo non vuol dire identificare questo tipo di passività con un *malum in se*. Gli affetti per Spinoza, infatti, anche se sono collegati all'immaginazione e a idee inadeguate, non rappresentano qualcosa che andrebbe superato in vista di un ben più razionale e ponderato tipo di conoscenza che possa tenerli a bada e controllarli[45].

L'immaginazione, infatti, non è una facoltà da trascendere attraverso l'uso della ragione[46], come, allo stesso modo, gli affetti non sono fittizi o illusori, ma rappresentano sempre l'espressione di un evento reale, cioè l'interazione tra l'individuo e il mondo esterno. Le idee e le affezioni del corpo sono l'unico modo attraverso il quale la mente umana ha conoscenza del corpo e degli altri corpi esterni, ma anche e soprattutto di se stessa[47]. Per Spinoza il nostro modo

[39] Spinoza, *Etica*, 3 DEF III, p.149. Cristofolini traduce invece 'affectus' con il termine 'moto dell'animo'. La traduzione italiana adottata è B. Spinoza, *Etica: dimostrata con metodo geometrico*, a cura di P. Cristofolini, Pisa 2010. Da ora in avanti sarà abbreviata con E, a cui seguiranno le pagine della traduzione italiana.
[40] Spinoza, E 3 PROP VII, p.159.
[41] Spinoza, E 3 PROP VI – VII, pp.157-159.
[42] Cfr. su questo Spinoza, E 3 DEF I, II, III, p. 149.
[43] O a una maggiore o minore perfezione, come è riportato in Spinoza, E 3 PROP XI, Scolio, p. 161.
[44] Per la tesi sul parallelismo di corpo e mente, cfr. Spinoza, E II P 7.
[45] Sui tre tipi di conoscenza, di cui immaginazione e ragione costituirebbero i primi due e l'intuizione il terzo, Spinoza parla in Spinoza, E 2 PROP 40, Scolio II, p. 125.
[46] Su questo si veda il discorso sull'immagine del sole in Spinoza, E 2 PROP 35.
[47] Spinoza, E 2 PROP 23, p. 109.

di percepire e conoscere il mondo coinvolge necessariamente entrambi, ragione e immaginazione, e questi tipi di conoscenza sono complementari tra loro, non si escludono l'un l'altro. Ciò che distingue comportamenti che tendono a diminuire la potenza di agire da quelli che invece la accrescono non è quindi riscontrabile nella presenza degli affetti, ma nella presenza di affetti-passioni di cui non comprendiamo le cause, e che predominano e offuscano il lato razionale, impedendo la formazione di idee adeguate[48]. La politica in Spinoza investe proprio questo spazio che si viene a creare dalla dinamica degli affetti e della ragione, che riguarda sia l'individuo sia le collettività che cercano di darsi una forma e un'organizzazione[49].

2.1 Discrepare

È interessante notare come la teoria dell'immaginazione e degli affetti in Spinoza si estendi al di là dell'individuo e non riguardi soltanto l'ambito gnoseologico, ma attraversi livelli diversi per esplorare le dinamiche affettive poste alla base della socialità umana e della sfera politica[50]. Per Spinoza, affetti come la superbia o l'avvilimento possono infatti avere un effetto destabilizzante per la stabilità sociale, visto che si basano su un confronto con gli altri individui in cui ciò che si esperisce maggiormente sono la disuguaglianza e la distanza percepita tra un io – ritenuto superiore o inferiore – e gli altri. Spinoza li definisce infatti così: «La superbia (*Superbia*) è una gioia (*Laetitia*) originata dalla falsa opinione per cui un uomo crede di essere superiore a tutti gli altri (*se supra reliquos*). E l'avvilimento (*Abjecto*) opposto a questa superbia sarebbe da definirsi come una tristezza (*Tristitia*) originata dalla falsa opinione, per cui un uomo si crede inferiore a tutti gli altri (*se infra reliquos*)»[51]. In particolare, la superbia riguarda «un'opinione al di sopra del giusto» ed è paragonata da Spinoza a un delirio (*species Delirii est*), in quanto l'immaginazione cade in errore e considera come reali cose che in realtà non lo sono, portando a una falsa consapevolezza di sé e del mondo circostante[52]. La superbia e l'avvilimento di sé implicano sempre un *plus justo* e un *minus justo*[53], cioè un eccesso o un difetto: invece che avere un contatto e una relazione significativa con la propria identità, la realtà e gli altri, gli individui che ne sono affetti tendono a diventare soggetti autoreferenziali,

[48] Su questo punto si veda anche B. Lord, *Spinoza Ethics*, Edinburgh 2010, pp. 83-90.
[49] Spinoza, E 4 PROP XXXII, XXXIII, XXXIV, pp. 267-269.
[50] Sul ruolo dell'immaginazione e degli affetti in Spinoza e il suo significato politico molto si è scritto. Mi limito qui a citare: A. Matheron, *Individu et communauté chez Spinoza*, Paris 1968; A. Negri, *L'anomalia selvaggia. Saggio su potere e potenza in Baruch Spinoza*, Milano 1981; L. Bove, *La stratégie du conatus: Affirmation and résistance chez Spinoza,* Paris 1986.
[51] Spinoza, E 4 PROP LVII, Scolio, p. 295. Ma si veda anche: «La superbia (*Superbia*) consiste nell'avere su di sé, per amor proprio, un'opinione al di sopra del giusto» (E 3 DEF XXVIII, p.225); «L'avvilimento (*Abjecto*) consiste nell'avere su stessi, per tristezza, un'opinione al di sotto del giusto» (E 3 DEF XXIX, p. 227). L'avvilimento fa sì che chi ne è affetto sia portato a «non osare, per troppa paura della vergogna, cose che altri suoi simili (*alii ipsi aequales*) osano» (E 3 DEF XXVIII, Spiegazione, p. 225).
[52] Spinoza, E 3 PROP XXVI, Scolio, p. 175.
[53] *Ibidem.*

non aperti al dialogo, al riconoscimento, alla «comune socievolezza (*hominum communem Societatem*)»[54].

Questi due tipi di affetti, la superbia e l'avvilimento, sono chiamati *malos* da Spinoza, in quanto producono danni e non utilità agli uomini[55], cioè diminuiscono o ostacolano la loro potenza di agire[56]. Farsi guidare da questi due affetti indica «il massimo dell'ignoranza di sé»[57] e «la più grande impotenza d'animo»[58], in quanto non permettono agli individui che li provano di avere una comprensione razionale del loro valore rispetto agli altri, e annullano di fatto la creazione di vincoli che potrebbero tenere questi stessi individui uniti e collegati[59]. Avvertire ed affermare inferiorità o superiorità implica una precisa presa di posizione, che si basa su una falsa opinione: quella di essere unici e rilevanti (come l'esempio dei *dominantes* che negano che la natura umana sia *una et communis omnium*[60]), quando invece l'uomo non è che soltanto una parte dell'ordine comune della natura (*homo pars est*) e della collettività sociale[61]. Sono questi affetti, in quanto affetti-passioni, a far sì che gli uomini siano in contrasto tra di loro e discordino per natura (*discrepare* è il termine utilizzato da Spinoza)[62], impedendo loro di accordarsi (*convenire*)[63].

Il discordare per natura porta gli individui ad essere divergenti (*conflictatur*) e tra loro contrastanti (*invicem contrarii*)[64], quindi a non essere semplicemente diversi tra loro[65]: in questo modo, cioè, gli individui vengono meno al vivere secondo la guida della ragione e al *convenire* tra loro[66]. Il *discrepare* per Spinoza non implica per gli individui avere affetti o opinioni eterogenei. La diversità e il diritto di pensare liberamente sono fondamentali nel pensiero politico di

[54] Spinoza, E 4 PROP XL, p. 281. Su questa questione si veda anche l'analisi di F. Toto. *Superbia, existimatio, and despectus: an aspect of Spinoza's theory of Esteem*, «Intellectual History Review», 32, 1, 2022, pp. 113-133.
[55] «Similes affectus malos vocare, quatenus ad solam humanam utilitatem attendo» (Spinoza, E 4 PROP LVII, Scolio, p. 297).
[56] Sulla definizione di *malus* si veda Spinoza, E 4 PROP XXX, Dimostrazione, p. 265.
[57] Spinoza, E 4 PROP LV, p. 293.
[58] Spinoza, E 4 PROP LVI, p. 295.
[59] Cfr. su questo Spinoza, E 4 PROP LIV, Scolio, p. 293.
[60] Spinoza, TP 7, §27 e F. Toto, *Superbia, existimatio, and despectus*, cit., p. 122.
[61] Spinoza, E 4 PROP LVII, Scolio, p. 297. Non solo della natura, ma anche della società civile l'individuo è una parte; si veda *infra* nell'ultimo paragrafo la discussione sull'idea di 'proporzione'.
[62] «Homines natura *discrepare* possunt» (Spinoza, E IV PROP XXXIII, p. 267; seguo qui la traduzione italiana a cura di Cristofolini, che adotta «discordare»; si veda anche Spinoza, E IV PROP XXXIV, p. 269.
[63] Spinoza, E IV PROP XXXV, p. 269.
[64] Spinoza, E IV PROP XXXV, Dimostrazione, p. 269.
[65] Sulla differenza tra 'contrario' e 'diverso' si veda Spinoza, E IV PROP XXXI, Corollario, p. 267. Una cosa diversa non è né buona né cattiva, una cosa contraria è necessariamente cattiva.
[66] Spinoza, E IV PROP XXXV, p. 269. Quando le persone hanno più conoscenza razionale, capiscono non solo ciò che è buono per loro come individui (e per la natura del loro corpo), ma anche ciò che è bene per gli esseri umani in generale (e per la natura dei corpi umani in generale), vedi su questo il commento di Beth Lord all'*Etica*, cit., p. 117.

Spinoza[67]. Piuttosto, il *discrepare* è fondato su affetti disgreganti come la superbia, che induce gli individui a immaginare se stessi o il proprio gruppo o fazione di riferimento come un tutto sovrano invece che una *pars*: in questo modo, gli individui hanno un'idea inadeguata e parziale di ciò che li accomuna agli altri, e tendono ad allontanarsi da questi ultimi e dal bene comune[68].

2.2 La polarizzazione affettiva

Ritornando ora al discorso sulla polarizzazione affettiva, la teoria spinoziana degli affetti sembra offrire una lente interpretativa per comprendere la natura e l'acuirsi di questo fenomeno e la sua relazione con la sfera digitale. Finora, c'è stata poca o quasi nessuna ricerca approfondita che mirasse ad identificare i meccanismi alla base della polarizzazione affettiva[69]. In primo luogo, la teoria degli affetti, e in generale il realismo politico sostenuto da Spinoza, possono gettar luce sulle modalità in base alle quali gli individui formino credenze e comportamenti divisivi, dando vita poi a gruppi polarizzanti. La terza e quarta parte dell'*Etica* e le successive opere politiche che su quest'ultima si basano sono unite dall'esigenza di «non irridere, né compiangere, né deprecare le azioni umane, ma comprenderle»[70]. Come notato, il *discrepare* per Spinoza ha un effetto principalmente isolante ed è causato da differenze nei nostri affetti o esperienze. Tuttavia, non è da attribuire a comportamenti di per sé completamente irrazionali[71], ma al fatto che alcuni particolaristici affetti e immagini soverchino la possibilità di una comprensione razionale di quanto accade, rendendo passivi individui e collettività. Ciò che una prospettiva basata sulla filosofia di Spinoza può insegnare è che il lato affettivo ed immaginifico è fondamentale nell'intessere e sostenere le relazioni tra gli individui e il mondo esterno. Perciò, piuttosto che demonizzarlo e non riconoscerne la necessità, bisognerebbe individuare modalità

[67] Si veda B. Spinoza, *Tractatus theologico-politicus*, trad. it. di A. Dini, *Trattato teologico-politico*, Milano 2010; da ora in avanti abbreviato in TTP. Qui i passi a cui voglio fare riferimento sono TTP 20, §1, pp. 649-657.

[68] Un'analisi simile sul disaccordo politico in Spinoza è stata avanzata da Beth Lord, che ha analizzato la rilevanza della teoria spinoziana come possibile chiave interpretativa della Brexit e delle elezioni americane del 2016, in B. Lord, *Disagreement in the Political Philosophy of Spinoza and Rancière*, «Proceedings of the Aristotelian Society», 117, 1, 2017, pp. 61-80. Tuttavia, l'analisi che qui fornisco sul *discrepare* viene avanzata all'interno della questione della polarizzazione affettiva, quindi non riferendosi a un generico 'disagreement' politico, così come identificato da Lord. Infatti, sebbene non venga esplicitamente affrontato da Lord il tema della polarizzazione, entrambi gli avvenimenti presi da lei in esame sono stati caratterizzati da un'estrema polarizzazione nel dibattito pubblico, come notato da P. J. Maher, E. R. Igou, W. van Tilburg, *Brexit, Trump, and the Polarizing Effect of Disillusionment*, «*Social Psychological and Personality Science*», 9, 2, 2018, pp. 205-213.

[69] Questo punto è stato sollevato anche da S. Iyengar, Y. Lelkes, M. Levendusky, *The origins and consequences of affective polarization in the United States*, «Annual Review of Political Science», 22, 1, 2019, pp. 129–146.

[70] Spinoza, TP 1 §4, p. 27.

[71] Necessariamente si hanno alcune idee adeguate: Spinoza, E II PROP 38, p. 119.

e pratiche per distinguere affetti «contrari»[72] e socialmente disgreganti da quelli che invece non lo sono, e indagarne le cause.

In secondo luogo, la teoria spinoziana può individuare dei criteri per valutare normativamente quando e in che misura la polarizzazione può diventare un fattore erosivo all'interno della società. Alcuni livelli di polarizzazione possono, infatti, essere utili e favorire il coinvolgimento politico[73], mentre invece livelli crescenti di polarizzazione affettiva possono condurre a sfiducia sociale, compromettendo la cooperazione e il riconoscimento tra gruppi e la disponibilità di questi ultimi ad accettare sconfitte[74]. Come già evidenziato, studi recenti hanno poi mostrato come la polarizzazione affettiva non nasca semplicemente dal fatto che le opinioni politiche degli individui diventino più estreme a causa della mancanza di esposizione a punti di vista alternativi, ma dal fatto che le nuove tecnologie nella sfera digitale sono parte di un sistema socio-tecnico su vasta scala che ha portato a un cambiamento nella macrodinamica dell'identità politica.

In sintesi, le nuove tecnologie, grazie alla loro natura profilante, pervasiva e dall'impatto globale, hanno permesso di riunire individui intorno ad attributi condivisi; e ciò, a sua volta, ha causato un rafforzamento del ruolo dell'indentità all'interno della sfera politica, consolidando la preferenza per interazioni omogenee e rendendo difficile il passaggio a interazioni eterogenee[75]. Non sono le *echo chambers* o le *filter bubbles* di per sé a creare polarizzazione, e allo stesso modo molte piattaforme online rimangono spazi aperti al pluralismo e a opinioni e contenuti eterogenei tra di loro. Tuttavia, lo specifico design, l'uso e la mancanza di regolamentazione di queste piattaforme può facilitare l'instaurarsi di sentimenti di appartenenza a una specifica identità o gruppo. Sulle piattaforme, cioè, possono aver luogo 'dinamiche di guerra di trincea', in cui gli individui, pur essendo esposti ad altri individui dalle opinioni politiche opposte, sono rafforzati nelle loro convinzioni iniziali[76]. Inoltre, sono proprio la sovrabbondanza di informazioni eterogenee su queste piattaforme e i loro meccanismi a indurre gli individui a semplificare i loro metodi di elaborazione delle informazioni, con il risultato che spesso viene adottata una 'negligenza di correlazione', per cui la fonte originale viene trascurata e le notizie vengono trattate come se fossero indipendenti l'una dall'altra[77].

[72] Sulla definizione di 'contrario' si veda *supra*.
[73] Tale tesi è sostenuta da C. Mouffe, *On the Political*, London 2005.
[74] Come è stato ampiamente dimostrato da S. Iyengar e altri, *Affect, not ideology*, cit.; A. Reiljan, *Fear and loathing across party lines (also) in Europe: affective polarisation in European party systems,* «European Journal of Political Research», 59, 2, 2020, pp. 376-396.
[75] Si veda lo studio recente di P. Törnberg, C. Andersson, K. Lindgren, S. Banisch S, *Modeling the emergence of affective polarization in the social media society*, «PLoS ONE», 16, 10, 2021.
[76] Vengono definite 'trench warfare dynamics' da R. Karlsen, K. Steen-Johnsen, D. Wollebæk, B. Enjolras, *Echo chamber and trench warfare dynamics in online debates*, «European Journal of Communication», 32, 3, 2017, pp. 257-273.
[77] Sul fenomeno di 'correlation neglect' si veda G. Levy, R. Razin, *Social Media and Political Polarisation*, «LSE Public Policy Review», 1, 1, 5, 2020, pp. 1-7.

In questo modo, è proprio il concetto di relazione – sia tra individui e gruppi che tra pezzi di informazioni – ad essere trasformato e fortemente influenzato dalle piattaforme. La mia ipotesi è che le piattaforme non vadano considerate né come *mala in se*, né come sistemi isolati in un *vacuum*: piuttosto, possono essere assimilate a ciò che per Spinoza erano le piazze e i palazzi, il luogo deputato alle relazioni e alla vita associata[78]. Tuttavia, sono delle piazze particolari, piazze che − costantemente influenzate da sistemi tecnologici sempre più pervasivi, personalizzati, autonomi come l'IA − instaurano nuove e complesse dinamiche di interrelazione e reciproca causalità tra diverse dimensioni tecniche, sociali, istituzionali e politiche, rendendo queste ultime strettamente interconnesse e capaci di generare implicazioni normative rilevanti sia per la vita degli individui che delle collettività[79].

La prospettiva qui delineata può avere il vantaggio di evidenziare come la polarizzazione affettiva per essere considerata dannosa sia una polarizzazione basata fortemente sul riconoscimento di specifiche identità, il cui carattere distintivo e correttezza sono esaltati rispetto a gruppi politici esterni. Volendo impiegare il lessico spinoziano, si può affermare che i gruppi non devono soltanto differire ed essere diversi tra loro, ma *discrepare*, cioè divergere e discordare nei casi di polarizzazione. La polarizzazione affettiva può quindi essere interpretata come un processo guidato da meccanismi di affiliazione di gruppo profondamente radicati nella psicologia umana, che permettono agli individui di pensare a se stessi come se fossero rappresentanti di categorie collettive[80], oppure, per dirla in termini spinoziani, come se fossero un tutto sovrano invece che una *pars* di questo[81].

Infine, tale prospettiva può aiutare a comprendere e valutare le conseguenze politiche di una crescente polarizzazione affettiva, questione su cui finora non c'è né molta attenzione né consenso[82]. Per Spinoza il pericolo per la conservazione dello stato, e delle istituzioni in genere, è interno, proviene cioè dagli stessi cittadini che, guidati da una natura prevalentemente passionale, possono preferire i propri interessi privati al diritto pubblico[83]. Nelle dinamiche di polarizzazione

[78] Si veda Spinoza, TP 1 §5, p. 29.
[79] Per esempio, la dinamica di interrelazione tra sfera digitale e il fenomeno di polarizzazione è più complessa di quanto sembri, tanto che ci si interroga ora sull'impatto che la stessa polarizzazione possa avere sull'uso dei social media e non viceversa. Tra questi il recente saggio di M. Nordbrandt, *Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties*, «New Media & Society», 2021. Sull'interrelazione tra diverse dimensioni (instituzionale, socio-tecnica, tecnica) e su come i sistemi di IA costituiscano un peculiare tipo di sistema sociotecnico 'multi-agent' che richiede delle ulteriori norme tecniche per essere regolato si veda I. van de Poel, *Embedding Values in Artificial Intelligence (AI) Systems*, cit.
[80] È una delle caratteristiche fondamentali della polarizzazione affettiva evidenziata da L. Huddy, L. Mason, L. Aarøe, *Expressive partisanship: Campaign involvement, political emotion, and partisan identity*, «Am Polit Sci Rev.», 109, 2015, pp. 1-17.
[81] Sull'idea di parte e tutto si veda *infra*.
[82] È quanto è stato notato da J. N. Druckman, S. Klar, Y. Krupnikov e altri, *Affective polarization, local contexts and public opinion in America*, «*Nat Hum Behav*», 5, 2021, pp. 28–38.
[83] Si veda su questo quanto detto da Spinoza in TTP 17.

affettiva, quando la distanza percepita tra i gruppi è grande, i dibattiti sono vinti sulla base delle identità e non su argomenti e prove empiriche, provocando conseguenze negative, come l'aumento della disoccupazione e del divario socio-culturale tra i cittadini[84]. In Spinoza alti livelli di affetti-passioni che inducono sentimenti di diseguaglianza possono dare vita ad identità o fazioni omogenee e divisive, che rischiano di avere un impatto negativo sulla tenuta della società e delle istituzioni, di ridurre la possibilità che «gli uomini possano vivere in concordia ed aiutarsi», e di impedire il «comune consenso», elemento proprio della *civitas* o cittadinanza che ha stabilito regole comuni di vita[85].

Nella sfera digitale, la necessità di ottimizzare tempo e risorse è in costante crescita insieme alla dipendenza da strutture e reti associative che cercano di bilanciare interessi pubblici e privati[86]. In filosofia della tecnologia, per esempio, molti studiosi hanno manifestato una crescente preoccupazione per l'impatto che grandi corporazioni private guidate da interessi di mercato possono avere sui contesti pubblici e la ricerca pubblica[87]. Inoltre, sia l'opacità dei sistemi di IA che determinano il modo in cui le informazioni sono classificate sul news feed o timeline degli utenti, sia la mancanza di trasparenza su come questo processo venga gestito dalle piattaforme online, hanno sollevato non poche preoccupazioni sull'esacerbarsi di disuguaglianze e asimmetrie di potere[88]. La giustizia per Spinoza consiste nel «non avere alcun rispetto per le persone, bensì considerare tutti uguali, e tutelare ugualmente il diritto di ciascuno»[89]: cioè, ogni cittadino − sebbene abbia risorse e possibilità diverse dagli altri in accordo allo sviluppo del proprio *conatus* − ha pari diritto nello stato civile, e come tale deve essere tutelato nelle sue relazioni e nella sua partecipazione politica. Perciò, nel paragrafo conclusivo vorrei analizzare brevemente come alcuni concetti della filosofia di Spinoza possano non soltanto fornire maggiore comprensione e consapevolezza di fenomeni come quello della polarizzazione affettiva online, ma anche suggerire dei rimedi contro le loro possibili implicazioni negative.

3. Conclusioni. Verso alcune regole comuni per la sfera digitale

Come si è cercato di dimostrare in questo articolo, essenziale dovrebbe essere la comprensione della complessa e costante costruzione affettiva degli utenti. Nella filosofia politica di Spinoza, lo stato civile per promuovere armonia sociale cerca di re-orientare gli affetti di individui e collettività quando questi

[84] L. Mason, *Uncivil Agreement: How Politics Became Our Identity*, Chicago 2018.
[85] Cfr. Spinoza, E 4 PROP XXXVII, Scolio II, p. 277.
[86] Si veda su questo H. Bernholz et al., *Digital Technology and Democratic Theory*, cit.
[87] In questa direzione vanno i lavori di Tamar Sharon sulla 'Googlization of Health Research': T. Sharon, *The Googlization of health research: From disruptive innovation to disruptive ethics*, «Personalized Medicine», 13, 6, 2016, pp. 563-574.
[88] Sull'opacità e la mancanza di trasparenza si veda il saggio di F. Pasquale, *The Black Box Society*, Cambridge 2015.
[89] Spinoza, TTP 16, §4, p. 537.

sono prevalentemente irrazionali, con «comuni regole di vita» e leggi che contrappongano affetti più forti e contrari a quelli che li rendono passivi[90]. Questo processo all'interno della sfera digitale dovrebbe mirare a prevenire *imaginationis deliria,* cioè deliri dell'immaginazione basati su una comprensione parziale della realtà e motivo di superstizione[91], e promuovere, al contrario, consapevolezza e alfabetizzazione mediatica tra gli utenti, consentendo loro di esprimere al meglio le loro capacità per accedere e comprendere i contenuti online, nonché di contestarli quando questi sono in contrasto con un ambiente digitale sicuro per tutti gli utenti. Un affetto come quello dell'indignazione (*indignatio*), descritto come promotore di uno stato di ostilità tra i cittadini[92], potrebbe per esempio venir re-orientato e impiegato per la contestazione verso forme di ingiustizie all'interno degli spazi digitali, traducendosi in specifiche pratiche e meccanismi che tutelino trasparenza e responsabilità in relazione alle nuove tecnologie[93]. Per quanto riguarda poi il caso specifico di riduzione della polarizzazione, bisognerebbe trovare modalità condivise per correggere le percezioni errate degli utenti, quando inaccuratamente percepiscono gli altri individui e gruppi come estremamente diversi e distanti da loro, e ri-orientare affetti disgreganti come la superbia verso forme di riconoscimento e giustizia[94].

Inoltre, per Spinoza le differenze all'interno della società non vengono annullate in vista di un'ideale di uguaglianza essenziale, o morale[95]. Gli individui mantengono il loro *conatus* e i loro interessi distintivi, ma cercano di convergere assieme per promuovere l'utilità reciproca, cioè la preservazione e sviluppo del *conatus* di tutta la collettività[96]. Se si adotta questa prospettiva nella sfera digitale, ciò significa che le istituzioni e i vari gruppi all'interno della società, con diversi livelli di potere e diversi valori, dovrebbero cooperare per raggiungere soluzioni normative ottimali, ad esempio fornendo strumenti di co-regolamentazione. Per esempio, il *Digital Services Act,* che è uno tra i recenti tentativi di regolare la sfera digitale da parte dell'Europa, ha adottato un approccio graduale alle piattaforme online, con l'intento di proporre regole diverse per le diverse dimensioni della piattaforma di riferimento[97]. Tale approccio graduale ha inoltre

[90] Tale processo è illustrato in Spinoza, E 4 PROP XXXVII, Scolio II, p.277; si veda anche Spinoza, TTP 20 §6. Sulla questione che soltanto un altro affetto può dominare un affetto contrario si veda Spinoza, E 4 PROP VII, p. 245. Sul re-orientamento degli affetti e il ruolo dello stato civile molto si è scritto, un esempio: S. Steinberg, *Spinoza's Political Psychology: The Taming of Fortune and Fear,* New York 2018.
[91] Si veda Spinoza, TTP, Prefazione, p. 41.
[92] Spinoza, E III PROP 22, Scolio, p. 173.
[93] Sul re-orientamento dell'indignazione rimando a Spinoza TP 7 §2, p. 111.
[94] Questa questione è stata discussa nello studio empirico di D. J. Ahler, G. Sood G. *The parties in our heads: misperceptions about party composition and their consequences*, «J. Politics », 80, 3, 2018, pp. 964-81.
[95] Si veda su questo l'interessante analisi sul concetto di *equality* in B. Lord, *Spinoza, Equality, Hierarchy*, «History of Philosophy Quarterly», 31, 1, 2014, pp. 59-77.
[96] Si veda Spinoza, E 4 P XXXI, p. 265.
[97] Cfr. *Proposta di Regolamento del Parlamento Europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE*, COM (2020) 825 final, 15 Dicembre 2020.

sottolineato l'esigenza di avere una chiara supervisione pubblica sull'operato delle piattaforme, pur consentendo regolamentazioni flessibili[98]. Tuttavia, se si guarda la recente proposta di regolamento sull'IA, si può notare come anche questo ultimo tentativo in realtà promuova troppa discrezionalità su come bilanciare i diritti fondamentali degli utenti, visto che questo compito sembra essere affidato prevalentemente ad attori del settore privato e a meccanismi di autovalutazione, oscurando il ruolo fondamentale che invece potrebbero avere i cittadini e le organizzazioni della società civile in possibili processi di co-regolamentazione[99].

Infine, in Spinoza è essenziale il concetto di proporzione[100]. Per Spinoza «nessuno trasferisce il proprio diritto naturale a un altro in modo che in seguito non sia più consultato, ma lo trasferisce alla maggior parte di tutta la società (*Societatis partem*) della quale è membro, e in questo modo tutti rimangono uguali, come lo erano prima nello stato di natura»[101]. Sia nello stato di natura che nello stato civile, gli individui e le collettività sono, infatti, caratterizzati dalla proporzione secondo Spinoza: cioè, sono parti di un tutto più grande (*Societatis partem*), in cui non rinunciano al loro diritto naturale, ma lo condividono, e la loro potenza d'agire è proporzionata a quella degli altri[102]. Il compito, quindi, di un sistema di governance nella sfera digitale dovrebbe essere quello non di evitare i conflitti e le differenze tra gli utenti, ma cercare piuttosto di evitare che tali conflitti degenerino in un'erosione delle istituzioni democratiche, e nella compromissione della uguale partecipazione al potere politico da parte dei cittadini.

Nella sfera digitale, il *convenire* si potrebbe tradurre con l'istituzione di un ecosistema in grado di garantire diritti e regole comuni, che sia quindi capace di offrire maggiore protezione e sicurezza agli utenti e alla società in generale. Si sta ormai entrando, infatti, in una nuova fase di 'costituzionalismo digitale', la cui sfida principale sarebbe rappresentata dallo sviluppo di nuovi modi di regolamentazione e limitazione degli abusi di potere da parte di un sistema complesso che include governi, imprese, organizzazioni, individui o gruppi che hanno la capacità di creare contenuti, elaborare dati e guidare processi decisionali

[98] *Ibidem*.
[99] COM (2021), cit. Questa obiezione è stata sollevata anche da N. A. Smuha. *Beyond the individual: governing AI's societal harm*, «Internet Policy Review», 10, 3, 2021, pp. 1-32.
[100] Sul concetto di proporzione in Spinoza si veda l'approfondita analisi di Beth Lord: B. Lord, *Ratio as the basis of Spinoza's concept of equality*, in B Lord (a cura di), *Spinoza's Philosophy of Ratio,* Edinburgh, 2018, pp. 61-73; B. Lord, *Spinoza on natural inequality and the fiction of moral equality*, in S Ducheyne (a cura di), *Reassessing the Radical Enlightenment,* London, 2017, pp. 127-142; B. Lord, *Spinoza, Equality, Hierarchy*, cit.
[101] Spinoza, TTP XVI, §3, p. 533.
[102] Nella natura gli individui erano soggetti a «infinite altre leggi, che riguardano l'ordine eterno di tutta la natura, di cui l'uomo è una piccolissima parte» (Spinoza, TTP 16, §1, p. 521). Si veda anche l'Epistola 32 del 20 Novembre 1665, in B. De Spinoza, *Epistolario*, a c. di A. Droetto, Torino 1974, pp.119-120. L'epistolario spinoziano è nel IV volume degli *Opera* (*Opera,* herausgegeben von C. Gebhardt, 4 voll., C. Winters, Heidelberg s.d., 1925).

nella sfera digitale[103]. Pertanto, vista la rapidità e le peculiari implicazioni dei progressi apportati dall'innovazione, bisogna non soltanto comprendere i fenomeni complessi mediati dalle nuove tecnologie digitali, ma anche definire e esplorare dei principi in grado di guidare questa innovazione verso nuove forme di giustizia e responsabilità condivise, così come mezzi e pratiche per attuarli e rispettarli.

Marianna Capasso
Scuola Superiore Sant'Anna di Pisa
✉ marianna.capasso@santannapisa.it

Bibliografia

Ahler, D. J., Sood G. 2018. *The parties in our heads: misperceptions about party composition and their consequences*, «J. Politics», 80, 3, pp. 964–81.

Bail, C. A., Argyle, L. P., Brown, T. W. e al. 2018. *Exposure to opposing views on social media can increase political polarization*, «Proceedings of the National Academy of Sciences», 115, 37, pp. 9216–9221.

Barberá, P. 2020. *Social Media, Echo Chambers, and Political Polarization,* In N. Persily e J. A. Tucker (a cura di), *Social Media and Democracy: The State of the Field, Prospects for Reform,* Cambridge, Cambridge University Press, pp. 34-55.

Barnidge, M. 2017. *Exposure to political disagreement in social media versus face-to-face and anonymous online settings*, «Political Communication», 34, 2, pp. 302–321.

Bernholz, L., Landemore, H., Reich, R. 2021. *Digital Technology and Democratic Theory*, Chicago, University of Chicago Press.

Bonenfant, M., St-Martin, L. Iseut, L. 2020. *Crémier, Affected Data: Understanding Knowledge Production in Algorithmic Events*, «Global Media Journal, Canadian ed. Ottawa», 11, 2, 12, pp. 66-78.

Bove, L. 1986. *La stratégie du conatus: Affirmation and résistance chez Spinoza,* Paris, Vrin.

Capasso, M. 2022. *Manipulation as Digital Invasion: a neo-republican approach*, in F. Jongepier, M. Klenk (a cura di), *The Philosophy of Online Manipulation*, New York, Routledge, pp. 180-198.

[103] C. Padovani, M. Santaniello, *Digital Constitutionalism: Fundamental Rights and Power Limitation in the Internet Eco-System*, «International Communication Gazette», 80, 4, 2018, pp. 295-301; G. De Gregorio, *The rise of digital constitutionalism in the European Union*, «International Journal of Constitutional Law», 19, 1, 2021, pp. 41–70 e il più recente G. De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society*, Cambridge 2022.

COM (2021) 206 final, *Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione*, Bruxelles, 21.4.2021.

COM (2020) 825 final, *Proposta di Regolamento del Parlamento Europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE*, Bruxelles, 15.12.2020.

COM (2016) 288 Final, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni. Le piattaforme online e il mercato unico digitale Opportunità e sfide per l'Europa*, Bruxelles, 25.5.2016.

Cristofolini, P. 2009. *La scienza intuitiva in Spinoza*, Pisa, ETS.

De Gregorio, G. 2022. *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society*, Cambridge, Cambridge University Press.

De Gregorio, G. 2021. *The rise of digital constitutionalism in the European Union*, «International Journal of Constitutional Law», 19, 1, pp. 41–70.

De Lucia Dahlbeck, M. 2018. *Spinoza, Ecology and International Law: Radical Naturalism in the Face of the Anthropocene***,** London, Routledge.

Del Vicario, M., Bessi, A., Zollo F. e al. 2016. *The spreading of misinformation online*, «Proceedings of the National Academy of Sciences», 113, 3, pp. 554–559.

Druckman, J. N., Klar, S., Krupnikov, Y. e al. 2021. *Affective polarization, local contexts and public opinion in America*. «*Nat Hum Behav*», 5, pp. 28–38.

Druckman, J. N., Levendusky, M. S., McLain, A. 2018. *No need to watch: how the effects of partisan media can spread via interpersonal discussions*, «*Am. J. Political Science*», 62, 1, pp. 99–112.

Farrell, T., Fernandez, M., Novotny, J., Alani, H. 2019. *Exploring Misogyny across the Manosphere in Reddit*, In WebSci '19 Proceedings of the 10th ACM Conference on Web Science, New York, Association for Computing Machinery, pp.87-96.

Fossa, F., Schiaffonati, V., Tamburrini G. (a cura di). 2021. *Automi e Persone. Introduzione all'Etica dell'Intelligenza Artificiale e della Robotica*, Roma, Carocci Editore.

Garrett, R. 2009. *Politically motivated reinforcement seeking: Reframing the selective exposure debate*, «Journal of Communication», 59, 4, pp. 676–699

Gounari, P. 2022. *From Twitter to Capitol Hill:* Far-Right Authoritarian Populist Discourses*, Social Media and Critical Pedagogy*, Leiden/Boston, Brill.

Heatherly, K. A., Lu, Y., Lee, J. K. 2017. *Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites*, «New Media and Society», 19, 8, pp. 1271–1289

Huddy, L., Mason, L., Aarøe, L. 2015. *Expressive partisanship: Campaign involvement, political emotion, and partisan identity*, «Am Polit Sci Rev. »109, pp. 1–17.

Iyengar, S., Lelkes, Y. Levendusky, M. 2019. *The origins and consequences of affective polarization in the United States*, «Annual Review of Political Science», 22, 1, pp. 129–146.

Iyengar, S., Sood, G., Lelkes, Y. 2012. *Affect, not ideology: A social identity perspective on polarization*, «Public Opinion Quarterly», 76, 3, pp. 405–431.

Kalpokas, I. 2021. *Towards an affective philosophy of the digital: Posthumanism, hybrid agglomeproporzionens and Spinoza*, «Philosophy & Social Criticism», 47, 6, pp. 702-722.

Karlsen, R., Steen-Johnsen, K., Wollebæk, D., Enjolras, B. 2017. *Echo chamber and trench warfare dynamics in online debates*, «European Journal of Communication», 32, 3, pp. 257-273.

Langhi, R. 2019. *Un'introduzione critica sugli affect studies*, «Enthymema», 24, pp. 171-188.

Levy, G., Razin, R. 2020. *Social Media and Political Polarisation*, «LSE Public Policy Review», 1, 1, 5, pp. 1–7.

Lord, B. 2018. *Ratio as the basis of Spinoza's concept of equality*, in B. Lord (a cura di), *Spinoza's Philosophy of Ratio,* Edinburgh, Edinburgh University Press, pp.61-73.

Lord, B. 2017. *Spinoza on natural inequality and the fiction of moral equality*, in S. Ducheyne (a cura di), *Reassessing the Radical Enlightenment,* London, Routledge, pp. 127-142.

Lord, B. 2017. *Disagreement in the Political Philosophy of Spinoza and Rancière*, «Proceedings of the Aristotelian Society», 117, 1, pp. 61-80.

Lord, B. 2014. *Spinoza, Equality, Hierarchy*, «History of Philosophy Quarterly», 31, 1, pp. 59-77.

Lord, B. 2010. *Spinoza Ethics*, Edinburgh, Edinburgh University Press.

Maher, P. J., Igou, E. R., van Tilburg, W. 2018. *Brexit, Trump, and the Polarizing Effect of Disillusionment*, «*Social Psychological and Personality Science*», 9, 2, pp. 205-213.

Mason, L. 2018. *Uncivil Agreement: How Politics Became Our Identity*, Chicago, University of Chicago Press.

Massumi, B. 1995. *The Autonomy of Affect*, «Cultural Critique», 31, pp. 83-109.

Matheron, A. 1968. *Individu et communauté chez Spinoza*, Paris, de Minuit.

Messing, S., Westwood, S. J. 2014. *Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online*, «Communication Research», 41, 8, pp. 1042–1063.

Mouffe, C. 2005. *On the Political*. London, Routledge.

Muller, K., Schwarz, C. 2017. *Fanning the flames of hate: Social media and hate crime*, «Journal of The European Economic Association», 19, 4, pp. 2131-2167.

Nagle, A. 2017. *Kill All Normies: Online Culture Wars From4Chan And Tumblr To Trump And The Alt-Right*, Winchester, Zero Books.

Negri, A. 1981. *L'anomalia selvaggia. Saggio su potere e potenza in Baruch Spinoza*, Milano, Feltrinelli.

Nordbrandt, M. 2021. *Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties*, «New Media & Society».

Padovani, C., Santaniello, M. 2018. *Digital Constitutionalism: Fundamental Rights and Power Limitation in the Internet Eco-System*, «International Communication Gazette», 80, 4, pp. 295-301.

Pariser, E. 2011. *The filter bubble: What the Internet is hiding from you,* London, Penguin Press.

Pasquale, F. 2015. *The Black Box Society*, Cambridge, Harvard University Press.

Pasquino, G. 2008. *Democrazie perfettibili*, «il Mulino, Rivista trimestrale di cultura e di politica», 3, pp. 499-505.

Putnam, R. D. 2000. *Bowling Alone: America's Declining Social Capital*, In: Crothers, L., Lockhart, C. (a cura di), *Culture and Politics.* New York, Palgrave Macmillan, pp. 223-234.

Reiljan, A. 2020. *Fear and loathing across party lines (also) in Europe: affective polarisation in European party systems,* «European Journal of Political Research», 59, 2, pp. 376–396.

Ribeiro, H., Blackburn, M., Bradlyn, J., De Cristofaro, B., Stringhini, E., Long, G., Greenberg, S., Zannettou, S. 2021. *The Evolution of the Manosphere across the Web*, «Proceedings of the International AAAI Conference on Web and social media», 15, 1, pp. 196-207.

Rost, K., Stahel, L., Frey, B. S. 2016. *Digital Social Norm Enforcement: Online Firestorms in Social Media*, «PLoS ONE», 11, 6.

Sætra, H. S. 2021. *AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System*, «Sustainability», 13, 4, 1738.

Settle, J. E. 2018. *Frenemies: How Social Media Polarizes America*, New York, Cambridge University Press.

Smuha, N. 2021. *Beyond the individual: governing AI's societal harm*. «Internet Policy Review», 10, 3, pp. 1-32.

Sharon, T. 2016. *The Googlization of health research: From disruptive innovation to disruptive ethics*, «Personalized Medicine», 13, 6, pp. 563-574.

Shmargad, Y., Coe, K., Kenski, K., Rains, S. A. 2021. *Social Norms and the Dynamics of Online Incivility*, «Social Science Computer Review», 40, 3, pp. 717-735.

Spinoza, B. 1972 (I ed. 1925). *Tractatus politicus*, in C. Gebhardt, (a cura di), *Opera. Im Auftrage der Heidelberger Akademie der Wissenschaften*, C. Winter Universität, Heidelberg, 4 voll.; Spinoza, B. 2011. *Trattato politico*, trad. it. di P. Cristofolini, Pisa, ETS.

Spinoza, B. 1972 (I ed. 1925). *Tractatus theologico-politicus*, in C. Gebhardt, (a cura di), *Opera. Im Auftrage der Heidelberger Akademie der Wissenschaften*,

C. Winter Universität, Heidelberg, vol. III; Spinoza, B. 2010. *Trattato teologico-politico*, trad. it di A. Dini, Milano, Bompiani.

Spinoza, B. 1972 (I ed. 1925). *Ethica. Ordine Geometrico demonstrata*, in C. Gebhardt (a cura di), *Opera. Im Auftrage der Heidelberger Akademie der Wissenschaften*, C. Winter Universität, Heidelberg, vol. II; Spinoza, B. 2010. *Etica: dimostrata con metodo geometrico*, trad. it. di P. Cristofolini, Pisa, ETS.

Spinoza, B. 1974. *Epistolario*, a c. di A. Droetto, Torino, Einaudi.

Steinberg, S. 2018. *Spinoza's Political Psychology: The Taming of Fortune and Fear,* New York, Cambridge University Press.

Sunstein, C. R., 2018. *#Republic: Divided Democracy in the Age of Social Media*, Princeton, Princeton University Press.

Sunstein, C. R. 2001. *Republic.com*, Princeton, Princeton University Press.

Thaler, R. H., Sunstein, C. R. 2008. *Nudge: improving decisions about health, wealth, and happiness,* New Haven, Yale University Press.

Törnberg, P., Andersson, C., Lindgren, K., Banisch S. 2021. *Modeling the emergence of affective polarization in the social media society*. «PLoS ONE», 16, 10.

van de Poel, I. 2020. *Embedding Values in Artificial Intelligence (AI) Systems*. «Minds & Machines», 30, pp. 385–409.

Weinmann, M., Schneider, C., vom Brocke, J. 2016. *Digital Nudging*, «Bus Inf Syst Eng», 58, 6, pp. 433-436.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770053

Articoli/2

Algorithm and Learning

Ethical and Theoretical Hypothesis on Empathic Robotics and Social Robotics

Igor Pelgreffi 0000-0002-4490-009X

Articolo sottoposto a doppia blind peer review. Inviato il 31/08/2022. Accettato 17/09/2022.

The article explores the relationship between algorithms, machine learning and artificial empathy. The pivotal problem is whether and how artificial learning of emotions by artificial agents, guided by algorithms, is possible. The first question consists in an in-depth analysis of the concept of learning (through some references to anthropotechnics, to Aristotle, to Merleau-Ponty and to Bateson), showing that – even in the human field – learning involves the use of forms of automatic and procedural repetition, such as the body schema or the behavioural algorithm. The hybrid aspect (between technique and nature) of learning processes in a broad sense is therefore underlined. The second issue is emotional learning by artificial agents. Through the use of an extensive specific bibliography, especially in the IT-engineering field, the possibilities and limits of this possibility are shown, discussing the major knot, namely that the machine or the algorithm has no living body (*Leib*), but mechanical (*Körper*). In the artificial field, the repetition and learning of patterns appear to be very far from that of human beings. However, the article tries to stress the usual categories and to formulate hypotheses, highlighting the possibilities of heterogeneous hybridization between the repetition learned in the machine and its ability to complexify its own behaviours, making them much less dissimilar from those of natural agents. The third question is that of the different ways in which the 'affective' dimension of the machine can be interpreted, through a relational systemic approach (social robotics) in which artificial empathy takes shape only within the relationship with the other. Different topics are explored, from those of emotional grafting (emotional chip) to that of a form of autopoiesis in the direction of a form of artificial autonomy, today in a larval but not absent form, through references especially to Dumouchel and Guattari. In the concluding part, some hypotheses are formulated on a conceptual frame for managing the ethical issue of autonomy, potentially also on an emotional level, of artificial agents guided by algorithms, suggesting an attitude capable of thinking of an ethics of hybridization but at the same time a hybridization of the usual ethical attitudes, that goes beyond the opposition between roboethics and machine ethics.

***

1. Algorithm and Learning: a background

The aim of this paper is to analyse the relationship between algorithm and automatic learning in the field of artificial empathy, but also to explore a number of hypotheses, that expect to be verified with further studies. Algorithm will be meant as a device that processes information in an increasingly organic and holistic way, rather than linear or serial, as in the field of *deep neural learning*, therefore as a device capable of solving *totally new* problems by learning from the environment, operating with increasing complexity, up to the extreme level: algorithm *that controls parts of the algorithm itself*. We reach here, therefore, the topic of *operational self-reflexivity* of the calculation machine, if not that of *machinic autonomy*. This last represents an extreme level of the questions. It concerns the *vexata quaestio* of the distinction between functioning and behaviour. After all, this is the philosophically central knot: will artificial agents be able to emancipate themselves completely? Will programmed algorithmic systems become autonomous through self-learning processes? The theme of *artificial autonomy*, recently re-proposed by Luca Fabbris[1], seems to be the only approach that leads to a real thought of AI understood as self-generation both of its behavioural patterns then of *autopoiesis* of its own ends[2].

But a further question should be introduced: can artificial systems learn an emotion? Can they gain *empathic autonomy*? The question, unlike that of autonomy in the computational field, involves the question of the body: can artificial agents feel emotions, if they have no living body (*Leib*)? The answer seems to be negative. If 'thought' for the machine today seems the undecidable, 'having a body' seems the impossible. Nonetheless, this article will suggest some investigative hypotheses on empathic learning by working on the 'margins' of the body-machine relationship, that is on borderline issues, still not fully codified and not even fully clear in their very definition.

Two preliminary questions. The first. Although the decision-making skill [3] of an algorithm still remain in the field of predictability or preventive programming, albeit stochastic, they collide with an *uncanny* empirical reality: programmers are unable to fully explain the behaviour of the algorithm, which becomes similar to a *black boxes model*[4]. Proof of this is the fact that a specific field of

[1] L. Fabbris, *Il programma, la rete, il circolo: un approccio Io-fi all'Autonomia artificiale*, «aut aut», 392, 2021, pp. 63-76.
[2] About a definition of autonomy, cf. also W. F. Lawless et al (eds.), *Autonomy and Artificial Intelligence. A Threat or a Savior?*, Cham 2017; J. Liu et al (eds.), *Autonomy Oriented Computing. From Problem Solving to Complex Systems Modelling*, Kluwer Academic Publisher, New York 2005; M. Prokopenko (ed.), *Guided Self-Organization. Inception*, Heidelberg 2014.
[3] On the value of the 'algorithmic decision', cf. at least G. Tamburrini, *Etica delle macchine*, Roma 2000 and, with extensive bibliography, F. Fossa, V. Schiaffonati, G. Tamburrini, *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Roma 2021.
[4] E. Esposito, *Dall'intelligenza artificiale alla comunicazione artificiale*, «aut aut», 392, 2021, pp. 20-34.

research is born under the name of *Explainable IA*[5]. The shift (historical, material: empirical) from single algorithms to huge extended algorithmic systems, which are massive computational machines[6], has among its consequences the existence of outputs not entirely derivable from inputs. The loss of total intelligibility represents a drastic break in the classical-modern logic at the base of algorithmic programming: between artificer and artefacts a relationship begins to emerge that is no longer only univocal or linearly causal. All this evidently opens up to great problem in the ethical field[7], promptly addressed also by the international guidelines for the ethical programming of intelligent machines[8].

However, the heuristic emphasis entrusted to concepts such as transparency, explainability, explicability or understandability[9] seems today to be more of a 'defensive' than 'normative'[10] character with respect to a technological progression that confronts us with outcomes that are not entirely controllable. The understanding of an unexpected or divergent 'algorithmic event' occurs *ex-post*: it would therefore not be a real understanding, but a rational justification possible only after the 'decision' of the artificial agent has already taken place. According to Elena Esposito, in the case of «deep learning algorithms, [*this*] is a fundamental obstacle: if by explanation we mean a procedure that allows human observers to understand what the machine does and why, the undertaking is hopeless»[11].

The second question concerns the relationship between machine and emotion: can an algorithm learn emotions? If we accept a general conception of emotion as an «inner process aroused by an event-stimulus relevant to the

[5] Concerning *Explainable AI* concept, see S. Whatcher, B. Mittelstadt, L. Floridi, *Transaparent, Explainable and Accountable AI for Robotics*, «Science Robotics», 6, 2017; D. Weinberger, *Machines Now Have Knowledge We'll Never Understand*, «Wired», 18.4.2017; F. Doshi-Velez, B. Kim, *Towards a rigorous Science of Interpretable Machine Learning*, arXiv:1702.08608v2, 2017; T. Miller, Explanation in *Artificial Intelligence. Insights from the Social Sciences*, «Artificial Intelligence», 267, 2019, pp. 1-38.

[6] N. Bostrom, *Superintelligence. Paths, Dangers, Strategies*, Oxford University Press, Oxford 2014.

[7] S. Vallor, G. A. Bekey, *Artificial Intelligence and the Ethics of Self-Learning Robots*, in Lin et al, *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*, Oxford University Press, New York 2017, pp. 338-353.

[8] Such as: EU, *White Paper on Artificial Intelligence. A European approach to excellence and trust*, Bruxelles 2020; High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI*, EU Guidelines, Bruxelles 2019 OECD, *OECD Council Recommendation on Artificial Intelligence* (2021), OECD, Paris 2021 *e* OECD, *Principles on AI OECD*, Paris 2020.

[9] A. Ienca, E. Vayena, *The global Landscape of AI Ethics Guidelines*, «Nature Machine Intelligence», 1, 9, 2019, pp. 389-399.

[10] The fundamental study still remains M. Ryan, B. Stahl, B., *Artificial Intelligence Guidelines for Developers and Users*, «Journal of Information, Communication and Ethics in Society», 19, 1, 2021, pp. 61-86.

[11] E. Esposito, *Dall'intelligenza artificiale alla comunicazione artificiale*, cit., p. 28.

interests of the individual»[12], therefore a physical dynamic linked to changes in the body and its neuro-physiological state[13], the answer can only be negative.

What does the machine really learn through increasingly accelerated repetition cycles, in training and self-correction that is progressively effective and adaptive? It only learns to *register* stimuli, to *recognize* emotions, to *categorize* frames, even if they are complex or 'open': the machine has no body. This also explains why in the setting of studies on artificial empathy[14] the dichotomy between two types of emotions is often affirmed: internal emotions and external emotions, where the former would be emotions that arise internally to non-artificial agents (typically humans and animals) and that the machine records and decrypts; the second would be those reproduced mechanically by artificial agents, typically shown by them to human beings. The opposition would therefore be between «genuine human emotions and simulated robotic emotions»[15]. In the following we will show how one can, cautiously and on the basis of a series of empirical and laboratory findings, attempt to deconstruct the internal/external partition, if one follows a more interactive, open or 'organic' model of Learning in the empathic field and if one he is willing to take on more complex epistemological but also ethical models. Before reaching this point, however, let us dwell on some aspects of the concept of Learning.

2. Learning and repetition

2.1 Functioning or behaviour

Speaking about algorithm that learns, are we not anthropomorphizing, that is, assigning the ability of 'learning' to something that by its nature cannot learn, but rather only re-produce, copy or execute what has already been 'learned'? Are we not naively applying a human lexicon to an artificial agent without a living body (*Leib*) but with only a mechanical body (*Körper*)? And even if it is legitimate to speak of learning in a computational machine, what is the object of the process? Of course, the machine purchases new functions, new protocol

[12] See «Emozione», in «Enciclopedia Treccani on-line» [12.3.22].
[13] This article will not examine the different theories of emotion, nor the nuances around the emotion/feeling distinction. As a basic orientation, *feelings* will be understood as elements more displaced, compared to emotions, on a subjective level, which stabilize external emotional stimuli over time; *emotions* instead as mechanisms closest to the psycho-bio-chemical level of the organism's response to the environment. Emotions and feelings will be understood, that is, as distinct but not ontologically independent, according to a line present, among others, in Antonio Damasio. Cf. A. Damasio, *Looking for Spinoza. Joy, Sorrow and the Feeling Brain*, HMH, Montreal 2004.
[14] As a reference, much used in this article, take P. Dumouchel, L. Damiano, *Vivre avec les robots. Essai sur l'empatie artificielle*, Seuil, Paris 2016. See also D. A. Norman, *Emotional Machines* in Id., *Emotional design*, Basics Books, New York 2004 and S. Schmetkamp, *Understanding AI. Can and Should we Empathize with Robots?*, «Review of Philosophy and Psychology», 2020, 11, pp. 881-887.
[15] P. Dumouchel, *Vivre avec les robots*, cit., p. 26.

segments, new automatic reactions to environmental stimuli: new 'machine fragments', in the pragmatic sense of machinic operation. That said, can we really claim that the machine learns new behaviours? Here again, the problem of anthropomorphization looms: the term behaviour is widely used in scientific and technical-engineering literature, but on a philosophical level it denotes a problem. Perhaps it would be more correct to use the term *functioning*? But, even here, it is quite clear that the lack of a correct distinction between *functioning* and *behaviour* can make any ethical reasoning on Learning very unsteady, as its foundations are unstable. So let's try to go in the opposite direction: beyond the easy man/machine mirroring effect – typical in these topics – let's try to reflect on a technical, if not 'machinic', component *latu sensu* present in human learning.

Since ancient times, philosophical thought on the meaning of learning has tried to define or contain an over-determined experience. Learning as reminiscence, how to remember something already 'written' (programmed?) but forgotten, as for the ἀνάμνησις in the *Meno* of Plato? Or learning as a recording, in the 'memory of the body', of functions or behaviours, of motor-patterns halfway between natural and cultural, as the notion of *hexis* is sometimes read in Aristotle, for example in the *Nicomachean Ethics*? Let's take a leap in time: in contemporary theories of learning in Cognitive Sciences, a widespread concept understands learning as a modification and acquisition of knowledge or behaviour. On the philosophical level, the notable point of this conception is to think together, that is, on the same logical level, *modification* and *acquisition*. How is this possible? Generally speaking, if I acquire an object, I do not necessarily change the behavioural pattern. But it is also true that if I modify behaviour, it does not mean that it is also incorporating an object.

As a first approximation, then, the heart of the learning process lies precisely in the *hybridization* between modification and acquisition, between insertion (passive) and modification (active). Here too, we find (as in all definitions of learning) an over-determined element on the logical level, which however on the practical-concrete level synthesizes in itself its own form.

We cannot dwell on these aspects; however, whatever the definition assumed, in learning it cannot be excluded an important, if not decisive, reference to the figure of repetition. Hanging up means being able to manage a relationship with repetition. Repetition is present not only in the reification, but also in the process, that is, in the 'becoming other' of a scheme (cognitive, bodily, cultural): in every experience that leads to having learned the object, repetition is essential.

Learning processes shows a close relationship with that of reiteration. This is likewise observable by recovering some layers of our philosophical tradition, in particular on the themes of doubling, of the *habit* of performing an action, of *second nature* understood as deep incorporation of *procedural automatisms*. In fact, it has been a debated issue since Aristotle the repetition of behaviour or the repeated practice as the key to conceiving a structuring of the ethical character,

for example, which leads to being-social agents. In *Nicomachean Ethics*, the ethical value of habit (*hexis*) was that of a *repeated action* can lead to virtue (*arete*)[16]. The *hexis* is an *embodied* and *embedded* disposition which also has its own transformative power, for example of an action that is not immediately moral in an action that is structured and consolidated as ethics, in the practical domain and in the social use[17].

The achievement of automaticity in the repeated gesture means that the learned gesture 'becomes natural', embodied as a second nature: and all this already means speaking of automaticity, of activation of a passive form, that is of activation in the human agent, *mutatis mutandis*, of an algorithmic procedure.

2.2 Technique and nature

The boundary between nature and technique shows itself – in the construction of the agent as an entity halfway between natural and artificial – much less rigid. To a certain extent, if the *hexis* can be said to be natural, it is because *it completes* human nature, extending the 'virtual-algorithmic' element with which the man is 'naturally' equipped. Learning then also becomes synonymous of experiencing, of living an active/passive dimension (also typical of a highly evolved artificial algorithm which, as mentioned, can even self-modify in the Learning process). But this transition can take place, from the point of view of the anthropotechnical question – from Mauss to Sloterdijk[18] – only thanks to the repeated exercise. From this point of view, the human being also learns to act *by applying a program*, exploring new input-output combinations, selecting some solutions over others and, in that, habit is a 'shaping' character – it *gives shape* – of our being. Shaping should be read in the sense of a future behavioural frame that self-codifies and stabilizes, orienting future choices. Somehow man needs *routine*, not to deny but to cohabit with his nature as an automaton, so that in the exemption of 'lower' mechanical functions, linked to procedural automatism, 'good' actions are performed without friction, automatically: in a natural way, as if a new physics of the body or a new form of the substitute were in play. The theme of the Aristotelian *hexis* therefore denotes the matter of procedural automatisms, which are nothing more than executable programs by now acquired but whose 'artificial nature' we have forgotten, and this is visible in tracery in that long path which leads to the heart of that reflection on the technique of the twentieth century: the topics of repetition and difference in French Though (Deleuze, Derrida), of alienation (Adorno, Horkheimer, Pollock,

[16] See Aristotle *Nicomachean Ethics*, tr. by R. Crisp, Cambridge University Press, Cambridge 2000, 1144a 26-28.
[17] *Ibid.*
[18] M. Mauss, *Techniques of the body* (1934), tr. eng. B. Brewster, «Economy and Society», 2, 1973, pp. 70-88 and P. Sloterdijk, *Du mußt dein Leben* ändern. Über Anthropotechnik, Frankfurt 2009.

Benjamin), of man as a *naturally artificial being* or even *naturally technical*, in German philosophical anthropology (Gehlen, Anders, Plessner).

However, repetition is also one of the features of every technical object: basically, technique *is* repetition. The graft in us of procedural automatisms is, indeed, the greatest sign of our 'technicalization'. The latter is, in turn, what allows us to reach skills and levels of expertise, to be ourselves 'good algorithms', that is 'effective' and is not something that is opposed to a supposed natural genuineness of the human being. In the wake of some typical findings of contemporary French epistemology, from Canguilhem to Simondon up to Derrida, we could say that between technique and body, between repetition and nature, between protocol and living body, no clear separation should be understood, but rather a sort of original reciprocal overlap: «The relationship between physis and technique is not an opposition; from the very beginning there has been *instrumentalisation*. The term 'instrument' is inappropriate in the context of original technicality. However, a repetition prosthetic strategy inhabits the very movement of life: life is a process of self-replacement, the passing away of life is a *mechanike*, a form of technique. Not only, then, is technology not in opposition to life, but it haunts it (*hante*) from its first beginning»[19].

2.3 Learning to unlearn

It must be noted that if learning deals with repetition, it deals also with a kind of repetition that is not trivially identical or merely mimetic: learning is indeed repeating, but in some way it is a repeating differently, a reiterating diverging in a way that, superficially, we can define as 'not mechanical'. Learning authentically also means *learning to unlearn* as well as *learning to resist*[20] to our own automatisms: the strength of the built-in algorithm. But what do we mean by the expression 'not mechanical'? Is there really something not mechanical in learning? Partly yes and partly no, as we have already shown limited to the theme of *hexis*, of repeated action in the anthropotechnical exercise, and gradually we could go as far as the theme of adaptive reinforcement or that of the formation of *body patterns*.

In learning, repetition is intrinsic. But is it an identical or mutable repetition? These are over-determined questions. Whether there is a non-repetitive element, we don't know. Generally, we are led to reply positively, and perhaps the problem lies in understanding the reason for this apotropaic reflex to deny the coexistence of inert repetition and living repetition, of identity and difference in learning. We should say that neuroscience[21] has not yet established

[19] J. Derrida, *Nietzsche and the machine*, «Journal of Nietzsche Studies», 7, 1994, pp. 7-66.
[20] From this point of view, some recent studies by Houdé are important, such as O. Houdé, *Apprendre à résister*, Le Pommier, Paris 2017.
[21] See J. Bernacer, J. A. Lombo, J. I. Murillo (eds.), *Habits: plasticity, learning and freedom*, «Frontiers in Human Neuroscience», 2015 (on line); I. Testa, F. Caruana (eds.), *Habits. Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory*, Cambridge 2020.

certainty about all of this. Neuroeducation[22], in turn, reports solutions to this problem which, in the progress of its discoveries, gradually appear less and less certain or unambiguous. Phenomenology of the body, which have pointed, since Merleau-Ponty, to an idea of body learning intrinsically capable of producing repetition and difference, also does not completely solve the problem. Every organism, thanks to fundamentally repeatable physical patterns and structures, learns to diverge: learning «never consists in being made capable of repeating the same gesture, but of providing an adapted response to the situation by different means. Nor is the response acquired with regard to an individual situation. It is rather a question of a new attitude for resolving a series of problems of the same form»[23]. Note also: the greatest difference between the living organism and the machine would lie precisely in the operational and functional capabilities of the *Leib*, which while processing the input transforms it into 'other' than being the mere premise of an automatic reaction: it also diverges from its presumed datum identity: «the organism [...] cannot be compared to a keyboard on which the external stimuli would play and in which their proper form would be delineated for the simple reason that the organism contributes to the constitution of that form»[24].

All this because, in the phenomenological setting, we not only have but also are a body, a living body, a *Leib* that cannot be reduced to its *Körper* 'machinity'. But in reality, other readings of the phenomena of productive or generative learning are also possible that focus more on the fact that the living body is both mechanical (*Körper*) and not mechanical, that is, both mechanical and, at the same time, *more* than mechanical or *less* than mechanical. The *Körperschema* is the *sinolus* between scheme, mechanical protocol and adaptive plasticity to the environment, and we can therefore advance the hypothesis that the 'body-that-learns' is structurally not reducible, on a theoretical level, to one of the two regimes.

The body-that-learns is therefore operation and function, both vital and algorithmic. It is something qualitatively different from the expression of a simple mechanism, on the one hand, and from a field of 'freedom', of absolute release from physical-mechanical needs on the other. From this point of view, the body-that-learns is always an element of indeterminate mediation: an hybridization, a process halfway between Mechanism and Vitalism. The theoretical references could touch here on the issue of the intrinsically complex dimension of the repetition system in the living (but also in cybernetics) found in the works of Gregory Bateson in *Steps to an ecology of mind*. [25] The body-that-learns

[22] On neuro-education, see O. Houdé, *La neuroéducation: magie ou science? Cerveau & Psycho/ Pour la science*, «Chronique. L'école des cerveaux», 86, 2017, pp. 80-83 and F. Ramus, *Neuroéducation et neuropsychanalyse: du neuroenchantement aux neurofoutaises*, «Intellectica», 2018 1-2, 69, pp. 289-301.
[23] M. Merleau-Ponty, *The Structure of Behaviour* [1942], Beacon, Boston 1963, p. 124.
[24] Ivi, p. 13.
[25] See G. Bateson, *Steps to an ecology of mind*, Ballantine Books, New York 1971.

obviously acquires information, it 'feeds' on data; but it already does so by vitally interpreting those same data. But its real historical and material consistency lies in this ontological ambiguity. Bateson, as is well known, read cybernetics, the feedback and adaptive circuits of data processing systems (biological or artificial), in this very open perspective, in which ontological outcomes were never taken for granted, in a unitary theoretical framework that embraced in a consistent way the themes of learning, deutero-learning, unconscious learning and computational learning. In this line, it cannot then be excluded that even the elementary behaviour, or the pre-logical instinct, is in turn constructed or artificial, understood as a sequence of behavioural stages inherited in the body, and therefore basically an algorithm.

In short, learning means acquiring not only information, but 'plastic methods' for solving problems. The expression 'body-that-learns' is equivalent to that of 'data processor' to which, however, is added the idea of being also a powerful multiplier of incorporation of schemes and algorithms. Indeed, we can hypothesize that the body-that-learns, that is, the *technical operation* of *synthesis* that allows learning, is basically a form of hybridization between the scheme (even *Körperschema*) and the algorithm: one without the other, it makes no sense. And then it is worth remembering Simondon's words, on the value assumed by the concept of *technical operation*: «the technical operation is a pure operation that brings into play the authentic laws of natural reality; the artificial is aroused by the natural, not by the false or by the human taken for natural»[26].

Understood in this way, however, the 'open' and generative learning process abandons its anthropocentric quality, as even an artificial agent, within certain limits, can learn and participate in the *technical operation* of 'understanding the world'.

3. Emotions in the machine?

3.1 Emotion and learning

Can an artificial agent learn emotions? Let's start by seeing how the iterability component and the aspect of technical-nature hybridization seen so far, typical of human learning processes, are also very typical in the field of empathic and social robotics. And we remember how today, with different degrees, we are faced with robotic agents run by algorithms that learn to recognize, manage, reproduce stimuli attributable to emotions: sadness, joy, fear and their combinations[27]. Artificial empathy takes hold in very different fields: companion robots, care robots (elderly, temporary or permanent disabled people); robots for playing and for free time; robots to communicate in an increasingly effective and human-like manner; robots in the most varied activities: commercial, industrial, medical, sexual, military, educational. The set of these artificial agents interacts

[26] G. Simondon, *Sur la technique*, PUF, Paris 2014.
[27] For a broad and consistent review, see P. Dumouchel, *Vivre avec les robots*, cit., pp. 110-115.

with the human 'counterpart', simulating emotions, facial expressions, sounds, movements, postures increasingly similar to those of living beings, where repetition is crucial along the 'learning line' scheme, test, measure (or check), that becomes new scheme, new test, new measure, as confirmed in many studies, starting from those on the «affective computing» by Picard[28]. Again, something very similar to how humans also 'experience' the world.

Many ethical questions arise from the analysis of the historical-technological framework just mentioned. That a machine can, for example, imitate human emotions or recognize human emotions, raises some questions: why does it do this? For the benefit of whom (of the Alzheimer's patient who finds benefit from the stimulating and 'convincing' companionship of the empathic robot; or the robot factory itself; or the exemption from commitment by the patient's own family micro-network of social relations?). And again: how does it do it? What dangers and what advantages[29] does this technology of artificialization of emotion involve?

From a philosophical point of view, this kind of learning is about *representation*: a machinic apparatus recognizes a living thing by reducing it to an object, to a code, to memory. We are therefore still mainly on the representational or semiotic level – and certainly not corporeal or emotional – of learning. Here then returns the question, already foreseen, of the distinction between *internal* and *external* emotions[30], the first considered authentic, 'true', expressive of a 'subjective' psycho-physical reality; the second considered constructed, artificial, externalized by the machine without any feeling, without concrete expressive corporeality, deceptive[31].

But things are actually more complex than this framework might suggest. Many of these *intelligent artificial systems* begin to *autonomously* distinguish emotions. For the moment, these are necessarily elementary learning, which develop in ways that today are techno-physiological, biochemical or even bio-mechanical, little more than an artificialization of the reflex arc mechanisms. In other words, 'feeling affection' corresponds to a sort of small autonomous (but still mechanical) perception, to resume the Leibnizian lexicon of *small perceptions* as a zero degree of a process of subjectivation of being.

Within these limits, however, we could say – often without a real explainability – robots today *are learning to feel emotions*. And repetition plays its own role in the emotional learning process. This last, in its *initial* or *basal*

[28] See R. W. Picard, *Affective Computing*, M.I.T. Media Laboratory Perceptual Computing Section, Technical Report n. 321 (Boston 1995); L. A. Camras, *Expressive development and basic emotions*, «Cognition and Emotion», vol. 6, n. 3 and 4, 1992; I. R. Murray, J. L. Arnott, *Toward the simulation of emotion in synthetic speech*, «J. Acoust. Soc. Am.», vol. 93, feb 1993, pp. 1097-1108.
[29] N. Bostrom, *Existential Risks. Analysing Human Extinction Scenarios*, «Journal of Evolution and Tecnhology», 9, 2002.
[30] See D. Parisi, *Internal robotics*, «Connections Science», 16,4, 2004, pp. 325-338; Id., *Future robots. Towards a robotics science of human beings*, John Benjamins, London 2014.
[31] See P. Dumouchel, *Vivre avec les robots*, cit., pp. 91-132.

sequences, is nothing more than a form of coding: a) tracing certain sensitive data (images, tactile sensations, sounds) acquired by sensors within predetermined frames: categorizing the data (sad, cheerful, reassuring, etc.); b) replicate that type of emotion through complex purely mechanical dynamics. Thus far we find *passivity* (a) and *activity* (b). But there is no static stage, as the algorithm works immediately in the next level, that is that system of feedback on itself that allows the machine to understand that it has 'hit' or not the object (the target of 'correct emotion') with a suitably small error[32].

Precisely in this last stage a form of self-reflexivity seems to emerge as a *recognition* of having acted correctly. In turn, reinforcement in a very short time implements other 'affective acknowledgments', other simulated and tested emotions, other 'emotional behaviours'. But there is always – integrated – an inverse feature, as a sort of *reversibility* in the relationship: the machine must return a certain level of emotion to the human, producing the 'correct' emotions. Incidentally, this empathogenic quality raises, starting from Mori's studies in the 1970s, the ethical problem of excess machine-human proximity known under the name of *Uncanny valley*: androids increasingly similar to humans, in appearance, in behaviour, in emotional expressions[33].

Again: do all this mean that algorithms *learn to feel emotions*? That, that is, by repeating and complexifying the stimulus/response links, they gain a sort of *emotional autonomy* that even allows them to *produce emotion*, in the sense of an *empathic autopoiesis*? If learning deals with a form of repetition, can this feature be applied to the embodiment domain, that is, a 'progressive corporealization' of emotion in the machine? Let's make it clear: this is not the case today. The answer seems negative, or limited to the field of science fiction or cinema, from *Blade Runner* or *Matrix*, to literature, the machine that 'comes alive', which becomes autonomous from the automatic, in terms of feelings, as well as emotions. As in *Klara and the sun*, by Ishiguro[34], a totally 'subjective' novel: the robot Klara tells his story while it learns to feel emotions, to deceive, to suffer, to experience a true friendship with a human and, even, to desire.

3.2 Emotional chip

Shouldn't we ask ourselves why the answer seems negative to us? As to why, as already written in the introduction, if the 'thought' for the machine today seems an undecidable, 'having a body' seems the impossible (*res cogitans* versus an 'alive' *res extensa*)?

Let's start again. How can a machine feel emotions? Pausing a little longer in the field of Fiction, we cite the example of the *emotional chip* taken from the famed Star Trek series. There we find on the one hand Spock (human mother

[32] L. Pitt, L. Valiant, *Computational limitations on learning from examples*, «Journal of the ACM», 35, 1988, pp. 965-984.
[33] See M. Mori, *Bukimi no tami (The uncanny valley)*, «Energy», 7, 4, 1970, pp. 33-35.
[34] K. Ishiguro, *Klara and the sun*, Random House Large Print, London 2021.

and Vulcan father), who is cold and particularly clever, and on the other, Captain Kirk, who is often emotionally involved in choices and decisions: the crystalline reason of the former is dialectically opposed to the 'emotional reason' of the second. Now, the aforementioned dialectic between pure intelligence and bodily reason finds a sort of update in the following series, in the figure of Deputy Commander Data. Data is, in fact, a totally artificial android. The problem is then that of endowing it with an emotional component. To equip him with emotions, a chip is inserted into the circuits and therefore the algorithms that govern his behaviour.

In short, the idea is that emotion is, in turn, something programmable, 'algorthimizable' by means of a physical graft. Something that, however, seems to naively re-propose the mind-body split[35]. The example of the emotional chip, while being part of Science Fiction, is able to represent the basic idea underlying a whole series of empathogenic devices potentially developable in the future, that is an algorithm capable to read emotions but above all to give emotions: receptive passivity and productive activity, where the learned skill consists in incorporating a frame of emotions on an increasingly 'machinic' level.

But the fundamental problem of the emotional chip, as well as of all devices related to it, is that of its 'coming from outside', perfectly in line with the conceptual paradigm of an external emotion that would be ontologically and genealogically completely different from an internal emotion. It is a graft, where the artificial remains artificial, therefore manageable, programmable and controllable by man. In its being an external object, the chip confirms – as in a game of mirrors – that the machine remains a machine: only an 'other' element can animate the machine and give life: this quasi-theological aspect recalls other well-known scenarios, as in the dramatic dialogue in the *Blade Runner* film between the replicant Roy Batty and his human programmer, Tyrrel. The problem is that of *giving life* on the part of the creator: giving life to 'something' that obviously has no life and that is destined to remain forever only 'mechanical' but who, after realizing that his operational capacity is over, that is, that he has been programmed from the outset with a countdown that leads him to death, asks dramatically: «I want more life, father!»[36]. Is it a question of the human or of the replicant?

3.3 The hypothesis of artificial emotional autonomy

We then begin to ask ourselves whether a less 'dualistic' vision is possible or not in the theoretical field. Again: can the machine gain some kind of *emotional artificial autonomy* if it is, by definition, lacking of a body? We must be very careful here if we want to avoid falling back into the *doxa*, albeit millenary,

[35] However, note that there is a discussion on these aspects, as in R. Sekuper, R. Blake, *Star Trek on the Brain. Alien Minds, Human Minds*, Freeman, New York 1998 and D. A. Norman, *Emotional Design*, cit., p. 92 e ss.
[36] From the film: *Blade Runner*, by R. Scott, USA 1982.

which hypostatizes the existence of an ontological difference between machine and nature. The problem is now of a historical-material nature: it is 'between us', in its *uncanney* novelty and urgency.

Reversibly, even on the statute of the *Leib* some questions become possible today, in the line already mentioned in the analysis of the phenomena of *learning with repetition*: between mechanical learning and 'alive' learning, what exactly are the differences, if they really exist? In this sense, come empirical-experimental findings concerning the discussion of the dualistic algorithm-body separation deserve to be mentioned.

Part of them derives from an epistemological instance, that is, the need for a complexification of the cognitive aspects of robots in the context of the experimental development of *embodied* approaches. In this line, certainly as its extreme branch, the idea of recognizing robotic agents as *having a body*[37] has also made its way. This vision is obtained, precisely, through cross-disciplinary approaches that range from the perspective of *synthetic biology* to the concepts of *Organismically-inspired robotics*[38]. From this perspective, robots are seen as «complete agents»[39] as they are «endowed with a body» and, as such, considered capable of autonomously modelling a set of dynamic interactions between the body dimension and the environmental or systemic dimension[40]. This behavioural self-sufficiency, for now limited to specific fields, nevertheless represents a form of *primitive autonomy*. The decisive point of this research is the increasingly frequent adoption of 'systemic' and 'holistic' paradigms, an adoption that facilitates the epistemology of artificial empathy. The framework is that of very refined experiments, that of small differential iterations and minimal procedural integrations which, overall, implement and improve the 'sentient' algorithm. In the holistic-relational paradigm, *emotion in the machine* is therefore conceived as an emergency but always and only within a systemic dynamic, often resorting to the old ideas of Varela and Maturana or Bateson.

[37] On the question of the body in a robot, there are several studies, starting from the works of Ziemke, like T. Ziemke, *What's that thing called embodiement?*, in *Processing of the 25th Annual Meeting of the Cognitive Science Society*, Lawerence Erlbaum, pp. 1305-1310, 2003; Id., *The body of knowledge*, in L. Damiano, Y. Kuruma, P. Stano (eds.), *What can shyntetic biology offer to artificial intelligence (and viceversa)*?, «Byosystems», 148, 2016, pp. 4-11; Id. e R. Lowe, *On the role of emotion in embodied cognitive architectures: from organisms to robots*, in «Cognitive Computation», 1, 1, 2009, pp. 104-107.

[38] See E. Di Paolo, *Organismically-insipred robotics. Homeostatic adaptation and teleology beyond the closed sensorimotor loop*, in K. Murase, T. Akasura (eds.), *Dynamic Systems Approach for Embodiment and Sociality*, «Advanced Knowledge International», Adelaide 2003, pp. 19-42; T. Froese, *On the role of AI in the ongoing paradigm shift within the cognitive sciences*, in M. Lungarella et al (eds.), *50 Years of Artificial Intelligence, Lecture Notes in Computer Science*, vol. 4850, Springer, Berlin 2007, pp. 63-75; L. Damiano, P. Stano, *Synthetic biology and artificial intelligence. Grounding a cross-disciplinary approach to the synthetic exploration of (embodied) cognition*, «Complex Systems», 27, 3, 2018, pp. 199-228.

[39] P. Dumouchel, *Vivre avec les robots*, cit., p. 118.

[40R.] Pfeifer, C. Scheier, *Understanding Intelligence*, MIT Press, Cambridge MA 1999; R. Pfeifer, J. C. Bongard, *How the body shapes the way of we think*, MITPress, Cambdridge 2006.

In other words: the continuous self-corrections of the algorithm's course, the integration of the infinitesimal automatic and quasi-organic adjustments, make sense only if they are designed in constant coordination with the environment. There is therefore a need for an *interactionist approach*[41] through which to build artificial systems capable of *synthesizing* the two dimensions of emotions (internal and external). From this point of view, indeed, strictly speaking, there are no two types of emotions; what pre-exists is an *affective circuit*[42]. In other words, internal and external emotions are both necessary and derived from the construction of affective circuits, with the consequence that the distinction between natural systems for regulating emotions and algorithmic systems is becoming less and less definite.

Therefore, the artificial agent can 'learn to feel' and 'to be empathic' only if it is thought and designed within an intrasystemic dimension, autopoietic in the sense of the notion of *enaction*, that is, that form of activity/passivity as a figure of the agent within complex systems. The epistemological models of the 'radically incorporated' artificial mind, conceiving it as an *emerging entity*, therefore tend to overcome the dualism between an immaterial substance and an extended substance. Within this emergent paradigm, the artificial quality of emotion is determined as an emerging form from a «process of co-specification that dynamically connects the agent's nervous system, his body, environment and other agents»[43]. This dynamic is therefore the real protagonist of the automatic learning scenarios: it escapes the spatial logic of the separation between internal and external, laying the logical and epistemological basis for also overcoming the threshold between internal and external emotions, where the logical structure of *relationship* (between artificial and natural agents) is that of reciprocal intertwining or *embrication*, in the terms of Varela and Thompson[44].

Let us come back for a moment, however, to the theoretical value of the autonomy of an artificial agent, also recalling other methodological approaches. Rather widespread, it is the one that put into parallel the empathic learning by artificial agents with computational learning of cognitive procedures in the strict sense. In these parallels, another key issue often emerges, namely that of autonomy in its relationship with automatism. What, for example, would the learning algorithms characterizing AI lack today? Not the ability to calculate, nor speed or precision, but the ability to *give oneself a purpose* for one's own operations. In short, the question of the autotelic agent would be the heart of *artificial autonomy*, that is the real boundary to be reached and overcome in order to really talk about *algorithms that learn*. Once again, as was clear since Wiener's cybernetics, a parallel is imposed between the biological organism that learns and the program of the machine. Why is an organism not a machine,

[41] See P. Dumouchel, *Vivre avec les robots*, cit., p. 129.
[42] Ibid.
[43] Ivi, p. 141.
[44] E. Thompson, F. J. Varela, *Radical embodiement. Neural dynamics and conscoiusness*, «Trends in Cognitive Sciences», 5, 10, 2001, pp. 418-425.

Georges Canguilhem also wondered, in the era of the affirmation of DNA and cybernetics? For the reason that «the body has [...] a wider range of action than the machine. Compared to the machine, it has fewer purposes and more potential»[45]. The organism certainly acts by virtue of a program, of a 'biological and genetic algorithm'; but at the same time, it can also modify the program, it can question its own goal, deny it, translate it, postpone it. As Fabbris argues, we can then speak of authentic AI if and only if it is endowed with artificial autonomy (AA), meaning the autonomy of an algorithmic system capable of autopoiesis of its own ends: «a system is not only autonomous when it is able to reach, without any external intervention, a certain purpose, but also when it is able to give itself a purpose. Only this last capacity allows to distinguish an autonomous system from a purely automatic system»[46]. Fabbris therefore is, implicitly, along a line we have repeatedly evoked of a radical change in the point of view: «AI never existed, there have been only good imitations of descriptions of behaviours considered by us to be an expression of intelligence. AI will not exist until an AA is reached»[47]. Something similar could then be argued about the question of the learning of emotions, including the question of the body and the automatic system as an artificial organism, underlining the emergent, dynamic or generative aspect of this type of algorithmic artifice, if it is true that strictly speaking, in the field of *Guided Self-Organization* as well as of *Autonomy Oriented Computing*, «you cannot program an AA system […] but you can generate it by simulating the process through which it is identified»[48].

What derives from these examples drawn indifferently from the experimental (in vitro) and real (*in vivo*) domain? A confirmation that internal emotions and external emotions are no longer defined by excluding each other, but *one in the other*. The emotional aspect of artificial learning would not be understandable without a reference to this complex, organic-systemic framework, in which emotion shows itself as primarily a relational reality, and not just autochthonous and individual.

Artificial emotion is not a state, but an emergency, itself a continuous correction and adjustment of *everted* processes, open to relationship, interconnected to the environment with which the machinic agent interfaces (both the natural and the artificial environment). Artificial emotion is a function internal to a field of bodily signification: a field of forces, that is, of actions and feedbacks, of new meanings of the activity/passivity couple of an artificial agent that dialogues with the natural and gradually imitates it better and better, it

[45] G. Canguilhem, *Machine and Organism* [1952], in «Incorporations» (J. Crary, S. Kwinter eds.), Zone Books, New York 1992, pp. 45-69.
[46] L. Fabbris, *Il programma, la rete, il circolo*, cit., p. 64.
[47] Ivi, p. 76.
[48] Ivi, p. 65. On the complex definition of autonomy, cf. also W. F. Lawless et al (eds.), *Autonomy and Artificial Intelligence. A Threat or a Savior?*, Springer, Cham 2017; J. Liu et al (eds.), *Autonomy Oriented Computing. From Problem Solving to Complex Systems Modelling*, Kluwer Academic Publisher, New York 2005; M. Prokopenko (ed.), *Guided Self-Organization. Inception*, Springer, Heidelberg 2014.

integrates it (under the guise of schemes), pursuing a sort of *autonomy in the relationship*.

4. Traces of algorithmic sociality

4.1 Social robotics

To complete and better verify this vision, let's take a look at the 'social' side of artificial emotion, exploring the field of social robotics, a research field in which converge Human-Robot Interaction (HRI), Affective Robotics, Cognitive Robotics, Affective Developmental Robotics[49]. The original question of social robotics concerns how it is possible to insert, in our relationships and social lives, devices guided by algorithms, such as robots or other automatisms, in a plurality of contexts (assistance, surveillance, information, education, etc.). For our interests, the question is how to create social robots that have a relationship with humans that is 'credible' or 'likely', in terms of bodily and emotional reactions, guaranteeing the so-called 'social presence' of robots. The algorithm should be therefore 'human oriented', that is, capable of causing, in the perception of the bodies with which it interacts, sensations and 'affects' comparable with, if not indistinguishable from, those of human relations, in the sense of what could be called an *affective performance*[50].

In our approach, the reference to sociality is significant above all for its contribution to the *evolution* of artificial empathy: the progress in the ability of robots to decode and reformulate learned emotions, in fact, helps to produce (both theoretically and practically) artificial models of emotional processes that are increasingly close to the 'true'. But social robotics proposes again the ethical and philosophical problem of the emotional automaton in a different way. It is now clear that the programming of the machines themselves is conceived in order to implement self-learning and self-adaptation skills: we now speak of «creation of emotional regulation systems»[51] for robotic agents, therefore of a form of autonomy in establishing hierarchies, in making 'emotionally oriented' decisions, in making the algorithm more adherent to the synesthetic complexities of the environment in which it operates. At this point, the self-

[49] See K. Dautenhahn, *Methodology and themes of human-robot interaction. A growing research field*, «International Journal of Advanced Robotics», 2007, 4, 1, pp. 103-108; C. Breazeal, *Emotions and sociable humanoid robots*, «International Journal of Human-Computer Studies», 59, 1, 2003, pp. 119-155 and Id., *Toward sociable robots*, «Robotics and Autonomous Systems», 42, 3-4, 2003, pp. 167-175.
[50] A. Paiva (ed.), *Affective Interactions: Toward a new generation of computer interfaces*, Springer, Berlin 2000; T. Fong, I. Nourbakhsh, K. Dautenhahn, *A survey of socially interactive robots*, «Robotics and Autonomous Systems», 42, 3-4, 2003, pp. 146-166; H. Li *et al*, *Towards an effective design of social robots*, «Int. J. of Social Robotics», 2011, 3, 4, pp. 333-335. On the philosophical problems that such realities raise, for an overview see I. Pelgreffi, *Ambiente digitale, automatismi e corporeità*, in (F. Miano, L. Alici eds.) *L'etica nel futuro*. Orthotes, Napoli-Salerno 2020, pp. 389-399, and Id., *Bernard Stiegler e la critica della società automatica*, in B. Stiegler, *La società* automatica, Meltemi, Milano 2019, pp. 11-26.
[51] P. Dumouchel, *Vivre avec les robots*, cit., p. 117.

sufficient, autonomous and only *internal* character of a human body of emotion begins to lose its supremacy in the logical-epistemological model.

As a last consequence, we see then that social robots operate today through increasingly differential and differentiated strategies with the environment and with other agents, not only human, but potentially also robotic. On a level of radical empiricism, we can even glimpse the possibility of a robot-robot learning relationship, that is a relationship in the machinic field, considering that communication between machines takes place more and more often, directly or indirectly, via the internet. The relationship, that is, is more open and more indeterminate than those programmed with the machine *in vitro*, allowing for artificial systems that are much more similar to an *organism that learns*, rather than to a *simple automatic repetition machine*. On these future scenarios of unprecedented hybridizations, Guattari wrote:

> Leroi-Gourhan emphasised that the technical object was nothing outside of the technical ensemble to which it belonged. It is the same for sophisticated machines such as robots, which will soon be engendered by other robots. Human action remains adjacent to their gestation, waiting for the breakdown which will require its intervention: this residue of a direct act. […]. Curiously, in acquiring more and more life, machines demand in return more and more abstract human vitality: and this has occurred throughout their evolutionary development[52].

Guattari's point, also representative of many others, must be understood along a line of development, certainly in the 'irregular' manner that characterizes the author, of the French epistemological *coté* that ranges from Simondon, Lyotard, Derrida to Deleuze. It testifies to an urgency: it suggests not a trivial elimination of the human in the presence of the machine, but an *accomplishment of the human* that passes through a modification of the entire reasoning on the *learning abilities* of algorithms and robots, that is, on the 'sense' of this capacity in relation to the human being, as well as on their ways of self-reshaping their internet connections, their transmission and sharing protocols.

Algorithms are experimenting today, every second, with other ways of communicating: they 'invent' new languages, new codes that will not necessarily be shared with their 'creators'. From this point of view, the question of Machine Learning always refers, even when it is not explicitly mentioned, to the theme of evolution, to a sort of 'evolutionary process' of which Guattari speaks, which involves machine and human within a inextricable and over-determined association, a relationship that *becomes new* and non-programmable.

[52] F. Guattari, *Chaosmosis. An ethico-aesthetic paradigm* [1992], Indiana University Press, Indiana 1995, p. 36.

4.2 Affection as a mechanism? Heterogeneity and relationship

Among the theoretical sides of social robotics, we find the crucial possibility of conceiving the *affect* as a mechanical phenomenon. Does this mean that artificial sensory stimulation can lead to a proto-emotion? The most correct answer, for now, is: we cannot know. Certainly, for this to happen, a sort of 'social dimension' of the machine must be integrated into the dimension of the 'stand alone' sensor.

This means that the whole emotional field can be understood as a «continuous mechanism of inter-individual coordination»[53]. The cooperating side between different fields (between artificial and natural; between robotic body and human body) interprets emotion as an *empathic operational ability*. The emotion arises in the coordination-with, that is in the relationship-between, and it is no longer just the result of a movement of 'introversion' or of 'stand alone' computational re-elaboration, or of a doubling by the artificial agent. Giving praise to this particular sociability of the machine also means, in perspective, accepting that there are shares of autonomy in these relations between machines; it means also accept a concept of *machinic assemblage*, so to speak, which, as anticipated, brings the problem back to the great discussion on the emergence of forms, on saliences, on pregnancies, along a line that from the studies of René Thom and Jean Petitot leads to recent perspectives, such as those, which deepen and revive those of Deleuze and Guattari, by Alessandro Sarti and other philosophers and scientists[54].

Under the name of the sociability of algorithms, we are witnessing today a free game of recombinations of forms, the outcome of which is not predictable or programmable, being rather more similar to a *heterogeneous becoming of forms*: a continuous metamorphosis of the machinic form (which at the same time recalls the natural form). Some of the most advanced approaches in empathic social robotics postulate, in fact, a sort of *cognitive heterogeneity*[55], according to which there is a plurality of human minds, animal minds, but also robotic minds, which 'coexist' together. And this therefore characterizes the field of relationships in which the first trace of artificial emotion will eventually take shape: a field that is actually an intersection of heterogeneous fields but able to coordinate and communicate continuously. In this perspective, the difference between 'minds' is not ontological qualitative, but lies only in their being objects (or better cognitive systems) emerging from different environments, with different dynamics, with historically different relationships and interactions.

[53] P. Dumouchel, *Vivre avec les robots*, cit., p. 21.
[54] Line clearly visible since form F. Berardi, A. Sarti, *Run. Forma, vita, ricombinazione*, Mimesis, Milano-Udine 2008, and recently deepened and modified in A. Sarti, G. Citti. D. Piotrowski, *Differential Heterogenesis. Mutant forms, sensitive bodies*, Springer, Berlin 2022, research based on a dynamic morphology of the living in a post-structuralist approach. See also the old but fundamental study U. Fadini, *Principio metamorfosi. Verso un'antropologia dell'artificiale*, Milano 1999.
[55] P. Dumouchel, *Vivre avec les robots*, cit., p. 24.

In conclusion: the research on social robotics opens glimmers towards a different thought of the *affective-machinic dimension*. Among these, perhaps the most relevant is that of a reduction in the dichotomy between authentic human emotion and simulated robotic emotion[56]. And this reduction derives, let us note, from considerations that are not so much theoretical as empirical: social robotics explores the *interaction* between inside and outside, progressively deconstructing – in the practical level – the dichotomy between natural and artificial. Take, for example, some old research by Paul Dumouchel[57], over time and also recently resumed and articulated with Laura Damiano[58], which effectively describe the picture we are sketching:

> it is a relational approach to the theoretical characterization of emotions that it identifies in the affective exchange based on emotional expression a mechanism of inter-individual coordination and mutual determination of the inclinations of action. The central hypothesis is that what we commonly call 'emotions' are salient moments, or points of balance, in a continuous process of coordination. From this point of view, emotions must be thought of as common works, resulting from interactions, rather than as internal and private events[59].

Starting from that, the whole question of the passive-mimetic aspect finds its completion in these theories towards a more active dimension: activity and passivity, which are naturally intertwined in the living organism, also begin in the robotic field to train each other in increasingly complex modes, which signify forms of 'relational autonomy'.

The thesis of *affective coordination*, if taken to the end, completely redefines the historical, logical and theoretical terms of the problem of empathic robotics, leaving us to imagine that there may be a zone of *independence* in the emergence of emotions, on a socio-relational basis, in the robot, and therefore of *autonomy* in the algorithms that govern them. An area of which we still do not know much on the theoretical level, but with which we are resonating on the level of historical and practical knowledge.

[56] See R. Sparrow, L. Sparrow, *In the hand of machines? The future of aged care*, «Minds and Machines», 16 , 2, 2006, pp. 141-161; A. Sharkey, N. Sharkey, *The crying shame of robot nannies. An ethical appraisal*, «Interaction Studies», 11, 2, 2010, pp. 161-190 and Ids., *Granny and the robots. Ethical issues in robot care for elderly*, «Ethics and Information Technology», 14, 2, 2012, pp. 27-40; S. Tisseron, *Le jour où mon robot m'aimera. Vers l'empathie artificielle*, Albin Michel, Paris 2015.

[57] See P. Dumouchel, *Emotions. Essai sur le corps et sur le social*, «R. Ph. de la France et de l'Etranger», 187, (1), 1997.

[58] See L. Damiano, P. Dumouchel, *Artificial empaty, imitation and mimesis*, «Ars vivendi», 1, 1, 2011, pp. 18-31; L. Damiano, P. Dumouchel, H. Lehmann, *Towards human-robot affective coevolution. Overcoming oppositions in constructing emotions and empathy*, «International Journal of Social Robotics», 7, 1, 2015, pp. 7-18

[59] P. Dumouchel, *Vivre avec les robots*, cit., p. 27.

5. Beyond the dichotomy: an ethics of hybridization or a hybridization of ethics?

Going to the conclusions. On the one hand, we have seen that if it is true that the issue of learning, in the field of AI, is connected to a computational dimension, it is also true that something different happens when we turn to Learning applied to the empathic field. In fact, here the topic of the body is somehow central, in the sense of a (problematic) corporeity of the machine. But this *latu sensu* corporeal dimension acquires meaning only, as seen, in reference to the social and systemic quality of the algorithm: emotions can be learned by the algorithm if and only if the concept and experience of emotion artificial is understood on a social and holistic basis.

The hypothesis therefore, which concerns a possibility of reading the near future, is the following: only starting from this relational field we might speak consistently of *machinic learning of affects*, of an *affectivity in the machine* (both even at an embryonic stage of evolution) and other formulas we have encountered.

On the other hand, we had instead emphasized the fact that, reversibly, even in the human behaviour of learning there is an automatic, repetitive, machinic component in a broad sense. And we see also how the automaton 'in ourselves', is not only to be relegated to secondary, but rather to be recognized, accepted and developed, as a completion of our being in its most comprehensive sense.

From the above, it follows the need to explore the hypothesis of a hybridization dimension: in both sides of the learning process – in the technical-artificial domain *and* in the human-natural domain – we find isomorphism and reversibility, at least on the level of epistemological models. They are two 'opposite' sides of a process that is one, in itself.

However, the great problem in the ethical field remains the following: how to make these concerns on artificial empathic learning adhere to the usual ethical frames, that is, those currently in the field in the main international guidelines on machine learning?

A possibility is to hypothesize a different ethical attitude, which starts from a slow, patient and shared attempt to deconstruct the dichotomy between natural and artificial, for example between internal emotion and external or 'machinic' emotion. As Damiano and Dumouchel suggest in their essay on artificial empathy, the assumption of a less 'dualistic' perspective around the theme of emotions, «transforms the configuration of ethical issues raised by the relationships between substitutes and human beings»[60]. There is, moreover, a historical datum, namely the emergence with increasing experimental evidence that «when artificial emotions and empathy are considered part of a coordination mechanism, they become something very different from deceptive surrogates.

[60] Ivi, p. 28.

They become the occasion for a possible *co-evolution* of new social partners, capable of being integrated into the fabric of our relations»[61].

It is well known that since the first international conferences, such as that of Sanremo in 2004 (and the related *Euron Roboethic Roadmap*[62]), two lines have emerged around the problems of learning and autonomy[63]. One is concerned with managing automation from the point of view of man: the ethics of machines is an ethics that clarifies the responsibilities of the man or the 'creator'. The other line, minority and more radical, is instead willing to give space to the autonomy of the artificial agent: the machine can become a 'subject', a subject of law but also an ethical subject, with its responsibilities and rights[64]. On the one hand, the ethics of machines therefore remains the ethics of the engineers, programmer or designer, separating the human from his 'object'; an object, if anything, to be ethicized (and where therefore the instance of management and control, the instance also of responsibility, remains firmly in the hands of the human). On the other hand, one can imagine an ethics where it is accepted – in a dystopian or utopian way – that the machine can acquire, by *learning from itself in view of itself*, a total autonomy: an *autotelia of artificial operation*. Even in this second case, however, we find, inside the logical structure of the problem, a separation between the human being and the machine: one wishes to ethicize the object; the other to de-humanize ethics.

The concerns collected in this essay, which are nothing more than a preliminary reasoning, cautiously suggest the option of a third model of ethics: a sort of ethics of hybridization, but also a sort of hybridization of the aforementioned ethical attitudes. There are already traces of it, if we think for a moment of the great tradition of philosophical anthropology as an anthropology of technology and of its basic assumption that sees man as a *naturally technical being*, or of the line of reflection that belongs to Donna Haraway, since from his 1985 *A cyborg manifesto*, where the challenge was to understand that non-humans exist not only in the field of animals or plants, as a form of biological life, but also in the sphere of the 'animated' inorganic, capable – within the

[61] Ibid.
[62] G. Veruggio, *Euron Roboethic Roadmap*, Release 1.2, january 2007 (www.roboethics.org).
[63] The specific bibliography on the 'dichotomy' is very extensive. See P. Lin, R. Jenkins, K. Abney, *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*, Oxford University Press, New York 2017; S. G. Tzsafestas, *Roboethics. A Navigating Overview,* Springer, Berlin 2016; Vallor, S. and Bekey G. A., 2017, *Artificial Intelligence and the Ethics of Self-Learning Robots*, cit.. G. Veruggio, F. Operto, G. Bekey, *Roboethics. Social and Ethical Implications*, in Siciliano, Khatib, *Springer Handbook of Robotics*, Springer, Berlin[2] 2017, pp. 2135-60; M. Anderson, S. Anderson (eds.), *Machine Ethics*, Cambridge University Press, Cambridge 2011. N. Bostrom, E. Yudkowsky, *The Ethics of Artificial Intelligence*, in *The Cambridge Handbook of Artificial Intelligence*, K. Frankish, W. M. Ramsey (eds.), Cambridge University Press, Cambridge 2014, pp. 316-334.
[64] A vision that, we should remeber, often encounters difficulties among scholars of ethical questions. See W. Wallach, C. Allen, *Moral Machines. Teaching Robots Right from Wrong*, Oxford University Press, Oxford 2009, or even the sceptical considerations of Luciano Floridi, for example those summarized in L. Floridi, *On algorithms: ethical and epistemological questions. Interview with Igor Pelgreffi*, «Lo Sguardo», n. 34, 3, 2022.

limits already mentioned – of a proto-sociality or a quasi-sociality, which is not necessarily a pseudo-sociality.

The hypothesis sketched in this article outlines the possible shapes of an *ethics of relationship*, where the relational aspect *precedes* the position of *related values*: an ethics closely linked, in its form, to the co-evolution of agents (natural or artificial) but at the same time an ethic that calls and calls to responsibility towards our *ability to read* the whole of all the emerging dynamics, of the current transformations, of the *uncunny* scenarios of contamination between biological and artificial, at various levels.

This seems to be the great challenge, like a void space of understanding that is now open in front of us, but which perhaps needs to be investigated, by looking for new languages, new vocabularies, new theoretical figures: it will be a slow and patient construction towards something still undetermined. At that point the algorithms could really be thought of as having a life of their own: a *life different* from the one we are used to thinking, for millennia, but still a *form of life*, a *Lebensform* of those who are 'differently alive', of systems equipped with their surprising organization, not entirely programmable, but not for this unmanageable by the human community, from forms not all pre-contained in the initial program. And where the property of 'not being pre-contained' has to do with improvisational skills that also become, in some way still not codified, emotional and social capacities.

In short, the elaboration of an ethics capable of conceiving not only our relationship with the machine but also – in perspective – of the algorithms towards us could have some use in the near future (mixing somehow a top-down approach and a bottom-up approach[65]). But then it will also be necessary to admit that, perhaps, in this context the term ethics is no longer adequate: something different will be needed, which starts from what we currently mean by ethics but is able to proceed further, completing and expanding our ethical vision. Ethics changes its very meaning precisely because it comes into contact with a disturbing or simply enigmatic reality, which it cannot dominate. Perhaps a *synthetic ethics*[66] (Dumouchel and Damiano) or an *incremental ethics*[67]

[65] See also B. Christian, *The Alignment Problem. Machine Leaning and Human Value*, W. W. Norton & Company, 2020; B. Goertzel, *Artificial General Intelligence: Concept, State of the Art, and Future Prospects*, «Journal of Artificial General Intelligence», 2014, 5, 1, pp. 1-46.

[66] P. Dumouchel, *Vivre avec les robots*, cit., p. 30.

[67] Similar hypotheses have also been advanced by other scholars, for example I. van de Poel, *An Ethical Framework for Evaluating Experimental Technology*, «Science and Engineering Ethics», 22, 2016, pp. 667-686: F. Amigoni, V. Schiaffonati, *Ethics for Robot as Experimental Technologies. Pairing Anticipation with Exploration to Evaluate the Social Impact of Robotics*, «IEEE Robotics and Automation Magazine», 1, 2018, pp. 30-36. In particular, Viola Schiaffonati recently proposed the idea of an incremental ethics, based on the existence of a close connection between the over-determined nature of some experimental results of digital technologies, in particular of the more 'extreme' and not *explainable* algorithms, and a correlative ethical return characterized by a 'more advanced' approach than traditional ones. Ethics itself becomes less 'static' and more exploratory, in turn, progressive and precisely 'incremental' with respect to the field of observation of the phenomena to be 'ethicized': «the epistemic uncertainty that

(Schiaffonati), in which the ethical aspects are continuously integrated with epistemological or technical philosophy aspects.

All this could involve a complexification of ethics itself, in which it opens its models to greater interaction with epistemology: an *epistemologically oriented* ethics? Ultimately, deconstructing the internal/external, natural/artificial dichotomy is certainly also a risk; but not pursuing it, at least in philosophical reflection, does not involve an even greater risk? And that is the risk of finding oneself, in *automatic society* (Stiegler) or in the *infosphere* (Floridi), totally immersed, involved, subjectivated but also without conceptual tools, inadequate tools as they are 'detached' from the materiality of an experience that could be, in a few years or decades, strongly characterized and crossed by the digital, by the algorithm, by the artificial autonomy, albeit in the primitive stages of a slow evolution?

Igor Pelgreffi
Università degli studi di Verona
✉ igor.pelgreffi@univr.it

References

Primary Bibliography

Anderson M., S. Anderson (eds.) 2014. *Machine Ethics*, Cambridge University Press, Cambridge.

Aristotle 2000, *Nicomachean Ethics*, tr. by R. Crisp, Cambridge University Press, Cambridge.

Bateson G. 1972. *Steps to an ecology of mind*, Ballantine Books, New York.

Berardi F., Sarti A. 2008. *Run. Forma, vita, ricombinazione*, Mimesis, Milano-Udine.

Bostrom N. 2014. *Superintelligence. Paths, Dangers, Strategies*, Oxford University Press, Oxford.

Canguilhem G. 1992. *Machine and Organism* [1952], in «Incorporations» (J. Crary, S. Kwinter eds.), Zone Books, New York, pp. 45-69.

Christian B. 2020. *The Alignment Problem. Machine Leaning and Human Value*, W. W. Norton & Company.

D. A. Norman 2005. *Emotional Design*, Basic Books, New York.

Damasio A. 2004, *Looking for Spinoza. Joy, Sorrow and the Feeling Brain*, HMH, Montreal.

lies at the heart of exploratory experimentation can be translated from an ethical point of view in a form of incrementalism, where exploratory experiments play a role precisely in acquiring knowledge on how these experimental technologies operate in their contexts of use» (cf. V. Schiaffonati, *Computer, robot ed esperimenti*, «aut aut», 392, 2021, pp. 51-62, p. 59).

Derrida J. 1994. *Nietzsche and the machine*, «Journal of Nietzsche Studies», 7, pp. 7-66.
Dumouchel P., Damiano L. 2016. *Vivre avec les robots. Essai sur l'empatie artificielle*, Seuil, Paris.
Fadini U., 1999, *Principio metamorfosi. Verso un'antropologia dell'artificiale*, Milano, Mimesis.
Guattari F. 1995. *Chaosmosis. An ethico-aesthetic paradigm* [1992], Indiana University Press, Indiana.
High-Level Expert Group on AI 2019. *Ethics Guidelines for Trustworthy AI*, EU Guidelines, Bruxelles.
Houdé O. 2017. *Apprendre à résister*, Le Pommier, Paris.
Lin P., Jenkins R., Abney K. 2017. *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*, Oxford University Press, New York.
Liu J. et al (eds.) 2005. *Autonomy Oriented Computing. From Problem Solving to Complex Systems Modelling*, Kluwer Academic Publisher, New York.
Mauss M. 1973. *Techniques of the body* (1934), tr. eng. B. Brewster, «Economy and Society», 2, pp. 70-88.
Merleau-Ponty M. 1936, *The Structure of Behaviour*, Beacon, Boston.
OECD 2021. *OECD Council Recommendation on Artificial Intelligence* (2021), OECD, Paris.
Paiva A. 2000. (ed.). *Affective Interactions: Toward a new generation of computer interfaces*, Springer, Berlin.
Picard R. W. 1995. *Affective Computing*, M.I.T. Media Laboratory Perceptual Computing Section, Technical Report n. 321, Boston
Prokopenko M. (ed.) 2014. *Guided Self-Organization. Inception*, Springer, Heidelberg.
Sarti A., Citti G., Piotrowski D. 2022, *Differential Heterogenesis. Mutant forms, sensitive bodies*, Springer, Berlin.
Simondon G. 2014. *Sur la technique*, PUF, Paris.
Sloterdijk P. 2009. *Du mußt dein Leben* ändern. Über Anthropotechnik, Frankfurt 2009.
Tamburrini G. 2000. *Etica delle macchine*, Carocci, Roma.
Thompson E., Varela F. J. 2001. *Radical embodiement. Neural dynamics and conscoiusness*, «Trends in Cognitive Sciences», 5, 10, pp. 418-425.
Tzsafestas S. G. 2016. *Roboethics. A Navigating Overview,* Springer, Berlin.

Secondary Bibliography

Amigoni F., Schiaffonati V. 2018, *Ethics for Robot as Experimental Technologies. Pairing Anticipation with Exploration to Evaluate the Social Impact of Robotics*, «IEEE Robotics and Automation Magazine», 1, pp. 30-36.
Bernacer J., Lombo J. A., Murillo J. I.(eds.) 2015. *Habits: plasticity, learning and freedom*, «Frontiers in Human Neuroscience», Lausanne.

Bostrom N. 2002. *Existential Risks. Analysing Human Extinction Scenarios*, «Journal of Evolution and Tecnhology«, 9.

Bostrom N., Yudkowsky E. 2014. *The Ethics of Artificial Intelligence*, in *The Cambridge Handbook of Artificial Intelligence*, K. Frankish, W. M. Ramsey (eds.), Cambridge University Press, Cambridge, pp. 316-334.

Breazeal C. 2003. *Emotions and sociable humanoid robots*, «International Journal of Human-Computer Studies», 59, 1, pp. 119-155

Breazeal C. 2003. *Toward sociable robots*, «Robotics and Autonomous Systems», 42, 3-4, pp. 167-175.

Camras L. A. 1992. *Expressive development and basic emotions*, «Cognition and Emotion», vol. 6, n. 3 and 4.

Damiano L., Stano P. 2018. *Synthetic biology and artificial intelligence. Grounding a cross-disciplinary approach to the synthetic exploration of (embodied) cognition*, «Complex Systems», 27, 3, pp. 199-228.

Damiano L., Dumouchel P. 2011. *Artificial empaty, imitation and mimesis*, «Ars vivendi», 1, 1, pp. 18-31.

Damiano L., Dumouchel P., Lehmann H. 2015. *Towards human-robot affective coevolution. Overcoming oppositions in constructing emotions and empathy*, «International Journal of Social Robotics», 7, 1, pp. 7-18.

Dautenhahn K. 2007, *Methodology and themes of human-robot interaction. A growing research field*, «International Journal of Advanced Robotics», 4, 1, pp. 103-108.

Di Paolo E. 2003. *Organismically-insipred robotics. Homeostatic adaptation and teleology beyond the closed sensorimotor loop*, in Murase K., Akasura T. (eds.), *Dynamic Systems Approach for Embodiment and Sociality*, «Advanced Knowledge International», Adelaide, pp. 19-42

Doshi-Velez F., Kim B. 2017. *Towards a rigorous Science of Interpretable Machine Learning*, on line arXiv:1702.08608v2.

Dumouchel P. 1997. *Emotions. Essai sur le corps et sur le social*, «R. Ph. de la France et de l'Etranger», 187, (1).

Esposito E. 2021. *Dall'intelligenza artificiale alla comunicazione artificiale*, «aut aut», 392, pp. 20-34.

EU 2020. *White Paper on Artificial Intelligence. A European approach to excellence and trust*, Bruxelles.

Fabbris, L. 2021. *Il programma, la rete, il circolo: un approccio Io-fi all'Autonomia artificiale*, «aut aut», 392, pp. 63-76.

Floridi L. 2022, *On algorithms: ethical and epistemological questions. Interview with Igor Pelgreffi*, «Lo Sguardo», n. 34, 3.

Fong T., Nourbakhsh I., Dautenhann K. 2003. *A survey of socially interactive robots*, «Robotics and Autonomous Systems», 42, 3-4, pp. 146-166.

Fossa F., Schiaffonati V., Tamburrini G. 2021. *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Carocci, Roma.

Froese T. 2007. *On the role of AI in the ongoing paradigm shift within the cognitive sciences*, in M. Lungarella et al (eds.), *50 Years of Artificial Intelligence, Lecture Notes in Computer Science*, vol. 4850, Springer, Berlin, pp. 63-75.
Goertzel B. 2014. *Artificial General Intelligence: Concept, State of the Art, and Future Prospects*, «Journal of Artificial General Intelligence», 5, 1, pp. 1-46.
Houdé O. 2017. *La neuroéducation: magie ou science? Cerveau & Psycho/Pour la science*, «Chronique. L'école des cerveaux», 86, pp. 80-83.
Ienca A., Vayena E. 2019. *The global Landscape of AI Ethics Guidelines*, «Nature Machine Intelligence», 1, 9, pp. 389-399.
Ishiguro K 2021. *Klara and the sun*, Random House Large Print, London.
Lawless W. F. et al (eds.) 2017. *Autonomy and Artificial Intelligence. A Threat or a Savior?*, Springer, Cham.
Li H. *et al*. 2011. *Towards an effective design of social robots*, «Int. J. of Social Robotics», 3, 4, pp. 333-335.
Miller T. 2019. *Explanation in Artificial Intelligence. Insights from the Social Sciences*, «Artificial Intelligence», 267, pp. 1-38.
Mori M. 1970. *Bukimi no tami (The uncanney valley)*, «Energy», 7, 4, pp. 33-35.
Murray I. R., Arnott J. L. 1993. *Toward the simulation of emotion in synthetic speech*, «J. Acoust. Soc. Am.», vol. 93, feb., pp. 1097-1108.
Norman D. A. 2004. *Emotional Machines* in Id., *Emotional design*, Basics Books, New York.
OECD 2020. *Principles on AI OECD*, Paris.
Parisi D. 2014. *Future robots. Towards a robotics science of human beings*, John Benjamins, London.
Parisi D. 2014. *Internal robotics*, «Connections Science», 16,4, pp. 325-338
Pelgreffi I. 2019. *Bernard Stiegler e la critica della società automatica*, in B. Stiegler, *La società* automatica, Meltemi, Milano, pp. 11-26.
Pelgreffi I. 2020. *Ambiente digitale, automatismi e corporeità*, in (F. Miano, L. Alici eds.) *L'etica nel futuro*. Orthotes, Napoli-Salerno, pp. 389-399.
Pfeifer R., Bongard J. C. 2006, *How the body shapes the way of we think*, MITPress, Cambdridge.
Pfeifer R., Scheier C. 1999. *Understanding Intelligence*, MIT Press, Cambridge MA.
Pitt L., Valiant L. 1988. *Computational limitations on learning from examples*, «Journal of the ACM», 35, pp. 965-984.
Ramus F. 2018. *Neuroéducation et neuropsychanalyse: du neuroenchantement aux neurofoutaises*, «Intellectica», 1-2, 69, pp. 289-301.
Ryan M., Stahl B. 2021. *Artificial Intelligence Guidelines for Developers and Users*, «Journal of Information, Communication and Ethics in Society», 19, 1, pp. 61-86.
Schiaffonati V. 2021. *Computer, robot ed esperimenti*, «aut aut», 392, pp. 51-62.
Schmetkamp S. 2020, *Understanding AI. Can and Should we Empathize with Robots?*, «Review of Philosophy and Psychology», 11, pp. 881-887.
Scott R. 1982, *Blade Runner*, USA.

Sekuper R., Blake R. 1998. *Star Trek on the Brain. Alien Minds, Human Minds*, Freeman, New York.

Sharkey A., Sharkey N. 2010. *The crying shame of robot nannies. An ethical appraisal*, «Interaction Studies», 11, 2, pp. 161-190

Sharkey A., Sharkey N. 2012. *Granny and the robots. Ethical issues in robot care for elderly*, «Ethics and Information Technology», 14, 2, pp. 27-40.

Sparrow R., Sparrow L. 2006. *In the hand of machines? The future of aged care*, «Minds and Machines», 16 , 2, pp. 141-161.

Testa I., Caruana F. (eds.) 2020. *Habits. Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory*, Cambridge University Press, Cambridge.

Tisseron S. 2015. *Le jour où mon robot m'aimera. Vers l'empathie artificielle*, Albin Michel, Paris.

Vallor S., Bekey G. A. 2017. *Artificial Intelligence and the Ethics of Self-Learning Robots*, in Lin et al (eds), *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*, Oxford University Press, New York, pp. 338-353.

van de Poel I. 2016. *An Ethical Framework for Evaluating Experimental Technology*, «Science and Engineering Ethics», 22, pp. 667-686.

Veruggio G. 2007. *Euron Roboethic Roadmap*, Release 1.2, january.

Veruggio G., Operto F., Bekey G. 2017. *Roboethics. Social and Ethical Implications*, in Siciliano, Khatib, *Springer Handbook of Robotics*, Springer, Berlin[2], pp. 2135-60.

Wallach W., Allen C. 2009. *Moral Machines. Teaching Robots Right from Wrong*, Oxford University Press, Oxford

Weinberger D. 2017. *Machines Now Have Knowledge We'll Never Understand*, «Wired», on line 18.4.2017.

Whatcher S., Mittelstadt B., Floridi L. 2017. *Transaparent, Explainable and Accountable AI for Robotics*, «Science Robotics», 6.

Ziemke T. 2003, *What's that thing called embodiement?*, in *Processing of the 25th Annual Meeting of the Cognitive Science Society*, Lawerence Erlbaum, pp. 1305-1310.

Ziemke T. 2016. *The body of knowledge*, in L. Damiano, Y. Kuruma, P. Stano (eds.), *What can shyntetic biology offer to artificial intelligence (and viceversa)*?, «Byosystems», 148, pp. 4-11.

Ziemke T., Lowe R. 2009. *On the role of emotion in embodied cognitive architectures: from organisms to robots*, in «Cognitive Computation», 1, 1, pp. 104-107.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770055

Contributi/8

Algoritmi viventi tra Shannon e Simondon

Francesca Sunseri 0000-0002-1559-8557

Articolo sottoposto a doppia blind peer review. Inviato il 23/12/2021. Accettato il 08/06/2022.

Living Algorithms. Between Shannon and Simondon

Almost a century after Turing's machine was theorised, a formal definition of what an algorithm is has still not been found by scholars. The reasons given are varied and it is clear that this is a problem that is not easy to solve. Nevertheless, one of the least travelled paths is the one that envisages the hypothesis of an algorithm that does not exhaust its potential in automatism. Starting from the algorithm written and implemented by the mathematician Claude Shannon in the 1950s in a machine made up of a group of relays, a mouse-shaped sensor and a maze structured in squares, the following work aims to understand whether a certain creative and evolutionary capacity can be found in the algorithm, i.e. the possibility of going beyond the logic of automatism. To be able to trace these capacities in the algorithm means, in the first instance, to try to verify the extent to which it is able to modify its own strategy according to the stimuli it receives and, in the second instance, to understand whether this creative and evolutionary ability responds to reflex or instinctual logic according to the distinction made by the French philosopher Gilbert Simondon in his courses on psychology in 1964.

***

1. Il labirinto del topo Theseus

Nel marzo del 1951 presso la Josiah Macy Jr. Foundation a New York si tenne l'ottava conferenza sulla cibernetica alla quale erano stati invitati il sociopsicologo Alex Bavelas, il letterato Ivor A. Richards, lo psichiatra Lawrence S. Kubie, il fisico e neuroscienziato Donald M. MacKay, il medico Herbert G. Birch e l'ingegnere e matematico Claude Shannon. Era la prima Macy Conference a cui non partecipava il padre della cibernetica Norbert Wiener a causa di divergenti vedute rispetto al resto del *core group*[1] che presiedeva agli incontri.

[1] Del *core-group* facevano parte G. Bateson, J. H. Bigelow, G. von Bonin, L. K. Frank, F. Fremont-Smith, R. W. Gerard, M. Harrower, G. E. Hutchinson, H. Klüver, L. S. Kubie, P. Lazarsfeld, K. Lewin, R. Lorente de Nó, W. S. McCulloch, M. Mead, J. von Neumann, F. S. C. Northrop, W. Pitts, A. Rosenblueth, L. J. Savage, N. Wiener.

Claude Shannon aveva già partecipato alla Macy Conference l'anno precedente affrontando la questione della ridondanza della lingua inglese, parte significativa della sua teoria dell'informazione presentata due anni prima, nel 1948, con il saggio *A Mathematical Theory of Communication*[2]. In questa seconda occasione, Shannon illustrò il progetto, portato avanti nei Bell Laboratories, di una macchina che poteva risolvere i labirinti[3]:

> This is a maze-solving machine that is capable of solving a maze by trial-and-error means, of remembering the solution, and also of forgetting it in case the situation changes and the solution is no longer applicable[4].

Si trattava di un'idea che Shannon aveva sperimentato anche a partire da questioni discusse con Alan Turing nel 1943 all'ora del tè nella caffetteria dei Bell Labs: «la possibilità che le macchine imparassero a pensare»[5]. La risposta di Shannon fu la costruzione di un meccanismo costituito, alla sua superficie, da: 1. un labirinto metallico composto da una serie di venticinque quadrati 5x5 cm separabili tramite divisori mobili; 2. un sensore a forma di topo che poteva girare nel labirinto riconoscendo le pareti e il centro dei quadrati; 3. un obiettivo (*goal*) montato su un perno che poteva essere spostato in ogni quadrato. Sotto questo labirinto Shannon aveva posizionato un circuito fatto da settantacinque relè che, prima che la macchina venisse azionata, stavano in quella che il matematico stesso definì «nominal position»[6].

In pratica i diversi relè si trovavano fin dall'inizio in due possibili condizioni, quella in cui è consentito far passare la corrente oppure quella in cui il passaggio della stessa viene bloccato. Questa posizione iniziale, però, non era significativa per il sensore in quanto i relè non erano in quelle posizioni a partire da un percorso compiuto precedentemente dal sensore. Ciò implicava che il sensore non avrebbe seguito quella indicazione finché non eseguiva la verifica prevista e non capiva se aveva senso rispetto al suo obiettivo. Questa verifica avveniva tramite una strategia che lo stesso Shannon definiva complessa, ma non casuale: a partire dal quadrato in cui viene inserito, il sensore eseguiva il compito di verificare, per ogni quadrato, quale parete era aperta e quindi gli consentiva di andare avanti. Sotto la struttura del labirinto, il circuito di relè poteva memorizzare quale parete era aperta quando il sensore aveva esplorato quel quadrato. Quella che Shannon chiama *strategia esplorativa* seguiva una logica antioraria per cui il sensore partiva dal verificare la parete nord e poi procedeva nelle direzioni seguenti in senso antiorario finché non trovava il punto in cui poteva proseguire. Questa strategia era la prima applicata dal sensore quando veniva attivata la macchina. Una volta

[2] C. E. Shannon, *A Mathematical Theory of Communication*, «Bell System Technical Journal», vol. 27, pp. 379–423 (luglio), 623–656 (ottobre), 1948.
[3] Cfr. *Cybernetics: The Macy-Conferences 1946-1953: The Complete Transactions*, a cura di Claus Pias, Zürich, Diaphanes, 2016.
[4] Ivi, p. 474.
[5] J. Gleick, *L'informazione*, tr. it. di V. B. Sala, Milano 2021, p. 198.
[6] *Ibid*.

trovato l'obiettivo, se il sensore veniva riportato al punto di partenza, la strategia che applicava era differente. L'esplorazione ha consentito al circuito di relè di registrare le informazioni necessarie per ogni quadrato, quindi il sensore non doveva più esplorare, ma raggiungere l'obiettivo. Questa seconda strategia è detta da Shannon «goal strategy»[7] e si implementava nel momento in cui, una volta entrato in un quadrato, il sensore non doveva esplorare ogni direzione per scegliere quale prendere, ma riceveva il segnale dal relè posizionato sotto quel quadrato di seguire la direzione che aveva selezionato la volta precedente. La macchina aveva imparato il percorso trovato durante la strategia esplorativa e, in questo secondo momento, per raggiungere l'obiettivo impiegava meno tempo.

Durante il suo intervento alla Macy Conference, dopo aver mostrato come la macchina poteva eseguire queste due strategie differenti, Shannon modificò il labirinto spostando alcune pareti e riportò il sensore-topo al punto di partenza. Una volta raggiunto il primo quadrato, la cui istruzione associata al relè sottostante non era più corretta, «the machine decides that the maze has been changed or that it is in a circular loop, or something of that sort, and that the previous solution is no longer relevant»[8]. Il sensore tornava quindi alla strategia esplorativa per poter risolvere il nuovo labirinto che si trovava davanti.

Una volta conclusa la dimostrazione del funzionamento della macchina (che successivamente avrebbe denominato Theseus in onore del personaggio del mito greco che risolse un labirinto attraverso un filo d'oro donatogli da Arianna e uccidendo il mostruoso Minotauro che lo presidiava[9]), il fisico e filosofo Heinz von Foerster chiese a Shannon se, dinanzi a diverse possibilità di soluzione del labirinto, la macchina scegliesse sempre la soluzione migliore. Il matematico rispose che la macchina non sceglieva necessariamente la soluzione migliore, ma c'era un'alta probabilità che optasse per la più breve. Ma fu la domanda del matematico e statistico Leonard Jimmie Savage ad aprire una più ampia discussione: preannunciando si trattasse di un problema molto importante per ogni tipo di apprendimento umano, si chiedeva come la macchina potesse comprendere se un percorso fosse migliore di un altro se entrambi avrebbero portato al risultato. Shannon aveva appena sostenuto che la scelta del percorso più breve era soltanto altamente probabile, ma non certa; aggiunse che questa macchina non poteva valutare tutte le possibili soluzioni e scegliere la migliore perché era necessaria una quantità di memoria molto più cospicua oltre ad uno o più relè delegati alla comparazione delle soluzioni. Ciò che però Shannon ammette è che «it surely could be done»[10], bisognerebbe costruire una macchina più complessa di quella che stava presentando, ma si poteva fare. Determinante

[7] Ivi, p. 476.
[8] Ivi, p. 477.
[9] Il filmato pubblicitario dei Bell Laboratories che sfruttava la dimostrazione di Shannon iniziava con le parole: «In ancient Crete the Greek hero, Theseus, entered a labyrinth to slay the Minotaur, then found his way out with the aid of a golden thread. Today a new Theseus is adept at solving mazes – but telephon relays have replaced the golden thread».
[10] Ivi, p. 478.

per il funzionamento del Theseus originale era però, come Shannon avrebbe precisato nel video illustrativo prodotto successivamente dai Bell Laboratories[11], che il sensore si comportasse esattamente come una catena di relè in grado di trovare, per ogni connessione telefonica da effettuare, *una via in quel momento libera* tra tutte le altre in corso. La strategia costruttiva era dunque modellata sull'idea di una sorta di 'sopravvivenza' o 'riuscita' nel contesto di una rete simultanea di strategie concorrenziali.

In *The Cybernetics Moment: Or Why We Call Our Age the Information Age* di Ronald R. Kline vengono riportate alcune problematiche riscontrate dai curatori degli atti della Macy Conference del 1951. Heinz von Foerster, Margaret Mead, e Hans Teuber notarono come il sensore-topo di Shannon fosse affascinante per gli studiosi di cibernetica quanto lo era stata l'innovazione in campo missilistico che aveva consentito di rendere la contraerea più efficiente grazie a sistemi che aiutavano il pilota a capire dove colpire il bersaglio in base ad una previsione statistica della traiettoria di quest'ultimo. In realtà il topo di Shannon secondo loro non aveva una reale somiglianza con il topo vero[12], ciò che però ammettevano era che «these models are rather convincing facsimiles of organismic or social processes»[13]. Von Foerster, Mead e Teuber sostenevano che il modello creato da Shannon poteva servire per studiare il comportamento di organismi o di gruppi sociali, ma soltanto in quanto modello che poi doveva essere modulato in base ai reali processi di comportamento organico.

Pochi mesi dopo, nel luglio del 1951 Shannon redasse una nota per Edmund C. Berkeley in cui riportava il «computing program», l'algoritmo che aveva implementato sulla macchina Theseus. Era diviso in due fasi (strategia esplorativa e strategia dell'obiettivo) e una terza fase aggiuntiva per evitare i circoli viziosi:

Exploration strategy
Enter squadre x.
1. Read direction D in memory for square x.
2. Make note of direction D^I by which mouse has just enterd square x.
3. Rotate D by 90° counter clockwise, obtaining D_1
    a. If D, ≠ D^I, try D_1.
    b. If D, = D^I, rotate 90° more to D_2 and try D_2.
4. In either case put the new chosen direction D, or D_2 into memory in place of D.
5. Start motor.
    a. If mouse hits barrier, then reserve motor and return to center of square, and start step 1 again.
    b. If no barrier, then at the next center of a square:

[11] Il video è consultabile all'indirizzo: https://www.youtube.com/watch?v=nS0luYZd4fs
[12] R. R. Kline, *The Cybernetics Moment: Or Why We Call Our Age the Information Age*, Baltimora 2015, p. 51.
[13] H. von Foerster, M. Mead and H. Teuber, eds., *A Note by the Editors*, in *Cybernetics: Circular Causal and Feedback Mechanisms in Biology and Social Systems. Transactions*, Vols. 6–10, (New York, 1950-1955), p. XIX.

(i)If no goal, return to step 1.
(ii) If goal, ring bell and go to Goal Strategy (i.e., keep goal relay help up)

Goal Strategy
Enter square x.
1. Read direction D from memory for x.
2. Try this direction D.
    a. If mouse hits barrier, return to center of square, and rotate 90° to D, and try D_1.
    b. If no barrier, continue to next square, and return to step 1.

To avoid Circles – additional part of program
Let C be a counter which counts to 25.
1. When the mouse enters a new square, and no barrier, C advances one unit.
2. When mouse strikes goal, C is reset to 0.
3. If counter reaches the count 24 on goal strategy, then machine returns to exploration strategy[14].

Le tre strategie descritte da queste poche righe consentivano, secondo Shannon, a Theseus di eseguire i compiti essenziali per qualsiasi essere si possa definire intelligente: imparare da un'esperienza elaborando strategie di prova ed errore in grado di autocorreggersi e raggiungere un obiettivo evitando di cadere in qualche trappola. Ciò che la dimostrazione di Shannon alla Macy Conference mostrava è che se veniva eliminato l'obiettivo dal labirinto Theseus cercava di trovare ugualmente una soluzione per risolvere il problema che aveva dinanzi. L'algoritmo, ossia le istruzioni implementate da Shannon sulla macchina, era cioè aperto alla possibilità del fallimento. Al di là delle critiche e delle annotazioni avvenute a seguito della presentazione di Theseus alla Macy Conference, si può ipotizzare che Shannon avesse dato seguito alle ipotesi già esposte da Alan Turing nel testo *Computing Machinery and Intelligence*[15] sulla possibilità di costruire delle macchine dotate di una certa intelligenza, ossia una certa capacità di riconfigurarsi in base all'esperienza. Il *machine learning* era ancora agli albori, ma Theseus certamente aveva qualcosa di «troppo umano» per non inquietare chi lo osservava per la prima volta[16].

In seguito, esperimenti simili a Theseus sono stati portati avanti in campo cibernetico, a partire dal tentativo di Berkeley e del suo team formato da Ivan e Bert Sutherland: il loro risolutore di labirinti si chiamava Franken e l'obiettivo del gruppo era quello di creare un prototipo per venderlo su piccola scala. Il prototipo venne consegnato a Berkeley nel 1955 e due anni dopo si

[14] Il documento in cui viene riportato l'algoritmo di Theseus è presente al seguente link: http://cyberneticzoo.com/mazesolvers/1951-maze-solving-mouse-claude-shannon-american/
[15] A. M. Turing, *Computing Machinery and Intelligence*, «Mind», 49, 1950, pp. 433-460.
[16] Questa la reazione di Lawrence K. Frank quando Shannon mostrava il loop *nevrotico* in cui entrava il sensore-topo se si toglieva completamente l'obiettivo dal labirinto. In *Cybernetics: The Macy-Conferences 1946-1953*, cit.

tentò di venderlo, mentre Ivan Sutherland venne notato da Shannon stesso che supervisionò il suo lavoro di tesi[17].

Oltreoceano nel 1954, all'istituto di psicologia sperimentale di Oxford, Anthony Deutsch costruì quello che venne considerato il risolutore di labirinti più sofisticato dell'epoca poiché era in grado di trasferire ciò che aveva imparato su un labirinto in un altro topologicamente equivalente anche se di misura differente[18].

Negli anni Settanta, il tedesco Johan de Boer costruì un *computer maze*, ossia un computer con circa 100 chip della serie 74, impostati su 5 circuiti stampati con moltissimi cablaggi. In questo computer-labirinto si accendeva una luce per indicare la posizione di un topo immaginario che girava al suo interno. Il tutto era controllato da un computer che aveva una memoria di 1kbyte e in cui era dunque possibile caricare 64 istruzioni, ossia un algoritmo fatto da 64 passaggi[19].

In seguito si cominciarono ad indire contest tra topi elettronici in tutto il mondo. Il primo si tenne a New York City nel 1978 con il titolo *The Great Electronic Mouse Race*[20] e tanti altri ne seguirono fino ad arrivare ai giorni nostri: in Italia all'IEEE (Istitute of electrical and electronics engineers) di Padova un gruppo di studenti ha partecipato al Micro Mouse Contest che si tiene in tutti i centri IEEE del mondo[21].

Il Theseus di Shannon ha ispirato i successivi tentativi di realizzare delle macchine in grado di imparare, ma ha sollevato anche una questione riguardante i limiti che il topo elettronico aveva rispetto ad un reale essere vivente. In una logica comportamentista alla Skinner, l'algoritmo implementato da Shannon su Theseus poteva aprirsi realmente alla possibilità di raggiungere livelli più complessi di apprendimento o la sua apertura era limitata in principio dalle stesse scelte logiche fatte dal suo creatore? Dal punto di vista di una psicologia cognitiva alla Tolman, invece, che limiti potevano essere posti alla capacità del topo elettronico di creare, memorizzare e confrontare 'mappe' per essere in grado di raffinare progressivamente le proprie strategie esplorative? Era solo una questione di 'capacità di memoria' o necessitava dell'altro?

In breve, l'algoritmo di Theseus avrebbe mai potuto 'evolvere', cioè superare la semplice reiterazione della logica stimolo-risposta per arrivare ad una elaborazione complessa dei tentativi di soluzioni falliti e di quelli riusciti in modo da pervenire ad una soluzione *creativa* del problema?

[17] http://cyberneticzoo.com/mazesolvers/1953-franken-maze-solving-machine-ivan-and-bert-sutherland/
[18] http://cyberneticzoo.com/mazesolvers/1954-maze-solving-machine-j-a-deutsch-british/
[19] http://cyberneticzoo.com/mazesolvers/1970-3-computer-maze-johan-de-boer-canadian/
[20] http://davidbuckley.net/RS/mmouse/micromouse78.htm
[21] http://www.dei.unipd.it/~ieeesb/micromouse/index.html

2. Simondon: riflesso e istinto

Una strada per capire se queste domande abbiano in generale senso rispetto al tentativo di Shannon di andare oltre la semplice analogia e codificare le strategie di apprendimento di una macchina intelligente *in generale*, è quella di verificare se la distinzione fra i due livelli di funzionamento dell'algoritmo, quello semplicemente reiterativo e quello creativo-risolutivo, possa essere riportata a una distinzione effettivamente interna al vivente: quella tra comportamento *riflesso* e comportamento *istintivo*. Si tratta cioè di capire se anche all'interno del vivente la transizione tra i due comportamenti come indice di una complessità crescente dell'organismo sia il frutto di una stratificazione progressiva del primo che permette la sua evoluzione nel secondo, o se le due sfere siano destinate a rimanere separate e a prodursi non tanto per evoluzione quanto per *rottura* dell'una rispetto all'altra. Si tratta di una problematica sulla quale il filosofo francese Gilbert Simondon[22] ha insistito in diverse occasioni durante i suoi corsi di psicologia[23]. Tra febbraio e maggio 1964 Simondon ha tenuto un corso sull'istinto all'Università Sorbona di Parigi la cui introduzione si apriva con una disamina delle diverse teorie filosofiche che lo hanno delineato fino a giungere ad una distinzione tra quattro tipologie di condotte: il riflesso, l'istinto, l'apprendimento e l'intelligenza.

Nella parte introduttiva delle lezioni Simondon riteneva necessario ricordare all'uditorio che esistono diverse tipologie di condotte istintive ma che ciò che le caratterizza tutte è un certo fenomeno di amplificazione grazie al quale due sottoinsiemi anatomici, fisiologici, comportamentali o funzionali che prima facevano parte di due organismi si associano trovando un certo equilibrio che li rende in grado di controllare, dirigere, il tutto escludendo altre possibilità di condotta[24]: «l'instinct serait donc d'abord une *fonction de changement de régime* de la conduite, supposant un phénomène d'amplification»[25]. In questo senso, Simondon portava l'esempio dei protisti sostenendo che questi organismi non sono in grado di raggiungere il livello delle condotte istintive in quanto la loro

[22] Da notare che tra il pensiero di Simondon, la teoria di Shannon e la cibernetica di Wiener sussistono importanti interrelazioni. In particolare, il concetto di informazione delineato da Shannon, e poi ripreso da Wiener, è stato oggetto di critica da parte di Simondon in quanto mancante di una dimensione qualitativa. Inoltre, Simondon ha organizzato il convegno intitolato *Le concept d'information dans la science contemporaine* presso Royaumont nel 1962 in cui sono state discusse le principali tematiche della cibernetica. Cfr: P. Chabot, *La philosophie de Simondon*, Paris 2003; J. H. Barthélémy, J. C. Beaune, *Penser l'individuation Simondon et la philosophie de la nature*, Paris 2005: Xavier Guchet, *Pour un humanisme technologique: culture, technique et société dans la philosophie de Gilbert Simondon*, Paris 2010; H. Gauthier-Villars (a cura di), *Le Concept d'information dans la science contemporaine,* Cahiers de Royaumont, Paris 1965; G. Simondon, *L'Amplification dans les processus d'information*, in *Communication et information, Cours et conférences*, Paris 2010; G. Simondon, *Epistemologia della cibernetica*, «*aut aut*», *377 (Effetto Simondon*, a cura di V. Cavedagna e G. Piatti) 2018.
[23] Cfr. G. Simondon, *Deux leçons sur l'animal et l'homme*, Paris 2004 e G. Simondon, *Communication et information: Cours et conférences*, Paris 2010.
[24] G. Simondon, *Communication et information*, cit., pp. 251-252.
[25] *Ibid.*

struttura non ha una complessità tale da poter avere sottoinsiemi abbastanza differenziati e indipendenti. Sia a livello individuale che a livello sociale, la condotta istintiva è determinata da un brusco cambiamento delle condizioni di adattamento, come una catastrofe naturale oppure l'invasione di un territorio da parte di una specie differente, ma che determina anche un necessario riadattamento.

In generale, diceva Simondon, ogni tipologia di condotta istintiva (individuale, di specie o sociale) risponde a tre condizioni fondamentali: quelle percettive-sensoriali, quelle di ordine centrale e quelle motorie. La prima tipologia di condizioni riguarda tutti quei *pattern* scatenanti che comportano l'emergere di una condotta istintiva.

Per i nostri interessi specifici, è interessante analizzare ciò che Simondon affermava parlando delle condizioni di ordine generale. Egli sosteneva che queste implicherebbero l'attivazione di un intero organo che andrebbe ad intervenire come «un *relais* (corsivo nostro) entre le stimulus reçu et la commande motrice envoyée aux effecteur»[26]. Il relè avrebbe il compito di stabilire una connessione tra ciò che provoca l'eccitazione e la reazione alla stessa; essere cioè un collegamento attraverso il quale passa dell'informazione tramite l'amplificazione. A partire dalle condizioni percettive-sensoriali, si andrebbe dunque ad attivare questo *device* organico che avrebbe il compito di fare da medium tra lo stimolo ambientale e l'azione che viene emessa come risposta.

Questa terza condizione riguarda la motricità vera e propria e viene identificata da Simondon come il ruolo svolto dal recettore in un'operazione tecnica mediata dal relè: una volta ricevuto lo stimolo, il relè amplifica l'informazione verso il recettore che dovrà leggerla ed interpretarla.

Una volta presenti le tre condizioni, Simondon sosteneva che si possono realizzare tutte le diverse tipologie di condotte istintive. Ciò si lega ad una questione più generale, vale a dire alla difficoltà riscontrata all'interno della storia della filosofia, sin dagli Stoici, di pensare l'istinto come ciò che distingue l'animalità dall'umanità. Se l'animalità era contraddistinta dall'istintualità, l'umanità poteva essere definita superiore poiché aveva raggiunto la capacità di controllare l'istinto, pensato come reazione automatica, grazie alla razionalità. Simondon si poneva invece in netta opposizione a questa visione antropocentrica che aveva dominato la storia della filosofia a partire dal presupposto che «mieux vaut poser d'abord les problèmes épistémologiques et logiques»[27].

Porre la questione dell'istinto dal punto di vista epistemologico e logico è per Simondon il primo passo per affrontare il suo grande progetto di indagare

[26] Ivi, p. 254.
[27] Ivi, p. 258.

l'incidenza degli studi sull'istinto che si sono sviluppati prima e dopo lo Stoicismo[28], in particolare per quanto riguarda la questione dell'evoluzione[29].

Il primo punto del progetto simondoniano verte sulla differenza che è possibile riscontrare tra la definizione di istinto e quella di riflesso. In prima battuta egli presentava ai suoi allievi una prima possibile definizione di riflesso: se si applica agli animali superiori, «désigne une réaction de contraction des muscles ou de sécrétion des glandes»[30], se invece si applica a tutti gli organismi «les réflexes sont essentiellement des réponses automatiques à des excitations définies»[31].

La differenza tra queste definizioni non viene però riconosciuta da tutti gli studiosi. Tra le teorie che non contemplano un divario tra riflessi e istinto Simondon citava quella di Rabaud in cui l'istinto viene ridotto ai riflessi considerati una reazione motrice all'eccitazione ricevuta che deve essere generalizzata (investe l'intero organismo), incosciente, imperativa, orientata (verso l'organo che deve reagire), ha una soglia di eccitazione che deve essere raggiunta e, in alcuni casi, può essere modulabile. In questa prospettiva, il riflesso sembrerebbe avere tutte le caratteristiche che altri attribuirebbero agli istinti. Difatti, Rabaud sosteneva che i riflessi sono la base «permettant de rendre compte de tous les comportements plus complexes, tels que les tropismes et les instincts»[32].

Una prospettiva che consente, invece, di riscontrare la distanza tra riflessi e istinti è, secondo Simondon, quella dello psicologo francese Gaston Viaud[33] che non considerava i riflessi come delle vere e proprie condotte, ma come l'insieme delle risposte motorie che l'organismo ha a disposizione per rispondere alle eccitazioni che riceve. Questa prospettiva verrà poi ripresa dal neuropsichiatra tedesco Kurt Goldstein, il quale intendeva il riflesso come una reazione dell'organismo sottomesso ad un disadattamento che però si verifica in tempi molto più brevi rispetto all'intera azione di mediazione tra un adattamento, un disadattamento e un riadattamento[34].

Riflesso e istinto sono, quindi, per Simondon differenti in quanto il primo riguarda le possibilità che un organismo ha di rispondere ad un processo che prevede stimolo, ricezione, amplificazione e risposta, mentre le condotte istintuali sono questo stesso processo nel suo insieme come portatore di potenziale

[28] Simondon riporta una disamina delle maggiori teorie che si sono sviluppate nella storia del pensiero sull'istinto mettendo l'accento sulla visione naturalista degli Epicurei e sulla svolta apportata da Jean Baptiste Lamarck grazie all'idea che una condotta, anche istintuale, possa avere un ruolo nella strutturazione morfologica e fisiologica di un individuo. Cfr, ivi, pp. 258-260.

[29] Si tratta di un piano di ricerca che viene riportato in questo corso del 1964, ma in una nota viene sottolineato che Simondon è riuscito a portare avanti durante queste lezioni soltanto la prima parte (su cinque) prevista.

[30] Ivi, p. 276.

[31] Ivi, p. 278.

[32] Ivi, pp. 280-281.

[33] «Les réflexes sont des réactions à des stimuli physiques et non à des signes perceptifs» (ivi, p. 285).

[34] Ivi, p. 287.

evolutivo. Dire che le condotte istintuali sono l'insieme delle condotte di riflesso potrebbe, però, far cadere nell'errore di pensare ad una distinzione senza differenza. Nella prospettiva simondoniana, invece, le condotte istintuali sono il centro di controllo cibernetico dell'insieme dei riflessi che, potendo scegliere tra le varie risposte possibili, opta per quella più adatta finché le condizioni non mutano determinando la necessità di una scelta differente. Dunque, ciò che li distingue effettivamente consiste nella capacità di automatizzarsi, ma anche di disautomatizzarsi, delle condotte istintive che, invece, i riflessi non hanno. Questo perché un riflesso ha una capacità di realizzarsi molto rapida e che non comporta necessariamente un precedente processo di adattamento, cioè si tratta di un automatismo che non prevede una libertà reale iniziale. Una condotta istintiva prevede, invece, che, a partire da un disadattamento, si realizzi un processo di adattamento/riadattamento che renderà, a lungo termine, quell'istinto un comportamento che risulta automatico.

Conclusioni

Questa brevissima disamina del pensiero simondoniano sull'istinto consente già di valutare se l'algoritmo implementato da Claude Shannon nel topo Theseus rimanga al livello di riflesso o possa costituire la base per la successiva elaborazione di comportamenti istintivi. In prima istanza, Theseus reagisce al disadattamento seguendo alcuni passaggi obbligati che sono pre-iscritti (letteralmente: precablati) nel suo meccanismo di funzionamento sotterraneo. I passaggi dell'algoritmo sono certamente obbligati, ma non tutti necessari: Theseus può saltare alcune istruzioni se le indicazioni lo prevedono ((i) If no goal, return to step 1). Questo potrebbe consentire di affermare che Theseus possieda una limitata capacità istintuale, ossia di reagire liberamente alle modificazioni ambientali. Se l'istinto è infatti un processo che, a partire da *pattern* motori (i riflessi) limitati, sceglie quale utilizzare come maggiormente adatto per riadattare un organismo ad un ambiente modificato, allora Theseus, seguendo i *pattern* dell'algoritmo di Shannon, sceglie di volta in volta quale utilizzare per muoversi all'interno dei diversi quadranti del labirinto. Come accade per gli istinti osservati in campo biologico, anche gli istinti di Theseus cambiano se l'ambiente circostante necessita un suo riadattamento e rimangono invariati, quindi apparentemente puramente automatici, se il labirinto non viene modificato. Ciò contribuirebbe progressivamente a liberare da una certa 'casualità' di partenza le strategie rispetto alle quali il topo elettronico sperimenta una soluzione *dopo* l'altra qualora la prima si riveli impraticabile fino a trovare quella che, in quel momento, sia la via più *libera*.

Più complesso, sulla base delle tesi di Simondon sull'intelligenza e il suo rapporto con l'istinto, il tentativo di verificare se Theseus oltre a procedere secondo condotte istintuali possa essere considerato intelligente. E se, in ultima istanza, sia necessario o meno questo ultimo passaggio per definire il suo algoritmo potenzialmente *vivente*.

Nell'ultima parte del corso del '64, Simondon affermava che:

> l'action intelligente consiste souvent à poser le problème résolu: le sujet, l'organisme, intervient dans la situation, et c'est cette introduction qui, faisant changer de plan l'ensemble des données, permet la systématisation unifiante[35].

Questo tipo di sistematizzazione unificante avviene però anche nella percezione, come nel caso della visione[36]. Ciò che distingue la risoluzione del problema percettivo e quella di un comportamento intelligente è l'aspetto simbolico: «par abstraction et généralisation, la connaissance symbolique permet de combiner les symboles des objets sans qu'ils soient présents eux-mêmes»[37]. Attraverso i simboli, l'organismo che porta avanti comportamenti intelligenti può, in effetti, prevedere e organizzare le proprie azioni in funzione di un fine. Non basta, quindi, semplicemente una certa quantità di memoria che incamera tutte le varie possibilità applicabili, ma anche la capacità di trovare una soluzione che non risulta già memorizzata.

Accomunare le condotte intelligenti a quelle istintuali è possibile a partire da due casi: quello dei comportamenti di aggiramento degli ostacoli e quello dei comportamenti collettivi.

Per quanto riguarda il primo caso, gli psicologi inglesi Mc Dougall hanno portato avanti degli studi in cui sostenevano che gli organismi sviluppano una condotta di aggiramento degli ostacoli quando è presente la «*foresight*»[38], ossia la capacità di percepire la situazione e di prevedere. Diverso dal semplice *insight* al centro dell'elaborazione di strategie risolutive nel cognitivismo in particolare, questo tipo di comportamento predittivo era riscontrato dagli psicologici inglesi durante l'apprendimento di labirinti le cui parti erano visuali[39]. Gli individui che devono apprendere la struttura di un labirinto per imparare ad uscirne o per trovare un obiettivo devono poter percepire le condizioni spaziali in cui si trova e poter sapere già in anticipo le diverse mosse da svolgere per trovare una soluzione. Per restare nelle coordinante tracciate dal piano di Shannon, potremmo dire che il Theseus elettronico del 1951 procede 'alla cieca' (come il Teseo del mito, il cui labirinto era probabilmente immerso nel buio per aggiungere difficoltà a difficoltà) e dunque sfrutta un'applicazione abbastanza limitata della topologia (precisamente quella bidimensionale). Un Teseo in grado di 'vedere' potrebbe invece creare mappe topologiche tridimensionali in grado di dar conto di effetti inaspettati[40].

[35] Ivi, p. 314.
[36] Ivi, p. 315.
[37] *Ibid.*
[38] Ivi, p. 319.
[39] *Ibid.*
[40] Un esempio particolarmente divertente a questo link: https://youtu.be/TsJJEsZ6m-E

Per il secondo caso, Simondon riportava l'esempio dei necrofori studiati dall'etologo Jean-Henri Fabre e ripresi anche da Gaston Viaud[41], ma in generale considera tutti i comportamenti istintivi sociali che comportano divisione del lavoro come i casi in cui il rapporto tra istinto ed intelligenza è più evidente e i meno antropocentrici in quanto è possibile osservarli tra animali come le api o le termiti.[42] Simondon chiudeva la trattazione accennando all'ipotesi di definire queste condotte comportamenti «d'intelligence collective»[43] anche se sosteneva che il corso che stava tenendo non era il luogo adatto in cui indagarli e rimandava ad «avenir des recherches est peut-être dans l'étude de la perception individuelle et collective»[44].

Quest'ultimo concetto di intelligenza collettiva e la più argomentata idea simondoniana di una «realitè humaine»[45] contenuta in tutti le invenzioni umane potrebbero consentire di riscontrare nell'algoritmo di Shannon una certa condotta intelligente. Il comportamento del topo è infatti espressione tanto di una rete di connessioni funzionali rappresentata dall'intrico di relè sotto il piano del labirinto che ne costituisce 'memoria estesa' e 'intelligenza distribuita', quanto del sistema storico-evolutivo di invenzioni e innovazioni al cui interno la rete di connessioni viene progettata, costruita e fatta realmente funzionare. Come il Teseo del mito, il topo di Shannon non è mai realmente 'solo' nell'affrontare le difficoltà che gli si presentano. Esso è invece assistito dall'inizio da un progetto-espressione di una reticolarità allargata e un'intelligenza distribuita che ne fanno un proto-organismo evolutivo a (quasi) tutti gli effetti. Tra l'algoritmo di Theseus e la collettività ingegneristica in cui è inserito (di cui Shannon sarebbe l'effettore ultimo) e la comunità di parlanti telefonici che con i suoi esperimenti di reticolarità contribuisce a mettere in relazione, vi sarebbe una divisione/distribuzione dei compiti come quella presente nei comportamenti istintivi sociali di cui parlava Simondon.

Come per il Teseo del mito, è allora l'annunciarsi di un rischio comune che inquieta e spaventa rispetto al caso di una macchina (vivente e non vivente) costretta per l'eternità ad aggirarsi in un labirinto di scelte e decisioni non sempre necessariamente binarie o garantite: il rischio di essere lasciati soli senza la possibilità di collegarsi ad altro, senza più *golden thread* o contatti (anche elettrici) a cui appoggiarsi per ritrovare la strada.

Possiamo dunque pensare che quando gli astanti alla dimostrazione di Shannon definivano l'algoritmo di Theseus «troppo umano» lo facessero con quel misto di fascino e terrore che sperimenta chiunque venga messo dinanzi a una prova per risolvere la quale è necessario assumere sulle proprie spalle tutto il peso dell'evoluzione strategica, tecnica e collettiva fino a quel momento.

[41] G. Simondon, *Communication et information*, cit., pp. 321-322. Cfr. anche J.-H. Fabre, *Souvenirs Entomologiques*, Paris 2009.
[42] *Ibid*.
[43] *Ibid*.
[44] *Ibid*.
[45] G. Simondon, *Del modo di esistenza degli oggetti tecnici*, Napoli-Salerno 2020.

Che è esattamente quello che accade agli algoritmi con cui cerchiamo oggi di funzionare e far funzionare le nostre società. Vista con gli occhi di Simondon, e dunque oltre la sola logica stimolo-risposta-regolazione-controllo definita dalla cibernetica di prima generazione contro la quale egli insorgeva già dal 1953[46], la tragica vitalità della macchina di Shannon era difficilmente occultabile.

Francesca Sunseri
Università degli studi di Palerno
✉ francesca.sunseri@unipa.it

Bibliografia

Barthélémy J. H. 2005. *Penser l'individuation Simondon et la philosophie de la nature*, Paris, Harmattan,

Chabot, P. 2003. *La philosophie de Simondon*, Paris, Vrin.

Couffignal, L. (éd. par), 1965. *Le Concept d'information dans la science contemporaine,* Cahiers de Royaumont, Paris, De Minuit.

Fabre, J. H. 2009., *Souvenirs Entomologiques. Etudes sur l'instinct et les moeurs des insectes*, Paris, éd. Robert Laffont.

Foerster (von), H., Mead, M., and Teuber, H. (ed. by) 1950. *A Note by the Editors*, in *Cybernetics: Circular Causal and Feedback Mechanisms in Biology and Social Systems. Transactions*, Vols. 6–10, (New York: Macy Foundation, 1950–1955).

Gleick, J. 2015. *L'informazione*, tr. it. di V. B. Sala, Milano, Feltrinelli.

Guchet, X. 2010. *Pour un humanisme technologique: culture, technique et société dans la philosophie de Gilbert Simondon*, Paris, PUF.

Kline, R. R. 2015. *The Cybernetics Moment: Or Why We Call Our Age the Information Age*, Baltimora, Johns Hopkins University Press.

Pias, C. (a cura di) 2016. *Cybernetics: The Macy-Conferences 1946-1953: The Complete Transactions*, Zürich, Diaphanes.

Shannon, C. E. 1948. *A Mathematical Theory of Communication*, «Bell System Technical Journal», vol. 27, pp. 379–423 (luglio), 623–656 (ottobre).

Simondon, G. 2004. *Deux leçons sur l'animal et l'homme*, Paris, Ellipses.

Simondon, G. 2010. *Communication et information, Cours et conférences*, Paris, La Transparence.

Simondon, G. 2018. *Epistemologia della cibernetica*, in «*aut aut*», *377*, Effetto Simondon, V. Cavedagna e G. Piatti (a cura di), Milano, Il Saggiatore.

Simondon, G. 2021. *Del modo di esistenza degli oggetti tecnici*, trad. it. a cura di A. S. Caridi, Napoli-Salerno, Orthotes.

Turing, A. M. 1950. *Computing Machinery and Intelligence*, «Mind», 49, pp. 433-460.

[46] G. Simondon, *Epistemologia della cibernetica*, cit.

Sitografia

http://cyberneticzoo.com/mazesolvers/1951-maze-solving-mouse-claude-shannon-american/
http://cyberneticzoo.com/mazesolvers/1953-franken-maze-solving-machine-ivan-and-bert-sutherland/
http://cyberneticzoo.com/mazesolvers/1954-maze-solving-machine-j-a-deutsch-british/
http://cyberneticzoo.com/mazesolvers/1970-3-computer-maze-johan-de-boer-canadian/
http://davidbuckley.net/RS/mmouse/micromouse78.htm
http://www.dei.unipd.it/~ieeesb/micromouse/index.html
https://www.youtube.com/watch?v=nS0luYZd4fs
https://youtu.be/TsJJEsZ6m-E

¶ Sezione Quarta

Algoritmo e società

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770074

Interviste/2

Questioning Algorithms
Interview with Igor Pelgreffi

Andrew Feenberg 0000-0003-4850-0402

Andrew Feenberg, one of the greatest living philosophers of technology, addresses the question of the algorithm starting from a comparison with some of his typical topics, in particular his vision of technology as a social construction derived from works like *Critical Theory of Technology* (1991) and *Questioning Technology* (1999). Automation in work is one of the logical centres of the evolutionary and historical movement that today leads to the primacy of what may be called 'algorithmic governmentality'. After an analysis of Marx's thought in the *Capital* and in the *Grundrisse*, the question of the Jacquard loom and of Babbage is also touched upon in the line of a capitalism of machine automation, which deskills and eliminates labour at many levels, including the intellectual one. Feenberg also dwells on other theoretical references, such as his concept of 'technical code', that of Foucault's *dispositif* and above all the possibility of re-actualizing Marcuse's thought, in the key of a critique of technology today. Furthermore, Feenberg explores the question of the loss of the *Lebenswelt* as a symptom of the one-dimensionality of the human being due to extensive digitalization in the historical-social world, and the future possibility of emancipation.

***

Igor Pelgreffi – *Professor Feenberg, I am asking you about the topics in question (algorithm: genealogy, theory, critique) as you have a long experience on the one hand on the subjects of philosophy of technology, but also historical and socio-political aspects of technology, on the other. The visual angle of this interview, therefore, will be precisely that of a look that takes into account your many years of work, already in two very important books released in the 1990s, such as* Critical Theory of Technology *(Oxford University Press, 1991) and* Questioning Technology *(Routledge, 1999). It is true that there is still no explicit talk of algorithms there, and still not much of the progressive trend towards global algorithmization in work processes. And it is true that, in those works, you did not intensively speak of digital technologies. And yet, it seems to me that there are contents or theoretical tools that – and this is what I ask you to begin with – can perhaps be applied to the current situation, and hence to some analysis of the algorithm.*

*It seems to me, for example, that in those books you can find a topic of this kind: technologies, in a sense, 'go by themselves'. Auto-matically (*autos*). A bit like the* global automatic machine *or the* automatisches Subiekt *mentioned, among others, by Marx in* Das Kapital. *There was already the idea of this hyper-functioning or meta-functioning, which tends to be more and more automatic (in the processes of integral and systematic automation, which the digital will later realize historically, albeit with some resistance layer).*

But, having admitted this, then we need an idea of technology, and an ability to read technology, which allows us to resist 'by staying internal'. In the dialectic between the technocratic model of control *and the* democratic model of communication*, of which you spoke in* Critical Theory*, the path of a* social construction of technology *emerged as a concrete possibility. But, as I said, this was just one of the possible examples. So, before dealing with the topic of* algorithmic governmentali*ty, which as you suggested is perhaps a good starting point, I asked you for a first reflection on the above.*

Andrew Feenberg – The social construction of technology is at work in both the technocratic and the democratic models of technology. All technology is socially constructed, although who constructs it may differ significantly. A general feature like the wheel or the gear can be incorporated into a variety of technical systems, with different social consequences, depending on the demands of the social actors able to influence the design process. Automatic functioning is such a general feature. It can play a role in a technical system organizing democratically or technocratically, but that role will be different in each case.

All this goes back to Marx's considerations on technology in *Capital* and in the *Grundrisse.* Marx observes that the combined and coordinated labour of groups of workers can exceed the contribution each would make working individually. Capitalism is based on the maximization and appropriation of this excess by the capitalist who is not only the organizer of groups of workers, but also in possession of the tools they use and therefore able to modify those tools in accordance with his interests. The capitalist occupies a peculiar position in the division of labour as both the technical coordinator of the work and the exploiter of the work group. The tools of the work group are gradually modified in function of these roles. The tools must enable control of labour during the workday despite the lack of interest of the workers in production, and that involves reducing effort and skill to a minimum.

In this capitalist context, automation reduces the worker to a slave of the machine. The intellectual forces of production are transferred to the automatic functioning of the machine and the worker left with simple mechanical operations the machine cannot perform. Unskilled workers who can demand little and are easily replaced correspond to this type of automatic functioning. In the most developed case, those operations consist of nothing more than supervision and maintenance, performed by a small cadre of workers for a vast array of machines. The ideal of the workerless factory is asymptotically achieved.

Marx believed that this condition would be incompatible with capitalism. If surplus value is extracted from unpaid labour, the ultimate success of capitalism in eliminated labour would be its downfall. Exactly how this would occur Marx does not say, but he argues that full automation would only be compatible with a socialist society in which most workers' roles were reduced to tending the automatic machines. The qualifications of work would rise in these circumstances as workers were charged with second order intellectual tasks requiring scientific knowledge. Marx did not specify the degree of work democracy involved in this system, although Engels suggested that democratic control would be desirable within practical limits.

There is much debate over whether Marx was a social constructivist in the full sense of the term. He certainly believed that types of production means were intrinsically suited to specific economic relations. And he showed in the case of capitalism that social actors, capitalists, constructed technology in conformity with their needs. But he did not and no doubt could not imagine a similar process of socialist reconstruction of technology by workers under socialism. Instead, he speculated about the future of automation as I have explained, thereby skipping the whole process of technological transformation implied in a constructivist conception of the socialist transition.

Be that as it may, what has actually happened as capitalism has survived its predicted demise is rather different from Marx's schema. I want to mention two important changes with respect to Marx's 19th century. Automation under capitalism has increased in intensity and scope far beyond Marx's expectations. The deskilling of labour has proceeded in ever more domains. What started out as an industrial strategy has morphed under neo-liberalism into a way of organizing bureaucratic work in both business and government. Furthermore, the drastic reduction of labour in many highly automated processes has not had the effects Marx foresaw and seems compatible with capitalism.

In this context, algorithms have played an especially important role. As is well known, it all began with the Jacquard loom. This was one of the most important innovations in the capitalist automation of work. Babbage saw the possibility of achieving similar results in the calculation of navigational tables and received a subsidy from the British government to create a computer for that purpose. Had he succeeded, the application of automation to all sorts of bureaucratic occupations would have begun in the 19th century. There is a science fiction novel that imagines the results. In reality, Babbage was unable to make his 'difference engine' work, and it was only after World War II that his vision was realized. That achievement is the basis of the radical extension of automation beyond industry to domains such as education and government. This is what is radically new about contemporary capitalism.

In these bureaucratic domains neo-liberal management operates in accordance with the capitalist imperatives of deskilling and eliminating labour even in non-capitalist and public sectors of the economy. The social construction of automated technology is still subject to the limitations of the capitalist division

of labour, which excludes democratic control and input into technological decisions, and it extends into many other areas of social life. It now becomes apparent that the essence of capitalism is not simply private ownership of the means of production, but more fundamentally the organization of every sort of cooperative activity from 'above', from a position that both coordinates and exploits the activities done 'below'. We can call the most advanced forms of automation under capitalism 'algorithmic governance' only so long as we keep in mind the specific social construction that depends on this organization.

The socialist construction of technology and work organization now comes more clearly into focus than in its original Marxian version. The principal issue is no longer the tyranny of the market but rather the construction of technology.

Under socialism, as Marx conjectured, automated processes would not be designed to exploit the labour force and control social activities. They would not be organized from above but controlled from below. The technical functions of management and engineering would be coordinated with collegial and democratic processes involving the whole work group and the surrounding community. A wide distribution of intellectual skills would be essential in such a system. The end of capitalism would mean a new design for automated systems and a new role for labour.

Igor Pelgreffi – *Thank you for this long analysis of the issue, focusing on Marx philosophy, but also crossing through the Jacquard loom or Babbage, that is the historical and somehow genealogical aspects of algorithm. Let us go more in detail in the concept of algorithmic governance or algorithmic governmentality. The previous analyses, based precisely on your idea of* social construction of technology *and on the centrality of practices in determining the even logical structures of technical repetition (if you agree, one can call them 'pre-digital algorithms'), are able to read still today the phenomena related to the algorithm? And in what, instead, do they show some limit?*

Andrew Feenberg – Well, I think that by the concept of 'pre-digital algorithms' you may mean what I have called the 'technical code'. This is the systemic design principle which accommodates technology to the social order. Deskilling is such a technical code. It formats the design of a whole range of technologies according to a single principle. It is in a sense a kind of analogue version of an algorithm encoded in economic culture rather than in a computer…

Igor Pelgreffi – *If we go through this political-sociological point of view again, it seems to me that the analogy that you sketched between the algorithm and your concept of* technical code *(very interesting, in this context of analysis) can be useful. Can you tell me more about what you mean by 'technical code'?*

Andrew Feenberg – Technical design translates social demands as specifications. For example, the social demand for clean air is 'translated' into

the catalytic converter on the tailpipe of automobiles. The purely technical 'meaning' of clean air is embodied in that device. The general principles of translation I call the 'technical code.' This is the ideal type which signifies the social meaning of technical specifications. Some codes regulate whole domains of technological design. Deskilling is an example. It is a desideratum for all sorts of technologies. If we think of an algorithm as modelling a domain through computer operations, then we can see the technical code as performing a similar function at the level of the culture of engineering.

Igor Pelgreffi – *I mean, probably, the important question today is the following: if the algorithm, which can also be considered as* a matrix, *as 'the' matrix of the* administered world *(taking up the old concept of Adorno), is already there, it works (very well!), and we are somehow crossed (and perhaps even* constructed*) by it, how can we find a key to resist it, to defer its effects, to modify (or transform) it? How can we have a* critical experience *– if any – of/in the algorithm?*

Andrew Feenberg – I believe Adorno and Marcuse to a lesser extent overestimated the ability of advanced capitalism to modify the structure of everyday experience. They both saw in consumer society a force able to repel critique, but in reality as we know critique continues to be a factor in social life.

Marcuse acknowledged this under the influence of the *New Left*, unlike Adorno. The key problem of critique in critical theory has to do with consciousness of potentiality. We see this consciousness appear in many social movements today, for example, the movement against climate change. That creates tensions within the society that have significant consequences. It should be obvious, but unfortunately is not to some critical theorists, that the world has changed in response to that consciousness in the last 75 years. This is not to say the revolution is on the horizon, but theoretically that is not the issue, which concerns the role of negation in the system.

Igor Pelgreffi – *In your opinion, in addition to what you have just told me, some analysis by Foucault (author on whom, I remember, you have worked anyway) is also usable, for example his concept of* dispositif, *a little re-declined or adapted, and therefore the theme of a 'disciplinary society' which would be, in some way, 'regulated' by an algorithmic logic, and which therefore is not so distant in the current state of affairs?*

Andrew Feenberg – Foucault and Marcuse have more in common that is usually believed. Foucault, like some other intellectuals I could name (Zizek for example), identified Marcuse with various clichés of the New Left such as the revolutionary virtues of sex. In reality, Marcuse's concept of 'repressive desublimation' is similar to Foucault's theory of the instrumentalization of sex by the capitalist system. One need only read a few chapters into *One-Dimensional Man* to avoid a stupid mistake. The concept of dispositive, which

in English would be translated as 'apparatus', includes not only devices but also the associated practices. A *dispositif* is socially inflected and technically effective at the same time. This is a constructivist critique of technology, again similar to Marcuse's position. Where they differ is in Marcuse's more positive utopianism.

Igor Pelgreffi – *Thanks. Just to come back to one theme that we touch upon in the analyses of this monographic issue of* Lo Sguardo, *I would like to ask you: in your opinion, what kind of relationship exists between the body, bodies, corporeity in general, on the one hand, and algorithm, algorithmic governmentality on the other? Here, I mean that, somehow, algorithmization* affects *the bodies, it modifies the psycho-physical world, for example, and also ends up in the sphere of the unconscious, even more so than in the 'subject' sphere. After all, on these aspects, authors like Marcuse still have a lot to say, to stay on those of the Critical Theory of the Frankfurt School, on which you has worked for a long time. Of course, it is perhaps a Marcuse that you have spent reading, on the question of technique and life, also in a stimulating critical comparison with other theoretical paradigms, including Heidegger (see even just your book* Heidegger and Marcuse. The Catastrophe and Redemption of Technology, *Routledge 2004).*

Andrew Feenberg – The question of the body appears in several ways in the contemporary context. People more and more consider their bodies as devices to be operated under optimal conditions with the help of various technological instruments and practices. This goes along with the psychological phenomenon in which people view their whole life from the standpoint of a technological operator. These are incursions of the model of the computer into the *psyche*, which have the effect of reducing the body to a mechanical device.

There is an enormous outpouring of advice on the Internet covering every aspect of the body-machine and huge sums are spent on fake remedies supposed to maintain it in good condition. People wear digital watches that monitor their heart rate and the number of steps they walk in the course of the day. One could argue that these practices are quasi-algorithmic in that they standardise bodily operations in accordance with a set theme. What is lost is all the complexity of the first person standpoint on existence, the *Lebenswelt,* now stripped down to a mechanical residue. These are symptoms of the triumph of one-dimensionality. It is interesting to note that Marcuse was already sensitive to these issues and placed great emphasis on the recovery of sensuous lived bodily experience in opposition to the 'plastic' perfection of the marketable body advertised in television commercials.

Igor Pelgreffi – *Thank you Professor Feenberg. A final question: future scenarios? With respect to the increasing pervasiveness of algorithms, what do you 'feel', or what do you 'imagine', for society, but also for the technology itself? And again: what future do you 'imagine', reversibly, for critical knowledge?*

Andrew Feenberg – I wish I could say I see socialism in the future but I have no reason to believe it will happen any time soon. So what do I see? I am impressed by the rise of critique and contestation in the last 50 years. I grew up in the 1950s when critique was confined to a small coterie of intellectuals. The conformist fifties in the US were the background to Marcuse's notion of one-dimensionality. We have come a long way since then. It may be hard to grasp how far if you did not live in that earlier world. The emergence of widespread critical consciousness belies the silly notion that critique has 'run out of steam'. The reality is one of constant dissensus, both on the right and the left. This seems to lead in two incompatible directions: a new type of fascism based on hatred of 'elites', as a consequence of neo-liberalism; or a return of the welfare state with many new aspects concerning identity and the environment. This latter outcome is not yet socialism but it at least opens the future to human possibilities.

Of course the society in both cases will be organized by algorithmic governmentality, which has the potential to co-opt change and innovation. That was already Marcuse's idea of one-dimensionality. The situation today is less original than one might imagine. The important point is that contestation is far more widespread than in the past and that opens up unprecedented potentials.

Andrew Feenberg
Simon Fraser University
✉ feenberg@sfu.ca

Igor Pelgreffi
Università degli studi di Verona
✉ igor.pelgreffi@univr.it

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770082

Contributi/9

Governare il comportamento

Le radici comportamentiste del capitalismo digitale

Marco Dal Pozzolo 0009-0001-5022-717X

Articolo sottoposto a doppia blind peer review. Inviato il 23/12/2021. Accettato il 08/06/2022.

Governing behavior. The behaviorist roots of digital capitalism

The economy of digital capitalism is based on extraction, profiling and manipulation of behaviour. The logic of this system is not entirely new: it is based on the assumptions of behaviourism, particularly in its Skinnerian version. The economics of behavioural surplus is indebted to Skinnerian psychology both in terms of its philosophical approach and because it exploits behavioural technologies developed by the American scientist in his experiments. Interpreting digital capitalism in the light of radical behaviourism puts the notion of behaviour at the centre and provides a critical perspective on both epistemological assumptions and the pathologizing effects of algorithms.

***

1. Il comportamento come materia prima

Sono molti gli studi contemporanei che identificano nel tracciamento e negli strumenti di previsione gli elementi più potenti e innovativi della nuova economia digitale. L'oggetto di queste tecniche è il comportamento umano, la nuova materia prima da cui l'economia digitale estrae valore economico. Le prassi abitudinarie e gli schemi d'azione corporei sono ormai terreno da cui estrarre profitto, espandendo e raffinando la capacità di catturare le regolarità comportamentali. A livello di queste regolarità, che fluttuano tra la consapevolezza e la dimensione inconscia dell'azione soggettiva, si innesta una logica di tracciamento e sfruttamento dei dati. Google a partire dall'inizio degli anni 2000 ha aperto la strada a una nuova logica di sfruttamento dei dati comportamentali, logica che poi è diventata una strategia di profitto comune a molti attori sociali ed economici. Questa svolta deriva essenzialmente dall'intuizione di riutilizzare i dati di scarto prodotti dagli utenti durante le ricerche: se fino a un certo punto della sua storia Google aveva usato i dati per

rendere più efficienti i propri algoritmi, successivamente li utilizzerà anche come materia prima per elaborare strumenti predittivi sul comportamento. Si tratta di un passaggio cruciale per la nascita di quello che Zuboff chiama capitalismo della sorveglianza: dal 2002 Google ha individuato nell'advertising e nella vendita dei dati a terzi la vera fonte di guadagno e di espansione del capitalismo digitale, laddove prima gli algoritmi di Google consumavano internamente tutto il valore creato[1] per migliorare il motore di ricerca. Si è passati dal feedback chiuso del reinvestimento dei dati a un sistema aperto che coinvolge la cessione dei dati a terzi come fonte di profitto[2]. Questo modello è stato successivamente implementato da Facebook e da altri attori. Cruciale per inquadrare il problema è la nozione di surplus comportamentale, ovvero lo sfruttamento dell'eccedenza di dati finalizzata alla profilazione dei singoli individui, dei loro comportamenti e delle loro interazioni[3]. In altre parole i dati sono usati per affinare gli algoritmi dell'intelligenza artificiale che li estraggono dalla vita degli individui; ma servono anche e soprattutto a profilare le persone al fine di produrre predizioni sui comportamenti futuri. Pilastri di questa logica sono appunto l'estrazione dei dati e la profilazione individuale.

Portando la logica dell'estrazione al suo estremo si può sostenere che l'essere umano stesso sia diventato materia prima: le vite nella loro globalità, sorvegliate e monitorate continuativamente, sono diventate terreno di estrazione del surplus comportamentale, secondo una logica di progressiva colonizzazione di spazi che non erano ancora stati oggetto di cattura[4]. Il tracciamento automatico delle interazioni umane si è esteso fino alla penetrazione degli spazi intimi grazie all'Internet delle cose. L'operazione centrale di questa logica estrattiva è quella di renderizzare, ovvero trasformare una vasta gamma di comportamenti in dati, mirando a ridurre lo scarto tra esperienza e quantificazione. Una breve panoramica permette di constatare la moltiplicazione degli strumenti di cattura del comportamento. La renderizzazione del corpo è stata notevolmente ampliata grazie agli smartphone e i loro sistemi di localizzazione che costituiscono pattern di mobilità; ma si pensi anche ai dati biometrici prodotti grazie alle centinaia di migliaia di applicazioni dedicate alla salute che monitorano parametri fisiologici, attività fisica e qualità del sonno[5]. In molti casi queste tecnologie sono frutto di quella che Dominique Cardon definisce «quantificazione del sé», la tendenza crescente degli individui ad affidare agli oggetti tecnici la verifica della coerenza tra le proprie risoluzioni e il comportamento effettivo[6]; con l'automonitoraggio

[1] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, Milano 2019, p. 79.
[2] Cfr. ivi, pp. 313-314.
[3] Cfr. ivi, pp. 92-96.
[4] Cfr. N. Coudry and Mejias, *The Costs of connection*, Standford 2019 cap. 1. Gli autori, muovendo da una prospettiva marxista, intendono la colonizzazione non in senso metaforico, ma come analogia strutturale con il colonialismo storico. Sostengono inoltre che la natura umana sia diventata una forma di astrazione sociale nel capitalismo dei dati, esattamente come lo è il lavoro nel capitalismo industriale.
[5] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., pp. 256-368.
[6] Cfr. D. Cardon, *Che cosa sognano gli algoritmi*, Milano 2016, pp. 63-64.

gli individui introiettano un'attenzione ai dati che è economicamente produttiva e moltiplicano gli ambiti in cui la propria esistenza diventa flusso di tracce[7]. Gli orizzonti dell'estrazione sono stati poi stati ancora allargati dai sensori installati negli oggetti smart, compresi sensori e assistenti vocali che catturano la dimensione sonora dell'informazione (terreno su cui Google, Amazon e Microsoft si contendono il dominio)[8], tecnicamente grazie alla miniaturizzazione dell'elettronica che ha avuto una fortissima accelerazione negli ultimi decenni[9]. Entro i limiti strutturali del ragionamento analogico, il parallelo tra capitalismo digitale e colonizzazione storica degli spazi extraeuropei permette di evidenziare l'estensione progressiva del primo su zone sempre più vaste, ancora scarsamente normate e ricche di materia prima da estrarre, in questo caso i dati, e suscettibili di essere ulteriormente ampliate. Al di là di ogni ricavo commerciale che la vendita di queste tecnologie consente, esse sono essenzialmente finalizzate all'estrazione massiva di dati comportamentali, business fondamentale delle aziende che le producono[10].

L'estrazione è però solo un aspetto dell'economia del surplus comportamentale: è una premessa essenziale, ma non sufficiente se non fosse integrata dalla profilazione personale[11]. La profilazione è essenzialmente finalizzata alla previsione dei comportamenti futuri il cui mercato è al cuore del nuovo capitalismo digitale. Nella storia di Internet varie tecniche di calcolo si sono susseguite e intrecciate al fine di *organizzare* l'informazione, ma una in particolare è divenuta centrale nell'ultimo decennio, al punto da aver soppiantato o marginalizzato altre tecniche[12]. Questa è basata sulla registrazione delle tracce degli utenti e sulla costruzione di un profilo comportamentale del singolo, grazie all'uso di algoritmi ad apprendimento automatico (*machine learning*). Una profilazione progressivamente più raffinata si basa spesso sull'associare l'utente ad altri individui che hanno avuto comportamenti analoghi o che hanno compiuti scelte simili: la previsione del comportamento di un utente è fondata essenzialmente sul passato degli individui che gli somigliano[13]. Questa logica, ad esempio, è concretizzata chiaramente dagli algoritmi di Netflix e Amazon, sfruttando la proprietà di co-occorrenza[14]. Anche Facebook e Google modellano un profilo personale dell'individuo, sulla base delle preferenze, delle interazioni,

[7] Cfr. F. Bocci, U. Zona, *La rete come una Skinner Box. Neocomportamentismo, bolle sociali e post-verità*, «Media education», vol. 9, 1, 2018, pp. 69-71.
[8] Cfr. N. Coudry and Mejias, *The Costs of connection*, cit., pp. 47-48; S. Zuboff, *Il Capitalismo della sorveglianza*, cit., pp. 269-284.
[9] Cfr. S. Quintarelli, *Capitalismo immateriale*, Torino 2019, pp. 43-44.
[10] Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, Milano 2018, pp. 14-17.
[11] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., p. 214.
[12] Per un approfondimento di queste tecniche, che non è possibile delineare in questa sede, Cfr. D. Cardon, *Che cosa sognano gli algoritmi*, cit., pp. 11-25.
[13] Cfr. ivi, pp. 25-26.
[14] Cfr. L. Laura, *Storia degli algoritmi*, Roma 2019, pp. 61-68. L'approccio che cerca di prevedere i comportamenti dell'individuo sulla base delle scelte passate di altri utenti è definito collaborativo. Questo approccio non esclude approcci basati invece sulla similitudine degli oggetti acquistati; anzi, spesso si le due strategie si integrano.

dei contenuti consultati, catturando le regolarità del comportamento. In particolare per Google, che monetizza solo se l'utente visualizza una determinata pagina (per esempio un contenuto pubblicitario), è essenziale riuscire a stimare accuratamente il comportamento futuro della persona[15]. Questo sistema si regge sugli effetti di rete prodotti dal tracciamento che non si limita a catturare l'informazione per fini interni, ma autorizza anche siti affiliati a farne uso. I dati ricavati dai cookies di un certo sito inviano informazioni a decine di server, rendendo più agevole l'operazione di profilazione da parte delle aziende[16]. Per comprendere la logica della profilazione è rilevante la distinzione tra segnali e tracce: i primi sono i contenuti, le ricerche, le interazioni degli utenti; i secondi sono registrazioni contestuali di azioni (geolocalizzazione, velocità di lettura etc.). Le piattaforme, che hanno a disposizioni grandi masse di informazione, incrociano molto spesso queste due tipologie di dati per produrre previsioni più precise[17]. Il geotargeting ha permesso un salto di scala, in particolare a Google, consentendo di associare il monitoraggio del comportamento sul web a quello nel mondo fisico. La geolocalizzazione non solo ha migliorato le previsioni, ma ha aperto nuove possibilità di manipolazione del comportamento, come si vedrà meglio in seguito[18]. La profilazione comprende infine anche l'affective computing, ovvero la capacità degli algoritmi di ipotizzare lo stato emotivo della persona incrociando comportamento non verbale, espressioni facciali e segnali vocali. Grazie all'utilizzo di modelli psicologici il comportamento non verbale e le micro-espressioni assumono un valore predittivo sullo stato emotivo dell'individuo, una delle frontiere su cui l'intelligenza artificiale può espandere ancora il territorio dell'estrazione[19]. Ma anche le emozioni sono in realtà inferenze statistiche costruite a partire dalla materia prima che regge tutta la struttura: il comportamento. Tutti i dati, anche i più raffinati e complessi, si riferiscono a un'analisi comportamentale.

Da tutto ciò emerge chiaramente l'assoluta centralità del concetto di comportamento: si tratta di un concetto non univoco, polisemico, di cui il capitalismo digitale si serve con un'accezione specifica e orientata. Certamente la nozione di comportamento integra la questione dell'automatismo, degli schemi corporei appresi e delle prassi incorporate; non si dà, in altre parole, comportamento senza disposizioni apprese e consolidate nella ripetizione[20]. Ma l'interpretazione di questo sistema di disposizioni non è neutra e può essere declinata in termini divergenti: il concetto di comportamento può rimandare a

[15] Cfr. ivi, pp. 69-73.
[16] Cfr. D. Cardon, *Che cosa sognano gli algoritmi*, cit., pp. 28-30; cfr. L. Laura, *Storia degli algoritmi*, cit., pp. 73-75.
[17] Cfr. ivi, pp. 51-59.
[18] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., pp. 256-262. L'impossibilità di renderli anonimi contribuisce alla dimensione strategica dei location data.
[19] Cfr. ivi, pp. 297-305. Si pensi, per esempio, allo schema FACS costruito per inferire le emozioni a partire dalla misurazione delle espressioni facciali.
[20] Per un approfondimento teoretico di questi temi, Cfr. I. Pelgreffi, *Filosofia dell'automatismo*, Napoli-Salerno 2018, in particolare capitoli 4 e 5.

una forma o una struttura, una totalità dinamica dotata di senso e in relazione dialettica con il proprio ambiente, come vedremo nell'ultimo paragrafo. Sotteso all'economia del surplus comportamentale vi è invece un modello sostanzialmente diverso: il comportamento è una concatenazione di gesti, puntualmente rinforzati dall'ambiente nel corso del tempo e assolutamente determinati dall'ambiente stesso[21]. Si tratta dell'interpretazione fornita dal comportamentismo radicale, in particolare nella sua declinazione skinneriana che, come vedremo, ha profonde implicazioni con il capitalismo digitale. Proprio sulla base di questa concezione, e come conseguenza di un'azione sostanzialmente unilaterale, parcellizzata e deterministica dell'ambiente sul comportamento, dobbiamo aggiungere la manipolazione quale terzo imperativo dell'economia del surplus comportamentale. Al di là del concreto interesse economico a modificare il comportamento al fine di produrre regolarità prevedibili, la manipolazione del comportamento è una dimensione strutturale del nuovo capitalismo digitale: dal momento che l'ambiente di vita dell'individuo contemporaneo è perennemente trasformato dall'interazione con gli algoritmi, seguendo una prospettiva comportamentista radicale anche il comportamento muterà significativamente in risposta. Assistiamo dunque a processi che implicano la modificazione del comportamento su larga scala e configurano anche una sorta di sperimentazione a cielo aperto sugli individui[22]. Dove passa la linea di confine tra estrazione di un comportamento dato e cattura di un comportamento prodotto? Quali sono gli effetti di queste operazioni sui corpi e sulle vite? Prima di rispondere a queste domande è necessario approfondire il modello teorico che governa l'uso degli algoritmi nella nuova economia digitale, ovvero il comportamentismo radicale skinneriano.

2. Algoritmi e comportamentismo radicale

«Siamo nelle mani della macchina algoritmica che ci vuole come piccioni nella gabbia»[23], «oggi la rete è una gigantesca *Skinner box*, governata da algoritmi che fagocitano senza sosta una mole impressionante di dati»[24], «Skinner ebbe l'amara sorte di prevedere, prima ancora che esistessero gli strumenti necessari a realizzare la sua visione, sia la possibilità di un potere strumentalizzante, sia il Grande Altro in grado di renderlo operativo»[25], «Now what I want to argue here is that modern workers, and indeed all of us, in an environment suffused

[21] Cfr. R. Boakes, *Da Darwin al comportamentismo*, a cura di Marco Poli, Milano 1986, pp. 376-383 e pp. 528-531.
[22] Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit., p.19; Cfr. F. Bocci, U. Zona, *La rete come una Skinner Box. Neocomportamentismo, bolle sociali e post-verità*, cit., pp. 59-61.
[23] C. Mazzucchelli, *Prigionieri della macchine algoritmica siamo come piccioni dentro alla gabbia di Skinner* (blog «Solotablet», 20.12.2020).
[24] F. Bocci, U. Zona, *La rete come una Skinner Box. Neocomportamentismo, bolle sociali e post-verità*, cit., p. 65.
[25] S. Zuboff, *Il Capitalismo della sorveglianza*, cit., pp. 396-397.

with AI, can end up sharing the predicament of Skinner's pigeons»[26], «We're entering the age of Skinnerian marketing»[27]. Queste sono solo alcune citazioni di intellettuali, ricercatori o giornalisti che associano Skinner al capitalismo digitale. Si tratta di un riferimento ricorrente, che tuttavia rimane in molti casi un'associazione vaga, anche se suggestiva. Quest'associazione merita invece di essere approfondita perché fornisce profondità teorica alla logica economica del surplus comportamentale e ne è una delle matrici più significative. L'approccio skinneriano ispira il nuovo capitalismo digitale almeno su due piani: il primo è legato al concetto di tecnologia comportamentale, come orizzonte teorico generale e in quanto fondamento di un progetto sociale; il secondo si riferisce a specifici meccanismi psicologici scoperti da Skinner che gli algoritmi sfruttano estesamente, in particolare il concetto di condizionamento operante e di rinforzo a intervalli variabili. Per entrambi i piani è possibile evidenziare applicazioni del capitalismo digitale che si dimostrano in evidente continuità con la concezione skinneriana.

Innanzitutto bisogna inquadrare quella skinneriana come una delle versioni all'interno della galassia comportamentista, celebre movimento scientifico a cui ha dato inizio John Watson negli anni dieci del Novecento. I lavori di Watson sono uno dei punti di riferimento della ricerca di Skinner, soprattutto per aver posto il comportamento come unico oggetto osservabile e misurabile da parte della psicologia scientifica. Altri riferimenti imprescindibili dello psicologo americano sono le ricerche sperimentali di Edward Thorndike (primo a studiare gli agenti di rinforzo nell'apprendimento) e il concetto di condizionamento elaborato dal fisiologo russo Ivan Pavlov. Negli anni trenta, a partire dalla confluenza di questi riferimenti, Skinner formulerà un suo protocollo di ricerca e il concetto di condizionamento attivo, o condizionamento operante, che è il suo apporto più originale e duraturo alla psicologia comportamentista[28]. La versione di Skinner è definita comportamentismo radicale perché si tratta di un'interpretazione rigida e rigorosa dei principi chiave sui quali si è sviluppata la psicologia comportamentista; si differenzia infatti dal lavoro di Watson essenzialmente sul piano delle innovazioni sperimentali e dell'ambizione a strutturare un progetto sociale, come vedremo, più che sul piano metodologico. Si poneva tra l'altro come una reazione nei confronti di altre interpretazioni del comportamentismo che assumevano elementi finalistici nella spiegazione del comportamento e riconoscevano una dimensione di intenzionalità e significazione ai soggetti agenti[29]. Durante la sua lunga carriera Skinner riuscì a imporre decisamente la sua prospettiva, sia grazie ai successi sperimentali, sia perché trasformò la

[26] J. Dahaer, *Escaping Skinner's Box: AI and the Era of techno-superstition*, «Philosophical Disquisition», 15.10.2019.
[27] B. Davidow, *Skinner Marketing: we're the Rats and Facebook limes are the Reward,* «The Athlantic», 10.6.2013.
[28] Cfr. R. Thompson, *Storia della psicologia*, Torino 1972, pp. 242-246.
[29] Cfr., ivi, pp. 246-250; cfr. R. Boakes*, Da Darwin al comportamentismo*, cit., pp. 524-531. In particolare Skinner si contrapponeva alla versione eterodossa del comportamentismo fornita da Tolman.

sua metodologia in un progetto filosofico e sociale, al punto che nel 1971 il Time lo definì il più influente psicologo vivente[30]. È impossibile sopravvalutare l'influenza di Skinner come scienziato e intellettuale, sopravvissuta alla decadenza del comportamentismo a partire dagli anni settanta del secolo scorso; lo psicologo americano è infatti tuttora centrale nella ricerca neuroscientifica contemporanea ed è presente nel discorso pubblico come il fantasma mai tramontato di una teoria totale dell'umano. Questa teoria è innanzitutto una concezione del comportamento fondata sul nesso stimolo-risposta, secondo un modello antropologico radicalmente riduzionista: l'integralità dell'umano è spiegabile in termini di risposta a stimoli esterni e persino pensiero e linguaggio sono interpretati integralmente in questo paradigma[31]. Il rapporto tra ambiente, composto essenzialmente di stimoli parcellizzati, e individuo è di causalità unidirezionale; l'ambiente condiziona e seleziona il comportamento, sia prima sia dopo le risposte date dall'organismo[32]. Questo presupposto consente a Skinner di teorizzare una sostanziale omologia tra ambiente naturale-sociale e laboratorio: «La differenza tra situazione di laboratorio e condizioni naturali non è una differenza importante»[33]. Secondo lo psicologo americano, come accade in laboratorio, l'ambiente sociale può essere manipolato e tramite esso il comportamento può essere modificato intenzionalmente[34]. Dal momento che «il comportamento può essere modificato mutando le condizioni di cui è una funzione», si apre un vasto campo di applicazione per le tecnologie comportamentali, come le chiama lo stesso Skinner, ovvero tecniche concepite per progettare comportamenti, supposte eticamente neutre[35]. Su questo piano Skinner si spinge al di là dei confini della psicologia comportamentista e teorizza una prospettiva politica fondata sulle tecnologie di controllo del comportamento: scompare la partizione rigida tra scienza e tecnica e si perde il confine tra scienza di base e scienza applicata, perché l'una non è isolabile dall'altra[36]. Questa prospettiva politico-sociale non può essere semplicemente ridotta a una fantasia di stampo totalitario: anzi, nelle intenzioni dell'autore un'ampia esplorazione delle tecniche di controllo del comportamento sarebbe un'arma in mano alle democrazie per prevenire i dispotismi e le contingenze aleatorie della storia[37]. Ma soprattutto si tratta di un altro tipo di potere perché esso non è interessato a imporre principi o a trasformare le anime, bensì a modificare, predire e controllare il comportamento governando le variabili che lo circondano[38]. Si tratta di una differenza sostanziale, perché il controllo esercitato dalle tecnologie del comportamento agisce

[30] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., p. 378.
[31] Cfr. B. F. Skinner e W. Correll, *Pensare ed apprendere*, Roma 1974, pp. 20-25. Il comportamento è per Skinner autonomo anche rispetto alla fisiologia, almeno sul piano epistemologico.
[32] Cfr. B. F. Skinner, *Difesa del comportamentismo*, Roma 2006, pp. 84-85.
[33] B. F. Skinner, *Oltre la libertà e la dignità*, Milano 1973, p. 185.
[34] Cfr. ivi, p. 31.
[35] Ivi, pp. 175-176.
[36] Cfr. B. F. Skinner, *Studi e ricerche*, Firenze 1972, pp. 334-337.
[37] Cfr. ivi, pp. 69-70.
[38] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., pp. 370-371 e 377.

essenzialmente al di sotto della consapevolezza dell'individuo ed è indifferente alla dimensione morale e ideologica del convincimento. Con la sua proposta Skinner non guarda tanto al suo Novecento, ma piuttosto immagina una società a venire in cui previsione e manipolazione del comportamento saranno, a suo parere, strumenti di progresso. Questa visione ha alcuni corollari rilevanti che lo pongono in forte discontinuità con l'impianto politico-morale prevalente della modernità: libertà e autonomia sono pure illusioni, frutto della cecità rispetto ai fattori determinanti dell'ambiente[39]; i giudizi e i valori non sono altro che espressioni di effetti di rinforzo, dal momento che bene e male si distinguono solo sulla base delle classi di risposte incorporate[40]; infine vi è nei testi skinneriani una critica ripetuta a punizione e repressione come modelli di apprendimento, poiché stimolano reazioni avversive, di aggressività e di evitamento, e sono meno efficaci rispetto ai condizionamenti indotti dalla manipolazione dell'ambiente (rispetto ai quali è molto più difficile esercitare resistenza)[41].

Una volta delineati i presupposti del comportamentismo radicale, appare forse più chiaro perché questa teoria può essere considerata fondamento al contempo filosofico, psicologico e antropologico del nuovo capitalismo digitale. L'idea di una società come laboratorio a cielo aperto su cui sperimentare ha in Skinner il suo principale teorico e nell'economia del surplus comportamentale la sua applicazione più sviluppata[42]. Un esempio di tecnologia del comportamento particolarmente impressionante è la sperimentazione compiuta da Google tramite l'applicazione Pokémon Go: si tratta di un gioco nel quale gli utenti si muovono nel mondo fisico alla ricerca dei Pokémon (personaggi di un noto cartone animato giapponese) che possono catturare con il proprio smartphone, una volta raggiunto il luogo fisico in cui la mappa indica la presenza del Pokémon. L'obiettivo del gioco è catturarne il più possibile per poi farli combattere contro i Pokémon di altri giocatori. Grazie a questo gioco l'azienda californiana ha potuto da un lato estendere la mappatura di Google Maps a zone non ancora renderizzate grazie alle immagini captate dalla fotocamera degli smartphone dei giocatori (violando probabilmente le normative sulla privacy); dall'altro ha venduto a esercizi commerciali il diritto di avere esche virtuali, i Pokémon, con cui attrarre i giocatori nel proprio locale, in modo che consumino o acquistino i loro prodotti, una volta sul luogo[43]. Con Pokémon Go, Google ha sperimentato come muovere le persone orientandone le motivazioni: il gioco

[39] Cfr. B. F. Skinner, *Oltre la libertà e la dignità*, cit., pp. 46-51.
[40] Cfr. ivi, p. 125.
[41] Cfr. B. F. Skinner, *Studi e ricerche*, cit.; B. F. Skinner, *Oltre la libertà e la dignità*, cit., pp. 27-28 e pp. 78-81. Le tecnologie comportamentali skinneriane sembrano ricalcare i tratti principali del potere positivo e produttivo della biopolitica foucaultiana, teorizzato per esempio nella lezione finale di *Bisogna difendere la società*: come nella teoria foucaultiana, Skinner rifiuta le tecniche di potere essenzialmente repressive. Cfr. M. Foucault, *Bisogna difendere la società*, Milano 2009, pp. 206-234.
[42] Cfr. F. Bocci, U. Zona, *La rete come una Skinner Box. Neocomportamentismo, bolle sociali e post-verità*, cit., p. 60; cfr. D. Cardon, *Che cosa sognano gli algoritmi*, cit., p. 51.
[43] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., pp. 325-336.

fornisce incentivi a mettere in atto determinati comportamenti (in questo caso recarsi in uno specifico luogo) nel contesto di una competizione collettiva; i risultati di questa capacità di manipolazione del comportamento possono essere venduti a terzi. Si tratta di un caso paradigmatico di modificazione del comportamento, proprio nell'accezione in cui la intendeva Skinner: se in molti casi la manipolazione del comportamento avviene esclusivamente online, in questo e in altri casi, grazie al geotargeting, la riconfigurazione dell'ambiente esercita un potere di indirizzo direttamente nel mondo reale. È solo un esempio emblematico dell'influenza che possono avere queste tecniche sull'orientamento dei comportamenti. Esattamente come negli esperimenti proposti da Skinner, la ricerca di una ricompensa attiva e rinforza un certo comportamento, aggirando la consapevolezza dell'individuo. Non a caso Zuboff definisce lo scienziato computazionale Alex Pentland, ex consigliere di Google e del World Economic Forum, lo «Skinner del ventunesimo secolo», proprio perché ha esteso la portata delle tecnologie comportamentali grazie al potere degli algoritmi[44]. È facile constatare le stessa mire di controllo del comportamento tra l'utopia di Skinner e le operazioni delle grandi piattaforme, seppure solo per queste ultime la finalità sia primariamente il profitto; d'altronde se l'obiettivo è quello di profilare e vendere strumenti di previsione, niente assicura un risultato migliore della capacità di condizionare massicciamente il comportamento nelle direzioni desiderate. A proposito di questo, va peraltro constatata un'ulteriore analogia con la metodologia skinneriana. Infatti lo psicologo americano ha sempre insistito sullo studio sperimentale del singolo organismo, perché in questo tipo di ricerche si manifesterebbero fattori e tendenze che negli studi basati interamente su campioni statistici rimarrebbero celati[45]. La profilazione individuale del nuovo capitalismo digitale riporta in qualche modo al centro l'idea di modificare e prevedere il comportamento del singolo individuo, mettendo in primo piano le tendenze proprie del singolo, mentre nelle tecniche di ricerca delle scienze sociali queste sono storicamente sempre risultate marginali[46]. Quest'ulteriore analogia rafforza l'idea del comportamentismo radicale come referente teorico privilegiato per interpretare la logica dell'economia digitale[47].

Oltre alla dimensione più propriamente teorica del comportamentismo radicale, anche gli aspetti sperimentali del lavoro di Skinner sono interessanti nell'ottica di studiarne le influenze sul nuovo capitalismo digitale. È istruttivo in questo senso descrivere le pratiche sperimentali di Skinner, il concetto di condizionamento operante e le scoperte correlate ad esso, per mostrarne le relazioni con la governamentalità algoritmica. Un ruolo centrale è occupato

[44] Cfr. ivi, pp. 433-438.
[45] Cfr. B. F. Skinner, *Studi e ricerche*, cit., pp. 12-13.
[46] Cfr. D. Cardon, *Che cosa sognano gli algoritmi*, cit., pp. 54-57.
[47] Ovviamente tenendo in debito conto i limiti di ogni confronto. È evidente che le tecniche statistiche sono essenziali per il capitalismo digitale, a partire dagli algoritmi stessi: l'attenzione alla manipolazione del comportamento individuale si colloca in un contesto ben diverso da quello del laboratorio skinneriano.

dalla famosa gabbia di Skinner, un dispositivo sperimentale inventato dallo psicologo comportamentista e divenuto poi celebre. Si tratta di uno spazio ristretto e chiuso (isolato acusticamente) in cui è costretto un organismo modello animale, tipicamente un topo o un piccione. All'interno della gabbia è presente un manipolandum, una leva o un pulsante, che produce eventi di rinforzo nel momento in cui l'animale entra in contatto con esso (un esempio di rinforzo è il cibo ad esempio, se l'organismo ne è privato). Di solito è presente anche uno stimolo di discriminazione (un suono, una luce), che funge da *starter* dell'esperienza, e un quadro elettrico per controllare gli stimoli[48]. L'obiettivo è di condizionare il comportamento dell'animale o inducendo in lui l'associazione tra l'interazione con il manipolandum e la ricompensa, oppure facendo in modo che associ la ricompensa a un determinato comportamento: per esempio ogniqualvolta il piccione compie una certa rotazione lo sperimentatore può fornirgli una ricompensa, in modo da innescare una risposta indotta (o più precisamente una classe di risposte) da parte dell'animale, grazie alla ripetizione degli eventi di rinforzo. Con un oscillatore lo sperimentatore misura la frequenza di risposta, la variabile essenziale dell'esperimento che identifica la capacità di apprendimento dell'organismo[49]. Alla base di questo modello di esperimento c'è il concetto di condizionamento operante che differisce sostanzialmente dal condizionamento classico di tipo pavloviano. Infatti se il condizionamento pavloviano identifica una risposta fisiologica allo stimolo (condizionato e incondizionato), rispetto alla quale l'individuo rimane sostanzialmente passivo, il condizionamento operante prevede sempre un'azione da parte del soggetto, un'azione che sarà oggetto del rafforzamento. In altre parole nel modello skinneriano non si tratta semplicemente di associare una risposta ad uno stimolo (come nei celebri esperimenti dello scienziato russo); si tratta piuttosto di plasmare un comportamento incentivando alcune risposte e inibendone altre[50]. Questo aspetto è centrale nel modo in cui gli algoritmi agiscono, perché è proprio nel quadro della manipolazione di un comportamento in azione che essi esercitano potere. Secondo Skinner i rafforzatori del comportamento sono di vario tipo: quelli di cui si è detto finora sono rinforzi positivi, ma ve ne sono anche di negativi, che inducono cioè una risposta per deprivazione (di luce, ad esempio)[51]. I rinforzi si distinguono poi tra primari e generalizzati: i primari rafforzano un'occasione specifica (cibo su un affamato, acqua su un assetato…); i rinforzi generalizzati non dipendono invece dalla condizione immediata. Per esempio il denaro è un rinforzo generalizzato; anche la ricerca di attenzione e di approvazione lo sono e questo ha direttamente a che fare con la logica del

[48] Cfr. B. F. Skinner, *Studi e ricerche*, cit., pp. 9-10.
[49] Cfr. ivi, pp. 10 e B. F. Skinner, *Scienza e comportamento*, Milano 1971, pp. 70-78.
[50] Per una più ampia analisi di Skinner sulle differenze tra condizionamento operante e condizionamento pavloviano, cfr. B. F. Skinner, *Studi e ricerche*, cit., pp. 639-643; Id., *Scienza e comportamento*, cit., pp. 91-93; cfr. anche I. P. Pavlov, *I riflessi condizionati*, Torino 2020, pp. 222-244.
[51] Cfr. ivi, pp. 100-101.

capitalismo digitale[52]. Al di là delle necessarie distinzioni, l'aspetto essenziale è che il rafforzamento avvenga successivamente alla risposta del soggetto, che sia ripetuto nel tempo e che il comportamento sia concepito come funzione degli eventi di rinforzo[53].

La scoperta del condizionamento operante è l'architrave attorno alla quale ruota tutta la teoria skinneriana. Non sempre però si analizza la finezza delle diverse tipologie di rinforzo in funzione di una variabile cruciale, anche per i meccanismi indotti dagli algoritmi: l'estinzione. L'estinzione misura la resistenza della risposta appresa in mancanza di rinforzo ed esprime quindi il radicamento della risposta stessa nel comportamento. Il valore dell'estinzione dipende da una serie di variabili, come il numero di rinforzi, la motivazione dell'individuo, lo sforzo per emettere la risposta (per esempio la difficoltà di interagire con il manipolandum)[54]. Ma dipende anche dal tipo di programma adottato nella fase di apprendimento e questo è sicuramente uno degli aspetti più innovativi della ricerca di Skinner. La distinzione essenziale tra i programmi di apprendimento è fra i programmi continui, che rinforzano la risposta ogniqualvolta venga emessa, e programmi intermittenti di rinforzo, in cui la risposta è rinforzata solo in alcune occasioni, in modo intermittente appunto. A loro volta i programmi intermittenti si possono distinguere tra questi i programmi ad intervallo fisso (in cui il tempo che intercorre tra un rinforzo e l'altro è sempre identico) e i programmi a intervallo variabile, in cui gli intervalli variano con un margine di aleatorietà, in genere attorno a un valore medio[55]. Su questo aspetto si innesta la scoperta più innovativa e, in un certo senso, controintutiva: è più difficile per lo sperimentatore estinguere una risposta condizionata da un programma intermittente e a intervalli variabili, piuttosto che una risposta appresa da un programma continuo. Skinner ha constatato che, sottoponendo un piccione a un programma di rinforzo intermittente, sono stati necessari fino a diecimila risposte prive di rinforzo per ottenere l'estinzione. Il condizionamento prodotto da rinforzo intermittente necessita di più tempo per essere incorporato dall'individuo, ma una volta integrato nel comportamento perdura più a lungo, dimostrandosi più resistente[56]. Anche in situazioni in cui ci sia un certo grado di casualità nelle occorrenze degli eventi rinforzanti si danno condizionamenti stabili, accompagnati in alcuni casi da comportamenti superstiziosi che hanno interessato Skinner[57]. L'approccio metodologico di Skinner non contempla l'attribuzione di una causalità specifica a questo comportamento: lo scienziato americano si limita a registrare e studiare nel dettaglio questo tipo di comportamenti, certificando che la variabilità, lungi dal costituire un

[52] Cfr. ivi, pp. 103-109.
[53] Cfr. B. F. Skinner, *Studi e ricerche*, cit., pp. 572-574.
[54] Cfr. ivi, pp. 25-26.
[55] Cfr. ivi, pp. 19-20.
[56] Cfr. B. F. Skinner, *Scienza e comportamento*, cit., pp. 107-108 e pp. 128-132.
[57] Sul tema Cfr. B. F. Skinner, *Studi e ricerche*, cit., pp. 577-580. Cfr. anche J. Dahaer, *Escaping Skinner's Box: AI and the Era of techno-superstition*, cit. Dahaer parla significativamente di *techno-superstition*, ma non è possibile in questa sede sviluppare questo punto.

ostacolo all'apprendimento, può moltiplicare la forza del condizionamento. Con queste ricerche il comportamentismo radicale skinneriano fa emergere alcune vulnerabilità del comportamento, umano e animale, sulle quali tecniche di manipolazione del comportamento possono fare leva[58].

A questo punto una serie di rapporti tra le tecniche sperimentali di Skinner e le tecniche di estrazione-profilazione permesse dagli algoritmi è intuibile, ma è importante in questa sede esplicitare e prolungare queste implicazioni. Il condizionamento operante è lo schema di fondo con cui agiscono le tecniche di modifica del comportamento. Come detto, la modificazione del comportamento secondo regolarità, possibile tramite la ripetizione di schemi comportamentali indotti dagli algoritmi, è pienamente funzionale ad elaborare previsioni più raffinate. Indubbiamente la logica alla base del nuovo capitalismo digitale (estrazione, profilazione, associazioni tra utenti simili…) implementa gli aspetti chiave del modello skinneriano: innanzitutto opera inducendo forme di automatismo aggirando la consapevolezza dell'individuo[59]; in secondo luogo agisce tramite stimolazioni ripetute in modo da indurre un certo comportamento nell'individuo, rinforzandone le inclinazioni; infine produce condizionamenti inducendo un comportamento attivo nell'individuo, come accade nel condizionamento operante. Questi aspetti certificano come l'economia del surplus comportamentale operi sullo stesso piano del comportamentismo radicale. Si pensi poi al già citato ruolo dell'attenzione nella psicologia di Skinner, che è un esempio di rinforzo generalizzato, ma svolge in realtà un ruolo ancora più strategico: «L'attenzione, perciò come forma di comportamento incoativa o preliminare è una specie di autocontrollo, che rende possibile reagire ad uno stimolo in modo tale che le successive forme di comportamento risultino con molta probabilità gratificate»[60]. L'attenzione, secondo Skinner, contribuisce a strutturare gli altri comportamenti ed è in un certo senso precondizione per operare modifiche comportamentali. Il ruolo centrale della cattura dell'attenzione nell'economia digitale è forse talmente evidente da non aver bisogno di segnalazione e la letteratura su questo tema è ampissima[61]. Anche sotto questo aspetto, in ogni caso, il comportamentismo radicale risulta essere il modello psicologico più coerente per inquadrare le tecniche del nuovo capitalismo digitale.

Se estendiamo le implicazioni anche alla questione del rinforzo a intervalli variabili, allora la relazione si fa ancora più stretta e definita. Come sottolinea Tristan Harris, in passato esperto di etica del design in Google e protagonista del documentario *The social Dilemma*, gli algoritmi agiscono sfruttando varie tipologie di vulnerabilità psicologica per aumentare la presa sul comportamento (indirizzare implicitamente le scelte, sfruttare il bisogno di reciprocità e di approvazione sociale, programmare interruzioni ad esempio); fra queste occupa

[58] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., p. 395.
[59] Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit., p. 20.
[60] B. F. Skinner e W. Correll, *Pensare ed apprendere*, cit., p. 27.
[61] Sull'economia dell'attenzione cfr. Y. Citton, *Pour une écologie de l'attention*, Paris 2014.

un posto di rilievo la progettazione di ricompense a intervalli intermittenti e variabili[62]. Gli algoritmi delle piattaforme sfruttano il fatto che ricompense a intervalli variabili producono più additività, ovvero risposte condizionate più stabili, per indirizzare il comportamento degli utenti. Si pensi, per esempio, ai social media e al ruolo delle notifiche nello stimolare un uso compulsivo della piattaforma: tempi, modalità, selezione delle interazioni sono continuamente ricalibrati dagli algoritmi dell'intelligenza artificiale sulla base dei dati ottenuti sull'utente. Molte piattaforme progettano probabilmente una variazione intenzionale di vari parametri con cui vengono presentate le notifiche, al fine di potenziare il condizionamento operante. Ma la variabilità degli intervalli è anche strutturale a causa della stessa conformazione degli algoritmi: trattandosi di algoritmi adattativi, eseguono occasionalmente test automatici per constatare le risposte ed applicare eventualmente ad altri utenti le stesse variabili. Questo produce una strutturale aleatorietà degli stimoli che è funzionale a un condizionamento più efficace; in altre parole gli algoritmi adattativi agiscono con un margine di sperimentazione intrinseca e d'altronde non smettono mai di replicare i propri schemi per adattarsi sempre meglio agli utenti[63]. Si tratta dello stesso tipo di meccanismo psicologico sfruttato dal gioco d'azzardo e, analogamente ad esso, l'obiettivo è vincolare gli utenti il più a lungo possibile sulle piattaforme, stimolando comportamenti additivi[64]. In questo senso le similitudini tra il comportamento del piccione negli esperimenti di Skinner e quello di un utente di un social media sono più profonde di quello che si potrebbe sospettare: su entrambi i soggetti la variabilità degli stimoli agisce come un moltiplicatore nel condizionare il comportamento.

Dal punto di vista neurobiologico la dopamina, neurotrasmettitore legato a uno dei più importanti sistemi a proiezione diffusa del cervello, gioca un ruolo strategico, confermando le intuizioni di Skinner[65]. Sean Parker, membro del team di sviluppo di Facebook, ha ammesso che il loro obiettivo era quello di indirizzare una piccola scarica di dopamina di tanto in tanto nell'utente, ovvero un rinforzo variabile[66]; anche Chamath Paluhapitiya, in passato vicepresidente di Facebook, ha parlato di «circoli viziosi di feedback a breve termine alimentati dalla dopamina»[67] e d'altronde l'analogia con i meccanisimi psicologici del gioco d'azzardo, usata anche da Harris, fa implicitamente riferimento al circuito dopaminergico[68]. È ormai appurato dalla letteratura contemporanea che il

[62] Cfr. T. Harris, *How technology is Hijacking your mind – from a magician and Google design ethicist*, «Thrive Global», 18.5. 2016.
[63] Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit., cap. 1.
[64] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., p. 466; Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit, cap. 1; Cfr. T. Harris, *How technology is Hijacking your mind – from a magician and Google design ethicist*, cit.
[65] È bene ricordare che questo livello di spiegazione del fenomeno è del tutto estraneo alla metodologia di Skinner e va sottolineato che la gran parte delle sue ricerche sono precedenti a questi studi nell'ambito delle neuroscienze.
[66] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., p. 467.
[67] J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit., cap. 1.
[68] Cfr. B. Davidow, *Skinner Marketing: we're the Rats and Facebook limes are the Reward,* cit.

sistema dopaminergico svolge un ruolo chiave nella ricerca della ricompensa, più ancora che nella componente edonistica (correlata maggiormente con altri tipi di molecole come oppiacei endogeni ed endocannabinoidi). In sostanza si assiste a una grande stimolazione del circuito dopaminergico nell'atto della ricerca di una ricompensa, in anticipo rispetto all'attivazione dei neuroni motori: il processo è del tutto compatibile con il comportamento prodotto dal condizionamento operante skinneriano sui suoi organismi modello, indotti a ricercare i pattern comportamentali che garantiscono loro le ricompense. Le ricerche neuroscientifiche degli ultimi decenni hanno verificato che il sistema dopaminergico si attiva con maggiore intensità proprio in corrispondenza di stimoli a intervallo intermittente, fornendo di fatto una base neurobiologica ai risultati di Skinner[69]. Si ipotizza che il soggetto sia maggiormente stimolato ad attivare i neuroni motivazionali alla ricerca di uno schema regolare proprio laddove non riesce a individuarlo a causa di eventi aleatori. Come nel caso delle slot machine, si sviluppa una compulsione a ricercare la ricompensa (per esempio un like), alimentata dall'incertezza[70]. Questo breve riferimento ai correlati neurobiologici del comportamento motivazionale dimostra ulteriormente la rilevanza contemporanea del comportamentismo skinneriano (tanto che lo studio dei meccanismi neuronali correlati all'apprendimento skinneriano è oggi un'importante branca dalle neuroscienze), oltre a mostrare un altro piano di connessione tra esso e le tecniche dell'economia digitale. Ma questi riferimenti aprono prospettive particolarmente interessanti anche perché la dopamina è un neurotrasmettitore fortemente implicato nei fenomeni di dipendenza e abuso e, come abbiamo visto, l'economia del surplus comportamentale produce strutturalmente nuove forme di additività. Il tema delle dipendenze e degli abusi è una problematica rilevante che concerne le conseguenze di determinate tecnologie comportamentali; costituisce una questione cruciale da sviluppare, non senza averla prima inquadrata in una prospettiva critica più ampia e strutturata dei presupposti comportamentisti.

3. Prospettive critiche

Una volta mostrate le profonde implicazioni sussistenti tra il pensiero di Skinner e il nuovo capitalismo digitale è possibile proporre alcune prospettive critiche che mirano ai presupposti comuni di questi due sistemi. Si tratta di cercare chiavi di lettura critiche che, dal punto di vista teorico e operativo, si situino al livello del comportamento e della corporeità, accettando la sfida sullo stesso piano in cui si declinano i poteri finora descritti. Due proposte di critica

[69] Per una panoramica sulle funzioni della dopamina cfr. R. Kong, *Dopamine the action of life*, Londra 2021. Non si deve inferire una relazione di causalità tra il sistema dopaminergico e i condizionamenti del comportamento studiati da Skinner. Si tratta di correlati neuronali e l'identificazione di un nesso causale rappresenta un passaggio epistemologicamente indebito.
[70] Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit., cap. 1; cfr. D. Yuhas, *Why social media makes peoples unhappy*, «Scientific American», 20.6.2022.

correlate e coerenti con questo proposito si possono ricavare da un utilizzo preciso, ma inevitabilmente anche atipico, di aspetti del pensiero di Maurice Merleau-Ponty, Georges Canguilhem e Kurt Goldstein[71]. La prima prospettiva è di ordine essenzialmente epistemologico e concerne i presupposti del concetto di ambiente e l'equiparazione skinneriana tra laboratorio e ambiente sociale; la seconda si basa invece sul potenziale patologizzante delle tecnologie comportamentali del capitalismo digitale ed è legata alla questione dell'ambiente tematizzata dalla prima prospettiva.

Provando a tracciare la prima linea di critica risulta innanzitutto necessario difendere una spiegazione coerente di cosa significhi estendere pratiche prima erano confinate al solo laboratorio, giungendo all'estrazione ubiqua di informazioni dall'essere umano[72]. Una versione ideologica di questo processo insiste unicamente sui benefici sociali della condivisione della conoscenza e dell'informazione o ancora riprende l'idea che un certo grado di sorveglianza dei comportamenti sia necessario ai fini dell'aumento di conoscenza[73]; queste narrazioni valutano positivamente le nuove risorse informative che aumentano sapere e controllo sui comportamenti delle persone, secondo il diffuso slogan per cui gli algoritmi ci conoscono meglio di noi stessi[74]. Occorre ribadire che si tratta di letture apertamente parziali e ideologiche che celano aspetti centrali dell'economia del surplus comportamentale. Da una parte nascondono sistematicamente la logica di profitto alla base dell'estrazione e della profilazione, ammantando i propri obiettivi di finalità progressive; dall'altra evitano di mettere in primo piano la dimensione di manipolazione strutturale dei comportamenti alla base delle loro tecnologie. Nel pensiero skinneriano la continuità tra laboratorio e ambiente sociale implica un'assunzione del carattere operativo e manipolativo della sperimentazione: come in laboratorio l'ambiente è riconfigurato in funzione dello studio di determinati comportamenti, così è legittimo applicare lo stessa idea alla vita degli individui. Come è stato già argomentato, questa modificazione strutturale dei comportamenti è centrale nella nuova economia digitale, ma non è mai riconosciuta. Una volta ribadito che gli

[71] È bene precisare che si tratta di prendere spunto da questi pensatori in chiave concettuale, piuttosto che individuare nessi storico-filologici. Infatti tutti e tre i pensatori si riferiscono al comportamentismo in chiave critica, ma la versione skinneriana non è l'oggetto privilegiato della loro attenzione. Il primo Merleau-Ponty, a cui si fa riferimento qui, conosce il comportamentismo principalmente grazie alla tesi di André Tilquin *Le bahviorisme*, in cui si tratta prevalentemente di John Watson (Cfr. A. Tilquin, *Le behaviorisme*, Paris 1942). Anche Canguilhem parla preferibilmente di Watson e di Tolman, salvo eccezioni di testi tardi come *Le cerveau et la pensée* in cui è citato Skinner, (Cfr. G. Canguilhem, *Oeuvres complètes Tome V*, Paris 218, pp. 895-932). Infine Goldstein, almeno all'altezza de *La Struttura dell'organismo*, non poteva accedere agli scritti di Skinner, per ragioni cronologiche (Cfr. K. Goldstein, *La struttura dell'organismo*, Roma 2010). I concetti di questi autori sono dunque da adattare rispetto ai presupposti skinneriani.
[72] Cfr. J. Lanier, *Dieci ragioni per cancellare subito i social network*, cit., cap. 1, p. 17.
[73] Cfr. S. Zuboff, *Il Capitalismo della sorveglianza*, cit., p. 218 e p. 237. Zuboff si riferisce anche a pubblicazioni di personaggi di primo piano della Silicon Valley, come Schmidt e Cohen.
[74] Cfr. N. Coudry and Mejias, *The Costs of connection*, cit., prefazione.

imperativi di estrazione e profilazione implicano una manipolazione massiva e strutturale del comportamento, conviene chiedersi cosa queste tecnologie dicano sul nostro comportamento. Se assumiamo che il nuovo capitalismo digitale agisce modificando i comportamenti e strutturando un nuovo campo d'azione artificiale per l'individuo, risulta impossibile difendere una visione meramente conoscitiva o neutrale degli algoritmi. In questo senso sono estesi e moltiplicati gli interrogativi che Georges Canguilhem poneva alla pratica sperimentale nelle scienze della vita: come si giustifica la validità della sperimentazione se essa si svolge in un ambiente artificiale che, inevitabilmente, perturba il comportamento dell'individuo? Come si può costruire un ambiente ad hoc che non finisca per normalizzare gli aspetti più imprevedibili del comportamento?[75] L'epistemologia che soggiace all'ipotesi comportamentista non prende in considerazione che una perturbazione sostanziale dell'ambiente in cui si svolge il comportamento ha una forte capacità di modificazione dello stesso, al prezzo di non riuscire a stabilire se i risultati ottenuti siano generalizzabili o frutto di una manipolazione contingente. In un breve testo, non a caso tradotto da Canguilhem stesso, Kurt Goldstein evidenzia la specificità dell'epistemologia delle scienze biologiche rispetto alle scienze fisiche: le prime devono fare i conti con l'attività creativa dell'organismo, con la sua tendenza ad imporsi sul proprio ambiente; inevitabilmente, l'isolamento di alcuni processi e la manipolazione dell'ambiente producono effetti distorsivi[76]. Se queste problematicità valgono per la sperimentazione animale in laboratorio, a fortiori sono pertinenti se applicate alla manipolazione del comportamento umano che comporta un numero di variabili ancora superiore. Gli algoritmi non ci conoscono meglio di quanto ci conosciamo noi, ma innanzitutto manipolano il contesto in cui ci muoviamo producendo pattern di comportamento e schemi corporei nuovi che poi cercano di prevedere. Un'epistemologia come quella comportamentista, modellata ancora sulla sperimentazione della fisica classica, non tiene in conto la complessità dell'interazione tra l'individuo e il proprio ambiente e si presta perfettamente a nascondere gli aspetti manipolativi delle tecnologie del comportamento[77]. Questa semplificazione del contesto sperimentale è consentita da una certa interpretazione dell'articolazione tra ambiente e comportamento che, come detto in precedenza, si limita al modello parcellizzato e deterministico stimolo-risposta-rinforzo. Vi si può contrapporre un'altra concezione del rapporto tra individuo e ambiente pensato come una dialettica aperta, un dibattito nel quale l'individuo co-costruisce l'ambiente che abita[78]; vivere significa organizzare il proprio ambiente, costituire uno spazio di

[75] Cfr. G. Canguilhem, *La connaissance de la vie*, Paris 2009, pp. 42-43 e p. 49.
[76] Cfr. G. Canguilhem, *Oeuvres complètes tome IV*, Paris 2019, pp. 1161-1164.
[77] Nella lezione del 21 marzo 1979 di *Nascita della biopolitica*, Michel Foucault parla di tecnologie ambientali per influenzare il comportamento associate alla psicologia americana e sempre più rilevanti per le politiche neoliberali. Si tratta più che altro di una traccia di ricerca, ma istituisce un punto di contatto interessante tra la governamentalità neoliberale e il comportamentismo, proprio sulla base di un certo modo di manipolare l'ambiente. Cfr. M. Foucault, *Nascita della biopolitica*, Milano 2015.
[78] Cfr. K. Goldstein, *La struttura dell'organismo*, cit., p. 81 e p. 106.

significazione: l'ambiente non si dà mai a prescindere dal senso e dall'esperienza pragmatica che l'individuo ne fa[79]. Anche Maurice Merleau-Ponty nei suoi primi testi aderisce a questo tipo di concezione del rapporto individuo-ambiente; la cosa è particolarmente interessante perché il filosofo francese ripropone il concetto di comportamento, ma lo risemantizza profondamente rispetto alla prospettiva comportamentista. Il concetto di comportamento è utile perché supera la dicotomia tra fisiologia e psicologia ma, lungi dall'essere una collezione di classi di risposta, è una struttura dinamica in cui l'azione degli stimoli non è mai isolabile. Lo stimolo è inserito in una significazione vitale, si integra in un complesso di norme incorporate e dotate di senso[80]. Bastino questi riferimenti per constatare che non c'è nulla di scontato nell'interpretazione che il comportamentismo radicale e, implicitamente, il capitalismo digitale danno dei concetti di ambiente e comportamento. Da una diversa prospettiva, le tecnologie comportamentali risultano tanto efficaci, quanto intrinsecamente manipolatrici, poco solide epistemologicamente e dal contestabile apporto conoscitivo.

Si potrebbe però rilevare che questa tipologia di critica mira solo ai presupposti epistemologici e non alle conseguenze che queste tecnologie hanno su corpo e comportamento delle persone. Sarà dunque rilevante aggiungere un'altra linea di critica orientata in questo secondo senso, capace di far emergere una certa tendenza alla patologizzazione dell'individuo. Se, come sostiene Merleau-Ponty nei corsi sulla Natura, il comportamento comprende la dimensione fisiologica e quella istintiva, oltre alle posture preferenziali dell'individuo[81], allora le tecnologie comportamentali concerneranno la corporeità e il suo rapporto con l'ambiente nel loro insieme. Si è già accennato alla questione della dipendenza, oggetto di un dibattito sempre più vivo relativamente all'uso di Internet, sia sul piano clinico sia su quello della ricerca empirica. Alcune ricerche hanno associato la dipendenza da Internet a quella per il gioco d'azzardo, a causa di alcuni aspetti analoghi nella sintomatologia; se la sintomatologia della dipendenza richiede criteri stringenti, vi è un'ampia fascia di disturbi meno debilitanti, ma ben più diffusi, che sono catalogati come abusi o come comportamenti eccessivi legati all'uso Internet[82]. Questi disturbi, seppur non ancora inseriti nel DSM5, sono sempre più studiati e la letteratura ne segnala la frequente correlazione con elevati livelli di stress, in quanto associate a strategie di evitamento dello stesso stress[83]. I meccanismi di dipendenza e abuso sono legati al funzionamento sregolato del sistema

[79] Cfr. G. Canguilhem, *La connaissance de la vie*, cit., pp. 187-197.
[80] Cfr. M. Merleau-Ponty, *La structure du comportement*, Paris 2013, p. 33 e pp. 166-168. Tutti i pensatori citati condividono una reinterpretazione del concetto di ambiente nel senso della *Umwelt* dell'etologo Jacob Von Uexküll. Per alcuni elementi della genealogia di questo concetto Cfr. W. Feuerhan, *Du milieu à l'Umwelt: enjeux d'un changement terminologique*, «Revue philosophique de la France et de l'étranger», Tome 134, 2009/4;
[81] Cfr. M. Merleau-Ponty, *La Nature*, Paris 1995, p. 220.
[82] Cfr. A. Radicchio, *Aspetti psicologici della dipendenza da Internet*, Milano 2018, cap. 1; cfr. D. Courbet, *«Addictions» et comportements problématiques liés à Internet et aux réseaux sociaux. Synthèse critique des recherches et nouvelles perspectives*, «ESSACHESS. Journal for Communication Studies», vol. 13, 1(25)/2020, p. 218.
[83] Cfr., ivi, p. 220.

dopaminergico e a una produzione di dopamina sistematicamente più elevata, caratteristica in comune con le dipendenze da sostanze[84]. Come già segnalato, i meccanismi di manipolazione del comportamento usati dalle piattaforme lavorano proprio su questo sistema, in coerenza con le tecniche skinneriane. Alcuni studi recenti mostrano un funzionamento sregolato del sistema dopaminergico (soprattutto del nucleo accumbens) nei soggetti che abusano dei social network e che sono influenzati dai meccanismi di ricompensa dei like e delle notifiche[85]. La ricerca empirica su questi temi è ancora limitata, data la recente esplosione del fenomeno, ma fornisce già alcune indicazioni sul peso del nuovo capitalismo digitale nell'ambito dell'abuso di Internet, coerentemente con il funzionamento delle tecnologie del comportamento impiegate. Sul segmento degli adolescenti vi sono poi alcuni studi che correlano l'uso dello smartphone a forme di dipendenza e di deficit di attenzione; anche in questo caso le evidenze sono destinate ad aumentare significativamente nei prossimi anni[86]. Questi disturbi sono l'altra faccia delle tecnologie del comportamento, sono i segni impressi sui corpi e sulle abitudini delle persone, ma è convinzione di chi scrive che il piano empirico non possa essere sufficiente da solo. Bisognerebbe forse pensare le dipendenze, gli abusi e altri disturbi come manifestazioni estreme di un sistema che ha tratti strutturalmente patologizzanti e che per i suoi stessi presupposti, non solo de facto, restringe la normatività vitale degli individui. Nella teoria canguilhemiana della salute e del patologico la prima è la capacità dell'individuo di strutturare nuove norme vitali, «è un margine di tolleranza nei confronti delle infedeltà dell'ambiente»[87], ovvero è una potenza propulsiva di variazione e di scarto dal semplice equilibrio dei valori fisiologici. Il patologico invece non è assenza di norma, bensì una norma inferiore che non tollera alterazione, è una perdita di capacità normativa[88]; quest'idea è in assoluta continuità con la teoria di Goldstein, per cui lo stato patologico è una restrizione del proprio ambiente di vita e una diminuzione della propria potenza sull'ambiente[89]. In fondo l'abuso si manifesta proprio come diminuzione dei margini di azione sull'ambiente (in termini di socialità, di opzioni di scelta, di percezione di stimoli alternativi) e incanala di fatto l'esistenza verso una progressiva perdita di controllo sullo

[84] Cfr. A. Radicchio, *Aspetti psicologici della dipendenza da Internet*, cap. 1; cfr. R. Mela, *La dipendenza da Internet*, 2007, cap. 3.
[85] Cfr. J. Montilly, *Shot de dopamine: ce que Facebook fait au cerveau de mon amie Emilie*, «L'Obs», 22.12.2019; cfr. D. Courbet, *«Addictions» et comportements problématiques liés à Internet et aux réseaux sociaux*, cit., pp. 221-222. Quest'ultimo studio mostra percentuali variabili di incidenza dei social network sulle dipendenze, a seconda delle ricerche, ma in ogni caso consistenti.
[86] J. Cho, C. K. Ra, M. D. Stone et al., *Association of digital media Use with subsequent symptoms of Attention-Deficit/Hyperactivity disorder among Adolescents*, «Jama», 320 (3), 2019, pp. 255-263; cfr. B. E. Custer, M. A. Lapierre, P. Zhao, *Short-term longitudinal relationship between smartphone use/dependency et psychological well-being among late adolescents*, «Journal of Adolescent Health», Volume 65, November 2019, pp. 607-612. La questione dell'economia dell'attenzione e dei deficit dell'attenzione correlati meriterebbe una trattazione a sé.
[87] G. Canguilhem, *Il Normale e il patologico*, Torino 1998, pp. 161.
[88] Cfr., ivi, p. 148.
[89] Cfr. K. Goldstein, *La struttura dell'organismo*, cit., pp. 353-354.

spazio di vita dell'individuo e sul proprio comportamento. Questa mancanza di controllo sembra essere un effetto strutturale delle tecnologie di cattura dell'attenzione e di indirizzamento del comportamento: sono costruite per ridurre gli spazi di autonomia dell'individuo, autonomia che, come si è visto a più riprese, in realtà negano per principio. D'altronde è una derivazione della epistemologia a loro sottesa e di una lettura del tutto deterministica dell'ambiente che condanna l'individuo a doversi adattare al contesto. Ecco allora che la minaccia all'autonomia dell'individuo può essere intesa in un senso più radicale rispetto ai discorsi che mettono al centro l'attacco alle libertà del cittadino e alle sue garanzie costituzionali: si tratta di intenderla come una restrizione della normatività, come la costruzione di un ambiente impoverito, su cui l'individuo ha meno controllo e decisamente meno capacità di variazione del proprio comportamento. Va insomma intesa come una manipolazione degli schemi corporei e delle prassi vitali, il cui prezzo è un'esposizione crescente a contesti patologizzanti. Da questa prospettiva anche i disturbi clinici assumono un significato più profondo e diventano strumenti decisivi per strutturare un discorso critico, articolato sulla dimensione del patologico.

Conclusione

Il percorso tracciato ha provato a fornire una lettura della nuova economia digitale sulla base della sua logica interna e dei suoi presupposti scientifici e filosofici legati al comportamentismo radicale; ha infine delineato alcune linee di critica diverse da quelle solitamente percorse. Proprio quando il comportamentismo sembrava aver definitivamente esaurito la propria capacità innovativa, esso si rivela più potente e vivo che mai, come una filosofia-scienza realizzata nel capitalismo digitale. Attorno al concetto di comportamento si gioca tutta la questione, dal lato della sua interpretazione e conseguentemente da quello della sua manipolazione. Se si accetta di porre la questione dal punto di vista della concettualizzazione e del governo del comportamento, anche la postura critica ne è per forza condizionata. In altre parole questa prospettiva impone uno scarto rispetto agli impianti critici che riguardano temi quali la riduzione di libertà civili, la tutela della privacy, o ancora la proprietà dei dati e gli accentramenti monopolistici. Si tratta di piani di analisi del tutto legittimi, e probabilmente anche necessari, ma in questa sede si è fatto il tentativo di mettere in questione presupposti ed effetti delle tecnologie comportamentali all'altezza delle strutture psichiche e corporee attraverso cui si manifestano. Inevitabilmente si è trattato di delineare prospettive critiche che rispettino questi criteri, piuttosto che svilupparle in senso compiuto. Resta molto da pensare, ma è difficile negare che l'articolazione tra capitalismo digitale e comportamentismo radicale metta in luce aspetti rilevanti della questione e fornisca una chiave di lettura feconda. Si tratta senza dubbio di materiale con il quale costruire nuove mappe del mondo a venire.

Marco Dal Pozzolo
Université de Bourgogne
✉ marco.dal-pozzolo@bourgogne.fr

Bibliografia

Fonti

Canguilhem, G. 2009, *La connaissance de la vie*, Paris, Vrin.
Canguilhem, G. 2018, *Œuvres complètes Tome V*, Paris, Vrin.
Canguilhem, G. 2019, *Œuvres comlètes tome IV*, Paris, Vrin.
Canguilhem, G. 1998, *Il Normale e il patologico*, Torino, Einaudi.
Foucault, M. 2009, *Bisogna difendere la società*, Milano, Feltrinelli.
Foucault, M. 2015, *Nascita della biopolitica*, Milano, Feltrinelli.
Goldstein, K. 2010, *La struttura dell'organismo*, Roma, Giovanni Fioriti.
Merleau-Ponty, M. 2013, *La structure du comportement*, Paris, PUF.
Merleau-Ponty, M. 1995, *La Nature*, Paris, Seuil.
Pavlov, I. P. 2020, *I riflessi condizionati*, Torino, Bollati Boringhieri.
Skinner, B. F. 2006, *Difesa del comportamentismo*, Roma, Armando.
Skinner, B. F. 1973, *Oltre la libertà e la dignità*, Milano, Armando.
Skinner, B. F. e Correll, W. 1974, *Pensare ed apprendere*, Roma, Armando.
Skinner, B. F. 1971, *Scienza e comportamento*, Milano, Franco Angeli.
Skinner, B. F. 1972, *Studi e ricerche*, Firenze, Giunti.

Studi

Boakes, R. 1986, *Da Darwin al comportamentismo*, a cura di Marco Poli, Milano, Franco Angeli.
Bocci, F. e Zona, U. 2018. Zona, *La rete come una Skinner Box. Neocomportamentismo, bolle sociali e post-verità*, «Media education», vol. 9, n.1.
Cardon, D. 2016. *Che cosa sognano gli algoritmi*, Milano, Mondadori.
Cho J., Ra C. K., M. Stone D. et al. 2019, *Association of digital media Use with subsequent symptoms of Attention-Deficit/Hyperactivity disorder among Adolescents*, «Jama», 320 (3), 2019.
Citton, Y. 2014, *Pour une écologie de l'attention*, Paris, Seuil.
Coudry, N. and Mejias, U. A. 2019, *The Costs of connection*, Standford, Standford University Press, Il Saggiatore.
Courbet, D. 2020, *«Addictions» et comportements problématiques liés à Internet et aux réseaux sociaux. Synthèse critique des recherches et nouvelles perspectives*, «ESSACHESS. Journal for Communication Studies», vol. 13, 1 (25)/2020.

Custer B. E, Lapierre M. A., Zhao P. 2019, *Short-term longitudinal relationship between smartphone use/dependency et psychological well-being among late adolescents*, «Journal of Adolescent Health», Volume 65, November 2019.
Dahaer, J. 2019, *Escaping Skinner's Box: AI and the Era of techno-superstition*, «Philosophical Disquisition».
Davidow, B. 2013, *Skinner Marketing: we're the Rats and Facebook limes are the Reward,* «The Athlantic» (10.6.2013).
Feuerhan, W. 2009, *Du milieu à l'Umwelt: enjeux d'un changement terminologique*, «Revue philosophique de la France et de l'étranger», Tome 134, 2009/4.
Harris, T. 2016, *How technology is Hijacking your mind – from a magician and Google design ethicist*, «Thrive Global» (18.5.16)
Kong, R., 2021, *Dopamine the action of life*, Londra, LAP Lambert.
Lanier, J. 2018 *Dieci ragioni per cancellare subito i social network*, Milano, Il Saggiatore.
Laura, L. 2019, *Storia degli algoritmi*, Roma, LUISS.
Mazzucchelli, C. 2020, *Prigionieri della macchina algoritmica siamo come piccioni dentro alla gabbia di Skinner*, «Solotablet» (20.12.2020).
Mela, R. 2007, *La dipendenza da Internet*, Create Space Indipedent publishing.
Montilly, J. 2019, *Shot de dopamine: ce que Facebook fait au cerveau de mon amie Emilie*, «L'Obs», 22.12.2019.
Pelgreffi, I. 2018, *Filosofia dell'automatismo*, Napoli-Salerno, Orthotes.
Quintarelli, S. 2019, *Capitalismo immateriale*, Torino, Bollati Boringhieri.
Radicchio, A. 2018, *Aspetti psicologici della dipendenza da Internet*, Milano, StreetLib.
Thompson, R. 1972, *Storia della psicologia*, Torino, Bollati Boringhieri.
Tilquin, A. 1942, *Le behaviorisme*, Paris, Vrin.
Yuhas, D. 2020, *Why social media makes peoples unhappy*, in Scientific American.
Zuboff, S. 2019, *Il Capitalismo della sorveglianza*, Milano, LUISS.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7770123

Contributi/10

Oltre il paradigma Smart City

Ontologia ed epistemologia dell'intelligenza urbana

Otello Palmini 0000-0003-0031-5558

Articolo sottoposto a doppia blind peer review. Inviato il 13/12/2021. Accettato il 08/06/2022.

Beyond the Smart City paradigm. Ontology and epistemology of urban intelligence

This article examines the concept of smart city in order to understand its philosophical implications. These implications, which concern urban ontology and urban epistemology, will be investigated by retracing the set of metaphors that innervate this model and understanding how these have changed in the contemporary technological context. Moreover, two conceptual sources of smartness will be analysed: the modernist architectural-urban space and the theory of cybernetic action which were respectively criticized by Henri Lefebvre and Hans Jonas. The article aims to contribute to the construction of a different relationship between technology and politics in smart urbanism and to promote a plural and articulated idea of urban intelligence.

***

Introduzione

Con il termine 'smart city' si indica il modello di *governance* urbana più influente degli ultimi vent'anni sia in termini di pervasività nella letteratura, sia per quanto riguarda gli investimenti economici compiuti[1]. A fronte di questa pervasività – teorica e pratica – è difficile trovare in letteratura una definizione univoca del concetto. Da un punto di vista qualitativo, considerando i principali indirizzi di definizione, è lecito affermare che questo modello riguarda l'implementazione nel contesto urbano di tecnologie dell'informazione e della

[1] Cfr. F. Bria, E. Morozov, *Ripensare la smart city*, Torino 2018, p. 18. Per una panoramica della diffusione del modello vedi A. Karvonen, F. Cugurullo, F. Caprotti, *Inside Smart Cities: Place, Politics and Urban Innovation*, Londra 2019. È necessario comprendere come questo modello sia il risultato di un cambio di paradigma decisivo, seguendo Picon «Our cities are on the verge of a radical transformation, a revolution in intelligence comparable in scale to the one that, in its time, brought about industrialisation. The smart city driven by digital technology, is poised to replace the typical networked city of the industrial era, whose success was built on hard infrastructure, from roads to water supply and sanitation systems, not only as a technological optimum but also as a social and political project» A. Picon, *Smart Cities. A Spatialised Intelligence*, United Kingdom 2015, p. 9.

comunicazione (ICT). Qui non si vorrà ripercorrere la breve e densa storia di questo modello ma concentrarsi su alcuni aspetti della sua prima formulazione, rintracciabile già alla fine degli anni Novanta e poi resa esplicita nella seconda metà della prima decade degli anni Duemila. Questo focalizzarsi sulle prime formulazioni del concetto sarà utile per evidenziarne una logica interna riguardante, nello specifico, la relazione tra tecnologia e politica che in gran parte innerva ancora la retorica e l'applicazione della *smartness* agli ambienti urbani.

Si vorrà dimostrare come questo particolare rapporto affondi le sue radici in un tempo decisamente anteriore alla rivoluzione digitale e in uno spazio concettuale assai più ampio di quello tecnologico. Queste radici importano, in primo luogo, un'ontologia dell'oggetto urbano, ovvero una particolare modalità di intenderne la composizione. Si vorrà mostrare come questa ontologia sia fondata su di una storia di modi metaforici di rappresentare l'urbano e su una teoria dello spazio che si dipana da una certa concezione della modernità. Si evidenzierà, inoltre, come in relazione a questa ontologia sia presente una corrispondente epistemologia in grado di strutturare la possibilità e la qualità dell'agire in questo spazio. Il modello smart city verrà quindi intenso come una traduzione di questo universo concettuale, di cui qui è offerta una parziale ma significativa analisi, nel linguaggio dei *Big-Data* e dell'Intelligenza artificiale (AI). Una traduzione che dona nuova vita al tradotto conferendogli caratteristiche specifiche. Qui si propone un inizio di critica, propriamente filosofica, che contribuisca a mettere in dubbio – anche attraverso alcune categorie e riflessioni mutuate dal pensiero di Henri Lefebvre e Hans Jonas – il rapporto tra tecnologia e politica che caratterizza questo modello di *governance* ponendo le basi per una visione più articolata dell'intelligenza urbana in era digitale.

1. Attraverso le immagini del paradigma *smart city*

La prima teorizzazione riguardo alla smart city è stata ad appannaggio delle grandi *corporations* che aspiravano ad entrare in questo mercato. Tra queste possono certamente essere ricordate IBM, Cisco e Siemens. Ad oggi IBM conserva il ruolo di leader nel settore sia da un punto di vista economico sia per quanto concerne la ricerca. Nell'ottica della *corporation* americana una città è *smart* se è *Instrumented*, *Interconnected* e *Intelligent*: dotata di sensori in grado di misurare i flussi e registrare la maggior quantità di eventi possibile nel contesto urbano, dotata di banche dati in grado di condividere tra loro le proprie informazioni e di sistemi intelligenti capaci di comprendere questi dati e favorire un processo decisionale razionale[2]. Le aree urbane avrebbero necessità stringente di queste «Capabilities that yesterday's leaders could not have imagined»[3] per due

[2] Cfr. Harrison et Al, *Foundations for smarter cities*, «IBM Journal Res. & Dev.», 54, 4, 2010.
[3] IBM, *Smarter cities, Smarter More Competitive Cities*, *Forward-thinking are investing in insight today*, New York 2012.

motivi: da un lato perché una larghissima parte della popolazione del pianeta vive ormai nelle città e questa tendenza è in crescita; dall'altro poiché dopo la crisi economica del 2008, le politiche di austerità e le relative ristrettezze di bilancio le amministrazioni sono obbligate a cercare sistemi per razionalizzare le spese, ridurre i costi e migliorare la distribuzione delle risorse. Nel contesto che abbiamo qui brevemente accennato il concetto di politica, di azione politica, di decisione politica viene toccato in maniera non frequente. Le implicazioni politiche del discorso sulla *smartness* vengono espressamente negate dai suoi promotori, che presentano il modello come squisitamente tecnico-tecnologico[4]. Questo lavoro vorrà portare alla luce alcuni meccanismi concettuali decisivi per la strutturazione del rapporto tra questo discorso sull'esperienza urbana e l'elemento politico.

Un primo aspetto decisivo riguarda le metafore attraverso cui la città viene pensata e rappresentata in questo contesto. La metafora non solo aiuta a strutturare un particolare campo del reale, ma influenza in maniera decisiva anche le modalità di azione e di intervento che quel particolare campo ammette, investe quindi tanto il campo ontologico che quello epistemologico[5]. Nel discorso sulla smart city si intrecciano almeno tre potentissime metafore che articolano, attraverso la loro intersezione, il campo delle possibilità nell'urbano: corpo, macchina e macchina intelligente. La relazione metaforica, ma anche materiale, tra città e corpo non nasce certo in epoca digitale. Per giungere all'altezza della declinazione che qui ci interessa dobbiamo riferirci alla rivoluzione epistemologica nella comprensione del corpo scaturita dallo studio *De Motu Cordis* di William Harvey del 1628. Harvey rideclina lo studio della circolazione sanguigna in termini di efficienza e meccanicità rompendo profondamente con la tradizione:

> Harvey made what seems in retrospect a simple discovery: the heart pumps blood through the arteries of the body, and receives blood to be pumped from the veins. The discovery challenged the ancient idea that the blood flowed through the body because of his heat […] Harvey believed circulation heated the blood, whereas the ancient theory supposed heat in the blood caused it to circulate[…] He pictured the body as a machine pumping life[6].

La prospettiva di Harvey – unita a quella di Thomas Willis (1621-1675) riguardante una simile interpretazione meccanicistica del sistema nervoso e a quella di Ernst Planter che fece lo stesso dal punto di vista della respirazione

[4] O. Söderström, T. Paasche, F. Klauser, *Smart city as Corporate Storytelling*, «City», 18, 2014, p. 317.
[5] «Those metaphors, which have evolved to reflect shifts in the zeitgeist, go along way in determining how we conceive of urban planning, urban form, governance, maintenance, citizenship, and so forth » S. Mattern, *City Is Not a Computer: Other Urban Intelligences*, Princeton 2021, p. 61. Cfr. S. Hnilica, *The metaphor of the city as a thinking machine. A complicated relationship and its backstory*, In *Architecture and the Smart City*, a cura di S. M. Figueiredo, S. Krishnamurty e T. Schroeder, Londra 2020, p. 68.
[6] R. Sennett, *Flesh and Stone. The Body and the City in the western Civilization*, New York 1994, p. 257.

– ebbe un profondo impatto sul modo di concepire la relazione tra corpo e città. La riflessione sull'urbano virò verso una concezione dello spazio cittadino come ambiente in cui regolare efficientemente i flussi al fine di rendere sano l'organismo. Inoltre il discorso di Harvey ha avuto il merito di rendere possibile un'ulteriore analogia, ovvero quella tra corpo e macchina. Come nota Richard Sennett, infatti, «While in his writings Harvey clung to the medieval Christian notion of the heart as an organ of compassion, by the time he published his findings he knew that it was also a machine»[7]. La concezione del corpo che contribuisce a informare la costellazione metaforica del discorso sulla *smartness* è dunque legata ad un avvicinamento epistemologico tra corpo e macchina capace di leggere entrambi come sistemi da portare alla massima efficienza-salute. Infatti se la metafora del corpo permettere di interpretare il problema della città in maniera binaria come una relazione tra malattia e salute, l'entrata nell'orizzonte simbolico della macchina ammette una concezione prettamente tecnica e meccanica dei problemi urbani[8]. Infatti:

> The connecting similarity between the living organism and a machine is the perfect coordination of a large number of individual parts. The origin of the word from the Greek organon (tool) makes it clear that living beings are thought of as a hierarchical structured systems consisting of several organs which, like a perfectly constructed machine […][9].

Dunque la città è interpretata come un qualcosa di composto da parti che non devono essere in conflitto tra loro in quanto, nei corpi, come nelle macchine, è proprio l'armonia tra gli organi e le componenti ad assicurare lo stato di salute-funzionamento dell'intero. L'azione sul corpo-macchina urbano sarà allora di tipo medico-meccanico, si costituirà come un riordinare l'armonia, il meccanismo.

L'ultima figura di questo trittico di metafore riguarda le macchine intelligenti. Segnatamente il riferimento è alle macchine cibernetiche capaci, attraverso meccanismi di feedback, di autoregolare il proprio comportamento, di bilanciare gli eccessi o le mancanze mantenendo una stabilità. Come noto la cibernetica ha quasi immediatamente straripato dagli angusti confini dello studio di particolari macchinari capaci di autoregolare il proprio movimento e ha cercato di imporsi in vari campi del sapere tra cui anche la gestione urbana. La sua ricezione più chiara in termini di architettura e pianificazione urbana è probabilmente *Urban Dynamics* di J. Forrester (1969). In questo lavoro – in riferimento al contesto tecnologico del suo tempo – Forrester tentò di

[7] Ivi, p. 258.
[8] «The metaphor of the city as a machine implies that instead of artist-architects, it is the engineer-architects who are necessary for the city. […] This formulation clearly expressed the conviction of modernity that a city – like every other problems – can be precisely defined scientifically and solved by technical means» S. Hnilica, *The metaphore of the city as a thinking machine*, cit., p. 70.
[9] Ivi, p. 73.

articolare un modello matematico capace di connettere tutti gli elementi centrali dell'esperienza urbana. Un impianto epistemologico finalizzato a strutturare dei modelli di pianificazione che dirigessero la macchina cibernetica urbana verso il massimo grado di efficacia. Dopo un iniziale entusiasmo il tentativo di Forrester fallì perché i suoi rigidi modelli non tenevano conto dell'evoluzione storica, delle differenze geografiche e delle complesse interazioni che strutturano quella che facendo nostro il suo linguaggio potremmo chiamare la dinamica urbana[10]. Tuttavia, nonostante questo fallimento applicativo, dal momento cibernetico in poi è stato possibile pensare non solo le città come corpi che vanno guariti o macchine che vanno riparate, ma che addirittura questi corpi e queste macchine potessero avere la capacità di ripararsi e gestirsi intelligentemente in maniera autonoma. Come notano Antonie Picon e Thomas Shay Hill da qui alla smart city il passo è concettualmente breve anche se tecnologicamente impegnativo. Seguendo i due studiosi:

> What seems to distinguish the smart city moment is not just the intensification of these features that likewise marked the earlier episodes – i.e., the vast expansion in scale and scope of the data available to contemporary urbanists, alongside the immense computational capacity modelers can use for their simulation[…] The assumption that seems to underlie many of these smart city approaches today is that our model do not simply approximate the reality: in integrating so much of the social and material world through the data set continuously fed into them, they are in fact precise numerical reflection of reality[11].

2. *From Impression to Knowledge*

Alla luce di questo breve accenno alla profondità storica della retorica sulla *smartness* è ora necessario comprendere come questo corpo metaforico venga riportato alla vita nell'epoca della rivoluzione digitale. Quello che avviene è, in primo, luogo, una re-ontologizzazione della città in termini di dati. L'idea, come abbiamo accennato, è che tutti i flussi di informazioni che costituiscono, nella metafora computazionale, la città possano essere registrati. Lo spazio urbano, per essere analizzato algoritmicamente, deve essere ridotto ad una superfice di equivalenza, ad un'ontologia piatta leggibile e interpretabile in linguaggio informatico. Questo è ritenuto possibile anche grazie a due di quelle che Adam Greenfield chiama «Tecnologie radicali»: lo *smartphone* e l'*Internet of Things* (IoT)[12]. Questi due elementi fungono da registratori in tempo reale di una enorme quantità di attività e di flussi di informazioni/azioni che accadono in ogni istante all'interno del contesto urbano e agiscono come traduttori della città reale in formato digitale. Lo spazio urbano – inteso sia come relazioni spaziali tra edifici sia come relazioni tra persone sia come

[10] Cfr. A. Picon, T. Hill, *Is the city becoming computable*? In *Architecture and the Smart City*, cit., p. 36.
[11] Ivi, p. 40.
[12] A. Greenfield, *Tecnologie radicali. Il progetto della vita quotidiana*, Torino 2017.

relazioni tra cose e persone – viene ridotto ad un'ontologia piatta e analizzabile.[13] A questo punto la città è veramente ridotta ad un campo univoco, ad un linguaggio unico che è possibile armonizzare in base ad un'idea di funzionalità complessiva (la salute del corpo, il funzionamento della macchina, la stabilità della macchina cibernetica). Tuttavia, Seguendo Greenfield: «Le persone che abitano, individualmente o riunite in comunità, sono portatrici di concezioni molteplici, concorrenti e altrettanto valide di ciò che è bene – ed è impossibile accogliere tutte contemporaneamente».[14]L'urbano in questo contesto non viene allora rappresentato come il risultato di un coagularsi di conflitti tra intelligenze diverse, relazioni di forza, interessi contrastanti e storicamente determinati, ma come l'esito di una costante evoluzione socio-tecnica verso il meglio, la maggiore efficienza, la migliore qualità della vita[15]. È indicativo che i cittadini – proprio quegli elementi attivi e in grado di interpretare-mutare le logiche dell'esperienza urbana – siano rappresentati in maniera passiva: invisibili, soggiogati, ridotti a sensore oppure condannati in quanto cittadini attualmente esistenti e proiettati in un'indefinita versione migliore e futura di sé[16]. Particolarmente interessante è l'unica di queste concezioni che in qualche modo prende in considerazione una qualche attività da parte del cittadino, ovvero quella che lo considera come un sensore in grado di inviare *feedback* ad un sistema già impostato, in grado quindi di contribuire alla stabilità e all'efficienza di un certo ordine, ma non di pensarne un'alternativa. L'intelligenza urbana è quindi la sola intelligenza cibernetica dell'intero e mai l'articolazione tra intelligenze situate che contribuiscono a generare un ecosistema complesso.

Se questa che abbiamo brevemente delineato è la struttura essenziale dell'idea di città re-ontologizzata dalla *smartness*, la questione tocca direttamente la dimensione del politico quando prendiamo in considerazione l'epistemologia che sorregge le possibili azioni in questo sistema. Guardando al discorso che qui stiamo sottoponendo ad analisi possiamo constatare che più che di impoliticità si tratta di un vero e proprio superamento del concetto stesso dell'agire politico. Un superamento che è basato su una coppia – *Big data* e Intelligenza artificiale (AI) – che dovrebbe assicurare una nuova oggettività, tecnologicamente ottenuta, delle decisioni. L'idea fondamentale è quella di ridurre la politica a ideologia, ovvero di interpretare l'agire politico come un agire costitutivamente viziato da presupposti parziali e strutturalmente reso ineffettuale dalla mancanza di una conoscenza obiettiva e completa della realtà. Il gesto politico sarebbe per natura inadatto ad agire per la salute del corpo urbano tutto, e per il funzionamento

[13] O. Söderström, T. Paasche, F. Klauser, *Smart city as Corporate Storytelling*, cit., p. 313.
[14] A. Greenfield, *Tecnologie radicali*, cit., p. 56.
[15] «Le smart cities vengono regolarmente presentate come l'apice di un'evoluzione dei centri urbani improntata all'innovazione tecnologica, e il controllo della loro crescita e ubiquità è vincolato al tasso dell'evoluzione dell'ingegno umano piuttosto che a fattori politici ed economici esterni» F. Bria, E. Morozov, *Ripensare la smart city*, cit, pp. 15-16. Cfr. O. Söderström, T. Paasche, F. Klauser, *Smart city as Corporate Storytelling*, cit., p. 316.
[16] Vanolo, A. '*Is there anybody out there? The place and role of citizens in tomorrow's smart cities*', «Futures», 82, 2016, pp. 26-36.

complessivo della città-macchina, in quanto viziato da un prospettivismo nei presupposti e da un'incompletezza informativa. L'incerta modalità dell'agire politico apparterrebbe ad un contesto tecnologico anteriore alla rivoluzione dei *Big data* che supererebbero il problema dell'incompletezza informazionale e precedente alla strutturazione di intelligenze algoritmiche e artificiali che eliminerebbero sia il problema della scarsa capacità computazionale umana sia quello della parzialità e del prospettivismo. Notiamo ad esempio come Rometty, ex CEO di IBM, considerasse la transizione dai precedenti modelli di *governance* a quello *smart* come una transizione «From gut-feeling and impressions to knowledge»[17], oppure come altre figure apicali dell'azienda intendessero questo tipo di modello come il «next logical step in governance» dato che finalmente sarebbe stato possibile avere una completezza informazionale sulla realtà urbana. Inoltre «It is claimed that can now be made on objective, comprehensive and rational evidence, as opposed to anectodes, intuitions, populist promises, or partial selective and ideological evidences».[18] Insomma la natura conflittuale e prospettica che informa l'ambito della decisione politica non è considerata un elemento ineliminabile della dinamica sociale, ma un momento nella storia progressiva dell'uomo, momento che deve la sua struttura plurale non all'effettiva presenza di diverse possibilità di sviluppo in seno alla società – che sono epistemologicamente inconcepibili in una logica binaria dal tipo salute/malattia o funzionamento/guasto – ma ad una sorta di infanzia tecnica dell'uomo. Le prassi, i discorsi che informano lo stadio iniziale del ragionamento sulla smart city hanno questo elemento di volontà di annullamento e di scioglimento dell'agire politico nella società. Nota Greenfield:

> Esiste una teoria implicita, una chiara posizione filosofica, persino una visione del mondo che sta dietro a tutti questi sforzi. Possiamo forse concepirla come una posizione tradizionale di positivismo logico, che ritiene, fra le altre cose, che il mondo sia in linea di principio perfettamente conoscibile, il suo contenuto enumerabile, e che ogni relazione possa essere significativamente codificata all'interno di un sistema tecnico senza preconcetti né distorsioni. Applicata alle questioni che riguardano le città, essa postula l'esistenza di un'unica risposta – universale e trascendente – a ogni bisogno umano individuale e collettivo, che questa risposta possa essere ottenuta algoritmicamente grazie all'operatività di un sistema tecnico fornito di input adeguati, e che possa essere formalizzata, sempre senza distorsioni, all'interno di una public policy […]. Ogni singolo passaggio di questo ragionamento è problematico[19].

Al di là dell'accostamento problematico con il positivismo logico possiamo affermare, seguendo questo lucidissimo passaggio, che il ragionamento è problematico in ogni suo passaggio e particolarmente nei suoi due punti fondamentali. Infatti, in primo luogo è un'illusione che i dati raccolti sui flussi

[17] O. Söderström, T. Paasche, F. Klauser, *Smart city as Corporate Storytelling*, cit., p. 312.
[18] S. Eren. *The Techno-Political Order of Smart City. An analysis through an Arendtian understanding of politics*, Twente 2018, pp. 8-9.
[19] A. Greenfield, *Tecnologie radicali*, cit., p. 54.

e sulle azioni, carne e sangue all'esperienza urbana, siano una loro restituzione oggettiva. I dati sono un prodotto: il risultato di un'attività per nulla trasparente di traduzione di fenomeni di vario genere in un qualsivoglia linguaggio, in questo caso quello informatico[20]. C'è sempre una scelta umana a monte della decisione che seleziona quali dati raccogliere, come farlo, con quale strumento e quindi con quale grado di approssimazione e traduzione. C'è insomma un contesto sociale e tecnologico che determina la produzione del dato. Inoltre è assurdo pensare di avere una completezza di dati su una città, pensare di avere un Digital Twin in scala 1:1 dello spazio urbano: questo sia perché dal punto di vista quantitativo questa risulterebbe un'operazione quantomeno problematica sia perché dal punto di vista qualitativo il dato non è la realtà, ma una sua traduzione-astrazione tecnologicamente e socialmente mediata[21]. Pensare, quindi, di considerare i *Big data* come un riferimento oggettivo in grado di superare dal punto di vista tecnico la parzialità dell'informazione che struttura i discorsi politici è nel migliore dei casi un eccesso di ottimismo tecnologico e nel peggiore un meccanismo ideologico.

Tuttavia il momento della registrazione del dato non è il solo momento di questo ragionamento in cui il contesto sociale, economico e politico viene tramutato in processo tecnico oggettivo. Infatti l'analisi di questi dati e la proposta delle soluzioni sono intrise della stessa logica. Ovvero l'intelligenza artificiale (AI) viene presentata come un metodo imparziale e oggettivo per prendere decisioni in merito ad un determinato problema. Essa sarebbe infatti un procedimento squisitamente tecnico e quindi, in questo universo di discorso, scevro da ogni parzialità politica e da ogni prospettivismo interessato. Anche in questo caso ci troviamo davanti ad una rappresentazione superficiale. In primo luogo gli importanti sviluppi recenti in materia di intelligenza artificiale sono legati a doppio filo alla grandissima quantità di dati che la nostra società è capace di produrre. Infatti, la grande capacità dell'AI di dare soluzioni a problemi estremamente complicati deriva da un cambio di impostazione nella sua architettura: dalla programmazione ad algoritmi procedurali al *machine learning*.

[20] «I dati sono comprensibili come una mappa della realtà, un dato (dal latino datum che significa 'dono') è una descrizione elementare, spesso codificata di un'entità o di un fenomeno, di una transazione, di un avvenimento o di altro. In quanto codifica, il dato è una riduzione della realtà analoga a quella che fa una mappa geografica» (P. Benanti, *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Bologna 2018, p. 120). «When we treat data as a 'given' (which is, in fact, the etymology of the word), we see it in the abstract, as an urban fixture like traffic or crowds. We need to shift our gaze and look at data in context, at the life circle of urban information, distributed within a varied ecology of urban sites and subjects that interact with it in multiple ways» (S. Mattern, *City Is Not a Computer: Other Urban Intelligences*, cit., capitolo 2, par. 4).

[21] «Data do not pre-exist their generation; they are not simply waiting to be collected or harvested in a technical, passive and objective manner. Instead, they are produced – that is, actively created via procedures and instruments of devising […] What data are generated, and how they are produced, handled and used, is the result of choices and decision by people […] The cooking of data does not take place in vacuum, but within context» (R. Kitchin, *Data Lives. How Data Are Made and Shape Our World*, Bristol 2021, capitolo 1, par. 7).

I procedimenti di *machine learning* permettono di analizzare e dare soluzioni a problemi in cui non siamo in grado di intuire, preventivamente, la logica di risoluzione. In questi procedimenti, infatti, i programmatori non danno già alla macchina il procedimento da utilizzare lasciandole solamente il peso quantitativo dal calcolo ma le forniscono una enorme quantità di esempi su cui 'allenarsi'[22]. Questi esempi altro non sono che dati, gli stessi dati che come abbiamo detto non si trovano in natura ma vengono prodotti in contesti tecnologici e sociali specifici. La capacità tecnica e sociale di produrre dati della nostra società è quindi condizione necessaria allo sviluppo degli efficacissimi sistemi di *machine learning*[23]. Gli algoritmi di *machine learning* vanno quindi allenati in palestre di dati prima di essere messi a contatto con le sfide in contesti reali e, già in questo primo passaggio, il modo di strutturazione di quelle che abbiamo chiamato 'palestre' è un procedimento altamente articolato e problematico[24]. A questo si aggiunge anche il fatto che il continuo aggiornamento di questi sistemi ha un costo computazionale ed economico elevatissimo[25]. A questi passaggi di ordine tecnico, che andrebbero indagati, e che già da soli metterebbero in crisi la pretesa di oggettività positivista citata da Greenfield, se ne aggiungono altri di tipo interpretativo. Infatti sia il procedimento attraverso il quale un'AI arriva alle proprie conclusioni sia la natura delle conclusioni stesse divergono profondamente rispetto al modo in cui ragionano gli esseri umani. Quel procedere per esempi e non per strutture logiche, a cui si è fatto riferimento, ha infatti profonde implicazioni sulla modalità di interpretazione che sulle soluzioni proposte. Alla causalità è stata sostituita la correlazione: invece di avere un modello teorico in base a cui strutturare le relazioni che connettono

[22] «Entrambe le categorie hanno pro e contro gli algoritmi basati su regole eseguono istruzioni scritte da esseri umani e quindi sono facili da capire. […] funzionano soltanto con problemi che l'uomo sa risolvere scrivendo una sequenza di istruzioni. Gli algoritmi di machine learning invece si stanno dimostrando particolarmente efficaci nell'affrontare problemi in cui non basta stilare una lista di istruzioni. […] Lo svantaggio è che consentire alla macchina di trovare da sé una soluzione, in molti casi sceglierà una strada che agli occhi di un essere umano risulterà incomprensibile» (H. Fry, *Hello World. Essere umani nell'era delle macchine*, Torino 2019, p. 20). Cfr. F. Rossi, *Possiamo fidarci dell'intelligenza artificiale?,* Milano 2019, pp. 57-60. Cfr. F. Zambonelli, *Algocrazia. Il governo degli algoritmi e dell'intelligenza artificiale*, Trieste 2020, pp. 33-50.

[23] Seguendo Francesca Rossi rispetto alle condizioni che hanno reso possibile il progresso esponenziale nel campo del *machine learning* «Fortunatamente, nel giro di pochi anni i dati disponibili sono cresciuti in maniera esponenziale. Per avere un'idea della velocità alla quale aumenta la quantità di dati disponibili, vi basti sapere che si sono generati più dati negli ultimi due anni che in tutta la storia dell'umanità» (F. Rossi, *Possiamo fidarci dell'intelligenza artificiale?,* Milano, 2019, p. 64).

[24] «La fase di apprendimento potrebbe essere stata, per dolo o per colpa, gestita in modo scorretto, per esempio, per favorire la capacità di gestire situazioni di parti specifiche trascurando di fargliene apprendere con dovuta adeguatezza altre di interesse generale» (F. Zambonelli, *Algocrazia,* cit., pp. 45-46).

[25] «Di solito vengono addestrati prima di essere messi in uso, e la loro capacità di apprendimento dopo viene congelata, perché impegnativa dal punto di vista computazionale […] Il congelamento dell'apprendimento fa si che i sistemi – oltre che a tendere a sbagliare in situazioni in cui hanno poca esperienza – non accumulino esperienza una volta terminata questa fase» (ivi, p. 45).

il tessuto della realtà queste sono individuate attraverso un'enorme quantità di esempi. In questo contesto epistemologico non si prende in considerazione la necessità di una relazione tra A e B, ma si constata moltissime volte che dove c'è A c'è quasi sempre B e tra le due allora si pone una correlazione[26]. L'algoritmo di *machine learning* ovviamente non si pone il problema dell'effettivo nesso di causalità tra A e B, ma un decisore politico che volesse ad esempio diminuire la frequenza di B intervenendo su A dovrebbe decisamente farlo. Questa differenza di funzionamento tra noi e le intelligenze artificiali è un ulteriore livello di articolazione e di problematicità della *Weltanshauung* positivista citata da Greenfield. Infatti, anche se controllassimo ed esplicitassimo tutti i presupposti tecnici di funzionamento e addestramento dell'AI, si porrebbe il problema dell'interpretazione delle soluzioni. Non è un caso, infatti, che a oggi l'approccio più sensato per strutturare la relazione tra AI e intelligenza umana sia di tipo collaborativo e non competitivo. Ovvero non abbiamo l'intelligenza umana che ragiona per impressioni fallaci, parziali e incomplete e l'AI che ha raggiunto l'oggettività della conoscenza, ma abbiamo due sistemi profondamente diversi tra i quali è necessario strutturare una relazione razionale di collaborazione. Abbiamo, infine, un ultimo livello di problematicità che riguarda il problema degli scopi, ovvero degli obiettivi che la macchina persegue o aiuta a perseguire. La selezione di questi scopi non è una scelta oggettiva e imparziale ma una scelta prospettica e politica. Seguendo Greenfield:

> In all my thinking about cities, it has frankly never occurred to me to assert that city have goals. (What is Cleveland's goals? Karachi's?) What is being suggested here strikes me a rather profound misunderstanding of what a city is. Hierarchical organizations can be said to have goals, certainly, but not anything as heterogeneous in composition as a city, and most especially not a city in anything resembling a democratic society[27].

Un ente complesso come una città o come una società in generale non è un qualcosa che possa avere degli scopi unici e univoci. Per quanto riguarda una città ha poco senso dire che essa debba funzionare, dato che al suo interno si svolge, auspicabilmente, un confronto su come essa debba farlo, sul per chi debba funzionare e in che modo. Ogni parte in causa in questo confronto ha idee diverse su cosa possa significare funzionare in un contesto sociale e urbano. La logica binaria che la retorica sulla smart city impone (salute/malattia, funzionamento/guasto, stabilità/instabilità, 0/1) è strutturalmente inadatta ad un fenomeno articolato come quello urbano. Da quanto detto dovrebbe essere chiaro come il rapporto tra tecnologia e politica sia pensato nei termini del superamento. Un superamento che affonda le proprie radici in un modello di interpretazione dell'urbano e su una precisa interpretazione del contesto tecnologico contemporaneo.

[26] Cfr. M. Carpo, *The Second Digital Turn. Design Beyond Intelligence*, Londra 2017, pp. 40-55.
[27] A. Greenfield, *Against the smart city: The City is Here for You to Use*, New York 2013, capitolo 2, par. 20.

3. Due modelli da indagare

Questa ontologia dello spazio e questa epistemologia hanno certamente degli antecedenti, delle fonti teoriche che, in contesti tecnologicamente e culturalmente differenti, hanno strutturato gesti e pratiche simili. Due di questi modelli, riconosciuti dalla letteratura sul tema, sono particolarmente interessanti in questo contesto: per quello che concerne la re-ontologizzazione dello spazio il modello è quello dello spazio architettonico moderno che arriva al suo apice in alcune espressioni urbanistiche del pensiero di Le Corbusier; invece, in riferimento alla re-epistemologizzazione dell'azione, il sistema di riferimento è il pensiero cibernetico. Quello che si vorrà fare qui è prendere in considerazione due critiche a questi modelli, che risultano particolarmente utili in questo contesto, per iniziare un lavoro di ripensamento teorico delle fonti concettuali del fenomeno della *smartness*. Per quanto riguarda l'insaturazione dello spazio architettonico moderno sarà presa in considerazione la critica di Lefebvre; in merito al pensiero cibernetico, invece, il riferimento sarà il pensiero di Jonas.

Una certa interpretazione del modernismo architettonico è riconosciuta da più parti come una delle fonti teoriche del discorso sulla *smartness*. Sennett considera la Carta di Atene e la pratica progettuale di Le Corbusier come l'antecedente teorico principale del modello smart city. Questo legame si radica in un preciso rapporto tra tecnologia e realtà urbana in cui la capacità di imporre un ordine propria del primo elemento è utilizzata per ridurre la confusione di uno spazio del secondo[28]. Sennett mette in luce la pericolosità di uno spazio urbano progettato per eliminare qualunque forma di attrito in continuità con l'imperativo dello 'user friendly' mutuato dal design tecnologico. L'idea è quella di un progettare prescrittivo in cui forma e funzione vengono unite nel tentativo, tutto modernista, di riordinare quella complessità, quei luoghi di possibile attrito che caratterizzano la vita urbana. Lo studioso americano nota che l'essere umano sviluppa la propria creatività e il proprio interesse per dei luoghi quando questi risultano articolati e non univoci: contesti che possano stimolare il suo pensiero, non privi di elementi problematici, capaci di fungere da punto d'avvio per una prassi creativa. Lo spazio univoco della smart city, d'altro canto, induce alla passività, alle possibilità prestabilite, all'ottundimento della capacità creativa. Anche Greenfield connette esplicitamente il discorso sulla *smartness* alle considerazioni di Le Corbusier sulla città.

> It's all there in Le Corbusier: the slate-wiping disregard for the existing urban fabric, the scientism, the surety that all is necessary for the human happiness can be captured in a formula. Compare his writings to any of the promotional materials we've examined, and you find the same claims that efficiency ought to be the single overriding concern of the urban designers, the same orientation toward the needs of

[28] «La Carta e il Plan Voisin sono alla base del concetto di smart city (città intelligente) in cui l'alta tecnologia tenta di ridurre la confusione e il caos tipici della vita in un logo complesso» (R. Sennett, *Costruire e Abitare. Etica per la città*, Traduzione di C. Spinoglio, Milano 2018 p. 95).

administration and the same belief that ideal urban form can be optimally specified and fixed for all time in serene perfection, Above all, we find the same essential hostility to the very notion of the urban[29].

Lo studioso americano nota come ci sia una distanza costitutiva tra il modo di progettare di Le Corbusier e l'effettivo svolgersi dell'esperienza urbana. Il progettare dell'architetto ha l'obiettivo di un formalismo funzionale – pensato in riferimento ai bisogni dell'uomo fisiologicamente inteso, ma non per quello storicamente e socialmente inteso – volto a sanare e a portare ad efficienza un tessuto urbano che non deve rassegnarsi al disordine malsano e non funzionale della città pre-moderna. Indagando le radici di questo rapporto conflittuale tra complessità urbana e ordine razionale mediato dall'elemento tecnologico possiamo risalire anche più indietro rispetto al modernismo architettonico collocandoci alla base di una certa interpretazione della modernità occidentale. Seguendo la riflessione di Federico Cugurullo, infatti, possiamo vedere nella *Nuova Atlantide* di Francis Bacon uno dei luoghi genetici di quel rapporto tra ordine e disordine mediato dalla tecno-scienza che abbiamo descritto: «New Atlantis can be understood as the first theory of experimental urbanism, capturing the force of modernity. In Bacon ideal city, urban developers fully control the built and the natural environment through technology to foster progress»[30]. Il riferimento a Bacone ci invita a riflettere sul fatto che in questo caso abbiamo davanti ad un'utopia concreta, ad un progetto politico che parte dal riconoscimento di una tendenza effettivamente operante e che viene individuata come un possibile punto di rotazione del medesimo. Non è un caso, allora, che questa tensione tra ordine e disordine mediata dalla tecno-scienza divenga, nella ricostruzione di Cugurullo, il cardine di quella modernità, articolatasi pienamente con la seconda rivoluzione industriale, che innerva di sé e che fa da fonte al discorso sulla *smartness*[31].

La lunga corsa di questa modernità è al centro delle riflessioni di Lefebvre sulla produzione dello spazio. In questo contesto ci si focalizzerà sulla teoria dello spazio proposta da Lefebvre nella convinzione che alcuni elementi portati alla luce dal filosofo possano essere estremamente stimolanti per una critica filosofica della re-ontologizzazione dello spazio della *smartness*. In primo luogo Lefebvre ci invita ad intendere lo spazio in maniera non cosale, esso non è cosa tra le cose, «esso avvolge le cose prodotte, e comprende le loro relazioni nella loro coesistenza e simultaneità[…] è il risultato di una serie di operazioni che non può ridursi a semplice oggetto»[32]. Lo spazio ha quindi la funzione fondamentale di dischiudere la possibilità delle relazione tra le cose, tra i corpi, tra le comunità diventando la struttura delle interazioni possibili tra gli enti. Una struttura che sta con questi enti in un rapporto di reciprocità: in cui gli enti stessi, con il loro

[29] A. Greenfield, *Against the smart city*, cit., capitolo 13, par. 4.
[30] F. Cugurullo, *Frankenstein Urbanism, Eco, Smart and Autonomous Cities. Artificial Intelligences and the End of the City*, Londra 2021, p. 50.
[31] Cfr. Ivi, pp. 48-60.
[32] H. Lefebvre, *La produzione dello spazio*, Milano 2018, p. 91.

agire, possono contribuire a modificare l'intelaiatura stessa che permette la loro azione[33]. Lo spazio, che va inteso allora con questa profondità di articolazione, è stato oggetto, nel moderno, di un approccio riduzionista. Scrive Lefebvre:

> Il riduzionismo si insinua sotto la pretesa di scientificità: si costruiscono modelli ridotti (della società, della città, delle istituzioni, della famiglia, ecc.) e ci si attiene ad essi. È così che lo spazio sociale si riduce allo spazio mentale, attraverso un'operazione 'scientifica' la cui scientificità maschera ideologia. I riduzionisti tessono l'elogio incondizionato del procedimento intrinseco alla scienza, poi lo trasformano in comportamento, e in seguito in sapere assoluto sotto la forma di una delle scienze (epistemologia)[34].

La pratica riduzionistica è qui individuata in una relazione con una certa scientificità ma anche con una intrinseca e decisiva carica politico-ideologica. Inoltre, questo elemento di re-ontologizzazione è già connesso alla pratica epistemologica utilizzata per indirizzare e rendere lecite, o illecite alcune tipologie di comportamento. Lefebvre riesce anche a isolare in maniera precisa lo scopo di questa pratica di spazializzazione nella trattazione di quello che definisce «spazio astratto», scrive il filosofo: «Questo spazio formale e quantificato nega le differenze, quelle derivanti dalla natura e dal tempo (differenze storiche)»[35]. Nello spazio omogeneo della modernità architettonica le differenze sarebbero negate, ricondotte a un'unica logica, depotenziate e mai prese in considerazione come radicali alterità. Lefebvre comprende come il vero obiettivo di questa re-ontologizzazione dello spazio sia la soppressione della categoria stessa di conflitto[36]. Anzi, meglio dire di una particolare forma di conflitto, ovvero quella in cui le parti si fanno portatrici di idee differenti e irriducibili di totalità. L'unico scambio di opinione concesso riguarda l'ottimizzazione, la maggiore o minore efficienza ma non i parametri, i presupposti su cui questa efficienza è misurata. Il punto è, insomma, una radicale depoliticizzaizone ideologica dello spazio che viene presentato come elemento neutro, logico, tecnico e, allo stesso tempo, è utilizzato come mezzo epistemologico-politico di gestione della società. In primo luogo Lefebvre allora invita ad interpretare propriamente lo spazio, a capire cosa è in gioco nel conflitto che lo innerva e che è possibile scatenare in esso. Come abbiamo visto lo spazio è interpretato come prodotto ma anche come mezzo di produzione, come risultato di un modo intendere il reale e al contempo anche come mezzo per darne una nuova interpretazione. Il primo

[33] In questo senso Lefebvre scrive di uno spazio che è sia prodotto sia mezzo di produzione «Prodotto che si utilizza e si consuma, è anche mezzo di produzione […] Questo mezzo di produzione, prodotto in quanto tale, non può separarsi né dalle forze produttive, dalle tecniche e dal sapere, né dalla divisione del lavoro sociale, che lo modella, né dalla natura, né dallo Stato e dalle superstrutture» (ivi, p. 102).
[34] Ivi, p. 120.
[35] Ivi, p. 60.
[36] «Lo spazio completamente omogeneo, perfettamente simultaneo, sarebbe indistinguibile. Sfuggirebbe l'elemento conflittuale, sempre risolto e nuovamente suggerito, tra simmetria e dissimmetria» (ivi, p. 201).

passo è insomma quello di non interpretare lo spazio urbano come un «dato di fatto», come un oggetto disponibile tutto insieme, simultaneamente, catturabile dal pensiero attraverso una sola logica, riducibile ad una interpretazione univoca. La città, lo spazio urbano meglio, è atto ma anche potenza, è il precipitato tridimensionale di certe pratiche ma anche la potenzialità virtuale di pratiche totalmente differenti. Si chiede il filosofo: «La coerenza non è forse l'ossessione di una società incoerente che cerca la propria via alla coerenza senza assumere la dimensione del conflitto ma, al contrario, rimuovendola e negandola in quanto tale?»[37]. Il processo attraverso cui la città è resa sistema tecnico-tecnologico coerente ha come oggetto, come posta in palio, la rimozione della potenzialità conflittuale della società, la dimensione del conflitto, della creazione della possibilità di riprogettarsi. È in questo contesto che, nell'analisi di Lefebvre, divengono centrali i concetti di opera, ludico e desiderio che altro non sono che espressioni per mostrare ciò che resta di residuale rispetto a queste analisi sistemiche e totalizzanti, un residuale che custodisce il possibile. L'opera viene interpretata come la possibilità per le comunità di plasmare, di appropriarsi dello spazio, di riconfigurarlo[38]; il ludico, nella sua accezione più profonda, è ciò che fuoriesce dall'ordine, è il tentativo carnevalesco del ribaltamento di un ordine dato, è la possibilità di una società diversa[39]; il desiderio diventa, in un contesto in cui la razionalità è quella del riduzionismo, il fuoco vivo capace di rompere lo schema imposto, la spontaneità di una volontà di uso e di appropriazione dello spazio. Con la creazione e il conflitto scompare anche la nozione di finalità, che viene ridotta a deduzione interna ad un processo tecnico, a derivazione secondo la logica dell'efficienza. Nelle parole di Lefebvre:

> La finalità, cioè l'insieme e l'orientamento dell'insieme, è frutto di decisioni. Dire che essa si deduce dalle operazioni stesse significa chiudersi in un circolo vizioso: darsi per scopo e per senso la ripartizione analitica. […] La nozione di sistema cela quella di strategia[40].

Proprio sulla questione della relazione tra finalità e sistema tecnico è utile fare riferimento al modello cibernetico e alla critica che di esso ha offerto Hans Jonas. Questo modello – che riguarda la presa in considerazione di macchine capaci non solo di sostituire la forza e l'energia umana ma anche una parte della loro facoltà di giudizio[41] – è da più parti riconosciuto, specialmente nella sua

[37] H. Lefebvre, *Il diritto alla città*, Prefazione di A. Casaglia, Verona 2014, p. 99.
[38] H. Lefebvre, *Spazio e politica. Il diritto alla città II*, Prefazione di Francesco Biagi, Verona 2018, p. 71.
[39] H. Lefebvre, *Il diritto alla città*, cit., p. 127.
[40] Ivi, p. 35.
[41] «La macchina è egualmente applicabile a un lavoro di carattere direttivo come al lavoro più pesante e meno qualificato, In tal modo i possibili campi in cui la nuova rivoluzione industriale sarà destinata ad estendersi appaiono già assai vasti e includono tutte le attività che richiedono giudizio di ordine meno elevato, esattamente come il lavoro che fu soppiantato dalla precedente rivoluzione industriale includeva tutti gli aspetti delle energie fisiche umane» (N. Wiener,

versione di primo ordine, come fonte del discorso sulla smartness[42]. Seguendo Anthony Townsend:

> Cybernetics took the idea of using sensing and feedback to optimize performance and extended to the universe generally. To Cyberneticians, everything – machines, organizations, cities, even the human mind – could be seen as a system, a balanced network of things connected by information flows. The component of every system, and the flows between them, could be represented as a set of equations that together could replicate the behaviour of the whole, they believed[43].

L'ipotesi cibernetica a cui qui si fa riferimento importa la volontà di mostrare una similitudine, e quindi una rappresentabilità o una sostituibilità, tra le macchine cibernetiche, ovvero quelle che possono adattare il proprio comportamento ad un obiettivo attraverso sensori e meccanismi di feedback, e sistemi complessi come l'organismo umano o urbano. Il risultato di questa nuova metafora, di questa nuova logica con cui interpretare la città è quello di intendere lo scopo o il fine di un'azione come un qualcosa di interno ad un sistema cibernetico. È su questo parallelismo – tra macchina cibernetica, organismo e organizzazione sociale – che ha come cardine il concetto di scopo che Jonas incardina la propria critica. L'idea fondamentale del filosofo è dimostrare che tra un siluro che corregge automaticamente la propria rotta in base alla capacità di percepire la propria posizione rispetto ad un bersaglio e un qualunque comportamento volontario umano non possa darsi similitudine. Semplicemente perché seguendo il pensatore tedesco: «In termini di semplice semantica possiamo dire che l'intera dottrina cibernetica del comportamento teleologico è riducibile alla confusione tra 'servire ad uno scopo' e 'avere uno scopo'»[44]. L'idea di Jonas è che la possibilità di percepire un oggetto-target e la capacità di orientare la propria azione ad esso – ovvero il fatto di avere sensori e di disporre di meccanismi di retroazione – sia certamente una condizione necessaria ad un'azione teleologica, ma non una condizione sufficiente[45]. Di più, il combinato disposto tra sensore e meccanismo di feedback è la condizione strumentale di possibilità per un'azione teleologica ma questa condizione nulla dice sullo scopo dell'azione. Per tornare all'esempio del siluro: lo scopo che esso ha di colpire il bersaglio non è assolutamente riducibile al meccanismo di percezione e retroazione che gli permette di colpirlo. Lo scopo è dato al siluro da una strategia militare tutta umana e completamente esterna a questo sistema, nei

Introduzione alla cibernetica. L'uso umano degli esseri umani, Introduzione di F. Ciafolani e Traduzione di D. Persiani, Torino 2012, p. 200).

[42] Cfr. A. Picon, *Smart Cities. A Spatialised Intelligence*, cit.; A. Picon, T. Hill, *Is the city becoming computable*? In *Architecture and the Smart City*, cit. A. Townsend, *Smart Cities. Big Data, Civic Hackers, and the Quest for a New Utopia*, New York 2013.

[43] A. Townsend, *Smart Cities. Big Data, Civic Hackers, and the Quest for a New Utopia*, cit., p. 75.

[44] H. Jonas, *La cibernetica e lo scopo: una critica*, a cura di M. Campanelli, con una introduzione di A. M. Iacono, Pisa 1999, p. 53.

[45] Ivi, p. 44.

termini di Jonas: il siluro non ha uno scopo ma serve a uno scopo. La macchina cibernetica è massimamente efficiente nel servire a uno scopo ma non ne ha uno, infatti «Le cose viventi sono creature del bisogno. Soltanto le cose viventi hanno bisogni e agiscono in base ad essi»[46]. Della macchina cibernetica si può dire che «L'intero congegno, giudicato dall'esterno, si comporta come se fosse indirizzato verso uno scopo»[47]. Interpretare l'azione umana e la società in base ad un modello come quello cibernetico significa, allora, ridurre la questione degli scopi di un organismo o di una società alla loro più o meno efficiente capacità tecnica di perseguirli. Come abbiamo visto significa ridurre la discussione sugli scopi a quella sull"efficienza dei mezzi, importa una rimozione più o meno cosciente della scelta dal valore che un dato meccanismo può o non può aiutare a realizzare. Conclude Jonas:

> In base a quello che affermano i cibernetici, la società è una struttura comunicativa per la trasmissione lo scambio e l'accumulo delle informazioni e sono queste di fatto che la tengono insieme. Mai prima è stata data una nozione di società più vuota di questa. Non si dice nulla sul tipo di informazioni, e perché dovrebbe essere così importante averle. Lo schema proposto non lascia spazio nemmeno all'emergere di tali domande[48].

4. Prospettive

La lettura di Lefebvre impone di rimettere in gioco lo spazio, di riproporre un pensiero oltre il riduzionismo, anche nella versione 4.0 della coppia *Big-Data* e Intelligenza Artificiale. L'indicazione di Lefebvre è chiara: è necessario ripartire dai residui. Ovvero dallo scarto del processo di riduzione, da ciò che non è stato mappato dal dato, da ciò che non è stato preso in considerazione, da ciò che nella nuova oggettività dei dati è stato reso letteralmente ciò che Shannon Mattern con una bella espressione chiama «Ontological and epistemological mud»[49]. Se la città è intesa come spazio astratto ciò che resta fuori – nei termini di Lefebvre – è la sfera dell'opera, del ludico, del desiderio. Ovvero resta fuori la possibilità di progettazione, di appropriazione dello spazio da parte delle comunità concrete contro la rappresentazione astratta dell'urbano. Lo spazio viene tecnicizzato e le comunità messe a funzione di un progetto che sotto la retorica tecnologica è schiettamente politico[50]. La critica dello spazio passa anche dal recupero della possibilità dello scopo. Passa dalla critica di Jonas e dal non vedere lo scopo

[46] Ivi, p. 59.
[47] Ivi, p. 48.
[48] Ivi, p. 60.
[49] S. Mattern, *City Is Not a Computer: Other Urban Intelligences*, cit., capitolo 1, par. 5.
[50] «Yet while purportedly impartial automated systems might seem to remove the inefficiencies and biases of human labour and deliberative processes, they ultimately impose their own encoded inequities and restrictive logics, like the hierarchical tree structure. They aim to merge the ideologies of technocratic managerialism and public service, to reprogram citizens as 'consumers' or 'users'. Filtering urban design and administration through algorithms and interfaces tends to bracket out those messy and disorderly concerns that simply 'do not compute'» (S. Mattern, *City Is Not a Computer: Other Urban Intelligences*, cit., Introduzione).

come qualcosa di interno al procedimento tecnico. La scelta delle comunità, delle varie intelligenze urbane deve essere libera, consapevole. In questo lavoro si è tracciata una possibile rotta per un'indagine critica del concetto di *smartness*. Un'indagine che ha considerato la genealogia storica delle metafore portanti di questo concetto, che ha condotto una critica dell'ontologia e dell'epistemologia della traduzione digitale della concezione della città come corpo-macchina-macchina intelligente. Questo impianto analitico-critico ha portato a una considerazione di come un approccio tecno-centrico alla relazione tra tecnologia e città corra il rischio di annullare la possibilità di una reale politica dello spazio e degli scopi. L'intero apparato di questo lavoro va contro ad una tendenza socio-tecnica-urbanistica che, seguendo Federico Cugurullo, possiamo definire come *Transurbanism*: ovvero l'idea che sarà la nuova tecnologia – la tecnologia cibernetica che sostituisce la nostra capacità di giudizio declinata nel contesto strutturato dalla coppia Big Data-AI – a risolvere i problemi di convivenza e di organizzazione urbana[51]. A questa impostazione va sostituito un diverso rapporto tra tecnologia e politica fondato in prima istanza sul fatto che la città e la società non possono essere interpretate come problemi tecnici. Come abbiamo visto buona parte di questo discorso ha al suo cuore la volontà di eliminare la conflittualità che innerva la società, ritenendo che la prima sia una possibilità evitabile nella seconda[52]. Questo rischio – che Mark Coeckelbergh inserisce tra quelli legati alla relazione tra AI e democrazia – è innervato dall'ontologia e dall'epistemologia la cui logica è stata parzialmente ricostruita in questo lavoro. Coeckelbergh oppone a questa logica un accostamento tra la critica all'intelligenza artificiale e il pensiero di Chantal Mouffe. Una prospettiva in grado non solo di pensare il conflitto come un qualcosa di strutturale al politico ma anche di valutarne positivamente il ruolo nel contesto democratico. Un contesto in cui ci sarà sempre una molteplicità di prospettive non riducibili ad una logica univoca e che dovrebbe abbracciare questo come elemento propulsivo del proprio essere:

> Conflict is the sign that democracies are alive. This is what she [Mouffe] calls the *agonistic* dimension of politics. […] There is no rational or objective ways of organizing

[51] «Just like transhumanism, transurbanism too approaches problems from a technological point of view, seeking to overcome them by means of superior technology. Such a technocratic problem-solving attitude does not fully work when it comes to the city. [This] book exposed the need for a radical reformulation of contemporary planning strategies, with the central aim of questioning the very ideas that form the core of cities. This need and this aim cannot be fulfilled by technology alone. They require substantial intellectual and philosophical inquiry which is not what transurbanism is about» (F. Cugurullo, *Frankenstein Urbanism,* cit., pp. 192-193).

[52] «AI may be used by those who seek to destroy the political itself: as a tool for Platonic philosopher-king or Habermasian democracy, it could be used to push a rationalist ad technosolutionist understanding of politics, which ignores the inherently agonistic dimensions of politics and risks excluding others point of view» (M. Coeckelbergh, *The Political Philosophy of AI*, Cambridge, 2022, p. 64).

society: such a solutions are also the result of power relations, Instead Mouffe proposes a agonistic pluralism: a system based on constructive disagreement[53].

Ciò che resta fuori dalla logica di interpretazione tecnocentrica e transurbanistica che abbiamo analizzato sono certamente delle soggettività, delle intelligenze delle materialità urbane; ma più di tutto ad essere annullata è la loro potenzialità politico-progettuale. Ontologicamente non riconosciute e epistemologicamente mutilate esse sono ridotte a materie di fatto prive di potenziale ludico, desiderante, teleologico e quindi politico. La direzione di ricerca che questa analisi vuole suggerire non riguarda solo una discesa filosofica nella storia concettuale della *smartness*, che pure resta necessaria; la volontà è quella di iniziare a contribuire ad una differente strutturazione dell'ontologia e dell'epistemologia del rapporto tra tecnologia e spazio urbano. Una prospettiva sulla filosofia del progetto che si opponga alla traduzione in versione AI di un riduzionismo tecnologico di vecchia data. L'intelligenza artificiale non può essere considerata l'unica intelligenza a cui ridurre la miriade di intelligenze storicamente e spazialmente stratificate che popolano le città: intelligenze umane intelligenze animali, vegetali e materiali non armonizzabili in una logica totalizzante che reclamano spazio vitale, o meglio che sono loro stesse questi spazi vitali[54]. Progettare, allora, potrebbe significare progettare l'espandersi e il comprimersi di questi spazi, tentare di comprenderne le logiche, le prospettive e le istanze. Progettare potrebbe voler dire comprendere come interpretarne i limiti, i margini di contatto con altri spazi e le potenzialità di questi contatti anche conflittuali. In questa prospettiva l'energia di uno spazio urbano non dovrebbe essere cercata nei centri di controllo e nei sensori che ne abilitano il funzionamento ma in queste faglie, in queste sovrapposizioni in cui realmente lo spazio diventa dinamico e apre possibilità di innovazione reale.[55] Pensare

[53] M. Coeckelbergh, *The Political Philosophy of AI*, cit., p. 72. Rispetto alla genealogia e alla logica di questa differente ontologia-epistemologia politica è interessante segnalare la riflessione di Roberto Esposito. Il filosofo italiano in un contesto non immediatamente riferibile alla critica della tecnologia struttura, attraverso una riflessione sulla prospettiva di Claude Lefort, una ontologia politica di valorizzazione del conflitto e che anzi lega essenzialmente la semantica del politico alla logica della conflittualità: «Il Politico non è altro da ciò – dalla risposta, o dal nome che le diverse società danno al 'fatto' della loro separazione. In questo senso, tutt'altro che un elemento esterno, è sempre ad esse immanente. È la coscienza della loro parzialità, dell'impossibilità di concepirsi come un tutto amalgamando le prospettive opposte che la abitano» R. Esposito, *Pensiero istituente. Tre paradigmi di ontologia politica*, Torino 2020, p. 170.

[54] «We cannot, as she [Gillian Rose] intimates, fall into the trap of thinking machinic intelligence is the only intelligence of the future or that human intelligence will simply be augmented by new mechanic intelligent systems. Our spaces are overly complicated by the constitutive post-human relations of socio-technological life to assert such reductions» (C. R. Lynch, V. J. Jr. Del Casino, *Smart Spaces, Information Processing and the Question of Intelligence*, «Annals of the American Association of Geographers», 2019, p. 6).

[55] Un approccio progettuale compatibile con il superamento di un approccio tecnocentrico e post-politico può essere individuato in P. Sendra, R. Sennett, *Designing Disorder, Experiments and Disruptions in the City*, London 2020. Inoltre esiste anche una geografia in espansione di progetto commentati in articoli che perseguono lo stesso intento di superamento. Cfr. C. McFarlan, O. Söderström, *On alternative smart city: From a technology-intensive to a know-*

questa ecologia dell'intelligenza urbana potrebbe aiutare a strutturare in maniera differente il rapporto tra tecnologia e spazio urbano, un modo che non avrebbe come obiettivo quello di superare il politico ma quello di approfondirne ed estenderne la logica all'interno del contesto dell'innovazione tecnologica. Seguendo Latour:

> Immaginare che un'ecologia politica di tale portata, già anticipata da tutti gli esperti, possa essere possa essere realizzata senza degli strumenti innovativi significa cercare il disastro. L'innovazione sarà assolutamente necessaria se dobbiamo adeguatamente rappresentare la natura conflittuale di tutte le cose che devono essere progettate – assumo qui il verbo 'rappresentare' nel senso più ampio, che include tecniche di rappresentazione scientifica, artistica e politica[56].

Otello Palmini
Università degli studi di Ferrara
✉ plmtll@unife.it

Bibliografia

Fonti

Jonas, H. 1999 *La cibernetica e lo scopo: una critica*, a cura di M. Campanelli, con una introduzione di A. M. Iacono, Pisa, ETS.

Lefebvre, H. 2014 *Il diritto alla città*, Prefazione di A. Casaglia, Verona, Ombre Corte.

Lefebvre, H. 2018. *La produzione dello spazio*, Milano, PGRECO EDIZIONI.

Lefebvre, H. 2018. *Spazio e politica. Il diritto alla città II*, Prefazione di Francesco Biagi, Verona, Ombre Corte.

Wiener, N. 2012. *Introduzione alla cibernetica. L'uso umano degli esseri umani*, Introduzione di F. Ciafolani e Traduzione di D. Persiani, Torino, Bollati Boringhieri.

ledge-intensive smart urbanism, «City», 21, 2017, pp. 312-328; S. Perng, S. Maalsen. *Civic Infrastructure and the Appropriation of the Corporate Smart City*, «Annals of the American Association of Geographers», 110, 2, 2019, pp. 507-515; J. Breuer, N. Walravens, S. Van der Graaf, I. Marien, *The right to the (smart) city, participation and open Data*, in *Architecture and Smart City*, cit., pp. 126-139; C. Charnock, H. March, R. Ribera-Fumaz, *From Smart to Rebel City. Wording, Provincialising and the Barcelona Model*, «Urban Studies», 58, 3, 2021, pp. 581-600; G. Trencher, *Towards the smart city 2.0: Empirical evidence of using smartness as a tool for tackling social challenges*, «Technological Forecasting and Social Change», 142, 2019, 117-128.

[56] B. Latour, *Un prometeo cauto? Primi passi verso una filosofia del design*, in B. Latour, *Le politiche del design. Semiotica degli artefatti e forme della socialità*, a cura di D. Mangano e I. Ventura Bordenca, Sesto San Giovanni 2021.

Studi

Benanti, P. 2018 *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Bologna, Marietti Editore

Breuer, N. Walravens, S. Van der Graaf, I. Marien, 2020 *The right to the (smart) city, participation and open Data*, in S. M. Figueiredo, S. Krishnamurty e T. Schroeder (eds), *Architecture and the Smart City*, London, Routledge. pp. 126-139.

Bria, F., Morozov, E. 2018. *Ripensare la smart city*, Torino, Codice Edizioni

Carpo, M. 2017. *The Second Digital Turn. Design Beyond Intelligence*, London, MIT Press.

Charnock, C., March, H., Ribera-Fumaz, R. 2021. *From Smart to Rebel City. Wording, Provincialising and the Barcelona Model*, «Urban Studies», 58, 3, pp. 581-600.

Coeckelbergh, M. 2022. *The Political Philosophy of AI*, Cambridge, Polity.

Cugurullo, F. 2021. *Frankenstein Urbanism, Eco, Smart and Autonomous Cities. Artificial Intelligences and the End of the City*, London, Routledge.

Eren. S. 2018. *The Techno-Political Order of Smart City. An analysis through an Arendtian understanding of politics*, Twente.

Esposito, E. 2020 *Pensiero istituente. Tre paradigmi di ontologia politica*, Torino, Einaudi.

Fry, H. 2019. *Hello World. Essere umani nell'era delle macchine*, Torino Bollati Boringhieri

Greenfield, A. 2013 *Against the smart city: The City is Here for You to Use*, New York, Verso.

Greenfield, A. 2017. *Tecnologie radicali. Il progetto della vita quotidiana*, Torino, Einaudi.

Harrison et Al. 2010. *Foundations for smarter cities*, «IBM Journal Res. & Dev.»,54, 4.

Hnilica, S. 2020. *The metaphor of the city as a thinking machine. A complicated relationship and its backstory*, in S. M. Figueiredo, S. Krishnamurty e T. Schroeder (eds), *Architecture and the Smart City*, Londra, Routledge.

IBM. 2012 *Smarter cities, Smarter More Competitive Cities*, *Forward-thinking are investing in insight today*, New York.

Latour, B. 2021. *Un prometeo cauto? Primi passi verso una filosofia del design*, In B. Latour, *Le politiche del design. Semiotica degli artefatti e forme della socialità*, a cura di D. Mangano e I. Ventura Bordenca, Sesto San Giovanni, Mimesis.

Lynch, C. R., Del Casino, V. J. Jr. 2019. *Smart Spaces, Information Processing and the Question of Intelligence*, «Annals of the American Association of Geographers».

Mattern, S. 2021. *City Is Not a Computer: Other Urban Intelligences*, Princeton, Princeton University Press.

McFarlan, C., Söderström, O., 2017. *On alternative smart city: From a technology-intensive to a knowledge-intensive smart urbanism*, «City», 21, pp. 312-328

Karvonen, A., Cugurullo, F., Caprotti, F. 2019. *Inside Smart Cities: Place, Politics and Urban Innovation*, Londra.

Kitchin, R, 2021. *Data Lives. How Data Are Made and Shape Our World*, Bristol, Bristol University Press.

Perng, S., Maalsen, S., 2019. *Civic Infrastructure and the Appropriation of the Corporate Smart City*, «Annals of the American Association of Geographers», 110, 2, pp. 507-515.

Picon, A., Hill, T. 2020. *Is the city becoming computable*? in S. M. Figueiredo, S. Krishnamurty e T. Schroeder (eds), *Architecture and the Smart City*, London, Routledge.

Picon, A. 2015. *Smart Cities. A Spatialised Intelligence*, United Kingdom, John Wiley & Sons.

Rossi, F. 2019. *Possiamo fidarci dell'intelligenza artificiale?,* Milano, Feltrinelli.

Sendra, P., Sennett, R. 2020. *Designing Disorder, Experiments and Disruptions in the City*, London Verso.

Sennett, R. 2018. *Costruire e Abitare. Etica per la città*, Milano, Feltrinelli.

Sennett, R. 1994. *Flesh and Stone*. *The Body and the City in the western Civilization*, New York, W. W. Norton & Company.

Söderström, O., Paasche, T., Klauser, F. 2014 *Smart city as Corporate Storytelling*, «City», 18.

Townsend, A. 2013. *Smart Cities. Big Data, Civic Hackers, and the Quest for a New Utopia*, New York, W. W. Norton & Company.

Trencher, G. 2019. *Towards the smart city 2.0: Empirical evidence of using smartness as a tool for tackling social challenges*, «Technological Forecasting and Social Change», 142, pp. 117-128.

Vanolo, A. 2016. *Is there anybody out there? The place and role of citizens in tomorrow's smart cities*, «Futures», 82, pp. 26-36.

Zambonelli, F. 2020. *Algocrazia. Il governo degli algoritmi e dell'intelligenza artificiale*, Trieste, Scienza Express.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771584

Contributi/11

L'algoritmo nelle trame del lavoro immateriale

Vincenzo Pallara 0009-0001-8755-8407

Articolo sottoposto a doppia blind peer review. Inviato il 24/11/2021. Accettato il 27/05/2022

The algorithm in the textures of immaterial labor

This paper discusses the meaning of the algorithm with regard to the current relations of social production, especially its functioning within the capital-labor relationship. Initially, in accordance with Antonio Negri's categories, the algorithm emerges as the extension of the computerization and digitization processes of the current production system, which goes beyond the walls of the factory, representing an investment of one's entire life, and which have as their production engine: knowledge, language and affect. Therefore, starting from this assumption, we use Antonio Negri's theoretical concepts, elaborated upon Foucault and Guattari's categories, which regard immaterial work as a hegemonic element necessary for the functioning of postmodern capitalism. Alongside intangible work, a new kind of work is examined, namely the digital in accordance with Casilli, Terranova and Mezzadra's categories. Regarding digital work, the close connection between the algorithm and the forms of extractivism of platform or network capitalism is evidenced. Finally, through the Marxist concept of General Intellect, the way the algorithm upsets some distinctions established by Marx emerges: a new configuration of labor subjectivity via the appropriation of fixed capital appears, then work, although increasingly autonomous from capital, can be considered as continuing under its command.

***

Introduzione

Attualmente, l'algoritmo è uno dei temi centrali del dibattito culturale. In ambito informatico, l'algoritmo è definito come complesso di operazioni numeriche volte alla risoluzione di problemi. Ma è possibile rintracciare nell'algoritmo un punto di vista diverso, che vada oltre la definizione generica proposta dalla scienza informatica? Effettivamente cos'è e soprattutto come funziona quando si interpone all'interno del rapporto capitale-lavoro? Per rispondere a questi interrogativi, l'analisi proposta in questo saggio si muoverà intorno ad un'impostazione materialistica di stampo post-marxista, con l'idea che un'indagine sull'algoritmo, a mio avviso, dovrebbe essere impostata sull'analisi dei fattori di produzione che ne dispieghino il significato e la sua funzionalità. In particolare, l'impostazione materialista a cui mi rifarò segue il

solco tracciato da quegli autori che a partire dalle lotte in Italia degli anni'70 in poi, hanno analizzato la trasformazione del capitale assistendo alla nascita del lavoro immateriale e digitale. All'interno di queste ricerche gli autori avanzano una tesi che può essere sintetizzata come segue: il capitale per contrastare le lotte operaie degli anni '70 avvia una vera e propria controrivoluzione, le fabbriche vengono destrutturate e delocalizzate, la produzione viene esternalizzata nel tessuto sociale, e s'introducono processi di automazione e robotizzazione[1]. Man mano che questa ristrutturazione del capitale prosegue, si assiste, assieme al cambio dei modi di produzione, a una mutazione del lavoro. Dalla produzione fordista basata sulla ripetitiva e meccanica catena di montaggio che rende l'operaio massificato e alienato, si passa a processi di produzioni post-fordisti, digitalizzati e informatizzati, dove la natura del lavoro – la composizione della forza lavoro – si trasforma diventando immateriale. Che significa immateriale? Significa che nella fabbrica post-fordista il lavoro diventa sempre più 'loquace'. Mentre la catena di montaggio fordista imponeva a tutti di lavorare in *silenzio* in maniera ripetitiva e monotona, ora il lavoratore produce attraverso il linguaggio, l'intelletto, la conoscenza e gli affetti. Non più composta da merci che proseguono l'assemblaggio finale, la catena di montaggio postfordista è intesa ora come un flusso di informazioni, dati e segnali che risalgono nei circuiti produttivi. Il linguaggio, la comunicazione e le informazioni (dati) diventano il perno centrale attorno a cui ruota tutta l'economia postfordista.

Inoltre, questi processi di trasformazione dei rapporti di produzione, evidenziano un ulteriore passaggio: il passaggio dalla 'società disciplinare' alla cosiddetta 'società di controllo'[2]. Citando Foucault e Deleuze si parta dai ben noti concetti foucaultiani di *biopolitica* e di *biopotere*. Il significato del primo termine sottolinea che la vita è messa al lavoro, 'a profitto', mediante una politica che organizzi le condizioni ed il controllo dello sfruttamento sociale sull'intera dimensione esistenziale per cui, in opposizione, si radica la resistenza e la capacità produttiva della nuova soggettività lavorativa. Per *biopotere*, invece, si intende una nuova figura di sovranità e di comando da parte del capitale sul lavoro. Di conseguenza, è a partire da questa impostazione post-operaista che, a mio avviso, è possibile definire meglio il concetto di algoritmo, poiché, il rischio è che esso divenga oggetto di una chimera metafisica. L'algoritmo è linfa necessaria al capitalismo post-industriale, in quanto, sebbene in forma immateriale esso

[1] Un tipico esempio di questa controrivoluzione viene riportato nell'articolo di Laura Fiocco del 1997, che analizza l'introduzione della tecnica *kanban* alla Fiat di Melfi come forma di riorganizzazione del lavoro al fine di reagire alle lotte operaie (cfr. L. Fiocco, *L'effetto kanban nell'organizzazione del lavoro alla FIAT di Melfi*, «Chaos. Quaderni di riflessione e dibattito politico-culturale», 1997, 10, pp. 23-45, http://www.arpnet.it/chaos/Fiocco.htm.

[2] Una precisazione. Il saggio si rifà all'interpretazione fatta dal filone post-operaista (in particolar modo Negri e Hardt) che aveva utilizzato il pensiero dei filosofi francesi, soprattutto di Michael Foucault e Gilles Deleuze, per analizzare le trasformazioni della società. In questo contributo i testi di riferimento sono: G. Deleuze, *Foucault*, tr. it. di P. A. Rovatti e F. Sossi, Napoli 2009; G. Deleuze, *Pourparlers*, QuodLibet, 2019; M. Foucault, *Nascita della biopolitica. Corso al Collège de France (1978-1979),* tr. it. di M. Bertani e V. Zini, Milano 2009.

rientra tra gli elementi necessari delle forze produttive materiali diventando il *medium* modificante non solo della composizione tecnica e politica delle soggettività lavorative ma, in ultima analisi, della loro stessa esistenza.

Alla base quindi dell'algoritmo ci sono i processi di informatizzazione e digitalizzazione dei sistemi di produzione che, trasformando la natura del lavoro (rendendolo immateriale) modificano la natura stessa del soggetto, la sua composizione corporea, in conformità alla macchina. Risulta fondamentale, al tal proposito, l'interpretazione di Toni Negri del passo marxiano sul *Frammento delle macchine* dei *Lineamenti fondamentali della critica dell'economia politica*. Ciò che Marx predisse in questo testo è la concettualizzazione di un Intelletto Generale capace di divenire forza produttiva immediata che, attraverso la sussunzione reale dell'intera società al capitale, permette nuove forme di estrazione di plus-valore dalla vita stessa. Partendo da questa definizione si possono identificare i modi di produzione postmoderni che, non solo portano alla modifica del rapporto tra lavoratore e macchina, ma implicano anche una parziale scomparsa della separazione tra lavoro necessario e sua riproduzione, tra tempo di lavoro e tempo di vita, e infine tra capitale fisso e sua riappropriazione da parte dei lavoratori.

Ricapitolando, percorreremo quattro direzioni che faranno emergere il significato sociale e la definizione dell'algoritmo. Il primo paragrafo, partendo dalle tesi di Negri e di Hardt, analizzerà le trasformazioni dei rapporti di produzione che, mediante la rivoluzione digitale e informatica, hanno pervaso la società andando oltre la fabbrica, investendo la vita stessa. Il secondo paragrafo, sarà incentrato sul novello concetto di lavoro cognitivo/immateriale; in questa disamina del lavoro immateriale, nel terzo paragrafo, attraverso le indagini di Tiziana Terranova, Antonio Casilli, Carlo Vercellone, Sandro Mezzadra e Renato Curcio si enucleeranno i diversi significati che sottendono al *Digital labour* – il quale, chiarifica i meccanismi di estrattivismo da parte dei giganti della Rete; infine, dialogando con le tesi di Negri, Pasquinelli e Fumagalli si ragionerà sulla auspicabile riappropriazione del capitale fisso da parte dei lavoratore, resa possibile dalle nuove forme del capitalismo informatizzato o digitale.

1. Dalla fabbrica alla società: la sussunzione reale

La rivoluzione informatica ha generato la democratizzazione, l'estensione massiccia degli strumenti tecnologici e digitali all'intero globo terreste intensificando, allo stesso tempo, i meccanismi di controllo e di valorizzazione da parte del capitale. Il capitalismo postmoderno nel suo estendersi *spazialmente* e intensificarsi *temporalmente* grazie alla digitalizzazione, alla tecnologizzazione, alla delocalizzazione e all'automatizzazione dei tessuti produttivi sulla sfera sociale riconfigura il tessuto urbano della metropoli dando luogo al fenomeno definito in termini marxiani *sussunzione reale*[3]: il capitale ha inglobato dentro di

[3] Come descritto ne *Il Capitale* e nel *Capitolo sesto inedito del Capitale*, per Marx la differenza tra sussunzione formale e reale corrisponde rispettivamente alla differenza tra plusvalore assoluto

sé tutti i rapporti sociali esterni ad esso. Non esiste più un 'fuori'. Si vive in una *fabbrica-società* dove:

> le regole specificamente capitalistiche dei rapporti produttivi e di sfruttamento, che si erano sviluppate dentro la fabbrica, si sono ore diffuse fuori dalle mura, permeando e definendo tutti i rapporti sociali[4].

Ne consegue che la fabbrica non può essere considerata più il luogo paradigmatico della produzione e della concentrazione del lavoro; i processi produttivi fuoriescono dalle mura della fabbrica investendo tutta la realtà.

> L'apparente declino della fabbrica come luogo della produzione non significa declino del regime e della disciplina di fabbrica, ma significa piuttosto che tale disciplina non viene più limitata a un luogo specifico interno alla società, ma si insinua in tutte le forme di produzione sociale diffondendosi come un virus. La società tutta è ora permeata dal regime della fabbrica, e cioè dalle regole specifiche dei rapporti di produzione capitalistici[5].

Il regime di fabbrica generalizzato, l'estensione dei processi lavorativi sull'intera vita sociale, attraverso il passaggio alla sussunzione reale, comporta una *socializzazione immediata delle forze produttive*. Non solo. Le macchine, la tecnologia (fondata sull'algoritmo) invadendo la realtà sussumono non più semplici lavoratori ma intere popolazioni tanto che «il macchinismo imbraga la vita»[6]. Tuttavia, se così è, quando questo sviluppo si compie, due *totalità sociali* si sovrappongono: il capitale (costante) che ha coperto l'intera realtà sociale e del capitale (variabile) che di questa realtà sociale e la sorgente di valorizzazione. […] non c'è più valore d'uso (e) non c'è, nemmeno natura – tutti i rapporti sociale (ovviamente quelli di produzione ma anche quelli di riproduzione e di circolazione) sono trasposti sul terreno dello sfruttamento – la *vita è sussunta nel capitale*[7].

Questa sovrapposizione tra due totalità sociali, una del capitale costante che si diffonde su tutta la dimensione sociale, e l'altra del capitale variabile che diviene la sorgente di valorizzazione con cui il capitalismo estrae plusvalore,

---

e relativo. Nel primo momento il capitale s'impadronisce del processo lavorativo solo formalmente, cioè riesce, si, a gestire e a controllare il processo produttivo, ma i processi lavorativi stessi sono ancora ancorati a forme antiche, precedenti, e quindi il capitale procede ad allungare la giornata lavorativa fino ad un *maximum*; mentre nel secondo momento – necessariamente conseguenziale al primo momento – il processo lavorativo è sottomesso omogeneamente al capitale, è sottomesso realmente ad un nuova tecnica, non più a quella antica. Di fatti la fase della sussunzione reale corrisponde alla fase della grande industria, caratterizzata dell'introduzione del sistema delle macchine che generano una riduzione al minimo della giornata lavorativa intensificando al massimo la produttività.

[4] M. Hardt, A. Negri, *Il lavoro di Dioniso. Per la critica dello stato postmoderno*, Roma 1995, p. 22.

[5] Ivi, p. 16.

[6] A. Negri, *Spunti di 'critica preveggente' nel Capitolo VI inedito di Marx*, «Uninomade 2.0», 27.8.2012.

[7] *Ibid.*

interpreta la vita come strumento di valorizzazione del capitale. Subentra, dunque, il concetto biopolitico di memoria foucaultiana atto ad introdurre una possibile definizione dell'algoritmo.

Al fine di rafforzare il suddetto concetto, è bene partire dalla memoria marxiana della sussunzione reale che, inglobando i rapporti sociali esterni al capitale nel capitale stesso, mette in crisi il cuore teorico del proprio pensiero, ovvero la teoria del valore-lavoro:

> Nella società capitalistica, il lavoro mostra una alternativa primaria e radicale, un'alternativa che permette di analizzarlo non solo come fatture costitutivo della società capitalistica, ma anche come negazione del capitalismo, come affermazione di un'altra società. Marx concepì la teoria del lavoro in due forme da due prospettive una negativa e una affermativa[8].

Dunque, oltre alla teoria economica classica che vede nel valore un prodotto del lavoro, Marx ne concepisce un'altra che «si distacca radicalmente dalle teorie capitalistiche perché prende in considerazione non tanto il processo capitalistico di valorizzazione quanto piuttosto i processi di autovalorizzazione (*Selbstverwertung*)»[9]. Allora accanto al processo di valorizzazione capitalistico nasce dunque un'altra autovalorizzazione: quella del lavoro vivo. Quest'ultimo si distacca e si sottrae dai circuiti di valorizzazione propri del capitale rivelandosi *negativo* e, allo stesso tempo, *affermativo* tanto che «rifiuta la sua astrazione nel processo di valorizzazione capitalistica e di produzione di plusvalore»[10] e, promuove, la realizzazione di una società alternativa[11]. Quindi, il lavoro vivo (immateriale) nella odierna produzione post-fordista, ha la possibilità immanente di separarsi dai circuiti della valorizzazione capitalista, in quanto esso è considerato «come figura antagonistica, come soggetto dinamico di rottura del sistema»[12], come forza-lavoro «relativamente indipendente dal funzionamento della legge capitalistica del lavoro. Ciò significa che l'unità di valore è primariamente identificata in rapporto al lavoro necessario, che non è una quantità fissa ma un elemento dinamico del sistema: il lavoro necessario è storicamente determinato dalle lotte operaie contro il lavoro salariato, volte a trasformare il lavoro stesso»[13]. Di conseguenza i concetti di valore e di lavoro si implicano a vicenda, facendo tramontare quel rapporto unidirezionale tra lavoro e valore: «non basta porre la struttura economica del lavoro come fonte della sovrastruttura culturale del valore; la nozione di struttura e sovrastruttura deve essere capovolta»[14]. Poiché l'attività lavorativa, «intesa come attività creatrice di

[8] M. Hardt, A. Negri, *Il lavoro di Dioniso*, cit., p. 14.
[9] Ivi, p. 15.
[10] Ivi, p. 12.
[11] «Il lavoro vivo e la forza che dall'interno costantemente determina non solo la sovversione del processo di produzione capitalistico ma anche la costruzione di un'alternativa» (*ibidem*).
[12] Ivi, p. 15.
[13] *Ibid.*
[14] *Ibid.*

valore, dipende sempre dai valori esistenti in un contesto sociale e storico»[15], essa viene riconosciuta e determinata dalle pratiche di lotta collocate storicamente e socialmente.

Ritornando ora alla definizione, se pur parziale, dell'algoritmo consideriamo che la produzione postmoderna nei sui elementi intriseci quali comunicazione, informazione ed affettività, riduce ad un *unicum* il piano strutturale e il piano sovrastrutturale; la fabbrica postfordista deve ascoltare il mercato e il desiderio dei consumatori, invertendo così il rapporto di preminenza dell'offerta sulla domanda, riducendo al minimo i costi, riorganizzando la produzione al fine di renderla 'snella con zero stock', di produrre *just-in time* e introdurre la tecnica del *Kanban*. Ne consegue che i momenti della produzione, riproduzione e circolazione non possono essere considerati separabili, ma inseriti in un unico circuito di valorizzazione in un totale appiattimento sul campo strutturale, ovvero nell'articolazione del campo di immanenza come campo biopolitico. Inoltre, il processo che sottomette la vita sociale alla logica produttivista del capitale può essere descritto nei termini tipicamente foucaultiani: come passaggio dalla società disciplinare alla società del controllo, in cui i meccanismi disciplinari propri della fabbrica vengono estesi in tutta la società, tanto che quest'ultima finisce per essere «governata esclusivamente dalle norme della produzione»[16].

> La disciplina è sia una forma della produzione, sia una forma del governo, di modo che la società disciplinare e la produzione disciplinare tendono a coincidere perfettamente. E in questa nuova società-fabbrica, le soggettività produttive vengono fabbricate come funzioni unidimensionali al servizio dello sviluppo capitalistico. Le figure, le strutture e le gerarchie della divisione del lavoro vengono massicciamente socializzate e minuziosamente definiti nella misura in cui la società civile viene assorbita dallo stato: le nuove norme della subordinazione e i regimi della disciplina capitalistica vengono estesi su tutto il sociale.[17]

Tale analisi foucaultiana è da estendere alla fase dell'accumulazione capitalistica moderna, ma a questa fase si deve integrare il passaggio alla società del controllo, che inaugura la produzione attuale. La società del controllo prevede che i meccanismi di comando divengono sempre più democratici, sempre più immanenti al sociale, e vengono distribuiti attraverso i cervelli e i corpi. La logica dell'inclusione e dell'esclusione, i comportamenti normali e quelli devianti prodotti dall'istituzione nella società disciplinare, ora, nella società del controllo vengono sempre più interiorizzati dai soggetti stessi:

> In questa società, il potere si esercita con le macchine che colonizzano direttamente i cervelli (nei sistemi della comunicazione, nelle reti informatiche ecc.) e i corpi nei sistemi di welfare, nel monitoraggio delle attività eccetera verso uno stato sempre più grave di alienazione dal senso della vita e dal desiderio di creatività la società del

[15] *Ibid.*
[16] M. Hardt, A. Negri, *Impero,* Milano 2001, p. 230.
[17] *Ibid.*

controllo può quindi essere definita come una intensificazione e generalizzazione dei dispositivi normalizzatori della disciplina che agiscono all'interno delle nostre comuni pratiche quotidiane[18].

Ecco che il concetto *biopolitico* diventa un'espressione paradigmatica del potere[19].

> Il biopotere è una forma di potere che regola il sociale dall'interno, inseguendolo, interpretandolo, assorbendolo e riarticolandolo. Il potere può imporre un comando effettivo sull'intera vita della popolazione solo nel momento in cui diviene una funzione vitale e integrale che ogni individuo comprende in sé riattiva volontariamente […]. La funzione più determinante di questo tipo di potere è quella di investire ogni aspetto della vita e il suo compito primario è quello di amministrarla punto il mio potere agisce dunque in un contesto in cui ciò che è in gioco per il potere e la produzione la riproduzione della vita stessa.[20]

Orbene, nell'attuale epoca post-moderna, la società è il luogo in cui il biopotere agisce attraverso le tecnologie. Alla resistenza del singolo, che riusciva ad opporsi alla disciplina del corpo e della coscienza individuale, ora, nella società del controllo il potere biopolitico pervade l'intero corpo sociale, in modo tale che quest'ultimo, viene compreso e assorbito nelle macchine del potere e fatto sviluppare nella sua virtualità. Insomma, il biopotere «si esprime mediante il controllo che raggiunge le profondità delle coscienze e dei corpi e, a un tempo, la totalità delle relazioni sociali»[21]. Quindi la sussunzione reale investe l'intero *bios* sociale mediante dispositivi di controllo e nelle modalità di disciplina proiettate, incorporate nei singoli individui. Se ne deduce che questo tipo di potere agendo sulla totalità della vita, non può essere unidimensionale, come sostenuto dalla scuola di Francoforte, che seguiva la linearità dello sviluppo capitalistico, ma deve agire su 'mille piani', in modo singolare e molteplice, come teorizzato da Deleuze e Guattari.

Le analisi fin qui condotte conduce ad una prima provvisoria definizione di algoritmo che si configurerà come un dispositivo biopolitico in grado di rafforzare i meccanismi di controllo e i processi di estrazione di (plus)valore del lavoro vivo da parte del capitale. Il controllo avviene attraverso la profilazione dei soggetti trasformati essi stessi in dati, i quali, vendono ceduti spontaneamente e continuamente sia negli ambiti lavorativi che nei momenti di svago in Rete.

---

[18] Ivi, p. 39.

[19] «Le enormi Corporation internazionali costituiscono la fabbrica che connette il mondo biopolitico […] e se strutturano e articolano direttamente territori e popolazioni […] distribuiscono direttamente la forza lavoro, allocano funzionalmente le risorse, organizzano gerarchicamente i settori della produzione mondiale. […] I grandi poteri finanziari industriali non producono solo merci ma anche soggettività: producono bisogni, relazioni sociali e cervelli; in altri termini producono i produttori […] la produzione biopolitica dell'ordine risulta immanente ai nessi di materiali della produzione del linguaggio, della comunicazione, e dei simbolismi che vengono sviluppate dalle industrie della comunicazione» (ivi, pp. 44-47).

[20] Ivi, p. 39.

[21] Ivi, p. 40.

Sottoposti in seguito a continua sorveglianza in guisa di prodotti venduti a terzi per fini pubblicitari e, in qualità di oggetti di studio del comportamento umano, gli stessi dati finiscono per direzionare il mercato ed essere direzionati per il mercato[22]. Infine, il capitale attraverso l'estrazione di dati e contenuti userà l'algoritmo come mezzo di valorizzazione del capitale stesso. Algoritmo diviene, dunque, uno duplice strumento biopolitico: da una parte, atto ad inseguire la logica produttivistica del capitale, dall'altra indispensabile mezzo di controllo del *bios*.

2. Il lavoro immateriale

Nel paragrafo precedente si è visto come il capitalismo odierno si sia imposto su tutta la sfera sociale e abbia modificato non solo la composizione tecnica e politica della forza-lavoro, ma anche il rapporto classico tra salario e capitale[23]. In questa congiuntura storica, il lavoro acquisisce la capacità essere *costitutivo*, distaccandosi sempre di più dai meccanismi del capitale. Quest'ultimo al contrario reso ormai parassitario, per mantenersi in vita, si impone attraverso dispositivi di comando e di controllo, di cui l'algoritmo è uno strumento centrale.

Ne risulta che «l'affermazione dell'economia informatizzata implichi necessariamente un mutamento nella qualità e nella natura del lavoro»[24] e tale cambio di paradigma economico finisca per ledere nel profondo la struttura antropologica e sociologica delle relazioni umani e del rapporto uomo-natura. Informazione e comunicazione, attualmente hanno «guadagnato un ruolo preponderante nei processi produttivi»[25].

L'Industria 4.0, la digitalizzazione, la robotizzazione e l'automazione apportano dunque una radicale trasformazione della natura del lavoro rendendolo immateriale e generando un cambio significativo nella relazione con lo strumento lavorativo in cui computer e smartphone si pongono come indispensabili 'prolungamenti' dell'individuo.

Il lavoro immateriale nasce e si sviluppa attraverso l'economia post-industriale, il che implica la fine della modernizzazione – segnata da un trasferimento massiccio della forza lavoro dal settore primario (agricoltura e miniere) verso il settore secondario (industria) – e sancisce l'inizio della post-modernizzazione o informatizzazione – «contraddistinta da un passaggio

[22] R. Fineschi, *Social e capitalismo crepuscolare (living in a box). Funzionamento e funzione dei social nelle dinamiche del capitalismo crepuscolare,* «La città futura», 30.10.2020.
[23] Si badi che mettere in discussione non significa negazione completa del rapporto salariale, non intendo dire che oggi il salario – come forma di pagamento di una somma di danaro del lavoro necessario – non esista più, ma intendo dire che l'economia informatizzata e digitalizzata produce un soggetto lavorativo che eccede questo rapporto. Produce un soggetto maggiormente precarizzato, marginalizzato, mobile e flessibile, che sottostà sempre più alle dipendenze del mercato del lavoro. Un soggetto, dunque, che vede deregolamentarsi i contratti come tutela dei diritti sociali.
[24] M. Hardt, A. Negri, *Impero*, cit., p. 271.
[25] *Ibid.*

dall'occupazione industriale al prevalere degli impieghi nei servizi (settore terziario)»[26].

Questo mutamento è stato avviato sin dall›inizio degli anni Settanta. I servizi coprono una vasta gamma di attività: dalle cure sanitarie alla pubblicità passando per l›educazione, la finanza, i trasporti e l'industria del divertimento. I posti di lavoro, per la maggior parte, sono temporanei e richiedono competenze fluidi e duttili ma l'aspetto più importante è che sono caratterizzati dal ruolo capitale svolto dalla conoscenza, dall'informazione, dall'affettività e dalla comunicazione[27].

Ponendosi come scopo la produzione non di mere merci materiali, durevoli, distaccate dal lavoratore stesso, bensì di «un servizio, un prodotto culturale, conoscenza o comunicazione»[28], il lavoro immateriale sceglie come strumento privilegiato della sua prassi il computer, strumento che «ha progressivamente riconfigurato le pratiche e le relazioni produttive e, più in generale, tutte le relazioni sociali»[29]. Mentre lavora, dunque, immerso nel suo mondo digitalizzato, il lavoratore trasforma il suo corpo. Si può parlare di una smaterializzazione del corpo in quanto il lavoratore produce attraverso uno scambio di informazioni-comunicazioni («Robert Reich definisce il tipo di lavoro immateriale associato ai computer e alla comunicazione come servizi simbolico-analitici»[30]) e conoscenza che permettono l'utilizzo intellettuale della forza-lavoro e non più un suo utilizzo fisico in senso stretto. Quindi si produce sempre più con processi intellettivi e linguistici e meno con dinamiche corporee. Non che l'intelletto e il linguaggio non appartengano al corpo, indubbiamente essi sono il loro prodotto, ma c'è una tendenziale alienazione delle proprie capacità e un evidente atrofizzazione del corpo durante le attività lavorative, soprattutto in abiti dove la presenza degli strumenti tecnologici è rilevante.

Le tecnologie informatiche e le macchine digitali «riducono sempre più la rilevanza delle distanze»[31], infatti, il corpo del lavoratore si smaterializza a causa di una stasi geografica imposta dal decentramento della produzione: l'avvento della informatizzazione sostituisce la catena di montaggio fordista con il 'Sistema di Rete', principale modello organizzativo della società post-industriale.

I lavoratori impegnati in un determinato processo produttivo possono infatti comunicare e cooperare anche da insediamenti molto distanti tra di loro dove la rete della cooperazione in sostanza non dipende più da un centro fisico o territoriale[32].

[26] Ivi, p. 267.
[27] «La tesi secondo la quale la modernizzazione terminata e l'economia globale un'economia post moderna informatizzata non implica che l'industria sia scomparsa o che abbia cessato di svolgere un ruolo importante anche nelle regioni più avanzate del globo» (*ibidem*).
[28] Ivi, p. 272.
[29] *Ibid.*
[30] Ivi, p. 273.
[31] Ivi, p. 277.
[32] *Ibid.*

Di conseguenza, la cooperazione risulta un elemento immanente del lavoro immateriale che implica, dunque, le interazioni sociali fino a rafforzare l'attività cooperativa tra i lavoratori stessi. Esiste un'immediatezza tra le singolarità e il comune. Ne risulta che, oggi, la cooperazione è imposta dall'*interno* della forza lavoro, non più dal padrone dall'*esterno*:

> gli aspetti cooperativi del lavoro immateriale non vengono cioè imposti organizzarti dall'esterno, come accadeva in altre forme del lavoro; nel lavoro immateriale *la cooperazione è completamente immanente alla stessa attività lavorativa.* Questo mette definitivamente in discussione la vecchia concezione del lavoro […] come capitale variabile, e cioè come una forza che viene funzionalmente attivata solo dal capitale; ormai, è il potere inerente alla cooperazione della forza lavoro (e, in particolare, del lavoro immateriale) che permette al lavoro di valorizzarsi. I cervelli e i corpi hanno ancora certamente bisogno di altri cervelli e di altri corpi per produrre valore; in questo caso, però, gli altri di cui si ha bisogno non vengono necessariamente fornite dal capitale e dalle sue capacità di orchestrare la produzione. Al giorno d'oggi, la produttività, la ricchezza e la creazione del surplus sociale sono determinate dalle forme della interattività cooperativa che corre lungo la rete dei linguaggi della comunicazione e dagli affetti[33].

Analizzando questo passo, vedremo in seguito come si strutturerà l'algoritmo, come il padronato, nei processi informatici di valorizzazione capitalistica, attraverso metodi di captazione estrarrà valore dai dati, ricaverà quel *plus* dal lavoro cognitivo, imponendo operazioni algoritmiche sopra la cooperazione delle soggettività lavorative che si muovono nei distretti digitali informatici, come una cappa che succhia valore sull'intellettualità di massa – per usare le parole di Virno. Allo stesso tempo, tuttavia ci accorgeremo che l'algoritmo, pilastro portante delle macchine digitali dell'economia del capitalismo postmoderno, rafforza la *separazione* e l'*autonomia* del lavoro 'vivo' rispetto a tali captazioni capitalistiche. Autonomia, che dal mio punto di vista, potrebbe rafforzarsi e tentare di attualizzarsi, attraverso l'instaurazione di un reddito sociale di base incondizionato. Ora vedremo come quest'estrattivismo venga rafforzato dai giganti della rete e della logistica attraverso il concetto di *Digital Labour.*

3. Il *Digital Labour* e i processi di estrattivismo della Rete

All'interno del lavoro immateriale s'inscrive invece la concettualizzazione di un nuovo tipo di lavoro nell'era digitale, il cosiddetto *Digital Labour*, in relazione alle nuove forme di estrattivismo del capitalismo delle piattaforme. Il Digital Labour permetterà di verificare in che modo le fabbriche dei social network riescano ad avviare nuovi processi di accumulazione attraverso il *data-mining*, ricavando valore dallo sfruttamento delle interazioni sociali, comunicative-linguistiche e affettive all'interno dei distretti virtuali. Allo stesso

[33] Ivi, p. 275.

tempo, risulterà evidente come l'algoritmo giochi un ruolo necessario nella raccolta di dati essendo alla base dell'economia delle piattaforme e fino a che punto contribuisca a rendere prospero il *Digital Labour*.

Il *Digital Labour* non presenta una univoca definizione, in quanto opera in varie *platform* e cambia in base alla piattaforma in cui agisce. Il termine '*Digital Labour*' fu utilizzato per la prima volta da Tiziana Terranova in un articolo del 2000[34], intendendo con esso lavoro *gratuito* e «apparentemente *libero* che una moltitudine di individui svolge attraverso e su Internet, spesso inconsapevolmente, a beneficio dei grandi oligopoli di Internet»[35]. Appare chiaro che tale definizione di lavoro gratuito tipico del consumatore che circola in rete non può considerarsi esauriente. Difatti, il sociologo Antonio Casilli la amplia presentendo il *digital labour* come lavoro che «include in sé l'elemento fisico, il movimento attivo del *digitus*, il dito che serve a contare ma anche quello che clicca sul tasto»[36]; in secondo luogo, include nella definizione «tutti quei lavori occasionali, quei contratti 'a zero ore' e anche quelle forme tradizionali di subappalto che lo sviluppo dell'automazione ha fatto esplodere, esacerbandone le logiche. […]. Dunque, dalle considerazioni dei due studiosi emerge che *Digital Labour* non può indicare soltanto il lavoro gratuito ma anzi designa un *continuum tra attività non remunerate, attività sottopagate e attività remunerate in modo flessibile*»[37]. In particolare, Casilli distingue tre tipologie di *Digital Labour* partendo dal modo in cui le piattaforme che estraggono valore sia dal lavoratore che dall'utente-consumatore: il *lavoro on demand*, il *microlavoro* e il *lavoro sociale in rete*.

Il *lavoro on demand* viene utilizzato dalle piattaforme *Uber*, *Foodera* o *Deliveroo* che rappresentano in generale la cosiddetta *Gig economy* (l'economia dei lavoretti). Il termine 'gig' indica l'intermittenza, la prestazione occasionale da parte dei lavoratori retribuiti attraverso il sistema del cottimo. I lavoratori cedono la loro prestazione, o per necessità materiali, o irretiti dalla falsa retorica della flessibilità e dell'autonomia come opportunità lavorativa (ad esempio il poter arrotondare qualche soldo nel tempo libero o poter lavorare 'quando si vuole'), retorica che si traduce invece in una costante disponibilità pretesa dall'impresa. Pertanto, l'inconsapevole «utente-lavoratore»[38] mentre svolge mansioni che garantiscono «servizi di trasporto, alloggio, spedizione, servizio alla persona, riparazione o manutenzione»[39]produce senza sosta dati. E scandendo tale produzione, l'algoritmo configura lo spazio e il tempo dell'attività dei lavoratori, definendo al contempo paghe e consegne. I corpi dei lavoratori vengono gestiti

[34] T. Terranova, *Free labor: producing culture for the digital economy*, «Social text», Volume XVIII, II, 2000, pp. 33-55.
[35] C. Vercellone, *Big-data e Free Digital Labour e big-data nel del capitalismo delle piattaforme: un nuovo estrattivismo?*, in (AA. VV.), *L'enigma del valore. Il Digital Labour e la rivoluzione tecnologica,* a cura di 'Effimera', Milano 2020, pp. 9-24, p. 16.
[36] Ivi, p. 48.
[37] Ivi. p. 50.
[38] Ivi, p. 85.
[39] *Ibid.*

e governati da una macchina digitale che, attraverso una combinazione appunto algoritmica, ridefinisce gerarchie ed estrae informazioni dall'impronta valoriale. Ad esempio, i *rider* o i tassisti di Uber, essi «accedono al loro potenziale lavoro 'loggandosi' sul proprio cellulare ad un'applicazione smartphone e, una volta connessi, le consegne vengono loro assegnate in modo automatico da un algoritmo»[40].

> il processo di lavoro, e soprattutto la performance lavorativa, appaiono dunque monitorati costantemente dall'algoritmo, che elabora in modo sistematico anche le valutazioni effettuate dai clienti[41].

Dunque, la raccolta di dati permette alle piattaforme di controllare la produttività del lavoratore e di vigilare sulle sue prestazioni. L'algoritmo funge da strumento per monitorare in che misura sia produttivo il lavoratore. Al contempo, il consumatore svolge il suo fondamentale ruolo rilasciando dati e contenuti valutativi che si riveleranno determinanti per la vita del lavoratore, il quale, rischia così di essere 'disattivato', ergo licenziato, se non sottostà ai criteri valutativi del consumatore o dell'azienda.

In conclusione, consideriamo un altro fattore determinante per il *lavoratore on demand*. Il tempo di attesa tra una consegna e l'altra, prende forma come *non-tempo di lavoro*, inteso come tempo non considerato retribuibile dalla piattaforma. Ne consegue, che la disponibilità ad 'attendere' del lavoratore non rientra nel calcolo del tempo di lavoro creando un'asimmetria tra il tempo di consegna e il tempo di attesa, configurando quest'ultimo come tempo gratuito rubato dalle piattaforme stesse.

> Il lavoratore intermittente di Foodera, JustEat, Deliveroo è una nuova figura esemplare del capitalismo digitale e dei suoi criteri di misurazione del tempo di lavoro, tempo ridotto all'atto circoscritto della prestazione. E ci stiamo soffermando su di essa poiché in essa si manifesta un dispositivo di appropriazione gratuita del tempo che in altre forme misure ritroveremo in qualunque altra manifestazione del lavoro digitale[42].

Davanti a tale problematica lavorativa ci troviamo ad osservare tre forme di estrazione del plusvalore da parte del capitale. In primo luogo, il capitale estrae plusvalore dal tempo di non-lavoro tra una consegna e l'altra; in secondo, l'estrazione avviene dal lavoro sottopagato in virtù delle politiche di deregolamentazione instaurate all'interno dei rapporti lavorativi. Infine, il ricavo del plusvalore risulta dall'estrazione dei dati, ad opera dell'algoritmo, nei confronti dei lavoratori e consumatori entrambi produttori di valore.

[40] E. Armano, D. Leonardi, A. Murgia, *L'ambivalenza della connettività dentro la macchina digitale. Una conricerca su soggettività e resistenza nel lavoro di piattaforma,* in *L'enigma del valore*, cit., pp. 107-121, p. 113.
[41] Ivi, p. 114
[42] R. Curcio, *La società artificiale. Miti e derive dell'impero virtuale*, Roma 2017, p. 81.

Proseguendo l'analisi delle varie forme di *Digital labour,* è doveroso soffermarsi sul *microlavoro* che consiste in una serie di micro-mansioni delegate agli utenti di portali come ad esempio Amazon Mechanical Turk. Generalmente, dato che i «lavoratori sono reclutati in massa si parla talvolta di lavoro della folla *(crowdwork)*»[43], e i committenti possono essere aziende, istituzioni pubbliche e individui privati.

> Questo tipo di digital labor e strettamente legato *Human-based computation*, ovvero la pratica che consiste nel delegare agli esseri umani operazioni che le macchine non sono in grado di eseguire da sole. Il microlavoro consiste nella realizzazione di piccole mansioni quali annotare video, smistare tweet, trascrivere documenti scannerizzati, rispondere a questionari online, correggere valori in un database […]. Gli utenti ricevono remunerazioni da pochi centesimi a qualche euro a singola mansione […] e le piattaforme prelevano una commissione su ogni transazione[44].

I compiti che i committenti esigono al 'turco meccanico' sono denominati *Human intelligence task* (Hit), compiti di intelligenza umana. Tali compiti «generano un *workflow,* (cioè) un flusso di lavoro in cui la partecipazione» della moltitudine intellettiva dei lavoratori «è sincronizzata, valutata e gestita da un sistema algoritmico di controllo. Questo sistema algoritmo amministra il gap tra il *workflow*, il flusso di lavoro ovvero la natura cooperativa del lavoro, la cooperazione, e il contributo dei lavoratori e delle lavoratrici come singoli: il lavoro vivo incarnato che è del tutto subordinato alla continuità e alla produttività del flusso»[45] del linguaggio informatico.

L'algoritmo, dunque, procede ad eseguire «una doppia astrazione del lavoro»[46]:

> da una parte estraendo specifici compiti dalle esperienze incarnate del singolo lavoratore e cancellando la soggettività all'interno del flusso, dall'altra riproducendo in modo frammentario ma cogente la misura del lavoro astratto come griglia normativa per la valutazione la remunerazione dell'attività umana[47].

Ci imbattiamo quindi nel concetto di alienazione marxiana descritto nei *Manoscritti del'44*: se da una parte i singoli lavoratori vengono esclusi dal flusso cooperativo, dall'altra, tale flusso non esisterebbe senza una combinazione di intelligenze.

Nelle micro-mansioni la forza-lavoro richiesta è prettamente di tipo linguistico-cognitiva, intellettiva, ed è accompagnata dal movimento del *digitus,* del dito. Il turco meccanico mette in rilievo il pensiero di Virno riguardo

[43] A. A. Casilli, *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?*, Milano 2020, p. 104.
[44] *Ibid.*
[45] S. Mezzadra, *Per la critica delle operazioni estrattive del capitale. Piattaforme digitali e cooperazione sociale*, in *L'enigma del valore*, cit., pp. 99-106, p. 105.
[46] *Ibid.*
[47] *Ibid.*

l'analisi delle varie forme della vita contemporanea[48]. Al lavoratore non viene richiesto un sapere specialistico o articolato, al contrario, basta che sia dotato semplicemente di una facoltà generale di linguaggio. L'attuale capitalismo informatizzato non mette a profitto le specifiche capacità linguistiche-cognitive, la conoscenza attualizzata e particolare, piuttosto, le *facoltà* di pensiero nella loro mera *potenzialità*. In questo modo, la conoscenza finisce per essere estirpata, frammentata e impoverita.

Infine, *Amazon Mechanical Turk* raggira scaltramente il rapporto lavorativo fino a rendere l'utente un consumatore che utilizza spontaneamente e in maniera ludica la piattaforma con un 'patto di partecipazione'. Risulta dunque arduo circoscrivere le micromansioni come lavoro a tutti gli effetti. La piattaforma riesce a valorizzarsi attraverso la cooperazione produttiva dei micro-lavoratori e la rivendita dei dati a terzi. Nel primo caso «quando un cliente pubblica una hit, il prezzo che paga ad Amazon è composto da una 'ricompensa' per i microlavoratori nonché da una commissione dovuta alla piattaforma»[49], che ammonta a circa il 20 e 40 per cento della somma versata ai Turker. Nel secondo caso, ogni piattaforma, rivende i dati dei suoi utenti al cliente che richiede un segmento specifico della popolazione, tale servizio aggiuntivo prende il nome di *Premium Qualification* e permette al cliente di decidere il genere, l'età, l'etnia e le abitudini dei Turker da utilizzare. È proprio in questo modo che si esplica l'utilizzo discriminatorio dell'algoritmo e che le piattaforme ricavano una rendita continua senza che i micro-lavoratori ne siano al corrente. Rendita che si fonda su politiche economiche fondate sulla delocalizzazione: le piattaforme *esternalizzano* le mansioni *direttamente al consumatore* che diviene linfa vitale per le piattaforme stesse.

> Ai consumatori si affidano delle operazioni di gestione dei dati e di calibrazione dell'algoritmo, per esempio taggando immagini o trascrivendo brevi stringhe di testo durante la navigazione quotidiana, e questo senza nessun scambio monetario ma attraverso applicazioni concepite appositamente per estrarre lavoro gratuito[50].

Pertanto, le operazioni compiute dai micro-lavoratori in cambio di una infima remunerazione «sono effettuate ogni giorno in maniera volontaria da centinaia di migliaia di utenti delle piattaforme gratuite. Insomma, una continuità tra questo lavoro esternalizzato, atomizzato […], e un altro lavoro, composto dalle stesse mansioni, dagli stessi gesti»[51], presentato come lavoro ludico; il microlavoro prende forma, dunque, nel terzo lavoro digitale: il *lavoro sociale in rete*.

[48] P. Virno, *Grammatica della moltitudine. Per una analisi delle forme di vita contemporanee*, Roma 2002.
[49] A. A. Casilli, *Schiavi del clic*, cit., p. 112.
[50] Ivi, p. 132
[51] Ivi, p. 136

Il lavoro sociale in rete, ovvero il lavoro di ognuno di noi compie navigando in rete, si svolge in piattaforme pubblicitarie come Facebook, Instagram o Google che rappresentano il paradigma più immediato per comprendere i modi di produzione e di accumulazione del capitalismo digitale.

Facebook lucra mediante la continua estrazione di dati che racchiudono i gusti, le abitudini, le inclinazioni e le relazioni delle identità virtuali che vivono nei distretti digitali. Questi dati sono raccolti per essere poi venduti a degli inserzionisti che li utilizzeranno per adibire vetrine virtuali per la vendita dei loro prodotti. Di conseguenza, il consumatore diviene un lavoratore produttivo, un *prosumer*, in quanto produce *valore di rete*, cioè dati digitali che poi vengono elaborati dalla tecnologia algoritmica per tramutarli in valore di scambio. Il consumatore produttivo diviene quindi il pilastro fondamentale il sistema produttivo capitalistico delle piattaforme. C'è da aggiungere che ogni volta che accediamo ai motori di ricerca e ai social network, i medesimi riescono «a far sottoscrivere agli utenti un contratto, iscritto implicitamente nelle condizioni d'uso, che può essere riassunto come segue»[52]:

> Se è gratuito, e perché in realtà siete voi non solo il prodotto, ma anche i lavoratori che, grazie ad un'attività collettiva, in apparenza libera e giocosa, mi permettete di produrvi e vendervi come una merce (fornendomi dati, i contenuti e, attraverso le economie di rete, la dimensione del mercato necessario per attirare gli inserzionisti)[53].

'Siamo' non solo produttori di valori di rete, ma diventiamo 'noi' stessi la merce venduta. Assistiamo, dunque, alla mercificazione totale della vita, ma quando si utilizza questo termine non possiamo attribuirgli un significato metaforico ma un valore prettamente economico. Valore che si struttura durante la riproduzione della vita, durante 'il tempo libero'. Siamo ben oltre la distinzione marxiana tra tempo di lavoro e tempo di vita, e tra lavoratore produttivo e improduttivo. Il lavoro gratuito del *prosumer* deve essere inscritto nel serbatoio del processo di valorizzazione dell'impresa-piattaforma. Il prosumer non ha coscienza di sé di produrre valore perché gli viene celato continuamente l'obiettivo delle piattaforme. La loro organizzazione è vicina alla descrizione elaborata da Gilles Deleuze di *società di controllo*, «cioè di un ambiente tecnico invisibile capace di seguirci ovunque e guidarci nelle nostre scelte, senza tuttavia mai dare l'impressione di costringerci veramente»[54]. Agiamo liberamente e gratuitamente al servizio dell'impresa in una sorta di *servitù volontaria*. Contribuiamo alla ricchezza delle piattaforme senza che questa ricchezza ritorni in termini di reddito.

In sintesi, attraverso queste analisi si è potuto osservare come alla base dell'economia delle piattaforme ci sia la macchina algoritmica, che trasforma i

[52] C. Vercellone, *Big-data e free Digital Labor del capitalismo delle piattaforme: un nuovo estrattivismo?*, cit., p. 16
[53] *Ibid.*
[54] Ivi, p. 21.

dati, contenenti conoscenza, informazione e individui in forma smaterializzata, in valore di scambio per gli inserzionisti. Il padronato si serve della cooperazione delle moltitudini cognitive per ricavare profitto. Siamo al cospetto di nuovi scenari di accumulazione capitalistica che hanno un impatto nocivo sulla natura[55] perché la rivoluzione e l'automazione algoritmica della produzione conduce ad un'estensione del lavoro che invade tutti i tempi di vita, i singoli gesti e le interazioni sociali degli individui.

4. Appropriazione del capitale fisso: ibridazione uomo-macchina?

Lo sviluppo del capitale fisso mostra fino a quale grado il *sapere sociale generale, knowledge,* sia diventato *forza produttiva immediata*, e quindi come le condizioni del processo vitale stesso sono passate sotto il controllo del *general intellect*, e rimodellate in conformità a esso. Fino a quale grado le forze produttive sociali sono prodotte, non solo *nella forma del sapere*, ma come organi mediati della prassi sociale del processo di vita reale[56].

Marx enuclea una tesi paradossalmente poco marxista: egli ha descritto in questo passo la caratteristica del nostro tempo. Il capitale fisso, sviluppandosi, produce un sapere sociale generale, che diventa forza produttiva immediata, e le condizioni del processo vitale della società sono controllate e modellate da un 'Intelletto Generale'. Quest'ultimo non è altro che capitale fisso, ossia la capacità scientifica oggettivata nel sistema di macchine. Dunque, Il motore del capitalismo contemporaneo risulta essere l'intelletto generale, cioè «una forma di intelligenza sociale collettiva creata con l'accumularsi della conoscenza della tecnica del sapere operativo»[57].

La preminenza del sapere, come principale forza produttiva dell'attuale produzione sociale, fa del tempo di lavoro una «base miserabile»[58], e relega il lavoro stesso, parcellizzato e ripetitivo, in una posizione residuale. In altri termini: le forze produttive assumono carattere *intellettuale* e questa precipua connotazione ridimensiona la ricchezza misurata attraverso la divisione del tempo di lavoro tra lavoro necessario e plusvalore. La legge del valore rischia così di depotenziarsi e sgretolarsi, in quanto l'intera società diviene produttiva, poiché, l'intelletto generale coincide con la sussunzione reale dell'intera società al capitale. La crisi del capitalismo, dunque, non è imputabile alla cosiddetta legge del valore e alla caduta tendenziale del saggio del profitto ma al divario sempre più ampio tra il lavoro vivo, che diventa autonomo dal capitale, e i processi di

[55] L'estrattivismo dei dati, considerati come il nuovo petrolio per le multinazionali, riconfigura anche l'assetto geopolitico e di accumulazione del capitale. Un esempio su tanti è lo sfruttamento di uomini e terra in Congo per estrarre materia prima come il coltan per fabbricare gli strumenti digitali quali ad esempio i computer e gli smartphone.
[56] K. Marx, *La scienza e le macchine*, Milano 2015, p. 23.
[57] M. Hardt, A. Negri, *Impero,* cit., p. 338.
[58] K. Marx, *La scienza e le macchine*, cit., p. 21.

misurazione che questo capitale stesso cerca di imporre al lavoro. L'algoritmo subentra come elemento di rottura, come prodotto del *general intellect*, come sua espressione fenomenologica. Nella produzione odierna il carattere cognitivo-linguistico, informazionale e comunicativo dell'intellettualità massificata, diviene l'elemento centrale uniformandosi al capitale fisso, in virtù del fatto che l'utensile, ovvero il mezzo di produzione, non è separato dal lavoratore ma è qualcosa che appartiene al suo corpo, più precisamente il suo cervello. In quanto utensile comune, il cervello, diventa la molla vitale dell'intero apparato produttivo postmoderno. Oggi a sostenere l'immaterialità del lavoro cognitivo c'è la cooperazione linguistica dell'insieme multitudinario alla cui base ci sono le facoltà generali del pensare. Il lavoratore non è più separato dai mezzi di produzione: è tutt'uno con essi.

Siamo ben oltre la dialettica dello strumento descritta da Marx ne *Il Capitale*, in cui il lavoratore aveva un rapporto subalterno e separato dallo strumento, il quale, viene utilizzato dai padroni come ricatto per estrarre plusvalore. Si può parlare allora di appropriazione di capitale fisso? Credo che la risposta possa darcela l'algoritmo. Quest'ultimo si configura come capitale fisso, come *macchina linguistica*, che si potenzia e rende autonomi, ma proprio la sua potenza e autonomia derivano dall'attività umana posta dietro. Dunque, la potenza dell'algoritmo può accrescersi solo attraverso il comportamento umano – un esempio vivido è la pratica del *machine learning* – mentre la sua autonomia risiede nella natura immanente della cooperazione del lavoro cognitivo-linguistico della moltitudine biopolitica[59]. Tuttavia, il capitale fisso, in senso marxiano, non ha la capacità di produrre autonomamente valore, ma ha bisogno sempre del lavoro vivo. La macchina – che traduce il linguaggio e la potenza cognitiva dell'uomo in bit informazionali e in codici numerici – non può esistere senza il lavoro vivo, anche al suo massimo grado di automazione raggiunta. C'è un rapporto di reciproca esistenza tra l'algoritmo e il lavoro vivo, tra il funzionamento del primo e la cooperazione dell'attività umana. Algoritmo e lavoro vivo si completano vicendevolmente. L'ibridazione uomo-macchina si è rafforzata «grazie soprattutto alla formulazione di algoritmi che sono in grado di evolversi continuamente sulla base degli input linguistici umani»[60]. Vediamo l'esempio posto da Matteo Pasquinelli riguardo l'algoritmo *PageRank*. Quest'ultimo ha la capacità di generare profitto attraverso la concentrazione di intelligenza in ogni *link*[61]:

[59] «La moltitudine post-moderna è un insieme di singolarità il cui utensile di vita è il cervello e la cui forza produttiva consiste nella cooperazione», A. Negri, *Kairòs, Alma Venus, Multitudo*, Roma 2000, p. 125.
[60] A. Fumagalli, *Metamorfosi del rapporto capitale-lavoro: ibridazione uomo-macchina*, «Effimera», 21.3.2017, p. 2.
[61] M. Pasquinelli, *L'algoritmo PageRank di Google: diagramma del capitalismo cognitivo e rentier dell'intelletto comune*, in F. Chicchi e G. Roggero (a cura di) *Sociologia del lavoro*, Franco Angeli, Milano 2009.

Ora, il *rank* di una web page è determinato dal numero e dalla qualità dei *links*, ed alta qualità significa un *link* ad una pagina che abbia essa stessa un alto *rank*. *PageRank* è dunque un meccanismo per incorporare il giudizio e il valore concesso dagli utilizzatori agli oggetti di Internet[62].

Se «le macchine industriali cristallizzano l'intelligenza passata in forma relativamente fissa e statica, (gli) algoritmi aggiungono continuamente intelligenza sociale ai risultati del passato, in modo da creare una dinamica aperta ed espansiva»[63]. Le macchine algoritmiche assorbono e incorporano continuamente intelligenza umana, c'è un flusso continuo generato da «legioni di uomini e donne davanti ai personal computer accesi in tutto il mondo»[64]. D'altronde a «differenza delle risorse naturali sommerse nella terra, gli input informazionali e i dati grezzi provenienti dagli utenti non sono uno stock», ma appunto un «flusso generato continuamente dalla loro attività»[65]. L'automazione algoritmica riesce a prosperare «grazie al valore prodotto da una forza lavoro»[66].

Dunque, la nuova generazione di macchine digitali si distingue dal macchinismo analizzato da Marx, divenendo l'interfaccia della vita e dell'intelligenze sociali, fuori e dentro i luoghi di lavoro, subordinando la vita ai meccanismi di valorizzazione capitalistica. Le macchine informatiche sono state, ormai incorporate, e la cooperazione del lavoro vivo, che dà corpo alle prestazioni linguistiche, non produce oggetti estrinseci e duratori, ma prosegue attraverso *attività senza opera*. Strumento e prodotto, mezzo e scopo sono identici, così come intenzione e realizzazione combaciano[67]. Tale analisi rimanda al concetto di soggettività macchinica degli autori francesi dei *Mille plateaux*, ovvero alla *combinazione attiva* tra la macchina algoritmica e le capacità linguistiche della forza lavoro. Precisamente, alla macchina tecnologica deve ampliare il concetto di *agencements machiniques*, cioè di assemblaggi macchinici, macchine assemblate e integrate tra loro e soggettività umane integrate a relazioni macchiniche, e macchine integrate dentro i corpi umani e nella società umana. Si può dire dunque che l'uomo postmoderno, è capitale fisso. Il vivente contiene in sé, allo stesso tempo, entrambe le funzioni: capitale fisso e capitale variabile.

Da un lato, l'attività umana e la sua intelligenza passate sono accumulate, cristallizzate come capitale fisso, ma dall'altro lato, rovesciando il flusso, i viventi umani sono capaci di riassorbire il capitale in se stessi e nella loro vita[68].

[62] A. Negri, *Appropriazione di capitale fisso: una metafora?*, «Euronomade», 2017 (3.3.), p. 5.
[63] *Ibid.*
[64] R. Ciccarelli, *Forza-lavoro, il lato oscuro della rivoluzione digitale*, Roma 2018, p. 24.
[65] C. Vercellone, *Big-data e Free Digital Labour del capitalismo delle piattaforme: un nuovo estrattivismo?*, cit., p. 9.
[66] R. Ciccarelli, *Forza-lavoro, il lato oscuro della rivoluzione digitale*, cit., p. 24.
[67] Cfr. P. Virno, *Lavoro e linguaggio*, in A. Zanini, U. Fadini (a cura di), *Lessico postfordista. Dizionario di idee della mutazione*, Milano 2001, pp. 181-185. *Lessico postfordista. Dizionario di idee della mutazione,* a cura di A. Zanini e U. Fadini, Milano 2001.
[68] A. Negri, *Appropriazione di capitale fisso: una metafora?*, cit., p. 3.

L'appropriazione del capitale fisso da parte delle soggettività lavorativa permette di comprendere lo statuto di dignità e di autonomia raggiunta dal lavoro. L'algoritmo dischiude forme di possibilità e di comprensione del nostro tempo. Nei processi produttivi post-industriale la cooperazione ha la possibilità di rendersi autonoma dai meccanismi di valorizzazione capitalistica, ma sottostà tuttavia ancora al suo comando. Dunque, si genera una situazione parossistica, dove «ad una continua (nel tempo) ed estesa (nello spazio) iniziativa autonoma produttiva, un'invenzione collettiva e cooperante, vengono subordinati all'estrazione di valore da parte del capitale»[69]. Si avverte una vera mutazione quando il rapporto tra processo lavorativo, in mano ai lavoratori, e il processo capitalistico di valorizzazione, «da sempre *formalmente* separati, cominciano ad esserlo *realmente*, affidando il primo all'autonomia del lavoro vivo, il secondo, al puro comando del capitale»[70].

L'algoritmo al quale si imputa la padronanza dei processi informatici di valorizzazione capitalistica è, non solo, come abbiamo visto, un dispositivo biopolitico volto al controllo e alla sorveglianza, attraverso la profilazione dei soggetti che navigano in rete trasformatisi in dati, ma anche una macchina, «che nasce dalla cooperazione dei lavoratori, dall'intellettualità logistica, e che il padrone impone sopra questa cooperazione»[71], sopra l'intellettualità massificata.

L'algoritmo si struttura dunque come una macchina padronale sul lavoro cognitivo e intellettuale. Tuttavia, dato il carattere cooperativo e autonomo della forza lavoro, e poiché i lavoratori si appropriano del capitale fisso, l'algoritmo diviene algoritmo conoscitivo, che verte cioè alla valorizzazione del lavoro capace di produrre linguaggi, di cui esso diverrà il *dominus*; ma tali linguaggi sono prodotti dai lavoratori, che ne posseggono la chiave e il motore cooperativo[72]. L'algoritmo e le tecnologie digitali, concludendo, dischiudono scenari di possibilità che consentono «l'autovalorizzazione della cooperazione sociale e della riproduzione della vita»[73], scenari distaccati, in qualche misura, dai circuiti di valorizzazione capitalistica.

Vincenzo Pallara
✉ vincenzopallara@yahoo.com

[69] A. Negri, *General Intellect e Individuo Sociale nei Grundisse marxiani*, «Euronomade», 20.5.2019, p. 4.
[70] *Ibid.*
[71] *Ibid.*
[72] *Ibid.*
[73] A. Negri, *Appropriazione di capitale fisso: una metafora?*, cit., p. 5.

Bibliografia

Fonti

Casilli, A. A. 2020. *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?*, Milano, Feltrinelli.
Ciccarelli, R. 2018. *Forza-lavoro, il lato oscuro della rivoluzione digitale*, Roma, DeriveApprodi.
Curcio, R. 2017. *La società artificiale. Miti e derive dell'impero virtuale*, Roma, Sensibili alle foglie.
Deleuze, G. 2009. *Foucault*, trad. it. di P. A. Rovatti e F. Sossi, Napoli, Cronopio.
Deleuze, G. 2019. *Pourparlers*, Macerata, QuodLibet.
Foucault, M. 2009. *Nascita della biopolitica. Corso al Collège de France (1978-1979),* trad. it. di M. Bertani e V. Zini, Milano, Feltrinelli.
Marx, K. 1976. *Il Capitale: Libro I, Capitolo VI inedito. Risultati del processo di produzione immediato*, intr. di G. Bedeschi, Roma, Newton Compton.
Marx, K. 2015. *La scienza e le macchine*, Milano, AlboVersorio.
Marx, K. 2018. *Il Capitale. Critica dell'economia politica. Libro I*, prefazione di E. Hobsbawn, a cura di A. Aiello, Roma, Editori Riuniti.
Napoleoni, C. 1972. *Lezioni sul Capitolo sesto inedito di Marx*, Torino, Boringhieri.
Negri, A. 2000. *Kairòs, Alma Venus, Multitudo*, Roma, Manifesto libri.
Negri, A. Hardt, M. 1995. *Il lavoro di Dioniso. Per la critica dello stato postmoderno*, Roma, Manifesto libri.
Negri, A., Hardt, M. 2001, *Impero,* Milano, Rizzoli.
Virno, P. 2001. *Lavoro e linguaggio*, in A. Zanini, U. Fadini (a cura di), *Lessico postfordista. Dizionario di idee della mutazione*, Milano, Feltrinelli, pp. 181-185.
Virno, P. 2002. *Grammatica della moltitudine. Per una analisi delle forme di vita contemporanee*, Roma, DeriveApprodi.

Letteratura Secondaria

Armano, E., Leonardi, D., Murgia, A. 2020. *L'ambivalenza della connettività dentro la macchina digitale. Una conricerca su soggettività e resistenza nel lavoro di piattaforma,* in *L'enigma del valore. Il Digital Labour e la rivoluzione tecnologica* (a cura di 'Effimera'), Casa della Cultura, Milano, pp. 107-121.
Fineschi, R. 2020. *Social e capitalismo crepuscolare (living in a box). Funzionamento e funzione dei social nelle dinamiche del capitalismo crepuscolare*, «La città futura» (30.10.2020).
Fumagalli, A. 2017. *Metamorfosi del rapporto capitale-lavoro: ibridazione uomo-macchina*, «Effimera» (21.3.2017).
Fiocco, L. 1997. *L'effetto kanban nell'organizzazione del lavoro alla FIAT di Melfi*, in «Chaos. Quaderni di riflessione e dibattito politico-culturale», n. 10, pp. 23-45.

Mezzadra, S. 2020. *Per la critica delle operazioni estrattive del capitale. Piattaforme digitali e cooperazione sociale*, in *L'enigma del valore. Il Digital Labour e la rivoluzione tecnologica* (a cura di 'Effimera'), Casa della Cultura, Milano, pp. 99-106.

Negri, A. 2012. *Spunti di 'critica preveggente' nel Capitolo VI inedito di Marx*, «Uninomade 2.0» (27.8.2012).

Negri, A. 2017. *Appropriazione di capitale fisso: una metafora?*, «Euronomade 2.0» (3.3.2017).

Negri, A. 2019. *General Intellect e Individuo Sociale nei Grunrdisse marxiani*, «Euronomade 2.0» (20.5.2019).

Pasquinelli, M. 2009. *L'algoritmo PageRank di Google: diagramma del capitalismo cognitivo e rentier dell'intelletto comune*, in *Sociologia del lavoro*, Milano, Franco Angeli.

Terranova, T. 2000. *Free labor: producing culture for the digital economy*, «Social text», Volume XVIII, Numero II, pp. 33-58.

Vercellone, C. 2020. *Big-data e Free Digital Labour nel capitalismo delle piattaforme: un nuovo estrattivismo?*, in (AA. VV.), *L'enigma del valore. Il Digital Labour e la rivoluzione tecnologica*, (a cura di 'Effimera'), Casa della Cultura, Milano, pp. 9-24.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771590

Articoli/3

Divenire classe

Per una critica del paradigma estrattivista applicato al capitalismo digitale

Roberto Ciccarelli 0000-0003-0788-725X

Articolo sottoposto a doppia blind peer review. Inviato il 07/01/2022. Accettato il 27/05/2022.

Becoming class. For a critique of the extractivist paradigm applied to digital capitalism

The article aims to investigate the historical reasons and the theoretical-cultural framework underlying the usual intellectual attitudes with which the theme of the pervasiveness of algorithms is addressed. In particular, the main topic is that of the 'algorithmization' of the historical-social world with attention to the phenomena related to work, the labour-power and the value extraction, within the field analysis of the economy of digital platforms. The main theoretical perspectives on this subject are analysed, i.e. the utopian, the disutopist and the extractivist one, with numerous insights into the specific literature, including that of surveillance capitalism (Zuboff, with Foucault in the background), a reconstruction of the so-called 'neoliberal revolution' of the post '68 years, and an investigation of the relationship between labour and capital as the logical centre of the critique of capitalism, according to the theories of Harvey and, basically, of Marx. Actually, a possible answer to the open questions in the first part can still be found in Marx's analysis of production, where the labour-power from the one hand is the working capacity sold on the market in exchange for a wage, on the other hand, it is the capacity of producing use values embodied in working subjects, and in a pre-logical sense in their 'corporeity'. Therefore, the essay examines the reason why it is necessary to take up this on-going contradiction, which allows us to conceive the metamorphosis and the change in the balance of power. The life of the human being thus appears structurally and historically to be crossed by new conflicts, starting with that between capital and digital labour, given that the assumption of equivalence between 'human nature' and 'data production' means imagining a form of subject caught in the work necessary to produce the data expropriated from capital.

***

Introduzione

Nel dibattito sull'automazione digitale esistono tre approcci: utopista, distopico e estrattivista. Queste prospettive possono essere usate in maniera intercambiabile sia a sostegno di una critica analisi del capitalismo digitale (e non solo di quello più conosciuto, in termini di cronache e *branding*,

delle piattaforme), sia in una sua critica radicale. La cartina di tornasole per riconoscere la parzialità e la capacità di sviluppare una prospettiva critica e una reale progettualità politica e trasformativa è il rapporto determinante tra capitale e forza lavoro, politica e natura.

Il dibattito in questione è basato sia sul capovolgimento della forza lavoro in 'capitale umano' da cui estrarre un 'surplus comportamentale', sia sulla trasfigurazione del valore prodotto dalla forza lavoro in dati. Questa duplice mossa è rilevata sia dalla critica dell'automazione digitale che dalla critica dell'estrattivismo anche se resta confinata a una forma di *ontologia informatizzata*, ovvero a un'idea che è il risultato di un'ipostatizzazione del capitale in una soggettività totalitaria che determina i destini del mondo in quanto manifestazione dell"Essere', cioè un concetto estenuato e obsoleto del modernismo reazionario filosofico che ispira anche una parte della critica del capitalismo digitale e in generale un'ampia parte della cultura contemporanea. In questo saggio mostreremo in maniera dettagliata, e alla luce della ricerca in corso[1], le ragioni storiche di questa concezione e una possibile alternativa.

1. La prospettiva (dis)utopista

Dalla crisi del 2008 è stato propagandato un discorso utopistico sull'automazione che ha legittimato il fatto che un'organizzazione del mercato del lavoro determinata da un'imponente ondata di riforme neoliberali avvenuta già nel decennio precedente era una realtà naturale basata sulla scarsità di posti di lavoro a tempo indeterminato, anche se questo dettaglio non è *mai* specificato da questi ideologi. L'attenzione si è soffermata in particolare sui lavori qualificati e ben pagati. La 'rivoluzione' in corso avrebbe portato a un aumento spettacolare della produttività e alla contemporanea diffusione della 'disoccupazione tecnologica' che annuncia un mondo senza lavoro[2]. Per un lungo

[1] Ho raccolto i risultati della ricerca in corso nella trilogia della forza lavoro in: R. Ciccarelli, *Forza lavoro. Il lato oscuro della rivoluzione digitale,* Roma 2018; Id. *Capitale disumano. La vita in alternanza scuola lavoro,* Roma 2018; Id. *Una vita liberata. Oltre l'apocalisse capitalista,* Roma 2022. Altri elementi utili sono emersi nella traduzione inglese di *Forza lavoro*, a cura di E. C. Gainsforth *Labour Power. Virtual and actual in digital production*, Londra 2021, in particolare nell'introduzione e nel primo capitolo in cui sono confluiti nuovi decisivi elementi. Rinvio anche a Id. *Il mito dell'automazione digitale*, in S. Bellucci (a cura di), *Ai-work. La digitalizzazione del lavoro*, Milano 2021; Id. *Il lavoro digitale*, in *La Grande Trasformazione. L'impatto sociale del cambiamento del lavoro tra evoluzioni storiche e prospettive globali*, a cura di Enzo Mingione, Annali della Fondazione Feltrinelli, LIV, Milano 2020.

[2] A titolo di esempio, ecco un piccolo florilegio pubblicato anche in Italia dalle maggiori case editrici: E. Brynjolfsson, A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Machines,* New York and London 2014; M. Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future,* New York 2015. Non va dimenticato il discorso strettamente connesso sul reddito di base svolto da A. Yang, *The War on Normal People: The Truth about America's Disappearing Jobs and Why Universal Basic Income is Our Future,* New York 2018. In questo caso si tratta di un'utopia libertariana molto diffusa tra i tecnofili e tra i sostenitori neoliberali del reddito di base che però non contempla una trasformazione dei rap-

momento l'opinione pubblica dei paesi capitalisti centrali è stata suggestionata, e irritata, da questa prospettiva che mescola le suggestioni apocalittiche sulla 'fine del mondo' con gli scenari prorompenti di una fantascienza coltivata dalle serie TV di un mondo pieno di automi che sostituiscono gli esseri umani o si ribellano alla loro sovranità. Non è accaduto, né accadrà nulla di quello che è stato profetizzato. Non solo le macchine resteranno al loro posto, ma il lavoro non è finito, la sua produttività non è aumentata, gli economisti continueranno a litigare sul concetto di 'disoccupazione tecnologica'. In compenso milioni di persone saranno messe al lavoro nella speranza di trovare un reddito o di integrarlo con il nuovo cottimo tecnologico creato dalle piattaforme digitali.

Nei casi delle tecno-(dis)utopie si dà per certo, coerentemente con il discorso teologico e finalistico che anima entrambe le prospettive, che i robot 'ruberanno' i lavori agli esseri umani. Nel caso specifico della tecno-distopia, cioè il rovescio dell'utopismo modernista tecnologico, si profetizza la scomparsa del lavoro in quanto tale, a cominciare dalle professioni come la medicina o il giornalismo e comunque gli impieghi più modesti. L'avvento del paradiso dei robot porterà l'inferno tra gli uomini anche se, per il momento, può 'aiutare' un'umanità generica a migliorare le 'performance' e la 'produttività' del lavoro, assicurando un aumento del benessere, non è mai però chiaro di chi. Di sicuro quello dei proprietari delle macchine e dei gadget pubblicitari che usiamo ogni giorno a cominciare da quelli presenti sugli smartphone o nei personal computer. È quello che stava accadendo nel momento in cui ho scritto questo saggio, cioè l'inizio del 2022. Apple aveva superato i tre triliardi di dollari di quotazione a Wall Street. Mai nella storia del capitalismo nessuna impresa ha raggiunto una quotazione simile, più alta del prodotto interno italiano. Il capitalismo digitale è in primo luogo l'effetto di una finanziarizzazione della vita nella quale si colloca anche la cosiddetta 'innovazione tecnologica'.

Perdere di vista questo legame strutturale significa accettare il discorso utopista in tutte le sue forme. Che sia un'utopia 'accelerazionista' (di 'sinistra', ma ci sono altrettanti casi di 'destra' a cominciare dal gruppo di teorici inglesi che ha rilanciato questo slogan tardo-futurista), o che sia una propaganda della Silicon Valley condita con profezie tecnologiche sulla storia dell'umanità, la conclusione è una sola: il lavoro precario non conta, non ha alcun significato, tantomeno un valore, ed è destinato a morire. Nessuno lo dice, ma tutti lo capiscono benissimo: la forza lavoro, le donne e gli uomini in carne e ossa, sono finiti. Il lato oscuro della fine del lavoro è occupato dall'apocalisse che colpisce tutti coloro che sono costretti a vendere la propria forza lavoro. Se il mondo è finito, non è solo inutile ribellarsi, ma è anche ingiustificato il desiderio di continuare a vivere. E il valore

porti di produzione, di proprietà e del mercato, oltre che dello stesso stato sociale, che sarebbero necessari per adottare, e rendere sostenibile, una simile misura dirompente a livello sociale e politico. Per una discussione sulle diverse opzioni politiche sul reddito di base cfr. A. Fumagalli, S. Gobetti, C. Morini, R. Serino, *Reddito di base. Liberare il XXI secolo,* Roma 2021; G. Allegri, *Il reddito di base nell'era digitale. Libertà, solidarietà, condivisione*, Roma 2018; P. Van Parijs-Y. Vanderborght, *Il reddito di base. Una proposta radicale*, Bologna 2017.

in borsa che straccia ogni record? Non è il prodotto anche degli esseri umani, ma degli automatismi impersonali che regolano gli algoritmi e operano in un universo separato che non ha nessun rapporto con l'attività umana.

Questa rappresentazione è l'esito di un'umanizzazione delle macchine e di una destoricizzazione del rapporto conflittuale e strutturale tra forza lavoro e capitale che sta alla base del capitalismo e si trasforma nei suoi diversi modi di produzione, che non vanno intesi in senso teleologico come fanno gli utopisti e i dis-utopisti. In questo caso la storia è stata concepita in base alla cronologia progressiva di 'rivoluzioni industriali' che si susseguono incessantemente dalla prima avvenuta nel XVIII secolo fino alla 'terza' (ma alcuni parlano di una 'quarta') che sta avvenendo oggi. Una simile interpretazione è uno degli effetti di una cultura teologica residuale eppure molto presente nel mondo della 'secolarizzazione'. Così come un tempo si pensava Dio in maniera antropomorfica, attribuendogli una personalità e una volontà umana nel governare le sorti di un mondo che ha creato a misura dei propri capricci, oggi gli *utopisti distopici* umanizzano le macchine e proiettano i propri desideri di onnipotenza negli algoritmi. In realtà, chi confonde l'intelligenza artificiale con l'ideologia dell'automazione totale parla di un ben più modesto *machine learning*: un apprendimento che le macchine realizzano mentre restano in tutto e per tutto dipendenti dall'agire e dal pensiero dell'essere umano. Ovvero: dalla sua forza lavoro. L'invisibilizzazione di questo dato strutturale, non troppo difficile da intuire anche nell'esperienza che facciamo tutti delle tecnologie digitali non è una rimozione, o l'occultamento di una 'verità', bensì un elemento fondamentale per permettere ai capitalisti di accrescere a dismisura il valore delle loro proprietà e la forza dei loro mezzi di produzione. Non si capisce nulla del nuovo modo di produzione digitale (che coesiste ed è immanente ad altri, e certo non li sostituisce) senza partire da questo assunto *politico*: il capitalismo è implicato in un rapporto sempre e comunque con la forza lavoro, ha tutto l'interesse a non riconoscerla e a strumentalizzarla per non pagare i costi della sua riproduzione, rendendo indiscutibili i suoi enormi profitti.

Nel pensiero della forza lavoro ho elaborato un altro modello interpretativo secondo il quale tale rivoluzione non cancella affatto il lavoro (la tesi modernista e romantica dei 'robot' che contraddistingue tutte le (dis)utopie tecnologiche), ma al contrario moltiplica il lavoro povero in tutte le sue forme coerentemente con uno specifico assetto del capitalismo nella sua versione neoliberale. Se è in atto un processo di sostituzione è quello che trasforma il lavoro subordinato in quello precario. tecnologia è soggetta ad un uso politico. Dunque l'automazione digitale svolge una funzione completamente diversa che può essere descritta nei termini seguenti:

- l'automazione digitale non cancella i lavoratori in carne ed ossa, ma rende invisibile la loro forza lavoro nella produzione;
- disloca il lavoro necessario in tutto il pianeta, soprattutto dove non è possibile osservarlo;

- nasconde il rapporto di subordinazione a un capitalista che possiede i mezzi della produzione: ad esempio, la piattaforma digitale scaricata sugli *smartphone*;
- moltiplica le occasioni per lavorare precariamente, sempre di più, in condizioni peggiori, al servizio delle piattaforme;
- non aumenta, né diminuisce la disoccupazione, ma gestisce quella esistente mettendo al lavoro precari e disoccupati al di là del rapporto di lavoro salariato;
- aumenta la produttività della forza lavoro, attraverso la continua attivazione degli individui che svolgono un lavoro *online* non riconoscibile dalle statistiche ufficiali sull'occupazione.

Queste sono le coordinate del discorso critico sull'automazione capitalista che persiste sebbene anche coloro che, adottando la vita (tecno)utopista, abbiano perso per strada l'idea marxiana per eccellenza: il capitale è un rapporto sociale di produzione, di potere e di vita. Da una decina d'anni, in coincidenza con la crescita dei primi movimenti dei lavoratori digitali in tutto il mondo (rider e driver per esempio), è avvenuta una svolta critica negli studi sul capitalismo digitale che ha permesso di iniziare a discutere ciò che sfugge, o è programmaticamente taciuto, tanto nelle apologie insipide sul digitale quanto nelle sue velleitarie critiche[3]. Viceversa la critica ha trovato una forza attingendo dalle lotte alcuni elementi di realtà, oltre che di verifica necessaria, anche se al momento non è riuscita ancora a rilanciare sia per l'evidente sproporzione delle forze in campo, sia per il contraccolpo provocato sulle stesse lotte dalle insidiose retoriche su un lavoro inteso come facoltà del capitalista umano che destina una quota del proprio tempo di vita alle piattaforme come se fosse un lavoratore autonomo (pur non avendo una partita Iva), o come se fosse uno studente, pur non avendo né l'età né i bisogni di ragazzi appena usciti dal liceo. Questa infantilizzazione di massa è un altro aspetto del discorso sul lavoro invisibilizzato ed è un elemento strutturale del funzionamento di moltissime piattaforme digitali, la parte commerciale di un dispositivo che collega aspetti diversissimi nello stesso modello di produzione.

[3] Tra i contributi decisivi alla critica del capitalismo digitale mi limito qui a citare, in ordine cronologico: R. Finelli, *Filosofia e tecnologia. Una via di uscita dalla mente digitale*, Torino 2022; J. E. Smith, *Smart Machines and Service Work. Automation in an Age of Stagnation,* London 2020; A. Benanan, *Automation and the Future of Work,* London and New York 2020; V. De Stefano-A. Aloisi, *Il tuo capo è un algoritmo: Contro il lavoro disumano*, Roma-Bari 2020; S. Abdelnour-D. Méda, *Les nouveaux travailleurs des applis,* Paris 2019; C. Vercellone *et alii* (a cura di), *Decode: data driven disruptive commons-based models,* Paris 2019; J. B. Ganascia, *Le mythe de la Singularité: Faut-il craindre l'intelligence artificielle?*, Paris 2019; A. Casilli, *En attendant les robots. Enquête sur le travail du clic*, Paris 2018; J. Prassl, *Humans as a service. The promise and Perils of Work in the Gig Economy*, Oxford 2018; N. Srnicek, *Capitalismo digitale. Google, Facebook, Amazon e la nuova economia del web*, Roma 2017; B. Vecchi, *Il capitalismo delle piattaforme,* Roma 2017; U. Huws, *Labor in the Global Digital Economy: The Cybertariat Comes of Age*, New York 2014; C. Fuchs, *Digital Labour and Karl Marx*, London 2013; T. Scholz (a cura di), *Digital Labor: The Internet as Playground and Factory*, London 2012. Andando ancora più indietro, non c'è dubbio che la critica del capitalismo digitale possa trovare molti elementi ispiratori in H. Ekbia, B. Nardi, *Heteromation and Its (Dis)contents: The Invisible Division of Labor between Humans and Machines,* «First Monday», vol. 19, no. 6, 2014*;* S. Zuboff, *In the Age of the Smart Machine,* New York 1989.

2. La prospettiva estrattivista

Nel caso dell'estrattivismo, quello di cui mi occuperò più estesamente in questo saggio, il problema della rivoluzione digitale è stato affrontato in maniera più sofisticata e diversamente problematica rispetto alle strade scelte dagli utopisti e dai disutopisti tecnologici. Innanzitutto è cambiato il registro: da quello allucinatorio della profezia si è passati all'iperrealismo. Dagli scenari di una trasformazione del pianeta in un macchinismo universale si è passati alla metafora dell'estrazione del petrolio. Questa immagine è stata trasposta anche nell'ambito specifico della produzione dei dati considerati il 'petrolio' da estrarre dalla carne umana che la produce attraverso particolari 'trivelle': le piattaforme che usiamo 24 ore su 24 per connetterci al mondo e gli altri, oltre che per lavorare. È lo stesso scenario raccontato in altri ambienti dove sono state adottate metafore come: 'la cultura è il petrolio' del XXI secolo. Siamo così proiettati in un universo mentale dove esiste un potere eminente e assoluto, una specie di monarca o di dittatura che violenta, stupra, sequestra per di più senza che i diretti interessati percepiscano (quando non sono conniventi con chi li sfrutta) l'esistenza di un simile concentrato di violenza. Eppure una 'trivella', sia pure nella forma accattivante e giocosa di una 'piattaforma', dovrebbe fare male. Una simile rappresentazione non va tuttavia intesa nei termini realistici del discorso estrattivista, ma appunto come una metafora di una concezione capitalistica dell'individuo il cui principale interesse è difendere l'integrità del suo corpo glorioso che coincide con un'unità impenetrabile e sacra di corpo e mente, di Io e Sé. Parliamo di quella formazione epistemica onnipresente nella morale e nella politica contemporanea che è stata rappresentata in termini filosofici come un trascendentale che oscilla tra la pienezza della proprietà di se stessi affermata da John Locke nel XVII secolo alla vuotezza di una 'nuda vita' affermata da Giorgio Agamben nel XXI secolo. Sono questi i termini di una soggettività 'libertariana', oggetto di speculazioni di origini diverse ma collimanti nella stessa rappresentazione del mondo all'interno del quale si è affermato anche il paradigma dell'estrattivismo nella discussione sul capitalismo digitale.

Ciò che unisce questa prospettiva al tecno-disutopismo è la rimozione della forza lavoro come elemento centrale, e oppositivo, in un rapporto sociale di produzione che è stato esteso attraverso un nuovo meccanismo dal capitalismo digitale. In questa concezione, in realtà, esiste un'idea precisa di forza lavoro considerata come una quantità fissa e misurabile, in termini marxiani una 'capacità di lavoro' determinata dal capitale non nei termini di 'manodopera' bensì come un sostrato biologico-psichico passivo e inerme e, allo stesso tempo, ricco di un'essenza insostituibile che coincide con la produzione di 'materiali grezzi' che l'industria estrattiva digitale sequestra e 'ruba' per incrementare il suo potere assoluto e i suoi sterminati profitti. Questa contraddizione in termini

attraversa molte teorie dell'estrattivismo, comprese quelle usate per spiegare il capitalismo digitale. La loro cifra comune è questa: pur avendo voluto politicizzare questa organizzazione della produzione e della società non sembra essere possibile riconoscere la politicità intrinseca di un sistema costruito sulla contraddizione tra forza lavoro e capitale. L'*impoliticità* del discorso estrattivista impedisce di prospettare soluzioni al dilemma di fondo: perché gli individui preferiscono obbedire liberamente a un meccanismo che li sfrutta? Senza volerlo e, anzi, partendo da una sua critica, queste teorie avvalorano la conquista di fondo dei suoi avversari: l'invisibilizzazione della forza lavoro di miliardi di persone che lavorano più di prima, e sempre peggio, anche su nuovi supporti ma hanno sempre meno potere nel rapporto costitutivo e conflittuale con il capitale.

Il limite impolitico di queste analisi impedisce di considerare il lavoro digitale nei termini di un rapporto sociale con il capitale, mentre invece è considerato solo come una relazione individuale con gli algoritmi. Il soggetto del lavoro digitale resta un individuo-consumatore, oppure una 'natura umana' obbligata a comportamenti contrari alla sua volontà 'autentica'. In questo caso si ritiene che senza il capitalismo digitale, e comunque con un capitalismo diverso, la 'natura umana' possa essere restituita a un ordine più rispettoso dell'essere umano. Questo non è possibile perché tutto il capitalismo, e non solo la sua forma digitale, ha trasformato la 'natura umana' nella propria forma ideale negando l'idea di un'essenza presupposta alle sue manifestazioni storiche. Dietro la natura alienata del lavoro non si nasconde un'essenza umana originaria, ma i rapporti sociali e di produzione che organizzano l'essere umano in uno strumentale regime di vita e produzione. Esiste anche un'altra tesi, correlata alla teoria dell'autenticità della natura umana, secondo la quale la vita, anche sulle piattaforme digitali, si darebbe integralmente, e senza alternative, come vita capitalistica. Questa conclusione nichilistica prospetta un sistema di dominio insuperabile. È necessario allora comprendere ciò che davvero il capitalismo digitale sfrutta, e ciò che permette di superare tale sfruttamento, contestualizzando le diverse interpretazioni esistenti. Da questa interrogazione dipende una rivoluzione culturale delle mentalità che potrebbe portare a un'alternativa al capitalismo digitale a partire dalla comprensione del suo funzionamento attuale.

Il cuore del capitalismo digitale non è il soggetto psicologico, la persona ridotta a una 'merce di se stessa' o l'algoritmo che impone all'individuo un comportamento programmabile e automatizzato. È la forza lavoro: una facoltà dell'essere umano, unica e insostituibile, l'elemento motore che è alla base della produzione del valore in un mutevole rapporto conflittuale con il capitale. Tale rapporto non è ontologicamente immutabile. Al contrario è storico, sociale e politico. Esso è configurabile nei termini di una contraddizione in atto nella quale si definisce la relazione tra il corpo e la mente di ciascuno e la vita sociale, politica ed economica in generale. Il capitalismo declina tale contraddizione secondo caratteristiche storiche, morali e culturali che si danno in un determinato momento e cambiano in un altro all'interno di un rapporto conflittuale tra capitale, forza lavoro, politica e natura. In sé la forza lavoro è la facoltà messa in

relazione capace di produrre i valori del mondo a cominciare da quelli di uso e di scambio, politici e morali. In essa può esprimersi una «personalità vivente»[4] di ciascuno al di là della nazionalità, del ruolo sociale, della condizione professionale e dello stato occupazionale. Questa facoltà è invisibilizzata, estraniata e reificata in una 'capacità di lavoro'. Così la forza lavoro, ridotta a una 'capacità di lavoro' mercificata, è presente in tutti i rapporti sociali di produzione esistenti, sia nei luoghi di lavoro sia nel tempo della riproduzione della vita. E tuttavia la forza lavoro non è mai riducibile alla sua mercificazione attuale. Non esiste un Capitale che decide, da solo e sovranamente, di generare il mondo e i suoi valori.

In questa cornice agiscono anche le piattaforme digitali che guadagnano enormi profitti grazie all'aumento esponenziale della produttività di una forza lavoro digitalizzata. La subalternità, e lo sfruttamento, ai quali quest'ultima è costretta sono forse la prova di un dominio assoluto a cui nessuno può sottrarsi? Non è affatto così. Questo sistema è fondato su una potenza che può condizionare, ma non possiede. I primi a saperlo sono i capitalisti digitali interessati a trarre un profitto dal modo in cui la forza lavoro si darà domani negli usi della sua facoltà di produrre i prossimi valori, e non solo quelli attivamente creati nel presente. Il progetto del capitale manca sempre di qualcosa. Per questa ragione è reinventato di continuo ed adeguato alla potenza agente della forza lavoro.

3. Il caso del capitalismo di sorveglianza

Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri di Shoshana Zuboff[5] è il concentrato ambivalente di tutte le teorie dell'estrattivismo e della conseguente critica del capitalismo qui definito nei termini di una 'sorveglianza' usata da imprese senza scrupoli, industrie militari, polizie ed eserciti di Stati che hanno oggettivamente favorito la crescita di un potere spaventoso che vampirizza con le sue 'trivelle' digitali ogni aspetto dell'esistenza umana, e di quella del pianeta. Questo capitalismo è descritto in cinque modi: 1. Un nuovo ordine economico che configura l'esperienza umana come una materia prima gratuita per pratiche commerciali nascoste di estrazione, predizione e vendita; 2. una logica economica parassita nella quale la produzione delle merci e dei servizi è subordinata a una nuova architettura globale della trasformazione comportamentale degli individui e delle masse; 3. una minaccia significativa alla natura umana nel XXI secolo così come il capitalismo industriale è stato per il mondo naturale nel XIX e XX secolo; 4. una violenta mutazione del capitalismo caratterizzata da una concentrazione della ricchezza, conoscenza e potere senza precedenti nella storia umana; 5. l'origine

[4] È la definizione di 'forza lavoro' data da Karl Marx, una delle più calzanti tra le diverse, non sempre conseguenti, che ha fornito, cfr. K, Marx, *Il capitale*, a cura di D. Cantimori, 1964, vol. I, p. 201; cfr. R. Ciccarelli, *Forza lavoro*, cit.
[5] S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma 2019.

di un nuovo potere strumentale che afferma il dominio sulla società e presenta una sfida impegnativa alla *democrazia di mercato.* La tesi estrattivista è descritta in questi termini: le metaforiche 'trivelle' dei vari Facebook, Google o Amazon trasformano l'esperienza da 'materiale grezzo gratuito' in valore economico. Tale materiale è estratto da un corpo, descritto come una 'carcassa', è raffinato, reso intelligente e trasformato in *dati comportamentali*. È un *processo di produzione* che ha lo scopo di produrre un *surplus comportamentale proprietario*. Il surplus è ottenuto attraverso processi di manifattura avanzata conosciuti come *intelligenza algoritmica* ed è potenziato attraverso la realizzazione di *prodotti predittivi* finalizzati all'anticipazione di ciò che vogliamo o desideriamo. I prodotti non sono scambiati sul mercato tradizionale dei *futures comportamentali*. Di solito questi *futures* sono titoli finanziari, contratti standard, e quindi negoziabili, scambiati in borsa sulla base di scadenze e prezzi prefissati. Il fatto che non esista una contrattazione tradizionale, innanzitutto tra i produttori diretti dei comportamenti e le aziende che estraggono e raffinano i dati, rende questa operazione arbitraria e speculativa.

L'economia dei *surplus comportamentali* non mira solo a soddisfare una domanda preesistente, ma a creare una nuova domanda e a farla riconoscere come naturale a chi ignora l'esistenza stessa della propria domanda. Questa è la principale innovazione realizzata da Google – il vero oggetto di indagine di Zuboff: «Google ha scoperto il modo per tradurre le interazioni non di mercato con i suoi utenti in un materiale grezzo eccedente finalizzato alle transazioni di mercato con i suoi veri clienti: i pubblicitari»[6]. È questo avanzamento tecnologico che ha permesso a questa azienda, come a tutte le altre digitali, di convertire gli investimenti in rendita. Questo ha cambiato il rapporto con gli investitori che hanno realizzato la possibilità di rendere monetizzabili i loro capitali di ventura. È stata l'invenzione del XXI secolo: il motore di ricerca di Google è un *mezzo di produzione* che allena i suoi utenti ad alimentare le capacità predittive dell'intelligenza artificiale. Google va considerata sia come un'interfaccia tra i pubblicitari e gli investitori finanziari, sia come un'interfaccia tra gli utenti che *lavorano* e i suoi algoritmi che li rendono produttivi per l'economia finanziaria in cui, senza saperlo, sono inseriti. Google non vende dati personali, ma previsioni su ciò che sentiamo, pensiamo, facciamo, scrive Zuboff. La vendita si basa sull'idea per cui l'essere umano è prevedibile, quantificabile, regolabile. Ciò avviene attraverso la definizione delle variabili di un comportamento che permette di ridurre i rischi di chi scommette sulla loro ripetizione. Questa operazione è fondamentale per chi deve vendere una polizza assicurativa, un farmaco, un'automobile, un corso universitario oppure un partito, un sito o un giornale. Un attore economico interessato a scambiare qualsiasi cosa non potrà mai più fare a meno di una simile tecnologia per commerciare un prodotto e scommettere sulla fortuna della propria impresa. E così faranno i singoli individui per programmare la loro esistenza.

[6] *Ibid.*

Il *ciclo economico del surplus comportamentale* è così composto: gli utenti delle piattaforme digitali sono trasformati in mezzi di produzione che fabbricano le previsioni sui loro comportamenti. Questi prodotti sono venduti ai clienti sui nuovi mercati dei *futures comportamentali.* Questi *futures* corrispondono a una *commodity*, il dato, che si aggiunge a quelli tradizionali del grano, oro, metalli, caffè. È su questa merce che gli investitori che finanziano la produzione di servizi predittivi e ottengono in cambio rendimenti che crescono proporzionalmente con il successo di tali servizi. Questa attività finanziaria non corrisponde a una valuta o un indice borsistico, ma condiziona il valore in borsa delle aziende che li creano. I capitalisti della sorveglianza possiedono aziende finanziarie, oltre che tecnologiche. Il ciclo produttivo rende evidente l'intreccio tra la *produttività dell'industria finanziaria* e la *finanziarizzazione dell'industria tecnologica*. La base comune di questa industria è il lavoro (e non solo il comportamento come sostiene Zuboff) degli utenti delle piattaforme digitali e le loro relazioni con la produzione e la riproduzione della società. La tecnologia è politica perché offre strumenti per potenziare il controllo, migliorare la previsione e indirizzare la decisione sulla vita delle persone.

Diversamente da altre teorie del capitalismo, quella della sorveglianza è fondata su un *ciclo di accumulazione continuo* che non avviene una sola volta per tutte. L'accumulazione dipende dallo sviluppo in tempo reale della vita stessa degli esseri umani intermediata dalle piattaforme digitali. Questa descrizione coincide con quella del filosofo e geografo David Harvey secondo il quale ci troviamo in un sistema di *accumulazione attraverso l'esproprio*. L'esproprio consiste nel sottrarre una serie di beni a costo vicino allo zero. Il capitalista vince quando riesce a guidare tale processo in territori nuovi e indifesi. È quello che ha fatto il capitalismo di sorveglianza. Il suo scopo è trasformare l'esperienza umana in *surplus comportamentale* sul quale il soggetto non ha alcun controllo. Tale *surplus* è considerato una *merce fittizia*, la quarta dopo la terra, il lavoro e la moneta secondo la tassonomia di Karl Polanyi.

L'espropriazione avviene attraverso la colonizzazione degli spazi di vita non commerciabili e la loro trasformazione in una vita capitalistica. Con una differenza: questo processo è continuo, include ogni azione, indipendentemente dai limiti spazio-temporali in cui è inserita. È il soggetto ad adattarsi, abituarsi e performare l'identità necessaria per rendere automatico il suo comportamento. Volontariamente si mette al servizio di un'economia algoritmica che mira a condizionare la sua forma di vita. Per questo il capitalismo di sorveglianza va inteso come un'*economia parassita*. Attraverso la progressiva colonizzazione del soggetto da parte del *potere strumentale* digitale si costruisce la mentalità e il lavoro produttivo secondo i criteri stabiliti. Non è soltanto uno sfruttamento praticato da un potere esterno, ma è l'adesione inconscia, e via via sempre più strategica, da parte del soggetto rispetto a ciò che lo sfrutta e nel quale intravede tuttavia l'unico orizzonte possibile. Se prima era estranea allo sfruttamento, ora lo considera la propria natura. Le piattaforme digitali sono uno straordinario strumento di trasformazione dell'individuo in *capitale umano* senza tra l'altro

renderlo del tutto cosciente del valore acquisito da tale capitale sul mercato dei *futures* comportamentali. Come *un parassita*, questa economia occupa il corpo e lo trasforma.

L'architettura globale di questo sistema è ispirata a un'economia di scopo: l'estensione colossale dell'attività di estrazione del valore dal virtuale al reale hanno una duplice funzione: da un lato, addestrano l'Io, l'umore, le emozioni e la personalità alla nuova disciplina performativa; dall'altro lato, cambiano il mondo in cui vive il *performer* in un ambiente tecnologico e interconnesso capace di favorire la raccolta e la produzione di nuovi dati. Obiettivo del *potere strumentale* è intervenire in maniera diretta, e senza mediazioni, sulle azioni delle persone al fine di valutarle, premiarle o punirle affinché seguano autonomamente gli obiettivi che permettono alla macchina di aumentare il suo valore. Secondo obiettivo: *rendere certa l'esecuzione e la programmazione dei comportamenti.*

L'imperativo *estrattivo* è accompagnato da quello *predittivo*. Tutto deve essere prevedibile e calcolabile, l'incertezza e il rischio vanno costantemente ridotti alla computazione. Il soggetto va trasformato in uno *strumento umano*. Lo può diventare solo se si convince liberamente che *l'automazione dei comportamenti coincide con la libertà assoluta*. In questa forma paradossale della libertà – libero è colui che è strumento in mano ad un altro – emerge la principale caratteristica della *libertà neoliberale* definita da Michel Foucault come *libertà liberogena:* produce libertà e la distrugge in nome della sicurezza. Una contraddizione in termini: un essere umano è libero quando la sua libertà è funzionale alla creazione del controllo sulla sua vita.

4. Ripensare la forza lavoro

La critica dell'economia comportamentale del capitalismo di sorveglianza fatta da Zuboff si ferma alla constatazione per cui il soggetto produttore di dati è un capitale umano e non una forza lavoro. Zuboff dà per scontato che la natura umana coincida con un'essenza, e che questa essenza dev'essere restituita alla sua autenticità, mettendo fine allo sfruttamento di cui la natura umana è vittima. Questa 'essenza' è un'idea astratta unica per tutto il genere umano che si riproduce in ogni individuo e resta identica a se stessa. Zuboff sembra così presupporre che senza il capitalismo della sorveglianza, o comunque con un capitalismo diverso, la natura umana possa essere restituita a un ordine più simile all'essenza dell'essere umano. Così non può essere perché *tutto* il capitalismo, e non solo la forma specifica di quello di sorveglianza, ha trasformato questa natura nella propria forma ideale e nega l'idea di un'essenza presupposta alle sue manifestazioni storiche. Dietro la natura alienata non esiste un'essenza umana incontaminata, ma i rapporti sociali e di produzione.

Evocare l'esistenza di una natura umana significa accreditare la tesi dei teorici neoliberali secondo i quali il capitale umano esprime la personalità del soggetto. In questo modo la natura umana coincide con ciò che la sfrutta: il capitale (umano). Per evitare simili esiti paradossali è necessario descrivere

diversamente ciò che il capitalismo della sorveglianza sfrutta e ciò che permette di superare tale sfruttamento. La risposta la troviamo nella filosofia della forza lavoro. Per Karl Marx la forza lavoro è un doppio: da un lato, è la capacità di lavoro venduta sul mercato in cambio di un salario; dall'altro lato, è la facoltà di produrre i valori d'uso incarnati nella 'personalità vivente' e nella 'corporeità' di ogni donna e uomo. Questa contraddizione in atto rende straordinariamente mutevole la vita dell'essere umano, soggetta a nuovi conflitti, a cominciare da quello tra capitale e lavoro che inizia sin dalla definizione della sua forma di vita. Quando allora sentiamo dire che il capitale umano oggi consiste nel produrre i dati e che questi dati sono *il nuovo petrolio* estratto dalle persone proviamo a rispondere che a questa convinzione sfugge l'idea che la forza lavoro, in quanto facoltà, può produrre *anche* i dati, ma non è riducibile alla mera capacità di produrre una sola merce.

Affermare l'equivalenza 'natura umana = produzione di dati', significa individuare esclusivamente il soggetto *nel lavoro necessario a produrre i dati espropriati dal capitale.* Questa reificazione è coerente con la rappresentazione capitalistica del lavoro: da un lato, sembra la fonte della ricchezza – e non lo è: lo è il capitale –; dall'altro lato, sembra la liberazione dallo sfruttamento, mentre invece ne è l'incarnazione. Nel capitalismo *il lavoro è sempre una merce, altrimenti non è lavoro*. Lo statuto ontologico dei dati prodotti e commercializzati dalle piattaforme digitali è invece l'espressione della duplice condizione della forza lavoro: da un lato, sono l'espressione di una facoltà irriducibile al ciclo produttivo, da cui questo stesso ciclo dipende; dall'altro lato, sono il prodotto di una capacità di lavoro – tra l'altro mai riconosciuta come tale – modellata intorno ai principi di produttività stabilita dal capitalismo della sorveglianza.

La funzione della forza lavoro è evidente: Facebook non esiste se nessuno di noi 'scrolla' il suo telefonino; Uber non esiste se nessuno guida una macchina o la chiama attraverso la app; Google non esiste se nessuno accetta di farsi profilare attraverso il motore di ricerca. I dati non sono l'origine della ricchezza, ma sono il prodotto della forza lavoro e dei rapporti sociali e di produzione in cui è inserita e da cui è estratto un valore. Il riferimento non è il dato in sé, ma la forza lavoro che produce il dato in un rapporto sociale di produzione. L'attività di estrazione non va considerata in maniera unilaterale, come avviene nell'estrazione del petrolio dove una trivella scava e una pompa estrae l"oro nero'. Questo è un modo per ridurre l'attività della forza lavoro a un unico comportamento produttivo, quello deciso da un potere assoluto che azzanna un trascendentale privo di soggettività e considerata un essere senziente privo di un pensiero e di un'azione che non sia quella predeterminata dalla macchina capitalistica.

Il capitalismo della sorveglianza vuole predeterminare, guidare e controllare a distanza sia le capacità che gli atti necessari alla produzione del valore che permette di accrescere i suoi profitti. Ciò non significa che controlli la forza lavoro e le sue potenzialità. Al contrario, il dispositivo è sempre in difetto e la sua straordinaria forza di innovazione è dovuta all'obbligo di approntare meccanismi predittivi che determinino il modo in cui la forza lavoro si configurerà domani.

Ma visto che nessuno può predeterminare il futuro, nemmeno un mercato cibernetico automatizzato, resta il fatto che la forza lavoro non è mai determinabile in partenza, né può essere codificata una volta per sempre in un 'capitale umano' e nemmeno in una 'natura umana'. La forza lavoro è la potenzialità dell'essere altrimenti ed emerge nella sua facoltà di collocarsi nel futuro. La forza lavoro è la facoltà che si infutura a partire dall'ora e dall'adesso in ogni atto materiale e intellettuale, nella produzione e nella riproduzione delle merci e delle relazioni, degli usi e delle loro contraddizioni.

5. Decostruire il mito del potere assoluto

Zuboff ha redatto una tavola comparativa dove distingue le caratteristiche del potere digitale da quelle del potere totalitario del XX secolo. Mentre il primo mira al possesso totale dell'esistenza, il secondo impone la certezza totale e sviluppa la proprietà dei mezzi per la trasformazione comportamentale dell'esistenza, per la computazione e il controllo dei suoi dati in un'ottica di subordinazione politica. Il potere strumentale non vuole dominare, ma *controllare a distanza e in maniera molecolare il libero sviluppo dell'esistenza.* La violenza di un simile potere è considerevole perché tende a presentarsi come democratico, mentre nasconde un apparato di controllo coercitivo. Tuttavia la nozione di 'potere strumentale' va precisata. Zuboff tende a considerarlo l'unico soggetto attivo nell'economia digitale, mentre rischia di ridurre il suo soggetto una 'natura umana' considerata come una 'foresta vergine', altre volte come 'oggetto'. Ne deriva l'impressione che tale potere sia assoluto, mentre all'opposto è condizionato dall'attivazione della forza lavoro dei suoi soggetti. La stessa Zuboff fonda il suo libro sulla storicità del potere e sulla trasformazione dei cicli capitalistici. L'idea della produzione di *futures comportamentali* è contraria all'immagine di un soggetto passivo, o semplicemente vittima di una macchinazione da cui risulta estraneo.

In questa economia il soggetto non è un 'oggetto', né tanto meno una 'foresta vergine'. Ciò che differenzia il nuovo potere da quello precedente è il fardello del soggetto neoliberale: la volontaria oggettivazione dell'Io finalizzata alla brandizzazione della persona e alla sua messa in produzione. Una strategia diversa dal desiderio di identificazione e conformità al comando di cui è stato accreditato il soggetto del totalitarismo. L'esperienza storica del nuovo soggetto è connessa alle macchine digitali che lo sfruttano sul mercato, non solo e non tanto al comando politico dello Stato. La ragione della crisi risiede nella libertà, non nel bisogno; nella volontà, non nell'obbligo; nella frustrazione, non nella coercizione. Queste osservazioni sono decisive per spiegare il neoliberalismo oggi. La sua agenda politica non univoca, e conflittuale, ha agevolato la creazione della gigantesca concentrazione di potere e ricchezza criticata da Zuboff. Una concentrazione accettata e talvolta persino rivendicata dai soggetti che sarebbero le sue vittime.

È il problema della politica contemporanea nata con la rivoluzione *neoliberale* a partire dalla metà degli anni Settanta del XX secolo, in una feroce

reazione al ciclo rivoluzionario mondiale che, in accordo con una ricca letteratura di studi storici e politici, abbiamo rappresentato nei termini di una 'costellazione', quella del Sessantotto[7]. La strutturazione di questa nuova 'rivoluzione passiva', o contro-rivoluzione preventiva e conservatrice-restauratrice, ha modificato anche le coordinate culturali e politiche del *liberalismo politico storico*. Oggi nel *nuovo liberalismo politico* definito con la categoria ampia e conflittuale di 'neoliberalismo' l'uso formale di concetti come libertà, mercato, democrazia o stato di diritto costituzionale non corrisponde mai alla loro realtà materiale, come scriveva Karl Marx già a metà Ottocento, a proposito del liberalismo economico borghese. In sé la 'democrazia' o la 'libertà' sono concetti polivalenti che alludono a un regime democratico regolato da un equilibrio tra l'interesse pubblico e privato, tra lo Stato e il mercato. Tuttavia questo equilibrio è sempre precario a causa del capitale. Capitale e democrazia sono politiche opposte che possono trovare una mediazione mai scontata. L'unica forma politica in cui la democrazia trova oggi uno spazio è la sua negazione: il mercato. Questo rovesciamento è in atto negli Stati Uniti, come in Russia o in Ungheria dove esistono capi di stato che teorizzano le presunte virtù delle 'democrazie illiberali'. Qualsiasi cosa siano tali democrazie, incarnano l'utopia del mercato e della concorrenza, a cui aggiungono il razzismo di stato. L'anti-liberalismo estremizza le premesse *neoliberali* di una società organizzata sul mercato e il suo darwinismo sociale. Zuboff spiega l'internità ideologica del capitalismo della sorveglianza allo 'spirito del tempo' neoliberale, ma sembra addebitare il nuovo autoritarismo solo a Google o Facebook, non al fatto che il neoliberalismo potrebbe arrivare anche a immaginare di sostituire la democrazia con una nuova forma di autoritarismo, e comunque con un tipo di governo ademocratico e tecnocratico che oscilla tra l'oligarchia e l''espertocrazia'. Il processo è velocizzato da queste aziende, ma non è stato causato solo da loro.

In questa prospettiva l'idea di un 'colpo di stato dall'alto', cioè di 'un sovvertimento della sovranità del popolo' compiuto dal capitalismo della sorveglianza, è parziale e profondamente equivoca. Presuppone il ripristino di una 'sovranità del popolo' in una democrazia dove il capitalismo contemporaneo non si adegua mai alle leggi, ma cerca di modificarle in funzione della propria egemonia, ricorrendo a uno *stato di emergenza continuo* per mantenere la concentrazione di ricchezze verso l'alto e aumentare le diseguaglianze economiche, politiche e culturali. Il problema della democrazia resta sempre quello del superamento del capitalismo. Ma questa non è certamente la questione di Zuboff che riconduce la critica della democrazia neoliberale al ripristino delle garanzie costituzionali della democrazia liberale contro un potere assoluto e incontenibile che sarebbe diventato tale usando tutte le leggi, e le contraddizioni, di quella democrazia che dovrebbe limitarlo in un futuro remoto.

[7] R. Ciccarelli, *Una vita liberata. Oltre l'apocalisse capitalista*, cit., primo capitolo: *Come siamo arrivati qui*.

La difesa di una *democrazia di mercato* contro il capitalismo di sorveglianza è una contraddizione. Da un lato, si argomentano le ragioni a difesa della democrazia costituzionale americana fondata sia sui diritti fondamentali della persona che sulla democrazia del mercato e, per questo, denuncia l'uso opportunistico del primo emendamento da parte degli ideologhi *cyberlibertariani*, dell'estrema destra e degli imprenditori della Silicon Valley. Dall'altro lato, si riconosce che la strumentalizzazione della 'democrazia del mercato' è la ragione stessa del capitalismo della sorveglianza ed è ciò che nega la mediazione tra democrazia e mercato.

Il capitalismo della sorveglianza, cioè l'apice di tutte le industrie estrattive del mondo, segna la fine della mediazione e afferma un'evidenza: *il mercato non è mai democratico*. Mark Zuckerberg o Eric Schmidt di Google si presentano come difensori della 'democrazia'. E tuttavia, sostiene Zuboff, affermano un totalitarismo in nome del mercato. In questo paradosso, la democrazia di mercato non è il rimedio, ma è la causa del regime attuale. Il problema non è solo logico, ma politico. La straordinaria capacità di Zuboff di analizzare il nuovo capitalismo rischia di configurare un soggetto vittima di una macchinazione universale che rende impossibile un ribaltamento dei rapporti di forza. Non può che essere così, restando sul terreno di una 'democrazia del mercato' da difendere contro il vampirismo dei capitalisti digitali. La stessa autrice sembra accorgersi, al termine del suo libro, dell'esistenza di un problema nel paradigma che ha adottato. Per questa ragione a un'altra idea della libertà rispetto a quella liberale articolata in una libertà negativa ('libertà da') e in una 'libertà positiva' ('libertà di'). Nella conclusione cita Hannah Arendt, nume tutelare della sua opera: *la libertà è una facoltà* , ovvero la sorgente segreta di tutte le attività umane. Per *facoltà della libertà* non si intende solo la capacità di compiere un'azione, o resistere contro un'azione altrui, ma attingere alla possibilità di nuovi inizi a cui tutti possono avere accesso perché *tutti possiedono la facoltà di iniziare una nuova vita al di là dei limiti imposti dal potere.* Non tutti sono capaci di farlo, perché il potere cerca di non lasciare mai una simile possibilità. La libertà resta comunque una facoltà dell'essere umano. La coscienza di una simile possibilità dà fiducia e forza. Non è un principio morale, ma un modo di vivere. Un'etica, non un partito preso ideologico. È, appunto, una facoltà ed è imprevedibile. È indocile, per questo il potere della sorveglianza cerca di controllarla ricorrendo a un mostruoso apparato tecnologico che tende a sostituire la sua facoltà con l'imperativo della prevedibilità e dell'(auto)controllo. Ma ciò non elimina il senso politico di questa facoltà: *il diritto al tempo futuro* (*The right to future tense*). È una definizione importante: la libertà è collocarsi in un mondo in divenire che non è *ancora così* ma che *può essere altrimenti*. È su questa facoltà *comune agli esseri umani* che si basa il *miracolo dell'inizio*. È un miracolo storico, non teologico: una discontinuità reale, non uno stato dell'anima. La facoltà della libertà non è una speranza, ma una prassi che consiste nel ribaltare i limiti e usare il reale a disposizione per altri fini. Il suo grado zero consiste nel riconoscere il modo in cui non vogliamo più vivere. Il passo successivo è comprendere come è possibile sconfiggere il potere

che occulta l'alternativa. Lo si può fare comprendendo la facoltà della libertà nella filosofia della forza lavoro. La forza lavoro è il centro della produzione del capitalismo di sorveglianza, ma è anche la facoltà della libertà, la capacità di produrre nuovi inizi e di rivendicare il diritto al tempo futuro.

6. Il problema dell'*accumulazione attraverso l'esproprio*

Abbiamo già segnalato come la teoria del capitalismo di sorveglianza sia stata concepita a partire dal concetto fondamentale coniato dal geografo marxista David Harvey, l'accumulazione attraverso l'esproprio (*accumulation by dispossession*). Alla luce della critica del concetto adottato da Shoshana Zuboff ora bisogna analizzarla più da vicino e a fondo. L'accumulazione attraverso l''esproprio' è una riformulazione problematica del concetto originario di Marx, quello dell'*accumulazione primitiva*[8]. Il problema di questa prospettiva è la rimozione del rapporto politico e sociale tra forza lavoro e capitale a favore di un rapporto di puro dominio ispirato alla violenza, la cifra stessa del paradigma estrattivista che caratterizza anche molte analisi sul neoliberalismo considerato un sinonimo di un capitalismo predatorio ormai privo di ogni forma di mediazione politica, la stessa che tuttavia si invoca quando si abbozza un'ipotesi di resistenza e limitazione del suo potere sterminato. Nella prospettiva marxiana l'interpretazione di Harvey è priva di una teoria dello sfruttamento alla quale ha sostituito una dell'esproprio. Lo stesso autore ne è consapevole e ha cercato di ripensare il proprio paradigma.

Harvey ha parlato di «una rivalutazione generale del ruolo continuo e della persistenza delle pratiche predatorie dell'accumulazione 'primitiva' o 'originaria' nella lunga geografia storica dell'accumulazione del capitale»[9]. Tale rivalutazione ha richiesto, a suo avviso, un cambio di terminologia. «Poiché sembra strano chiamare un processo in corso 'primitivo' o 'originale'», Harvey ha deciso di «sostituire questi termini con il concetto di accumulazione per esproprio»[10]. Da qui nasce anche l'uso che ha fatto Zuboff del concetto: l'idea di un esproprio continuo del materiale grezzo dei dati raffinato poi in 'futures comportamentali' è una parafrasi di Harvey. In termini marxiani questo è un paradosso poiché si considera la forza lavoro solo nella misura in cui è una 'merce umana'. Tutti i suoi 'comportamenti' non sono altro che il prodotto di un'immanenza al capitale anche quando non sono stati ancora 'estratti' e trasfigurati ontologicamente ('raffinati') in un'altra merce. Questo controsenso contraddice tra l'altro la dialettica tra mercificazione e demercificazione che tuttavia Harvey concepisce in termini *spaziali*, legati alla privatizzazione dei beni comuni, all'appropriazione

[8] Qui presento un'analisi incrociata dei seguenti volumi di David Harvey: *The Limits to Capital*. Londra 1982 (2003); Id., *The New Imperialism*, Oxford 2003; Id., *A Brief History of Neoliberalism*, Oxford 2005; Id., *Spaces of Global Capitalism*, Londra 2006; Id., *Seventeen Contradictions and the End of Capitalism*, Oxford, 2014.
[9] D. Harvey, *A Brief History of Neoliberalism,* cit., pp. 158 e ss.
[10] *Ibid.*

della terra da parte di latifondisti o multinazionali interessate all'agricoltura, all'acqua o alle materie prime, in generale ai problemi della redistribuzione dei redditi o della ricchezza. Se tutto è già contenuto teleologicamente nel capitale com'è possibile che quest'ultimo faccia partire il motore dell'accumulazione continua per esproprio? Se non c'è nulla *fuori* da un rapporto con il capitale, e tutto è mercificato al suo *interno*, non è possibile nemmeno prospettare una 'demercificazione', e dunque un divenire altrimenti da ciò che si è: sfruttati, e non solo spossessati di tutti gli attributi umani irriducibili allo stato di merce.

Rimuovendo l'idea per cui il capitale è un rapporto, e non un soggetto, Harvey nega fondamentalmente l'esistenza della politica, oggettiva il potere e lo colloca nel solo capitale. Viene così meno a una delle tesi fondamentali di Marx secondo il quale il capitalismo è fondamentalmente un sistema che presuppone la circolazione, e che il processo di circolazione, per quanto possa comportare la ridistribuzione della ricchezza tra i proprietari di merci, non può causare da solo la produzione di nuovo valore o di nuovo capitale a livello del sistema nel suo complesso. Per 'espropriazione' Harvey intende l'appropriazione di un valore presupposto, ma non materialmente prodotto e ottenuto tramite lo sfruttamento di una forza lavoro che resta una somma di individui dispersi determinati in maniera unilaterale dalla volontà di un 'capitale' in sé già dato che trasferisce un valore oggettivato da un capitalista all'altro, una quantità fisica che accresce la ricchezza. In questa descrizione si perde di vista l'idea di Marx secondo il quale il semplice trasferimento di beni da un capitalista all'altro non può causare un aumento della forma capitalistica di ricchezza a livello del sistema nel suo complesso. La classe dei proprietari capitalisti non può essere destinataria netta di beni, né può sempre defraudare se stessa pur essendo in permanente competizione al suo interno come al suo esterno. La somma dei valori in circolazione non può chiaramente essere aumentata da alcun cambiamento nella loro distribuzione ma dalla produzione di un nuovo valore attraverso una diversa organizzazione della produzione e dunque degli elementi che sono ancora esterni al suo circuito, a cominciare dalla forza lavoro e dal rapporto con il mondo e l'ambiente che muta a seconda delle circostanze e delle contraddizioni che non sono contenute nel capitale ma che il capitale cerca di sussumere volta per volta e al di là di un piano che non può essere predefinito in partenza. Questo problema è al centro della dialettica tra mercificazione e demercificazione (cioè il problema della sussunzione/liberazione) e contraddice l'idea di Harvey per cui il capitalismo moderno è un capitalismo basato prevalentemente sulla frode e sulla predazione e non anche sulla creazione del plusvalore formalizzato in rapporti politici e giuridici che legittimano e regolano lo sfruttamento sulla base di precisi rapporti di proprietà inseriti anche nelle carte costituzionali degli Stati di diritto.

È in ragione di questi problemi che Sandro Mezzadra e Brett Neilson hanno ripensato il problema dell'estrattivismo 'in maniera allargata' al fine di «individuare le logiche che animano […] una moltiplicazione ed eterogeneizzazione dello spazio

e del tempo in essa impliciti»[11], pensando l'estrattivismo *con* lo sfruttamento. Per farlo è tuttavia necessario tornare a pensare il capitalismo «come relazione sociale»[12]. «Non vogliamo ridurre le operazioni estrattive al prelievo di materiali grezzi e forme di vita dalla superficie, dalle profondità e dalla biosfera della Terra»[13], scrivono. Questa 'riluttanza' si spiega con l'impossibilità di considerare ogni dimensione della vita, così come ogni spazio della terra, come una metafora della *terra nullius*, il principio giuridico usato in molte società coloniali per giustificare spesso retrospettivamente l'occupazione delle terre dei nativi sulla base del fatto che esse venivano pensate come disabitate. L'analisi è condizionata dall'idea per cui il capitalismo sia un imperialismo che espropria le proprietà comuni possedute da popolazioni aggredite, in particolare nell'America Latina e in Africa, e sia l'espressione degli interessi capitalistici legati allo sfruttamento delle materie prime.

La metafora della *terra nullius*, come abbiamo visto nel caso dei processi di *data mining* e *data harvesting* descritti da Shoshana Zuboff, non può essere valida universalmente per ogni manifestazione dell'esistenza. Dato che non esiste una 'nuda vita' allora non esistono gli 'spazi vuoti' in cui un potere imperialista del capitalismo assoluto può installarsi sterminando le popolazioni o disintegrando i corpi e le coscienze con le sue operazioni 'estrattive'. L'aspirazione del capitalista di vedere una *tabula rasa* nel mondo si scontra con gli insuperabili ostacoli posti dalla realtà, e non solo da una qualche forma di resistenza organizzata. È sempre più difficile organizzare un circuito di produzione che garantisca una produttività e un profitto crescente in un sistema attraversato da crisi a ripetizione che incidono sulle catene globali del valore sempre più lunghe, e dunque sempre più fragili perché esposte agli imprevisti e alle reazioni delle crisi pandemiche, naturali, sociali o economiche nel mercato mondiale. Lo spazio geografico, come la soggettività della forza lavoro, è ineliminabile e enormemente complesso. La volontà di potenza che Harvey, e i teorici dell'estrattivismo, attribuiscono al Capitale inteso come Soggetto, non può eliminare tutto questo anche se sono in molti a temere, o a sperare, che un giorno possa farlo. Su questo punto, da diverse prospettive, i sogni dei tecno-(dis)utopisti sono l'incubo degli estrattivisti. Entrambi condividono un'immagine totemica del potere.

7. Il capitale è un rapporto sociale

Kate Crawford, una delle interpreti più interessanti del paradigma estrattivista adattato al capitalismo digitale, ha ripensato il capitalismo digitale come *capitale integrato* che connette estrazione, logistica e produzione materiale e immateriale oltre che finanziaria. L'autrice ha evidenziato il problema di tutte

[11] S. Mezzadra, B. Neilson, *Operazioni del capitale. Capitalismo contemporaneo tra sfruttamento ed estrazione,* Roma 2021.
[12] *Ibid.*
[13] *Ibid.*

le critiche dell'economia politica: «le forme di pensiero distopiche rischiano di immobilizzarci nell'azione e di impedire interventi urgenti»[14]. Queste distopie tra le quali abbiamo riscontrato caratteristiche comuni all'estrattivismo cancellano un elemento fondamentale per la comprensione del capitalismo contemporaneo: «Le macchine sono ansintotiche», cioè «non raggiungono mai la massima precisione». Questo fatto «giustifica un'ulteriore estrazione di dati dal maggior numero di persone tale da potere alimentare le raffinerie dell'Intelligenza Artificiale»[15] e porta alla trasformazione degli esseri umani da 'soggetti di idee' a 'soggetti di dati'. Ma, di nuovo, anche in questo approccio risulta ancora impensato *cosa produce l'asintoticità tra le macchine e il programma di dominio dettato dai proprietari dei mezzi di produzione.* La differenza che impedisce alla 'totalità' del capitale di chiudersi su se stessa, e di ridurre il mondo a una mercificazione universale, è la forza lavoro. È da essa che dipendono i modi in cui sono interpretati, raccolti, classificati, denominati i dati, le merci sono spostate.

Il concetto di 'asintotico' usato da Crawford è interessante perché rinvia all'idea di una *dimensione infinita nella finitezza di un rapporto sociale*. Tuttavia non coglie la specificità di questo rapporto. Asintoto, alla lettera, significa *essere privo di congiunzione*. È proprio della retta che si avvicina a una funzione ma non la tocca ed è tangente all'infinito. Nel caso del rapporto tra forza lavoro e capitale ci troviamo in realtà davanti a *una congiunzione disgiuntiva* che può essere rappresentata graficamente attraverso una figura topologica non orientabile come il nastro di Moebius[16]. Forza lavoro e capitale si congiungono e allo stesso tempo si disgiungono, opponendosi e integrandosi. Questo movimento infinito nella finitezza si svolge su un unico piano che intreccia una molteplicità nella singolarità e viceversa cambiando gli elementi mentre questi ultimi si intrecciano in ordine prodotto dalla differenza delle differenze. Il capitale e la forza lavoro sono sintesi disgiuntive di un movimento infinito in un sistema finito. Questa idea permette di superare, e criticare, le concezioni metafisiche di un capitale che procede sulla sua strada all'infinito, abbattendo da solo i propri limiti come se fosse un'entità autonoma dotata di una volontà indipendente dal mondo. Quando si dice che il capitale è una rivoluzione dei suoi limiti si intende la trasformazione continua dei suoi rapporti con la forza lavoro, la natura, la società e l'economia in cui essi si danno.

In questa trasformazione persiste anche il rapporto tra produzione e riproduzione e quello tra lavoro produttivo e improduttivo. Una forza lavoro deve sia produrre che riprodursi: per lavorare davanti a un personal computer, in fabbrica, o nei 'servizi' deve avere i soldi per pagarsi una casa, una scuola dove mandare i figli, un ospedale dove curarsi, e così via. Senza contare che in

[14] K. Crawford, *Né intelligente, né artificiale. Il lato oscuro dell'IA,* Bologna 2021, pp. 57 e ss.
[15] Ivi, pp. 130 e ss.
[16] Per la teoria della 'sintesi disgiuntiva' e la sua applicazione anche alla critica dell'economia politica del capitalismo si vedano almeno: G. Deleuze, *Logica del senso* (1969), Milano 1975; G. Deleuze-F. Guattari, *AntiEdipo*. *Capitalismo e schizofrenia* (1972), Torino 1975.

una scuola o in un ospedale c'è un'altra forza lavoro che va pagata per svolgere il lavoro che permette a un'altra forza lavoro di vivere in maniera salutare, imparando a praticare i linguaggi e i modi di vivere nel mondo, compresi quelli che sono necessari a creare dati o a spostare le merci. Su ciò che è 'produttivo' o 'improduttivo' per la forza lavoro e per il capitale si svolge un conflitto continuo: la 'produttività' rispetto all'obiettivo del profitto per il capitale non è la stessa rispetto all'idea di una vita dignitosa e liberata dall'imperativo di lavorare precariamente e con salari da fame. Senza contare che non passa giorno in cui il capitale cerchi di scaricare i costi della riproduzione della forza lavoro sulle spalle degli stessi lavoratori, oppure sullo Stato che deve rimediare anche ai suoi catastrofici fallimenti. Allo stesso tempo però ha bisogno di formare, curare e promuovere la forza lavoro perché non tutto può essere trovato sul mercato come una merce già prodotta. E dunque deve pagare le spese. Questo rapporto è diventato sempre più centrale da quando la sfera della riproduzione è intesa anch'essa come una 'produzione'.

Il processo è stato definito da David Harvey nei termini di 'riproduzione allargata' del capitale e indica uno dei principali argomenti di discussione nel marxismo contemporaneo. Ciò ha cambiato i confini tra la vita privata e quella pubblica. Basti qui pensare al ruolo devastante dei social network sulla soggettività performativa neoliberale. Oppure alla scuola, o all'università. Sono sempre state intese come 'spese improduttive', in particolare gli insegnanti sono stati disprezzati, i fondi tagliati, funzionano come quasi imprese mentre l'istruzione è gestita come un quasi mercato. Eppure, queste 'fabbriche' svolgono un ruolo fondamentale in un capitalismo concepito come produzione di 'capitale umano' in un mercato del lavoro costruito su bassi salari, poche tutele e povertà dilagante. Il conflitto è tale da determinare gli assetti dell'intera economia. Su queste basi si definisce, tra l'altro la natura della crescita, la direzione di un'economia 'stagnante' e persino la qualità del patto sociale vigente, oltre che la 'qualità' o la 'produttività' di un lavoro. Su queste basi va giudicato il potere di 'creazione' e 'perimetrazione' del mondo da parte del capitale. È un'indicazione che viene dallo stesso Harvey secondo il quale l'accumulazione per esproprio è connessa con la riproduzione allargata. Il problema è questo: il soggetto dell'una (l'accumulazione) e dell'altra (la riproduzione) *non è solo il capitale bensì il suo rapporto con gli elementi eterogenei che esso combina.* Nel capitalismo digitale, infatti, il 'capitale' è sempre un 'capitale aggregato'. Una simile aggregazione è attraversata, e fondata, su un rapporto aperto con i suoi elementi che, al loro interno, sono a loro volta composti. Al centro c'è sempre qualcuno che lavora e un altro che sfrutta.

Quando il potere del capitale è asimmetrico rispetto alla capacità di resistergli esso si sviluppa nel rapporto sociale con ciò che gli si oppone e con ciò che gli permette di diventare se stesso. Dunque il rapporto verticale procede insieme a uno orizzontale, va dall'alto al basso e viceversa. Non cala dall'alto come in sostanza è immaginato nel paradigma estrattivista. Questa riflessione ha una conseguenza molto pratica. Se parliamo di 'estrazione' di un 'valore' parliamo

sempre del lavoro svolto da chi realizza concretamente tale estrazione. Dunque l'atto di estrarre è anch'esso iscritto nel rapporto di lavoro, non è la manifestazione di un'azione sovrana che risponde solo a se stessa, cioè è assolta dal rapporto nel quale in realtà sussiste. Pensare invece che avvenga per opera di un soggetto eminente o sovraumano, come in fondo sembra che accada nell'estrattivismo, rischia di alimentare le leggende che circondano la produzione del valore e il suo rapporto contraddittorio sia con il capitale che con la forza lavoro. Nell'ambito del capitalismo digitale il problema potrebbe essere ripensato in questo modo: la potenza attribuita fittiziamente alla macchina è tanto più grande quanto più è profonda l'alienazione della forza lavoro e la sua adesione ai principi di un capitalismo che garantisce la sua libertà a condizione di sfruttarla. Oggi i termini dell'equazione sono separati da uno dei rapporti di potere più squilibrati mai registrati nel corso della storia recente e tuttavia non va persa di vista la relazione, né va sostituita con le figure retoriche come l'iperbole. Anche nel momento di massima *invisibilizzazione del rapporto sociale di produzione* con la forza lavoro il capitale (digitale) resta un rapporto sociale con ciò che intende negare e tuttavia resta il suo motore.

8. Divenire classe

Non considerare il rapporto tra forza lavoro e capitale come il centro della critica del capitalismo, svincolandolo dalla trasformazione della composizione del capitale, significa perdere di vista la politicità intrinseca del capitalismo e immaginare la politica in due modi opposti e paradossalmente simmetrici: da un lato, David Harvey tratteggia una politica dittatoriale gestita da un manipolo di capitalisti spietati e senza scrupoli che esercitano un potere sovrano di vita e di morte; dall'altro lato, indica una sfera separata che, a un certo punto (ma non è mai chiaro quando né come) interverrà per risolvere i problemi. Nel mezzo sembra esserci il vuoto in cui vagano miliardi di esseri umani sospesi tra la realtà di un potere sempre uguale a se stesso e le aspirazioni morali di chi sogna 'un altro mondo possibile'. La situazione è paradossale perché Harvey stesso, e con lui molti degli estrattivisti critici, cercano un metodo concreto per ridefinire il rapporto politico con il capitale alla luce delle molteplici resistenze esistenti che però non riescono a creare ancora una politica efficace e duratura. Tuttavia molte indicazioni che emergono da questo dibattito sono preziose per capire il modo in cui dall'inizio del XXI secolo i movimenti si concepiscono, si aggregano e cercano di strutturarsi nonostante le gravissime difficoltà imposte dall'organizzazione della produzione e dalla frammentazione sociale che essa comporta.

Il problema resta il divenire classe, cioè il soggetto della politica anticapitalista. La storia del pensiero critico degli ultimi 50 anni è stata attraversata da una critica dell'idea di un soggetto centrale nel rapporto con il capitale – la classe operaia professionale e anche 'di massa' – e dall'annuncio di una proliferazione di soggetti ulteriori rispetto al mero rapporto di subordinazione definito da un contratto di lavoro a tempo indeterminato. Su questa differenza si sono innestate

tutte le altre: di razza, di sesso e genere per esempio. Questa disseminazione non ha corrisposto a una generalizzazione della classe, bensì a un suo indebolimento politico facilitato sia dalla violenta campagna ideologica che ha accompagnato la contro-rivoluzione neoliberale per cui 'non esistono più le classi sociali' (tranne una: la classe media), sia da un drastico cambiamento dei rapporti politici che hanno indebolito l'iniziativa sia della classe operaia che degli altri soggetti rendendo illeggibile il concetto di 'classe' tra coloro che potrebbero usarlo in maniera completamente diversa da come è stato fatto fino agli anni Settanta del XX secolo.

Questa tensione è avvertibile anche in David Harvey quando ha spiegato la ragione per cui ha adottato la categoria dell'accumulazione per esproprio preferendola a quella 'classica' dello sfruttamento. A suo avviso la prima si riferirebbe a tutti gli *spossessati, o meglio agli oppressi*, la seconda *solo* agli sfruttati nel lavoro di fabbrica. Questa distinzione rischia però di non considerare una realtà elementare: sia pure non lavorando in una fabbrica è facile presumere che anche tutti gli oppressi abbiano un lavoro *e dunque* siano *anche* in un rapporto di sfruttamento. Lo stesso vale per chi lavora in fabbrica: tutti vivono nella società, e dunque sono protagonisti di rapporti di oppressione di razza, sesso o di altra natura. Faccio anche notare che la distinzione contraddice il carattere generale e astratto del processo di 'accumulazione per esproprio' senza contare quello di 'riproduzione allargata' del capitale in settori della società e della vita non ancora mercificati. Opporre momenti in realtà connessi nella parte diagnostica del potere, e non considerarli nella prassi politica significa cedere alla mentalità neoliberale che oppone i diritti in una guerra tra gli ultimi e i penultimi della divisione sociale del lavoro. Questo è il lavoro dell'estrema destra alleata al neoliberalismo che al momento sembra avere in mano l'iniziativa politica e simula un discorso 'sociale' mentre aizza il revanscismo, il razzismo e gli istinti proprietari anche della classe operaia 'bianca' con toni da 'guerra civile'.

La ragione di questo saggio è dimostrare che esistono gli strumenti, e le esperienze, capaci di scartare da questo insidioso terreno e rilanciare un'iniziativa politica egemonica. Da questo punto di vista, una volta chiarite le sue contraddizioni, il lavoro di Harvey e quello che ha ispirato la sua opera si inseriscono in un ampio dibattito in cui la critica dello sfruttamento del lavoro non è più anteposta a quella sessuale e nemmeno a quella della natura. Definire il lavoro alla luce dei rapporti sessuali o razzializzati, interpretare il sessimo e il razzismo come espressioni della violenza sociale del potere significa rompere le gerarchie esistenti e coniugare i conflitti in una classe oggetto di moteplici oppressioni e soggetto di possibili resistenze. Questo significa prospettare un divenire co-rivoluzionario tra soggetti differenti, come ha sostenuto lo stesso Harvey[17]. Il loro intreccio nella stessa condizione dimostra come la politica possa estendersi dalla contestazione della proprietà privata dei mezzi della produzione ai rapporti biopolitici del potere e alla lotta contro lo sfruttamento del vivente. Una

[17] D. Harvey, *L'enigma del capitale*, Milano 2010, pp. 228 e ss.

molteplicità capace di rendere politicamente virtuosa la classe è rara. La politica consiste nel rendere il più durevole possibile il divenire classe, individuando rapporti di forza che non replichino quelli dominanti ma li decentrino oltre i suoi attuali confini.

Roberto Ciccarelli
✉ ciccarelli.ciccarelli@gmail.com

Bibliografia

Abdelnour S., Méda D. 2019. *Les nouveaux travailleurs des applis,* Puf, Paris.
Allegri G. 2018. *Il reddito di base nell'era digitale. Libertà, solidarietà, condivisione*, Fefé editore, Roma.
Benanan A. 2020. *Automation and the Future of Work,* London and New York, Verso.
Brynjolfsson E., McAfee A. 2014. *The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Machines,* New York and London, W. W. Northon.
Casilli A. 2018. *En attendant les robots. Enquête sur le travail du clic*, Seuil, Paris.
Ciccarelli R. 2018. *Capitale disumano. La vita in alternanza scuola lavoro,* Manifestolibri, Roma.
Ciccarelli R. 2018. *Forza lavoro. Il lato oscuro della rivoluzione digitale,* DeriveApprodi, Roma
Ciccarelli R. 2020. *Il lavoro digitale*, in *La Grande Trasformazione. L'impatto sociale del cambiamento del lavoro tra evoluzioni storiche e prospettive globali*, a cura di Enzo Mingione, Annali della Fondazione Feltrinelli, LIV, Feltrinelli, Milano.
Ciccarelli R. 2021. *Il mito dell'automazione digitale* in S. Bellucci, a cura di, *Ai-work. La digitalizzazione del lavoro*, Jaca Book, Milano.
Ciccarelli R. 2021. *Labour Power. Virtual and actual in digital production*, Springer, Londra.
Ciccarelli R. 2022. *Una vita liberata. Oltre l'apocalisse capitalista,* DeriveApprodi, Roma.
Crawford K. 2021. *Né intelligente, né artificiale. Il lato oscuro dell'IA,* Il Mulino, Bologna.
De Stefano V., Aloisi A. 2020. *Il tuo capo è un algoritmo. Contro il lavoro disumano*, Laterza, Roma-Bari.
Deleuze G 1975. *Logica del senso* (1969), Feltrinelli, Milano.
Deleuze G., Guattari F. 1975, *AntiEdipo. Capitalismo e schizofrenia* (1972), Einaudi, Torino.

Ekbia H., Nardi B. 2014, *Heteromation and Its (Dis)contents: The Invisible Division of Labor between Humans and Machines,* «First Monday», vol. 19, no. 6.
Finelli R. 2022. *Filosofia e tecnologia. Una via di uscita dalla mente digitale*, Rosenberg&Sellier, Torino.
Ford M. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future,* New York, Basic Books.
Fuchs C. 2013. *Digital Labour and Karl Marx*, Routledge, London.
Fumagalli A., Gobetti S., Morini C., Serino R. 2021. *Reddito di base. Liberare il XXI secolo,* Momo edizioni, Roma.
Ganascia J. B. 2019. *Le mythe de la Singularité: Faut-il craindre l'intelligence artificielle?* Points, Parigi.
Harvey D. 1982. *The Limits to Capita*l. Verso, Londra.
Harvey D. 2003. *The New Imperialism*. Oxford University Press, Oxford. Harvey D. 2005. *A Brief History of Neoliberalism*, Oxford University Press, Oxford.
Harvey D. 2006. *Spaces of Global Capitalism*, Verso, Londra.
Harvey D. 2010. *L'enigma del capitale*, Feltrinelli, Milano.
Harvey D. 2014. *Seventeen Contradictions and the End of Capitalism*. Oxford University Press, Oxford.
Huws U. 2014. *Labor in the Global Digital Economy: The Cybertariat Comes of Age*, Monthly Review Press, New York.
Marx K. 1964. *Il capitale*, a cura di D. Cantimori, Editori Riuniti, Roma vol. I.
Mezzadra S., Neilson B. 2021. *Operazioni del capitale. Capitalismo contemporaneo tra sfruttamento ed estrazione,* Manifestolibri, Roma.
Prassl J. 2018. *Humans as a service. The promise and Perils of Work in the Gig Economy*, Oxford University Press, Oxford.
Scholz T. (ed.) 2012. *Digital Labor: The Internet as Playground and Factory*, Routledge, London.
Smith J. E. 2020. *Smart Machines and Service Work. Automation in an Age of Stagnation,* London, Reaktion Books.
Srnicek N. 2017. *Capitalismo digitale. Google, Facebook, Amazon e la nuova economia del web*, Luiss University Press, Roma.
Van Parijs P., Vanderborght Y. 2017. *Il reddito di base. Una proposta radicale*, Il Mulino, Bologna.
Vecchi B. 2017. *Il capitalismo delle piattaforme,* Manifestolibri, Roma.
Vercellone C. *et al* (eds.) 2019. *Decode: data driven disruptive commons-based models,* CNRS, Paris.
Yang A. 2018. *The War on Normal People: The Truth about America's Disappearing Jobs and Why Universal Basic Income is Our Future,* New York, Hachette Books.
Zuboff S. 1989. *In the Age of the Smart Machine,* Basic Books, New York.
Zuboff S. 2019. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, Roma.

Recensioni, discussioni e note

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771596

Recensioni

D. Cardon, Che cosa sognano gli algoritmi. Le nostre vite al tempo dei Big Data

Mondadori Education 2016

Marko Ćeranić

Qualche anno dopo la Conferenza di Dartmouth (1956), in occasione della quale vennero gettate le basi di quell'imponente e ambizioso progetto di ricerca che porta ancora oggi il nome di Intelligenza Artificiale, Herbert Simon, che ne fu protagonista indiscusso, ebbe a dire che «le macchine saranno in grado, entro vent'anni, di realizzare qualsiasi opera che un uomo possa compiere» (H. Simon, *The Shape of Automation for Men and Managment*, New York: Harper and Row, 1965; trad. it. nostra). Esattamente sessant'anni dopo, Dominique Cardon, nel suo *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei Big Data*, si assume l'onere di fare un punto della situazione: non soltanto i calcolatori non sono ancora nemmeno lontanamente equiparabili a un cervello umano sotto molti aspetti, in primis sotto quelli della creatività e della plasticità, ma spesso e volentieri falliscono nel portare a termine gli obbiettivi che vengono loro assegnati dai programmatori. Quali sono questi obbiettivi? Per chi e per quale motivo rivestono oggi un'importanza capitale? In che modo l'avvento dei Big Data ha rivoluzionato profondamente il nostro stare in società? Qual è il grado di controllo che gli algoritmi esercitano sulla nostra condotta e qual è, di rimando, il grado di libertà di cui godiamo rispetto ad essi? Queste sono solo alcune delle domande a cui Dominique Cardon tenta, nel suo testo, di dare delle risposte che possano orientare il nostro agire quotidiano rispetto a una realtà, quella degli algoritmi, che pervade il nostro quotidiano in maniera così impalpabile, e al contempo così invasiva, da restarci (quasi) del tutto sconosciuta.

Un algoritmo può essere definito come la successione finita di istruzioni in cui è scomposta la risoluzione di un dato problema; una volta inserite delle variabili (input) l'algoritmo applica tali istruzioni su di esse fintanto che non arriverà a calcolare l'esito esatto del problema alla cui risoluzione è stato designato. Nel primo capitolo del testo, Cardon suddivide gli algoritmi in relazione alla posizione metaforica che occupano rispetto ai dati che devono elaborare, e

ne distingue quattro tipologie: gli algoritmi possono stare *accanto* ai dati che devono elaborare, possono stare *al di sopra* di essi, *dentro* di essi, o *sotto* di essi. In ordine cronologico i primi a fare la propria comparsa nel mondo digitale delle *data sciences* sono gli algoritmi che si situano accanto al Web: deputati a calcolare la popolarità, o *audience*, dei primi siti online, nascono dall'esigenza dei loro creatori di conoscere il numero di visitatori che frequentano la loro pagina. Ciò che fa l'algoritmo in questo caso è tenere il conto degli accessi alla pagina, e la categoria può a sua volta essere suddivisa in algoritmi *user centric* e algoritmi *site centric*. Nel primo caso viene isolato un campione rappresentativo degli internauti, auspicabilmente più ampio possibile, e ne vengono registrati gli accessi. Tuttavia, questo metodo rileva ben poco dell'identità degli utenti, poiché di fatto non fa altro che stimare degli *accidenti*. Non sappiamo nulla, difatti, di chi si è seduto davanti al computer rappresentato, per quanto tempo è rimasto connesso, cosa ha fatto prima e che relazioni intrattiene con il resto della comunità virtuale. Diverso è il caso degli algoritmi user centric; questi ultimi, infatti, (Google Analytics ad esempio) sono in grado di dire esattamente l'indirizzo IP dal quale è stato effettuato l'accesso, la visita precedente dello stesso e per quanto tempo è rimasto connesso alla pagina. Come dichiarato dallo stesso autore, la tensione tra algoritmi che si servono di categorie isolate in precedenza da altre discipline, come ad esempio la sociologia o il marketing, e algoritmi che mirano invece a ricostruire il comportamento dei singoli individui senza doverli necessariamente 'etichettare' è un tema, se non *il* tema, che ci accompagna per tutta la lettura del libro, e solleva per giunta stimolanti riflessioni sulla natura della rappresentazione sociologica; la dialettica tra individuo e società, infatti, che percorre sotterraneamente la sociologia fin dalla sua fondazione, viene letta in questa chiave come opposizione tra un utente *amorfo*, di cui conosciamo tutte le abitudini ma non le generalità (da intendersi in senso lato, non come le generalità vere e proprie riportate, ad esempio, sui documenti d'identità), e la rappresentazione che ne viene data dagli analisti quando si ritrovano a dover incasellare queste tracce digitali in delle categorie sociali o socio-economiche (tuttavia non è detto che lo debbano necessariamente fare, poiché nell'ottica di una sempre maggiore personalizzazione dell'utenza, la categoria d'appartenenza degli internauti diventa via via sempre meno rilevante).

Quando si tratta di misurare invece l'autorevolezza di un sito internet, sapere quante volte viene visitato potrebbe non essere sufficiente. È per questo motivo che vengono introdotti gli algoritmi al di sopra del web. Ispirati dal sistema di classificazione delle riviste scientifiche, il principio su cui fondano il loro calcolo è quello della risonanza sociale; in poche parole, quanto più il numero di volte che un sito viene rinviato a un'altra pagina è alto, tanto più esso sarà autorevole, e se la pagina a cui viene rinviato è a sua volta stata rinviata un numero elevato di volte, lo sarà ancora di più. L'idea soggiacente a questa tipologia di algoritmo è che essi misurino quella porzione di attività sul web dell'utente che non è in alcun modo condizionata dal fatto che esso sappia dell'esistenza di questi algoritmi, poiché altrimenti potrebbe truccarli,

ed è ciò che una volta regolarmente accadeva. Molti siti, infatti, tentavano di guadagnare in autorevolezza affidandosi a strateghi del mercato digitale che pagavano delle pagine, o ne costruivano delle altre, che citassero i loro clienti. Oggi le maglie dei motori di ricerca più scaricati si sono strette sempre più attorno a questi siti 'fittizi', per evitare che a loro volta ne venisse minata la reputazione. Tuttavia, Cardon ci mette in guardia da altri due grandi rischi in cui incorrono gli algoritmi appartenenti a questa classificazione. Il primo è quello della centralizzazione dell'autorevolezza: i siti che si trovano al centro dei rinvii sono destinati ad aumentare sempre di più in autorevolezza, in quanto il fatto di essere quelli con il maggior numero di rinvii fa sì che ne ricevano degli altri, mentre siti che stanno in periferia potrebbero essere a loro volta molto validi, ma verranno esclusi dalla competizione in quanto poco conosciuti. Il secondo è l'effetto censitario delle misurazioni: in quanto queste tengono conto solamente della natura ipertestuale delle informazioni condivise, rimangono fuori tutte quelle informazioni che non possiedono un vero e proprio 'corpo', come ad esempio molte delle informazioni che vengono condivise sui social network: i *likes*, i commenti, gli *screenshot* sono tutti strumenti adottati dai frequentatori dei social per scambiarsi reciprocamente autorevolezza, ma questa tipologia di algoritmo non ne tiene conto. Per questo motivo vengono quindi introdotti gli algoritmi dentro il web.

Questi ultimi nascono per ovviare alle deficienze che gli altri algoritmi esibivano riguardo al loro modo di calcolare l'attività sui social network, e per certi aspetti si può dire che tra essi andrebbero annoverati anche gli utenti stessi, che lavorano in sinergia con gli algoritmi per misurare la propria attività online. Il principio a cui si ispira quest'altro tipo di algoritmo è che, in una comunità virtuale, si è tanto più influenti quanto più si è attivi e conosciuti. È il caso, per esempio, dei già sopracitati social network, ma anche dei siti che sfruttano le recensioni per aumentare la popolarità di attività che stanno al di fuori del web, come ristoranti, alberghi, agenzie turistiche, ecc. Ma anche questi algoritmi, ci mette in guardia l'autore, non sono privi di criticità. La prima riguarda la chiusura delle comunità virtuali: l'utente, circondandosi di persone che gli sono più affini per gusto, opinioni e simpatia, vive all'interno di una 'bolla', come la definisce Cardon, che gli preclude sempre più l'accesso a informazioni inattese o sorprendenti. La seconda riguarda invece l'asimmetria tra l'eterogeneità delle interazioni che possono avvenire tra gli utenti del web, e dei social network in particolare, e il numero limitato di reazioni con cui queste possono essere espresse; nelle arene social, infatti, spesso una data reazione, in mancanza di altro, può essere il riflesso delle intenzioni, dei desideri, delle sfumature espressive più disparate, e non può pertanto essere ricondotta a un significato univoco. Si pensi ad un like ad esempio. In una situazione in cui siamo impossibilitati ad esprimere con immediatezza il nostro dissenso (su Facebook, ad esempio, dove non vi è l'opzione 'non mi piace') potremmo addirittura ricorrere al like in segno di scherno, oppure di supporto nei confronti di una notizia che suscita il

nostro sdegno e che, in quanto tale, merita di essere diffusa affinché accresca il disprezzo popolare nei suoi confronti.

Vi è infine l'ultima categoria, quella degli algoritmi che operano al di sotto del web, i più diffusi oggigiorno e anche quelli più gravidi di considerazioni per quanto riguarda il rapporto tra l'attività online di un individuo e la sua rappresentazione. Questa particolare tipologia di algoritmo si avvale del *machine learning*, o apprendimento automatico, per indirizzare l'internauta su scelte future a partire dalle scelte che ha fatto in passato. Gli algoritmi al di sotto del web, infatti, tengono traccia di tutti gli spostamenti digitali effettuati, onde poi confrontarli con dati provenienti da bacini di utenza simili che possano suggerire i comportamenti da adottare in futuro. Ciò che rende questa classe di algoritmi di gran lunga più interessante, secondo Cardon, è il fatto che essa si svincola del tutto da rappresentazioni sociodemografiche a priori; gli algoritmi che operano dal basso non hanno bisogno di selezionare dei target, in quanto mirano a personalizzare sempre più l'esperienza di navigazione degli internauti, fino a renderla unica. E non si fidano nemmeno di ciò che gli utenti dicono di sé stessi o degli altri: i fatti, o meglio le loro tracce, parlano per loro.

Una volta illustrate, dunque, le quattro tipologie di algoritmo, l'autore dedica il capitolo successivo ai profondi cambiamenti che i Big Data hanno introdotto nel tradizionale modo di intendere la statistica, e alle ripercussioni che questo cambio di paradigma ha avuto sulla società. Tra questi, il più importante è senz'altro il passaggio da una concezione 'naturalistica' dei dati statistici, a una sua concezione strumentale. I dati statistici, infatti, lungi dall'essere ancora considerati delle verità inossidabili misurate a posteriori, saranno da intendersi piuttosto come gli indicatori probabilistici che guidano il comportamento di eventi che si devono ancora verificare. Gli individui partecipano di questa incertezza, e la mettono in conto quando devono prendere delle decisioni; in questo modo viene a crearsi un circolo ricorsivo tra l'algoritmo e l'utente: quest'ultimo, sapendo di essere costantemente tracciato, opera comunque delle scelte sul web che a loro volta vengono tracciate, e svelano qualcosa di più sulla sua identità digitale. Cambia in questo modo anche il valore epistemologico delle categorie sociodemografiche: esse non guidano più la conoscenza statistica in quanto, come abbiamo visto, il tracciamento individuale dell'attività online non ne tiene conto. A esse saranno da preferirsi delle 'istantanee' di eventi che esibiscono, tra loro, una correlazione contingente che viene costantemente rivista e ricalcolata, e che guida l'internauta in direzione di comportamenti e decisioni 'probabili' che, nel momento in cui diventano certi, confermano o disattendono la previsione fatta su di essi, mettendo a disposizione in questo modo nuove e preziose informazioni sulle preferenze dell'utente digitale in un circolo senza fine.

Il terzo e penultimo capitolo introduce una distinzione fondamentale tra segnali e tracce. Se i primi possono essere intesi come le attività esplicite, quindi consapevoli, che l'internauta dissemina sul suo percorso con le proprie attività (apprezzamenti, commenti, editing), le tracce invece sono i segni 'invisibili' della sua attività online, proprio quelli alla cui rilevazione sono preposti gli algoritmi

del quarto tipo, quelli che Cardon ha definito *al di sotto del web*. Il massimo dell'efficacia di calcolo nella predizione di un comportamento viene raggiunta quando tracce e segnali vengono incrociati, o meglio quando un'intensa attività di tracciamento aiuta a individuare una correlazione tra differenti segnali. Questi, a loro volta, vengono confrontati anche con gli output derivati dall'attività di tracciamento di altri utenti, in modo tale che utenti che hanno fatto scelte simili possano essere indirizzati verso gli stessi prodotti digitali. A questo punto l'autore introduce una delle riflessioni, a nostro avviso, più interessanti riguardo alla prevedibilità dei comportamenti individuali in relazione alle predizioni algoritmiche: gli algoritmi sono efficaci perché fondano le loro predizioni su comportamenti abitudinari, oppure la pervasività degli algoritmi è tale che indirizza i nostri comportamenti verso le stesse abitudini? Sono vere entrambe le cose. Da una parte, infatti, studi condotti dai ricercatori di Facebook hanno mostrato che sono gli stessi utenti, sulla base delle loro preferenze, a selezionare le community e le attività che li rispecchiano maggiormente, tendendo a escludere, invece, tutta quella fetta di informazioni che diverge dal loro pensiero e dai loro interessi. Se un profilo è, al contrario, originale e poliedrico, allora l'algoritmo proporrà una gamma di opzioni più vasta rispetto a quanto non farebbe con un profilo più 'ridondante'. L'algoritmo, tutto sommato, non fa altro che reiterare una gamma di proposte probabili sulla base dell'attività trascorsa degli internauti, adottando l'ipotesi che il nostro futuro non sarà altro che una riproduzione del nostro passato. D'altra parte, tuttavia, è anche vero che le grandi piattaforme mondiali (Apple, Google, Facebook, Amazon) si impegnano affinché vengano raccolti flussi di dati sempre più ingenti, che poi a loro volta verranno incrociati per ottenere predizioni sempre più precise anche nei confronti degli utenti più eclettici. Più che promuovere l'originalità e l'unicità delle persone, parrebbe che i raccoglitori di dati vogliano al contrario convogliarle in monopoli commerciali sempre più elitari. Come osserva giustamente Cardon, in conclusione al capitolo, queste grandi aziende vendono ai loro clienti la promessa di potersi emancipare dai rigidi schemi sociali in cui sono aggrovigliati, e di poter promuovere, con scelte che vengono proposte loro *ad hoc*, la loro individualità. Ma calcolando le ipotesi di scelte future sulla base di quelle passate, gli algoritmi non fanno altro che riproporre uno specchio deformato, per quanto esatto, della società, che non aggiunge nulla di più rispetto a quanto essa è già. Essi, anzi, agiscono piuttosto come un Leviatano deterministico che annulla qualsiasi possibilità di crescita degli individui, in quanto li riporta a un'immagine statica di sé: essi non sbagliano, non escono mai fuori dal tracciato, non fanno mai niente di *imprevedibile*.

Veniamo infine all'ultimo capitolo del libro. Nella parte conclusiva l'autore si impegna a far emergere in tutta la sua brutalità il contrasto netto tra ciò che gli algoritmi, metaforicamente, 'sognano' di fare per le società umane, e i loro fallimenti. Se da una parte essi vorrebbero che i meccanismi con cui si assegna la popolarità fossero aperti e trasparenti a tutti quanti, dall'altra, come abbiamo già visto, la popolarità non fa che attirare al suo centro altra popolarità, lasciando

fuori tutti quegli utenti, tutte quelle informazioni, tutte quelle pagine web che stanno ai margini dell'influenza mediatica. Se gli algoritmi sognano una società meritocratica, in cui i migliori emergono perché sono quelli 'democraticamente' più menzionati da altri utenti, d'altra parte quando questa gigantesca macchina di attribuzione del merito viene avviata, diventa difficile competervi se si è un neofita del web. Se poi un altro sogno accarezzato dagli algoritmi è quello di promuovere la nascita di organizzazioni spontanee e di zone autonome che sfuggano ai rigidi controlli delle istituzioni (si pensi anche solo alle Temporary Autonomous Zones teorizzate da Hakim Bey, venuto a mancare solo pochi mesi fa), in cui gli internauti si possano riconoscere come facenti parte di una stessa mentalità di condivisione degli spazi virtuali, dall'altra, come fa notare Cardon, la visibilità si è imposta come nuovo principio di gerarchizzazione dell'ordine sociale. Se poi a ciò si aggiunge il fatto che non tutti gli individui dispongono delle stesse risorse per la promozione di sé, il sogno si infrange ancora una volta contro la rivisitazione della diseguaglianza sociale che vige 'là fuori'. Infine, se il sogno supremo degli algoritmi è quello di rendersi al contempo invisibili e completamente a disposizione del genere umano, è anche vero che un lusso che ci concediamo sempre più raramente è quello di 'staccare la spina' e riprendere in mano le redini della nostra quotidianità, anche e soprattutto in quelle attività che farebbe l'algoritmo al posto nostro. In conclusione, il testo di Dominique Cardon si delinea come un utile prontuario per comprendere non soltanto le influenze più sottili che i Big Data esercitano sulle nostre esistenze quotidiane, ma anche e soprattutto per rivedere alcune categorie sociologiche classiche, e interrogarsi se esse possano considerarsi adeguate per descrivere i complessi assemblaggi che si vengono a creare tra algoritmi e individuo, e i profondi cambiamenti che questi apportano al nostro modo di vivere e intendere la società.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771598

Recensione

C. Barone (a cura di), *L'algoritmo pensante. Dalla libertà dell'uomo all'autonomia delle intelligenze artificiali*

Il pozzo di Giacobbe 2020

Remo Trezza

Il volume qui recensito, sicuramente di grande attualità, fa parte della collana dei Quaderni di *Synaxis* dello Studio Teologico 'S. Paolo' di Catania ed è stato curato da Christian Barone, docente di Teologia presso la Pontificia Università Gregoriana di Roma e presso lo Studio Teologico su menzionato.

Contributo 'delicato', che appassiona e determina una certa dose di *curiositas*, stimola la riflessione e cerca di scandagliare le problematiche relative all'avvento delle nuove tecnologie (I.A., algoritmi, *robot*, etc…), ponendosi in un dialogo di *ordo ordinans*, sempre più intersezionale con i principi fideistici della cristianità e dei valori etici fondanti la cattolicità.

L'*hortus conclusus*, ovvero la struttura del volume, caratterizzata – secondo chi scrive – da un divenire argomentativo mai stancante, ma sempre avvincente e dinamico, può essere sintetizzato come di seguito.

La premessa (pp. 5-19), del curatore dell'intera opera, si apre con una citazione – nell'opinione di chi recensisce – lapidaria. Seppur la '*quote*' sia di Isaac Newton ed il testo in italiano, si preferisce riportare la versione in lingua latina della frase, avente il medesimo significato, di qualcuno che di storia se n'è inteso, a motivo della più concisa rappresentazione della realtà: '*nos esse quasi nanos gigantium humeris insidentes*'. Ciò non fa altro che immergere il lettore in questioni *novae*, ma sempre *veterae*. Cosa c'è di notevolmente 'nuovo' nelle 'nuove tecnologie'? Punto nodale della premessa è fondato sul 'pensiero umano', che – specie nella nuova dimensione dell'innovazione – si scontra con la possibilità di essere sostituito – in quanto *homo sapiens* – dai procedimenti decidenti e decisori di automi. Rilevante, quasi come se il curatore dell'opera volesse portare con mano il lettore per le vie impervie delle intelligenze artificiali, è l'attenzione posta sui concetti di *machine learning*, neuroscienze, *neuroimaging*, *artificial neural*

networks, *deep learning* etc…, attraverso l'approccio definitorio dei quali, nasce un primo grande interrogativo: chi programma le macchine? Quali capacità deve avere il programmatore? E chi scrive aggiunge: quale responsabilità giuridica sarà imputabile al programmatore qualora vi sia una decisione non conforme alle variabili algoritmiche introitate?

Con le parole del curatore, anche chi scrive si chiede se gli algoritmi saranno capaci (se non lo siano già) di rivoluzionare il mondo sociale, in una dimensione antropologica del tutto mutata ed in una altrettanto dimensione fenomenologica del tutto trasformata. Una riflessione assai convincente è relativa ad una nuova visione della 'performatività'. Come può cambiare il mondo, che ha dei suoi linguaggi – alcuni dei quali totalmente fondati sulla naturalità – alla luce dell'artificiosità delle macchine? Quale orizzonte per questo tecno-mondo? È solo una 'elucubrazione fantascientifica', una 'fervida immaginazione trans-umanista' o qualcos'altro? Come deciderà una macchina? Chi immetterà nelle macchine i valori umani fondanti, i principi etici su cui si fonda la nostra essenza, affinché possano decidere? Il curatore, inoltre, solo nella parte conclusiva della premessa, ricorda che il volume – di per sé collettaneo – racchiude i contributi di eminenti studiosi offerti al Convegno di studi intitolato *Dio, macchine, libertà* celebrato dallo Studio Teologico San Paolo e dall'Università degli Studi di Catania.

Il primo capitolo (pp. 21-38) racchiude le osservazioni di Giuseppe O. Longo, docente emerito di Teoria dell'informazione presso il Dipartimento di Ingegneria e Architettura dell'Università degli Studi di Trieste, incentrate principalmente sul connubio uomo-tecnologia, sulle origini storiche dell'intelligenza artificiale e sulla 'metamorfosi paradigmatica'. La 'delega tecnologica', sulla quale si nutrono dubbi circa l'estensione possibile di rischi, annichilisce man mano l'agire dell'uomo, lo calpesta nelle sue attività intellettive, lo divora completamente, imbrigliandolo – alcune, ma ormai fin troppo spesse, volte – nel vortice dell'iper-tecnologismo. L'uomo viene sempre più colpito da una quantità – a dire il vero incalcolabile – di dati e di informazioni. Si potrebbe dire che si tratti di una vera e propria 'bomba informativa', che quasi sfocia nella c.d. 'datacrazia'. La parte più rilevante del primo capitolo è relativa al 'libero arbitrio', dalla cui personale concettualizzazione dell'Autore discende l'analisi del caso di Phineas Gage e dell'interessantissima teoria del 'determinismo'. Si passa, poi, a quella che l'Autore definisce 'vergogna prometeica', ovvero alla concezione che – secondo l'analisi di chi scrive – può essere sintetizzata così: l'uomo 'impotente' nei confronti della 'potenza' irraggiungibile della macchina. Verso la fine del suo analitico contributo, il professor Longo – in maniera del tutto condivisibile – affronta il tema dell'anonimato che in rete di certo non esiste. Ciò pone una serie di riflessioni 'serie' intorno al tema della protezione dei dati personali nella 'virtualità' che non va mai – salvo casi eccezionali – fatta prevalere sulla 'realtà'. L'analisi – si ribadisce, acuta – dell'Autore termina con una considerazione di certo non fondata sulla positività: «in questo *mare magnum* della comunicazione mediata dalla tecnologia digitale, rischiamo davvero di

naufragare […] qualunque grado di libera scelta ancora abbiamo, lo stiamo rapidamente perdendo». Ciò induce a pensare che l'uomo debba salvaguardare il suo essere uomo – senza se e senza ma – non facendosi 'schiavizzare' dagli automi, ma sfruttandone l'occasione per far sì che essi lo passano 'servire', 'aiutare'. Diventi la macchina un mezzo, ma non il fine. Vi sia 'funzionalismo', mai 'schiavismo'.

Il secondo capitolo (pp. 39-56), redatto da Paolo Arena, docente di Automatica presso il Dipartimento di Ingegneria elettronica e Informatica dell'Università degli Studi di Catania, è, dal punto di vista strettamente metodologico, ineccepibile. Si muove, infatti, secondo una 'triplice' direzione. La prima, è fondata sui '*robot* per l'uomo'; la seconda, invece, sui '*robot* tra gli uomini'; infine, la terza sui '*robot* con l'uomo'. È una concezione triadica di sicuro interesse, dato che l'Autore – al quanto appassionato della materia – discorre, in una maniera del tutto costruttivistica, di concetti tecnici, ma attraverso esemplificazioni – quasi paraboliche – che rendono la strada per la comprensione più fluida e chiara al lettore. L'argomentazione è incentrata sul lavoro/ausilio delle intelligenze artificiali, sulle protesi mediche e sui *robot* umanoidi di 'ispirazione biologica'. Si può dire che l'Autore si soffermi sulla c.d. 'biologia meccanica'. La presa d'atto è che si è di fronte ad una vera e propria 'rivoluzione industriale'. Condivisibile l'affermazione secondo cui «è compito delle istituzioni garantire l'uso etico di tali potenti strumenti». Viene in luce il 'nocciolo duro' del tema, ovvero la relazione eticità-meccanicità. Nelle sue conclusioni, l'Autore enfatizza che 'la spinta positiva della scienza e della tecnologia, innate nell'uomo al principale scopo di creare oggetti, manufatti, dispositivi e complessi sistemi artificiali per contribuire al miglioramento della qualità della vita, ben fa quindi da risonanza al perenne anelito dell'assemblea cristiana' per cui l'onnipotente Dio abbia affidato all'uomo l'opera della creazione e posto al suo servizio le immense energie del cosmo, il quale dovrà collaborare a un mondo più giusto e fraterno. Vien da chiedersi se le nuove tecnologie siano in grado di raggiungere una 'giustezza meccanica'. Oppure la giustizia è un concetto – non assolutamente astratto – rinvenibile solo nell'ontologia naturale dell'essenza umana?

Il terzo capitolo del lavoro (pp. 57-84), scritto da Salvatore Amato, docente di Filosofia del diritto presso il Dipartimento di Giurisprudenza dell'Università degli Studi di Catania, parte – in maniera avvincente, appassionata e appassionante – dal concetto di 'personalizzazione della persona', ovvero si interroga su come la materia sia mai potuta diventare intelligente nel corso della sua evoluzione. Domanda retorica quella dell'Autore? Dietro la personalizzazione vi è – *sine dubio* – la 'teocrazia'. Non si dimentichi «[…] a sua immagine e somiglianza li creò». Il *Logos* (la teologia *stricto sensu*) è per l'Autore la risposta al 'riduzionismo materialistico'. Partendo da un rifiuto della concezione poc'anzi delineata, Egli si sofferma sulle vicende assai complesse che investono il cervello umano, il quale – in buona sostanza – finirebbe per apparire solo «un fenomeno chimico, regolato dalla fisica. L'Io, il soggetto, la coscienza e le stesse singole decisioni non sarebbero altro che l'esito imponderabile delle

connessioni tra le reti neuronali, la conseguenza imprevedibile delle qualità emergenti di una molteplicità di particelle biochimiche». È questo ciò che si vuole raggiungere? Un'approssimazione dell'Io e una superficialità dell'Essenza? L'Autore – bisogna riconoscerlo, in maniera pregevole – si pone due domande di rilevante caratura riflessiva: «La nostra intelligenza sarà la loro (*delle macchine*) intelligenza? La nostra coscienza sarà la loro coscienza?». Sulla stessa scorta interrogativa dell'Autore, si può proseguire con un terzo punto di domanda: «La nostra dignità sarà la loro dignità?» In chiave 'evoluzionista', il contributo passa al rapporto tra Dio e caso, inter-relazionate con l'intelligenza artificiale. Da Goethe, l'Autore approda a Gianmarco Veruggio, padre della robotica italiana, passando per la c.d. '*translatio creativitatis*' di Sloterdijk e ritornando al *Logos* inziale, grazie al quale l'uomo, in relazione con Dio – *Summum Bonum* – non può che tendere al bene. La macchina, ci si chiede, se non 'sorvegliata' dall'uomo, tenderà mai al bene? Ma l'uomo, ancor prima della macchina, sa cosa sia il bene e sa come si fa a raggiungerlo? Attraverso un'esegesi etimologica della parola *daimon*, prendendo spunto – ma forse esagerando, ed è lo stesso Autore ad ammetterlo – dal *De Civitate Dei* di Sant'Agostino, Egli arriva a definire il *daimon* – nella sua radice greca – come 'scienza'. «La scienza gonfia, la carità costruisce». Ebbene, anche in tal caso, conosceranno le macchine l'emblema altisonante della carità? Sanno essere caritatevoli con l'uomo? Forse dovranno insegnare all'uomo la tecnica, ma a quest'ultimo toccherà – come principio indissolubile di moralità – insegnare loro la carità. Le intelligenze artificiali senza carità sono assimilabili – secondo l'Autore – al concetto di *default*. Ma a quale *default* si vuole fare riferimento? Ad un *default* tecnologico oppure ad un *default* esistenziale? In accordo con il pensiero dello scrivente, l'Autore, nelle battute finali del suo *iter* logico-argomentativo, ribadisce la 'funzionalizzazione' delle macchine e la loro 'inclinazione' al soddisfacimento dei bisogni dell'uomo, senza che questo ne diventi 'assuefatto'. Coerentemente, poi, Egli afferma che

> i giuristi affrontano il problema delle macchine intelligenti, cercando di configurare una sorta di personalità giuridica e dignità 'numerica', in analogia con la personalità e dignità umana, proprio quando vacillano i presupposti (libertà e responsabilità) dell'attribuzione della personalità giuridica agli esseri umani.

La scienza, per poter essere Vera e Giusta, deve essere una 'scienza caritatevole'. Siamo davvero liberi? O il nostro libero arbitrio, come affermato anche nel primo capitolo, soggiace ad un determinismo meccanicistico che ci imbriglia nell'immobilismo esistenzialista?

Il quarto capitolo del lavoro (pp. 85-102), di estrema importanza per chi scrive, in quanto relativo all'apparato normativo 'intelligente', è stato redatto da Giovanni Di Rosa, docente di Diritto privato e di Biodiritto presso il Dipartimento di Giurisprudenza dell'Università degli Studi di Catania. L'Autore fornisce una serie di indicazioni sul quadro giuridico – assai dirompente e soprattutto di matrice europea – relativo alle nuove tecnologie e – in special modo – alle responsabilità giuridiche discendenti dal malfunzionamento delle

macchine o, ancora di più, dalle errate decisioni frutto dell'articolato e assai complesso procedimento algoritmico. Si è d'accordo sul fatto che non vi sia una definizione univoca di *robot* sul versante tecnologico e sulla concezione dell'*homo tecnologicus* come 'figura ibrida'. Dal punto di vista del congegno responsabilistico, l'Autore ben sottolinea che le categorie della responsabilità variano a seconda che si tratti di *robot* tele-operati, *robot* autonomi, e *robot* cognitivi. Egli, inoltre, è palesemente contrario – si evince dalle marcate osservazioni – all'attribuzione della 'soggettività giuridica' ai *robot*, anche se le nuove indicazioni europee sembrano convergere verso una possibile adozione di una personalità elettronica. Inoltre, a tal proposito, si è proprio espresso il Comitato economico e sociale europeo, il quale si è detto contrariato ad una simile attribuzione. Il *robot*, seppur cognitivo, ha una cognizione talmente tanto elevata che possa ad esso essere attribuita addirittura una 'capacità giuridica e di agire'? Si dovrebbe quasi prevedere un codice civile degli automi? Se così fosse, ci si troverebbe davvero davanti all''affievolimento' della persona umana, intorno a cui ruota tutto il sistema di protezioni e di tutele del codice sostanziale, ma specie della Carta costituzionale. Molto interessante – seppur in chiave discorsiva e solo in alcuni punti risolutiva – è la panoramica sui risvolti della perpetrazione dei danni e delle conseguenze risarcitorie, quantificatorie e assicuratorie. La responsabilità degli automi, in definitiva, va ricercata nelle categorie giuridiche già esistenti o nella creazione di un apposito strumentario? Pare che l'idea dell'Autore – come quella di chi scrive – è di prevedere un sistema di responsabilità del produttore o, quanto meno, di una responsabilità del programmatore. Infine, il *software* dovrà essere considerato un bene mobile oppure no? I proprietari delle macchine, si aggiunge, dovranno dotarsi di un'assicurazione obbligatoria – esattamente come quella per le automobili – e, si spera, dovrà essere anche previsto un Fondo di garanzia per il risarcimento dei danni per quegli *owners* che non se ne siano dotati. Di certo, la materia civilistica in tal senso, collegata all'intelligenza artificiale, è ancora sotto la lente di studio dei più e ci si aspetta che i fatti dipendenti da strumenti elettronici possano trovare quanto prima certa e decisa tutela. *Ex facto oritur ius*. L'opinione personale è che bastano le categorie giuridiche esistenti per incasellare i 'fatti elettronici' nelle fattispecie ad essi collegabili, attraverso ovviamente l'adozione di uno sforzo interpretativo *ad hoc*.

Il quinto capitolo del saggio (pp. 103-116), scritto da Antonio Allegra, docente di Storia della Filosofia presso il Dipartimento di Scienze umane e sociali dell'Università per stranieri di Perugia, si apre con un titolo stimolante: Transhomo Deus. *Forme tecnoutopiche di reincantamento del mondo*. Lampante è l'idea per cui l'Autore abbia voluto discorrere circa il 'trans-umanesimo' che – nel mondo del divenire odierno – sempre più si affianca alla post-modernità e al post-umanesimo. Tale concetto è strettamente ancorato alla dimensione 'futuristica' e a quella dell''immortalità'. Non esiste il futuribile? Esiste solo il futuro? L'Autore – in completa sinergia con la *mens* riflessiva dello scrivente – afferma che «[…] si tratta di una narrazione teologica. Al suo centro c'è l'uomo, ma l'uomo nella misura in cui è capace di fare di se stesso qualcosa di diverso,

di più». Espressiva ed eloquente è l'immagine dell'‘antropocentrismo' a scapito del 'meccanocentrismo'. O meglio – secondo le riflessioni di chi scrive – ci si potrebbe augurare una concezione 'antropo-meccanica', ovvero quella fondata sulla relazione uomo-macchina, laddove il primo sia sempre 'agente' e mai 'succube' della seconda. Per l'Autore, le mitologie transumane sono l'immortale e il *cyborg*, al quale la prima è inscindibilmente connessa. Tutto si sintetizza nella domanda se sia possibile che l'uomo, anche nella sua corporalità o, meglio dire, nella sua materialità corporale, possa in qualche modo farsi aiutare dalle macchine per migliorare la sua esistenza. La risposta è certamente positiva, ma tutto ciò trova dei limiti specifici nella dignità umana che è e sempre sarà il baluardo su cui poggia l'impianto giuridico italiano. Di fronte a tutto ciò – in sintonia con l'Autore – si ribadisce che «il compito intellettuale non può che essere quello del richiamo inattuale alla cautela, alla misura, alla tradizioni *demodè* dell'umanesimo, insomma allo sguardo criticamente *ortogonale* rispetto al movimento».

Il sesto capitolo del lavoro (pp. 117-123), redatto da Andrea Vella, dottore di ricerca in Filosofia e Storia delle Idee presso l'Università degli Studi di Catania, esplora il campo – assai inusuale – dell'applicabilità alla macchina del senso della fede. Un automa può mai avere in sé il senso della religione? Domanda forte perché bisognerebbe che la macchina, oltre ad essere dotata di intelligenza, sia dotata anche di un'anima. Ma l'anima elettronica può mai esistere? In definitiva – in chiave assolutamente nuova e disincantata – l'Autore ritiene di poter rispondere a tali domande mediante tre risposte. La prima è quella di un «imperativo categorico *à la* Kant»; la seconda è ispirata alla «dottrina aristotelica del giusto mezzo», ovvero bisognerà dotare l'intelligenza artificiale di buon senso; la terza, basata sull'ipotesi di uno stretto legame tra religiosità e moralità, «sarebbe quella di dotare le macchine di un senso di religiosità». Infine – forse in un'accezione assai provocatoria – ci si chiede se dotare l'I.A. dei miracoli possa aiutare a sviluppare in essa la fede. Riflessioni sicuramente curiose, ma che non troveranno appiglio in una dimensione di 'assenza di animismo' di cui sono dotate le macchine. Potrebbe essere solo l'uomo di fede (chi?) a poter introdurre la fede (quale?) nella macchina, ma ci si rende conto che sono solo delle 'superfetazioni'.

L'ultimo capitolo del volume, il settimo (pp. 125- 134), si chiude con le riflessioni scritte di Giovanni Basile, docente invitato di Filosofia presso lo Studio Teologico San Paolo di Catania, il quale nel suo '*Cur homo machina*?' tenta di far comprendere la dimensione salvifica della fede, in una visione squisitamente teologica, ponendo in risalto il rischio del 'sovvertimento umano' e della 'sovversione religiosa' per mezzo delle intelligenze artificiali. L'Autore afferma che 'quello che stiamo vivendo con le I.A. non è altro che l'attuazione dell'anti-destino degli uomini, l'inatteso che si è fatto accadimento'. L'*homo faber* è uomo finito, è uomo che non si salva mediante le macchine (anche se crede il contrario), ma solo attraverso quell'Uomo che 'ci salva'. L'I.A. porterà mai alla

noia? L'I.A. ci guarirà? L'I.A. ci 'grazia'? Infine – per riportare delle osservazioni dell'Autore che molto convincono chi scrive –

> nel gioco di ruolo delle I.A. contro gli uomini, ne sono certo, […], saranno gli uomini ad avere la meglio. Anche perché saremo noi ad assegnare alle macchine una etica perfetta, una intelligenza perfetta, una moralità perfetta, persino probabilmente una fede perfetta, e sarà per questo che alla fine saranno le macchine a rendersi conto che gli uomini deludono poiché insaziabili […].

Secondo quest'ottica, in buona sostanza, l'uomo deve essere il 'controllore' delle azioni meccaniche, decidere con la sua testa in quanto avente una 'dignità intellettiva' che discende direttamente dalla 'dignità dell'essere' e perfezionare la macchina per il raggiungimento di una sua più serena e tranquilla esistenza. Tutto ciò – e si conclude – conferma anche quanto sostenuto dallo scrivente, ovvero che la responsabilità giuridica delle macchine è del programmatore (introiettore, il *deus ex machina*), a meno che non si tratti di meri malfunzionamenti (in tal caso potrà essere imputata al produttore).

La mia considerazione – davvero per chiudere – è un plauso a chi ha curato questo volume che, seppur piccolo nella quantità, trasuda qualità di riflessioni, qualità di stimoli culturali e qualità di insenature argomentative. Un volume che si innesta in un solco – ormai sempre più acceso – di iniziative, opinioni, tesi e teorie avallanti una posizione piuttosto che un'altra, ma in ottica sempre più sinergica e in una dimensione totalizzante di arricchimento del sapere scientifico.

Prendendo spunto dal sillogismo aristotelico e dal brocardo di Santi Romano, si può affermare che: *Ubi homo, ibi machina; ubi machina, ibi ius; ergo, ubi homo, ibi ius*. Alla fine è sempre l'uomo, soggetto dell''umanità', ad avere la responsabilità morale di regolare il *caos* sregolato della 'tecnicità' con il *cosmos* della 'giuridicità'.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771602

Note

Note di lettura su forza-lavoro, lavoro e algoritmo in R. Ciccarelli, *Labour Power. Virtual and Actual in Digital Production*

Springer 2021

Ubaldo Fadini

La recente pubblicazione di *Labour Power. Virtual and Actual in Digital Production* (Springer 2021), che riprende in parte l'indagine iniziata con *Forza lavoro. Il lato oscuro della rivoluzione digitale* (DeriveApprodi 2018), a cui sono da affiancare *Capitale disumano. La vita in alternanza scuola lavoro* (Manifestolibri, 2018) e *Una vita liberata. Oltre l'apocalisse capitalista* (DeriveApprodi 2022), consente la messa a fuoco di un percorso di ricerca tra i più significativi e stimolanti degli ultimi anni. Il suo autore, Roberto Ciccarelli, si impegna infatti su un motivo centrale di qualsiasi impresa teorica che voglia qualificarsi come radicalmente critica nei confronti degli assetti e delle configurazioni in trasformazione della nostra società di segno capitalista in cui la questione dell'algoritmo e della sua pervasività nel nostro mondo storico-sociale si inserisce a pieno titolo. Tale motivo è quello della forza lavoro da intendersi marxianamente (e – sullo sfondo – spinozianamente: come si sottolinea nelle prime pagine di *Labour Power*) soprattutto come potenza. Si sa come per la più consapevole formulazione di progetti di critica radicale si debba ancora tenere presente l'osservazione di Mark Fisher, che ricordava come sia sempre «più facile immaginare la fine del mondo che la fine del capitalismo». Al cosiddetto 'realismo capitalista', Ciccarelli contrappone un punto di vista analitico concretamente materialista, fondato pure sul recupero di elementi rilevanti di quelle tradizioni di pensiero che hanno accompagnato e cercato di favorire i tentativi di realizzazione di una nuova e alternativa idea di società, di solidarietà e libertà. Tale operazione di rilancio ha alla base una convinzione non di poco peso, cioè che nel postfordismo il lavoro non è più rivestito dagli ordini abituali, ben determinati a livello spazio-temporale, della produzione puntuale, 'passo dopo passo', in quanto gli strumenti più importanti del lavoro vivo, ad esempio i mezzi che consentono di agire comunicativamente *e* relazionalmente, sono

contenuti nel corpo vivente del soggetto di lavoro, con l'effetto conseguente che così non si smette mai di lavorare e che il lavoro cattura il soggetto per le 24 ore della giornata e per tutti i sette giorni della settimana.

Risulta in ogni caso ancora qui decisiva la distinzione tra forza lavoro e lavoro, che Marx coglie e sottolinea a partire dalla sua critica ai 'classici' dell'economia politica (da Smith a Ricardo): l'oggetto dello scambio tra il capitalista e l'operaio non è il lavoro, bensì la forza lavoro, la merce decisiva della società capitalistica in quanto il suo valore, il lavoro in essa oggettivato e che le consente di darsi e riprodursi, è cosa differente dal lavoro vivo, che si può trarre da tale merce. Come ricordava opportunamente Claudio Napoleoni, «il fondamento della produzione capitalistica sta nel fatto che da una determinata forza-lavoro avente un certo valore, il capitalista può trarre più lavoro di quello che è in esso oggettivato. Questo *pluslavoro* si oggettiva dunque in un *plusvalore* che, in quanto appartiene al capitalista, costituisce il suo *profitto*» (C. Napoleoni, *Lezioni sul Capitolo sesto inedito di Marx*, Bollati Boringhieri, 1972, p. 22).

Questa distinzione tra la forza lavoro e il lavoro è tradotta da Ciccarelli nel senso di individuare nella forza lavoro una potenzialità di esistenza, una facoltà di fatto sempre attiva che esprime al meglio i caratteri propri della vita umana, del suo combinato corpo-mente, a livello individuale e collettivo, e questo vale, in *Labour Power*, anche negli scenari sociologici aperti dall'estrema capillarità del lavoro digitale e dal 'potere' ambiguo degli algoritmi nei processi di produzione o *digital production*, come recita il sottotitolo. Sempre ricordando la lezione di Marx, che indica nella forza lavoro o capacità di lavoro il combinarsi delle «attitudini fisiche e intellettuali che esistono nella corporeità», nella «personalità vivente» dell'uomo, che messe per così dire in movimento producono valori d'uso di qualsiasi genere, l'autore di *Labour Power* osserva come proprio l'insieme di facoltà e capacità – la facoltà che consente di produrre valori d'uso di qualsiasi genere e la capacità di lavoro da portare al mercato, per la sua vendita – sia appunto la forza lavoro, la 'personalità vivente' del corpo umano. Tale combinato è da distinguersi dal 'capitale', poiché nell'atto della vendita della forza lavoro il soggetto si distacca certo dalla capacità ma ciò non significa che perda addirittura complessivamente la facoltà della forza lavoro così movimentata/mobilitata. È su questa base che si delinea non tanto una filosofia del lavoro, quanto una rinnovata riflessione sulla forza lavoro a partire cioè dalla questione non di cosa sia attualmente il 'lavoro' ma di cosa oggi possa una forza lavoro (questione ribadita nella sua centralità anche nelle pagine conclusive del testo: «*What Can Labour-Power Do?*»).

Ciccarelli mette soprattutto in evidenza come la forza lavoro sia insieme una forza produttiva consegnata al capitalista, all'utilizzo voluto da quest'ultimo, *e* una facoltà non riducibile pienamente a tale uso, riferito, come detto, alla produzione delle merci. A questo punto si può anche apprezzare uno degli sfondi dell'indagine, quello propriamente filosofico, visto l'utilizzo del tema deleuziano della sintesi disgiuntiva, di un'unità che non si dà come chiusa ma apre invece a delle articolazioni ulteriori, di fatto imprevedibili, proprio in virtù del suo porsi

come un rapporto connettivo *e* disgiuntivo, capace di differenti modulazioni, che riguardano pure gli elementi che lo compongono. Va qui ricordato che Ciccarelli è autore di testi importanti che a mio modo di vedere supportano gli sviluppi appunto più propriamente filosofici delle pagine sulla forza lavoro e sul mitologema così ormai ampiamente diffuso del cosiddetto 'capitale umano', che rimuove il fatto che quest'ultimo non è altro che la stessa forza lavoro ricompresa nella logica della valorizzazione del capitale: tra gli altri, *Immanenza e politica in Spinoza* (Roma, Aracne 2006) e l'essenziale *Immanenza. Filosofia, diritto e politica della vita dal XIX secolo al XX secolo* (Bologna, Il Mulino 2008).

In questo senso, la forza lavoro si mostra come *facoltà* di tutte le altre facoltà individuabili nel soggetto vivente. Si può anche dire così: la forza lavoro *è* il soggetto come facoltà capace di produrre qualsiasi valore d'uso e ciò significa che c'è un *venir prima* di carattere produttivo, ontologico, 'etico-politico' della forza lavoro rispetto alla forza lavoro considerata soltanto dal lato del suo essere forza produttiva 'di' capitale. È d'altronde vero che oggi tutte le facoltà dell'essere umano sono messe a valore, 'mobilitate', ma ciò avviene in modalità tali da consegnarle di fatto a un presente di loro alienazione e oggettivazione predefinita in un pacchetto dato/predeterminato di capacità. *Labour Power* descrive con puntualità ed efficacia tutte quelle politiche attive del lavoro, della formazione continua, del capitale umano che negli ultimi tempi hanno contribuito a rendere ancora più automatici, di fatto irriflessi e quasi-naturali, processi di (auto) sfruttamento che vengono presentati come inevitabili. Si vedano in particolare il quinto e il sesto capitolo: *The Dwarf of History* e *The Entrepreneurial Self*, soprattutto in riferimento agli sviluppi delle piattaforme digitali, del capitalismo delle piattaforme, e al loro protagonismo sempre più sfrenato.

Al di là dell'analisi dettagliata e preziosa dei meccanismi sempre più sofisticati di espropriazione delle facoltà, del divenire-uomo del capitale in ambito algoritmico-digitale, Ciccarelli delinea una ricerca sulla forza lavoro che non schiaccia quest'ultima sulla teleologia del lavoro stesso all'interno dei processi odierni di sua subordinazione alle dinamiche della valorizzazione capitalista. L'insistenza sull'eccedenza costitutiva della forza lavoro, sulla sua 'potenza', si traduce poi nella individuazione del carattere multiverso degli stessi rapporti di potere (classe, razza, sesso, etc.) che appunto coltivano la pretesa di restituire la soggettività unicamente in modalità rigidamente subordinate. Si sottolinea così che la forza lavoro come soggetto non può essere considerata pienamente contenuta nel lavoro e nella formula corrente del 'capitale umano' e dunque non va riferita al suo impiego produttivo *dato*. È la differenza specifica della forza lavoro, da considerarsi come forza produttiva a disposizione del capitale e fattore di suo superamento, a permettere di riflettere su possibili condizioni di pratiche originali di libertà, di emancipazione individuale *e* collettiva, da svilupparsi in direzioni politicamente, giuridicamente, eticamente, diverse da quelle presenti.

In effetti, la questione centrale che anima la ricerca complessiva di Ciccarelli è quella di pensare un processo di superamento della realtà/mondo dell'odierna soggettività neoliberale a partire da una rinnovata sperimentazione di una

potenza, la forza lavoro, che non può essere fatta rifluire all'interno del racconto della fine di tutte le fini. È in quest'ottica che si prende posizione a favore di un prospettivismo storico delle dinamiche di liberazione contro il riproporsi incessante delle illusioni di una apocalisse prossima ventura. Prassi della liberazione sociale e politica, quindi, irriducibile al soggettivismo e alla rappresentazione del mondo prevalenti anche nelle culture critiche contemporanee.

La potenza può essere liberata quando è concepita nella forma concreta, individuata e storica, della forza lavoro, contro la raffigurazione delle soggettività sotto la veste dei 'capitalisti umani', vera e propria parola d'ordine della 'rivoluzione passiva' (per ricordare Gramsci) appunto di segno neoliberale, da intendersi proprio come politica della passività organizzata.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771602

Discussione

La dialettica e il Bene

Discussione di Sarah Broadie, *Plato's Sun-like Good*

Cambridge University Press 2021

Lorenzo Giovannetti

Dialectics and the Good. A discussion of Sarah Broadie, *Plato's Sun-like Good*

In this paper, I discuss the key aspects of S. Broadie's Plato's Sun-like Good. In a nutshell, she argues that the nature of the Good in Plato's Republic is interrogative, i.e. it can only be grasped by raising the question `is it good?'. This leads Broadie to argue that dialectic is essentially practical reasoning and that its results are context-relative. In this paper, I introduce the key texts from the Republic and, after expounding Broadie's view, I raise some issues concerning her reading, particularly against the idea that dialectic is fallible qua context-relative, and against the view that being Forms receive from the Good is the same as their empirical implementation.

***

Introduzione

Nel suo recente e ultimo volume *Plato's Sun-like Good*, Sarah Broadie avanza una lettura piuttosto audace di uno degli snodi più complessi della filosofia platonica e conseguentemente della storia della filosofia, vale a dire la natura del Bene e il modo in cui essa è catturata dall'analogia con il sole. Il nucleo della lettura di Broadie consiste nel pensare il corrispettivo al livello della cognizione dell'idea del Bene in forma interrogativa: il modo in cui si può pensare il Bene, per Platone, è attraverso quella che Broadie chiama la G-question, 'is it good?'. In altre parole, movimentare a livello cognitivo la natura del Bene si esaurisce essenzialmente nell'interrogarsi sulla bontà di qualcos'altro, processo che sarebbe in grado di assicurare una conoscenza autentica di ciò su cui ci si chiede se è buono o no. Chiedersi se qualcosa è buono e approdare a una risposta positiva è la garanzia di un'apprensione autentica e totale di questo qualcosa. Conseguentemente, dato che la forma di conoscenza che si affatica alla ricerca della natura del Bene è la dialettica, per Broadie questa sarebbe nella sua essenza

la reiterazione della domanda sulla eventuale bontà delle elaborazioni fornite dalle proprie procedure definitorie e razionali, la quale domanda fornirebbe il fondamento ultimo e certo del successo della dialettica. Questo impianto interpretativo arriva a sostenere che la dialettica sia in ultima istanza un *practical reasoning*, vale a dire una forma di esercizio della razionalità che elabora delle risposte in grado di guidare l'azione situata in contesti specifici, la cui validità è però limitata a tali contesti. Prima di esplorare approfonditamente alcune tra le tesi principali del libro di Broadie e offrire più estesamente una discussione critica di quanto appena riportato è bene schematizzare i passaggi principali del testo platonico, ad uso del lettore e senza alcuna pretesa di esaustività.

1. L'analogia del sole, l'immagine della linea, la dialettica

I testi analizzati da Broadie sono essenzialmente quelli del libro VI, in cui il personaggio Socrate paragona il Bene e la sua funzione rispetto agli enti intelligibili al sole e alla sua funzione rispetto agli enti sensibili[1]; quelli in cui la varietà delle facoltà conoscitive dell'anima è collocata metaforicamente su un segmento di retta a sua volta diviso in sotto-segmenti[2]; e alcune battute dal libro VII, in cui, dopo l'illustrazione del noto mito della caverna, Socrate introduce il percorso educativo cui sono destinati i filosofi-governanti mirante all'acquisizione della scienza o tecnica dialettica, il cui fine ultimo è proprio l'apprensione del Bene[3]. Questi celeberrimi passi sono anticipati da un'introduzione sul ruolo dell'idea di Bene. Riporto qui, parafrasando e sintetizzando[4], gli scambi salienti di queste

[1] Inutile a dirsi, la letteratura critica sull'argomento è estremamente vasta e variegata. Per una contestualizzazione dello *status quaestionis* si veda R. Ferber, *Platos Idee des Guten*, Sankt Augustin 1989; F. Fronterotta, *ΜΕΘΕΞΙΣ La teoria platonica delle idee e la partecipazione delle cose empiriche. Dai dialoghi giovanili al* Parmenide, Pisa 2001, pp. 128-144, soprattutto pp. 137-139 n. 38, e M. Vegetti, To agathon: *buono a che cosa? Il conflitto delle interpretazioni sull'idea del buono nella* Repubblica, in Id., *Il potere della verità. Saggi platonici*, Roma 2018, pp. 113-133. È bene comunque qui riportare gli studi più rilevanti per l'interpretazione di Broadie (da ella presi in considerazione oppure no) e per la critica che qui si propone: G. Sillitti, *Al di là della sostanza: ancora su* Resp. *VI 509b*, «Elenchos», 1 ,1980, pp. 225-244; M. Baltes, *Is the Idea of Good in Plato's Republic Beyond Being?*, in M Joyal (ed.), *Studies in Plato and the Platonic Tradition*, Aldershot 1997, pp. 3-23; G. Santas, *The Form of the Good in Plato's Republic*, in G. Fine (ed.), *Plato*, Oxford 2000, pp. 249-276; M. Dixsaut, *L'analogie intenable: le Soleil et le Bien*, in Ead., *Platon et la question de la pensée. Études platoniciennes*, Paris 2000, pp. 121-151; F. Ferrari, *La causalità del bene nella* Repubblica *di Platone*, «Elenchos», 22, 2001, pp. 5-38; M. Vegetti, Megiston mathema. *L'idea del "buono" e le sue funzioni*, in Id. (ed.), *Platone. Platone. La Repubblica*, Vol. V, Napoli 2003, pp. 253-286; F. Ferrari, *L'idea del bene: collocazione ontologica e funzione causale*, in M. Vegetti (ed.), *Platone. La Repubblica*, cit., pp. 287-325.

[2] Anche in questo caso, gli studi sono innumerevoli. Si consideri il classico lavoro di Y. Lafrance, *Pour interpréter Platon. La ligne en* République *VI 509d-511e. Bilan analytique des études (1804-1984)*, Montréal-Paris 1987. Cfr. anche N. D. Smith, *Plato's Divided Line*, «Ancient Philosophy», 16, 1996, pp. 25-46 e F. Franco Repellini, *La linea e la caverna*, in M. Vegetti (ed.), *Platone. La Repubblica*, cit., pp. 355-403.

[3] Anche qui risulta imprescindibile la discussione di M. Vegetti, *Dialettica*, in Id. *Platone. La Repubblica*, cit., pp. 405-433.

[4] In quello che segue, mi baserò ampiamente sulle scelte traduttive di M. Vegetti, attenendomi però alla traduzione tradizionale del *to agathon* come 'il Bene', sebbene, come Vegetti asserisce

sezioni, concentrandomi su quelli essenziali per l'interpretazione di Broadie, sì da fornire gli strumenti per comprendere il senso della sua operazione esegetica.

1.1 Introduzione dell'idea del Bene

505a2-3: l'idea del Bene è la conoscenza di maggior importanza (*megiston mathema*) e solo in relazione ad essa le cose giuste, assieme alle altre, divengono utili e vantaggiose.

505d5-9: a differenza, ad esempio, delle cose giuste, per cui ci si potrebbe accontentare della mera apparenza, nel caso delle cose buone (*agatha*) tutti ricercano quelle che sono realmente buone.

506a4-7: rispetto alle cose giuste e belle (*dikaia te kai kala*), se si ignora in qual modo possano essere buone, nessuno potrà conoscerle adeguatamente (*gnosesthai ikanos*).

506d5 e seguenti: non è possibile a Socrate definire che cosa sia il Bene in sé (*auto ti pot'esti tagathon*) come è stato fatto nel caso delle altre virtù. Ci si deve accontentare della prole del Bene (cioè il sole) che è molto simile ad esso.

1.2 L'analogia del sole

507b8-9: le molte cose diciamo che sono viste (*horasthai*) e non pensate (*noeisthai*), le idee diciamo che sono pensate ma non viste.

507c e seguenti: la vista e il visibile hanno bisogno di un terzo elemento (*triton*) affinché ci sia realmente visione e questo elemento è la luce. La luce proviene dal sole. Il sole, dunque, è causa della vista ed è visto da essa stessa.

508b13-c2: come nella sfera dell'intelligibile il Bene sta al pensiero, così nella sfera visibile il sole sta alla vista e i suoi oggetti.

508c9-509a5: Come il sole illumina gli oggetti della vista rendendoli visibili, così l'anima pensa e conosce ciò che è illuminato dalla verità e dall'essere (*aletheia te kai to on*). Il Bene garantisce la verità agli oggetti conosciuti. Come nel visibile la luce e la vista sono simili al sole, senza essere il sole, così conoscenza e verità sono simili al Bene, ma non sono quest'ultimo che è ancora più degno di onore.

509b1-9: il sole non procura agli oggetti visti unicamente la facoltà di essere visti, ma anche la generazione (*genesin*), pur non essendo esso stesso generazione. Allo stesso modo gli oggetti conosciuti non derivano dal Bene solo l'esser conosciuti, ma ne traggono anche l'essere e l'essenza (*to einai te kai ten ousian*), pur non essendo il Bene un'essenza, bensì ancora al di là dell'essenza (*epekeina tes ousias*) superandola per dignità e potenza.

e conseguentemente adotta, una traduzione più fedele sarebbe 'il buono'.

1.3 L'immagine della linea

Di questa complessa sezione mi limiterò a riportare i passi chiave per l'interpretazione di Broadie. L'immagine, come è noto, consiste nel distribuire su una linea il visibile e l'intelligibile, per poi dividere ognuno di questi segmenti in due sotto-segmenti. Il punto qui rilevante è la distinzione tra i due sotto-segmenti dell'intelligibile.

510b4-511b1: coloro che si occupano di geometria, aritmetica e scienze simili pongono come ipotesi, ad esempio, il pari e il dispari, le figure, i tre tipi di triangoli, considerando queste cose come note e assumendole come ipotesi e non ritengono di doverne dar conto (*logon didonai*) e partendo da queste ne traggono delle conclusioni. In questo processo, costoro si servono di forme visibili (ad esempio un disegno di un quadrato) e su di esse conducono le dimostrazioni pur non pensando a queste forme visibili, bensì in vista delle forme a cui esse somigliano, come ad esempio il quadrato in sé. Questo tipo di conoscenza è incapace di superare il livello delle ipotesi e di procedere verso il principio.

511b2-c2: l'altra sezione dell'intelligibile è appresa dalla ragione (*logos*) attraverso la potenza della dialettica (*dialegesthai dynamei*). A differenza della prima, la ragione in questa sezione della linea non tratta le ipotesi come principi dati, bensì come punti di partenza per giungere a un principio non-ipotetico del tutto. Quando un tale principio è stato appreso, la ragione ne trae delle conseguenze conoscitivamente fondate, rimanendo sempre e unicamente sul piano delle idee.

511d8-e1: lo stato dell'anima che corrisponde al segmento più alto, che rimane unicamente al livello delle idee, è il pensiero puro (*noesis*); lo stato che corrisponde al segmento che assume le ipotesi e si serve di immagini è il pensiero discorsivo (*dianoia*).

Buona parte degli interpreti, tra cui Broadie (p. 20), associa il principio non-ipotetico del tutto, che troviamo a 511b2-c2, all'idea del Bene, che quindi deve fungere da principio di conoscenza, coerentemente a come esso è descritto nell'analogia del sole. La proposta originale di Brodie risiede proprio nell'offrire un'interpretazione innovativa di come il Bene, in quanto principio non-ipotetico, sia fonte di conoscenza rispetto alle ipotesi formulate dal pensiero discorsivo.

1.4 La dialettica

Quest'ultima selezione di passi si estende su buona parte del libro VII, tuttavia alcuni riferimenti all'idea del Bene sono ampiamente impiegati dagli interpreti nell'indagine sull'analogia del sole, risulta quindi appropriato riportare quelli più rilevanti.

517a8-c4: l'idea del Bene risiede all'estremo limite del conoscibile, tuttavia essa può essere vista, sebbene a fatica, e una volta vista si deve concludere che essa è la causa (*aitia*) di ciò che è retto e bello, signora e dispensatrice di verità e pensiero (*aletheian kai noun*), e che deve averla vista chi intenda agire saggiamente nelle faccende private e quelle pubbliche[5].

532a5-b2: con la ragione ci si muove verso ogni singola essenza (*auto ho estin hekaston*) e non ci si ferma prima di aver afferrato con il pensiero (*noesei*) l'essenza del Bene (*auto ho estin agathon*), che coincide con il termine stesso dell'intelligibile.

534b8-d1: chi conosce l'idea del Bene è in grado di delimitarla con il discorso (*diorisasthai toi logoi*) isolandola da tutte le altre, sforzandosi di confutare secondo l'essenza (*kat'ousian*).

540a-b: al culmine del processo educativo e un pieno allenamento nella dialettica, i filosofi saranno costretti a detenere il governo della città sulla base del fatto che hanno visto il Bene in sé (*idontas to agathon auto*) e sono in grado di servirsene come modello (*paradeigmati chromenous ekeinoi*).

2. L'interpretazione di Broadie

L'interpretazione di Broadie muove da un punto ampiamente condiviso: l'idea del Bene presenta una eccentricità rispetto alla cosiddetta teoria *standard* delle idee. Tra le altre cose, questo significa soprattutto che il ruolo o funzione svolto dal Bene non è riconducibile, o unicamente riconducibile, alla dinamica di partecipazione per cui un gruppo di cose empiriche che appaiono o sono qualificatamente *F* devono questa loro determinazione alla natura di *F* in sé (cfr. *Phaed*. 100c e seguenti). Sebbene in 1.1 si sia visto che in qualche modo il Bene è causa dell'essere buone di altre entità, funzione che ricorda la causalità *standard* appena ricordata, è innegabile che il Bene sia presentato (cfr. *supra* 1.2) come non definibile alla maniera delle altre idee, che esso sia fonte di verità, conoscibilità ed essere delle cose (e non quindi solamente di bontà) e che esso è misteriosamente descritto come al di là dell'essenza o dell'essere. Questa caratterizzazione dell'idea del Bene ha portato per esempio Mario Vegetti a collocare il Bene nell'ambito dell'essere e delle cose conoscibili (come visto in 1.4, esso è in effetti appartenente all'ambito dell'essere e oggetto di qualche forma di conoscenza) senza però che esso sia, alla maniera delle altre idee, oggetto di definizione (si parla qui del sintagma *logos tes ousias*, cfr. 543b3-6)[6]. In altre parole, si apprenderebbe la natura del Bene guardando a cosa il bene compie,

[5] Altri riferimenti alla visione del Bene sono a 518c-d; 519c-d; 526d7-e5.

[6] Una lettura per cui l'*idea tou agathou* non fosse un'idea nel senso ordinario è già ampiamente argomentata da G. Sillitti, *Al di là della sostanza: ancora su* Resp. *VI 509b*, cit. Una lettura secondo cui il Bene ha un'essenza ed è conoscibile (l'"al di là' si deve intendere come "non si limita ad avere un'*ousia*, ma ha delle funzioni speciali) è quella di M. Dixaut, *Métamorphoses de la dialectique dans le dialogues de Platon*, Paris 2001, pp. 94-102.

piuttosto che a ciò che esso è[7]. La comprensione del Bene risiederebbe dunque nell'apprensione della sua funzione causale, cioè nel riconoscimento del fatto che esso è fonte di conoscibilità delle altre idee e causa del loro specifico modo di essere[8], e che in qualche modo la potenza stessa del Bene è essenzialmente legata alla potenza del discorrere dialettico, quella procedura cioè, come abbiamo visto in 1.3, che coglie la verità nella sua forma più certa e autentica.

È proprio sul solco di questa lettura che si sviluppa l'interpretazione di Broadie: «What we learn [*scil*. through the sun-analogy] is something about the form of the good, namely that it has, or is, a power of making certain other things known» (p. 41)[9], come da ella apertamente riconosciuto (cfr. p. 42 n. 54 e p. 75 n. 107), approdando però ad esiti quanto mai originali, sebbene non privi di criticità. Il primo capitolo del volume, essenzialmente legata alle sezioni di testo menzionate in 1.1, serve da introduzione all'argomentazione di Broadie. Lo scopo dei filosofi-governanti è pragmaticamente connotato: si tratta di governare correttamente sulla base della conoscenza delle virtù necessarie ai cittadini affinché la giustizia sia implementata nella *polis* storica. Stabilire con assoluta certezza quale sia la natura di queste virtù è essenzialmente dipendente dalla natura del Bene, il quale costituisce, come abbiamo visto, il *megiston mathema*, il contenuto conoscitivo della maggior importanza. Nel resto di questo paragrafo seguirò la struttura dello studio di Broadie che corrisponde ad un'analisi della funzione epistemologica e della funzione ontologica del Bene, provvedendo ad evidenziare gli aspetti che risultano più importanti e in alcuni casi che necessitano di una discussione[10].

2.1 Epistemologia

La maggior parte del testo di Broadie consiste in un'analisi puntuale del testo e delle implicazioni dell'analogia del sole e dell'immagine della linea. L'autrice parte dall'assunto che lo scopo della dialettica sia determinare dettagliatamente come implementare le azioni virtuose nella vita della *polis*. Ciò si configura come un esame critico delle definizioni delle virtù che regolano i vari tipi di comportamento. Questo esame consiste essenzialmente nell'analisi di errori concettuali nella formulazione delle definizioni delle virtù e soprattutto nella connessione di queste definizioni con l'idea del Bene. Questa connessione

[7] Cfr. M. Vegetti, Megiston mathema. *L'idea del "buono" e le sue funzioni*, cit., pp. 275 e 282-283.
[8] È questa la posizione di F. Ferrari, *L'idea del bene: collocazione ontologica e funzione causale*, cit., secondo cui il Bene sarebbe la fonte degli attributi comuni a tutte le idee, come la perfezione, l'immutabilità, etc., posizione ad ogni modo non accettata da Broadie (pp. 155-159 e soprattutto p. 156 n. 33).
[9] A differenza di Vegetti, Broadie mi pare ritenga che la *dynamis* del Bene non sia oggetto di conoscenza, bensì sia messa in atto attraverso la G-question. Quindi, non solo, come Vegetti sostiene, il Bene è colto attraverso la sua funzione e la sua essenza non è oggetto di definizione, ma esso non è proprio oggetto di apprensione se non attraverso la sua applicazione.
[10] In maniera piuttosto pratica, Broadie fornisce una sintesi della sua proposta interpretativa a pp. 132-135.

risiede nel porre la G-question: 'questa definizione è buona?' nel senso di 'è bene compiere delle azioni in conformità a questa definizione?'. Questa domanda è analoga al ruolo del sole nella misura in cui è l'unica procedura che permette di illuminare in maniera certa una definizione di virtù e conseguentemente la correttezza e opportunità della sua implementazione, ed è questa domanda che permette di rimuovere le ipotesi infondate in un movimento di 'risalita', a differenza della matematica.

Questa lettura implica che gli intelligibili a cui il Bene dona verità ed essere non sono tutte le idee, ma si limitano unicamente a quelle delle virtù discusse precedentemente nella *Repubblica*, come la giustizia, il coraggio, la temperanza. Che il Bene sia paragonato al principio non-ipotetico significa per Broadie che il Bene non può essere trascurato nella misura in cui esso è il principio del valore e della desiderabilità[11]. In altre parole non è possibile pensare che ci sia qualcosa di importante e di valore che non sia legato al Bene e quindi essenzialmente alla domanda 'è buono?/è bene che…?'. Allo stesso tempo, la G-question è principio di conoscenza perché l'unico modo per darvi risposta è addurre delle ragioni che coincidono con l'esercizio del metodo dialettico. Se la risposta alla domanda rispetto alla definizione di una virtù 'sarebbe bene se si agisse in conformità a questa definizione?' è positiva allora si ottiene l'unica dimostrazione possibile che la definizione in questione rappresenti veritieramente la sua virtù. In questo senso il Bene è fonte di conoscenza e verità rispetto agli altri intelligibili, cioè le idee delle virtù. Questo impianto è fortemente radicalizzato da Broadie che si spinge ad asserire che l'idea del Bene è nella sua stessa natura *interrogativa* e pertanto essa non è un oggetto di indagine da parte del dialettico quanto piuttosto il suo strumento (*tool*) essenziale. Di conseguenza il Bene non è un paradigma verso cui guardare per ordinare la città (cfr. p. 54) e non è neanche già che soggiace all'uso del predicato 'buono' in genere.

La garanzia dell'applicazione della G-question è data dal fatto che i governanti hanno ricevuto un'educazione che ne sviluppasse in massimo grado le virtù morali e intellettuali. Il contrasto con la matematica che abbiamo visto in 1.3 è letto da Broadie nella maniera più radicale: Platone starebbe specificando che la matematica non è affatto un modello di ragionamento che la dialettica deve riprodurre ed eventualmente potenziare o inverare, ancor meno che il Bene in quanto principio non-ipotetico sia legato a procedure di carattere deduttivo, simili, sebbene superiori, a quelle delle tecniche matematiche o serva a fornire un fondamento alle ipotesi della matematica. Si tratterebbe di un tipo di ragionamento di genere completamente diverso, nello specifico un *pratical reasoning*. La differenza fondamentale è che la dialettica è essenzialmente legata

[11] A p. 51, Broadie fornisce questa argomentazione: «There is no possibility of interrogating the value of the method of the G-question by some further standard of value, for there is no further standard; and we cannot use the question to interrogate itself since if we distrust it enough to interrogate it, we cannot trust any answer it elicits, whether about other intelligibles or about itself». Il che è rappresentato dal fatto che il sole può essere visto solo dopo che esso stesso ha emanato la sua luce.

al tipo di contesto in cui si deve agire e in cui si intende implementare una certa virtù o norma. Dato questo obiettivo, una parte significativa dello sforzo di Broadie sta nell'enfatizzare gli aspetti che possono suggerire una differenziazione radicale tra matematica e dialettica operata da Platone.

Questo spiegherebbe anche perché non c'è una vera e propria illustrazione di un ragionamento dialettico: questo, infatti, richiederebbe l'essere collocati in una certa situazione per compiere i ragionamenti pratici del caso. Questa lettura è piuttosto audace e può offrire il fianco ad alcune critiche, come si vedrà a breve. Ad ogni modo, Broadie non disconosce completamente la natura ontologica delle idee (cfr. anche p. 53). Ella asserisce che le idee hanno una natura infinitamente complessa o quanto meno carica di implicazioni che non devono essere conosciute nella loro totalità dal dialettico, il quale è interessato unicamente a ciò che l'idea detta alla ragione per agire correttamente in un contesto dato. Il fatto che l'apprensione dell'idea non sia completo non significa che le conclusioni tratte dal dialettico siano provvisorie, esse sono piuttosto categoriche. In altre parole, una qualunque idea è colta nella sua unicità e immutabilità e la diversità delle conclusioni derivanti dalla sua applicazione è determinata dalla varietà dei contesti in cui si applica. Questo è connesso al fatto che di volta in volta per dire che qualcosa è ad esempio giusto in un dato tipo di contesto bisogna sempre fornire delle ragioni.

Vorrei ora sollevare solo alcune criticità e di contro-argomentazioni, tra le molte possibili, alla lettura di Broadie. Vorrei partire dalla tesi che il Bene è uno strumento che è conosciuto in quanto impiegato, cosa che sembra confliggere con molti passi, soprattutto quelli menzionati in 1.4. Broadie considera questa obiezione a pp. 55-65 e risponde in maniera innegabilmente ingegnosa, che però non risulta convincente: quando Socrate parla dei dialettici che vedono e conoscono l'idea di Bene in realtà ci sarebbe un passaggio a un meta-livello linguistico o secondo ordine logico per cui Socrate sta parlando del fatto che i dialettici si rendono conto pienamente, e si impegnano all'uso, della G-question. In altre parole, conoscere il Bene significa in quei passi aver piena consapevolezza del dover usare l'interrogativo del Bene come unico fondamento della conoscenza. Questa tesi è sviluppata soprattutto in relazione a 540a-b, dove si dice che i filosofi hanno visto il Bene in sé (*idontas to agathon auto*) e sono in grado di servirsene come modello (*paradeigmati chromenous ekeinoi*). L'interpretazione di Broadie risulta eccessivamente speculativa perché non c'è alcun riferimento ad accezioni diverse del termine Bene che dovrebbero corrispondere a ordini logici diversi o differenze uso/menzione[12]. Inoltre, i termini che compaiono nel passo, ad esempio il sintagma *to agathon auto* nella *Repubblica* e altrove è chiaramente riferito a una realtà in sé, sia essa il Bene o meno, come è anche chiarissimo da 532a5-b2, e non può certo indicare il precetto di adottare il metodo di applicare

[12] La stessa Broadie riconosce che la distinzione da lei proposta è forse troppo rigida, cfr. pp. 64-65.

il Bene come domanda. Lo stesso valga per il termine *paradeigma* che dovrebbe essere usato qui in un senso completamente alieno alla lettura tradizionale, senza alcuna indicazione in merito.

Questo risulta ulteriormente problematico se si tiene a mente che nel *Cratilo* (389 e seguenti) e *Resp*. X (596 e seguenti), al netto della complessa questione delle idee di *artefacta*, gli strumenti sono presentati come aventi una natura ben definita che svolge innanzi tutto il ruolo di modello, nel senso classico. Ancor di più, in entrambi i contesti, la conoscenza delle idee è associata all'uso che è essenzialmente regolato dall'apprensione di un oggetto, che è la natura dello strumento[13].

Ovviamente questa critica non è definitiva nella misura in cui la lettura di Broadie parte proprio dall'assunto che l'interpretazione, per così dire, *difficilior* dei passi riportati in 1.4 è quella giusta proprio perché si accorda con l'interpretazione data del resto. È proprio su questo che ora vorrei esprimere delle perplessità, che si condensano su due punti, uno esterno alla prospettiva di Broadie e uno interno.

A pp. 84-94, Broadie presenta la dialettica come un *practical reasoning*. Sintetizzando al massimo, la dialettica consegna delle risposte che sono conoscitivamente certe ma solo e unicamente rispetto a un determinato contesto. Il confronto con la matematica aiuta a comprendere questo punto. La matematica è detta essere *monotonic*. Se ci sono tre premesse A, B e C si ottiene la conclusione X. Se si aggiunge una premessa D, sarà eventualmente il caso che le quattro premesse diano come conclusione Y, ma questo non influenza la deduzione precedente. In altre parole, finché ci saranno le tre premesse si avrà una data conclusione, quando esse saranno quattro se ne darà un'altra e le due deduzioni sono parimenti valide e corrono in parallelo. Secondo Broadie invece la dialettica non è un'impresa conoscitiva di questo tipo. Se ci si chiede se un tipo di azione o norma è giusta, la dialettica fornirà una risposta assoluta e giusta fintantoché si considerino le premesse dettate da un certo contesto, ma variando il contesto allora non si avrà più una risposta soddisfacente, che dovrà essere integrata, rispetto alla nuova situazione empirica attraverso la reiterazione della G-question.

L'esempio è quello del restituire le proprietà di un altro. Tendenzialmente questa norma è razionale e giusta, ma in alcuni contesti, se il proprietario della spada è in uno stato di delirio e rischia di danneggiarsi, non è giusto restituire le sue proprietà. Per usare le parole di Broadie a p. 88: «The conclusion of a piece of mathematical knowledge is always as provisional as the premisses; the conclusion of a piece of dialectic is final for *this* occasion, although never for *all* occasions», e continua asserendo «It is therefore reasonable to say that the ruler, in relation to this situation, has knowledge of justice». Quindi conoscere un'idea come la Giustizia coinciderebbe con il conoscere quando un'azione o

[13] Su questo rimane fondamentale G. Cambiano, *Platone e le teniche*, Roma-Bari 1991, il quale sviluppa un'interpretazione della teoria delle idee dove l'aspetto pratico-normativo è fondamentale, senza abdicare all'aspetto onto-epistemologico (sull'idea del Bene si veda pp. 172-176.

una norma è effettivamente giusta unicamente in riferimento a un dato contesto. Questa interpretazione radicale porta Broadie (p. 95) a dire che: «none of this implies that the results of dialectic – the rulers' judgements on whether this or that is the right interpretation of some virtue-concept – are infallible», e questo perché la conoscenza della virtù in questione è fondata sul Bene, ma è sempre in riferimento a fatti empirici su cui ci si può sbagliare.

Questa interpretazione è sostanzialmente difficile da accettare poiché in contrasto con le indicazioni fondamentali che Platone dà sulla natura della conoscenza. Basti considerare la fine del libro V della *Repubblica* (476 e seguenti): ci viene detto che l'*episteme* è *anamartetos* (*tout court*), priva di errore, e quindi *deve* essere infallibile; che essa è riferita a *to pantelos on*, ciò che è pienamente, che è ben diverso dai fatti empirici che allo stesso tempo sono e non sono[14]; inoltre, risulta molto difficile che quando alla fine dell'immagine della linea Socrate parla del fatto che la dialettica rimane unicamente al livello delle idee (cfr. *supra* 1.3)[15], questo significhi valutarle rispetto a dei fatti empirici su cui ci si può sbagliare *tanto da rendere la dialettica fallibile*. Una distinzione che Broadie non sembra considerare è quella tra la conoscenza di un'idea e la capacità di interagire con la realtà empirica *sulla scorta* della conoscenza di quell'idea[16]. Le due nozioni sono ben distinte[17], e probabilmente il dialettico interagisce in maniera fallibile con la realtà sensibile, a causa della natura di quest'ultima, ma questa interazione non sembra essere ciò che Platone chiama conoscenza, sebbene conoscere le idee sicuramente fornisca quanta più solidità possibile nell'apprendere i caratteri dei fatti empirici. Anche chi tra gli interpreti consideri il raggiungimento dell'*episteme* come sottostante a condizioni troppo restrittive per le anime incarnate, riservandola a una vita ultraterrena, adottando quindi una lettura scetticizzante, non arriva mai a dire che l'*episteme* è fallibile, ma al contrario che dato che è infallibile essa non è raggiungibile in vita.

Per quanto riguarda la fissità ontologica delle idee rispetto ai contesti empirici, Broadie la rappresenta metaforicamente come segue (pp. 118-120): l'idea è simile alle funzioni matematiche, queste hanno un'identità ben precisa

[14] Cfr. F. Ferrari, *L'idea del bene: collocazione ontologica e funzione causale*, cit., p. 303: «Ogni tentativo di comprendere la natura e la funzione dell'idea del bene deve fare i conti con l'esigenza di spiegare il nesso di identità e differenza che lega questa entità al prodotto della sua azione causale, cioè appunto alla coppia formata da verità ed essere». Qui Broadie dà un senso eccessivamente empirico, e in quanto tale non platonico, del senso di 'essere'. Questo è chiarissimo da quanto segue nello studio di Ferrari: essere e verità costituiscono il *pantelos on*, cioè ciò che distingue ontologicamente le forme dagli oggetti empirici, come è dimostrato dal riferimento all'opinabile a 508d6-9.

[15] Su cui comunque si veda pp. 78-79.

[16] Cfr. anche p. 122 n. 179 dove Broadie pone la seguente domanda: «Does Plato assume too easily that knowledge is only ever of something pure and unmixed with its opposite, instead of allowing that, as well as perfectly unambivalent knowledge of perfectly unambivalent things, there is also chiaroscuro knowledge of chiaroscuro things?». La risposta genuinamente *platonica* sembra essere negativa.

[17] Per una chiara discussione di questa distinzione e del perché la seconda *non* è conoscenza nel senso platonico, si veda da ultimo J. Moss, *Plato's Epistemology. Being and Seeming*, Oxford 2021, pp. 122-131.

che fornisce delle risposte e si manifesta unicamente in relazione ai diversi argomenti, secondo lo stesso criterio o regola, argomenti che fuor di metafora sono i dati empirici dei diversi contesti[18]. A questa funzione, chiaramente giudicata più interessante da Broadie, la studiosa accompagna la possibilità di pensare le idee come isolate dall'empirico, che in ultima analisi le rende oggetto di devozione (sic). Per comprendere le riserve di Broadie verso il lato delle idee che le qualifica come isolate dai contesti pratici si veda il seguente passo (p. 50):

> Only when predicative 'good' plays this interrogating role do the forms, essences, or natures of the virtues come to be securely recognized as such, with the result (for the ideal city) that they are not just 'there' in the landscape of forms but pass to fuller actualization in impact on human life, becoming 'useful and beneficial' (505a2–4).

L'obiezione maggiore che sorge verso questa interpretazione è che come abbiamo visto in 1.2, se il Bene è origine di essere e conoscenza, e conoscenza è il riconoscimento della verità della definizione rispetto a un dato contesto empirico e pratico, allora l'essere di cui è origine il Bene sarebbe l'attualizzazione della virtù definita nell'esperienza. Eppure, sembra molto difficile che l'essere e l'essenza che il Bene conferisce agli enti intelligibili sia l'essere realizzati empiricamente, cioè che una data azione o norma sia ad esempio giusta, perché qui l'essere sarebbe lo stesso che la partecipazione di una cosa empirica all'idea, generando un cortocircuito chiaramente antiplatonico. Della lettura della funzione ontologica del Bene parlerò brevemente nel prossimo sotto-paragrafo. Ora vorrei concludere con la seconda criticità rispetto all'epistemologia del Bene che sorge dal testo di Broadie.

Come abbiamo visto, risulta chiaro che il Bene non sia una forma *standard* e che quindi l'esercizio della sua funzione e della sua causalità non sia coinvolto unicamente o primariamente in dinamiche di partecipazione o di imitazione da parte di altre entità[19]. Tuttavia, è nell'impianto stesso dell'interpretazione di Broadie, vale a dire quello il cui centro nevralgico è la G-question, che l'intelletto si interroga se una data cosa *è buona* e questo sembra proprio associare l'idea del bene a una dinamica partecipativa simile a quella tradizionale. Questo di per sé non è un problema, anzi ritengo con Broadie (ad es. p. 173) che sia essenziale riconoscere anche questo aspetto all'idea del Bene. Allo stesso tempo, Broadie specificando che l'apprensione cognitiva del Bene è esaurita dalla G-question sembra lasciare in sospeso il punto che ritengo essere tipicamente platonico: 'x è buono rispetto a cosa?' dove la risposta non può che essere ancorata alla natura

[18] L'interpretazione secondo cui le idee siano regole del pensiero di per sé non è affatto nuova, si veda P. Natorp, *Dottrina platonica delle Idee*, Milano 1999, sebbene la conclusione sia completamente diversa rispetto a quella di Broadie (cfr. pp. 242-243). Un aspetto che sembra corretto della lettura di Broadie è che la decodifica di questo modo di funzionare delle idee coincide con la necessità di elaborare razionalmente e quindi fornire argomenti razionali come base dell'implementazione di una norma nella realtà empirica.

[19] Come abbiamo riferito a più riprese, è questa la posizione di Vegetti, e anche quella espressa in M. Dixsaut, *L'analogie intenable: le Soleil et le Bien*, cit., pp. 144-145.

del Bene come fonte di informazione e non solo come domanda. In altre parole, inevitabilmente alla constatazione 'la tal cosa è *F*' deve seguire la domanda 'che cos'è *F*?' e da questo punto di vista non è data dall'autrice una soluzione convincente[20]. Quando ci si chiede se una certa definizione di giustizia è buona, dove, in caso di risposta positiva, si avrebbe anche garanzia che quella definizione è vera (questo, che è il nucleo dell'esegesi di Broadie è certamente ingegnoso e non privo di qualche plausibilità), è necessario avere *qualcosa* che stabilisca in maniera univoca la risposta. Questo secondo passaggio è totalmente assente in Broadie, la quale, coerentemente al suo impianto, ritiene che la risposta provenga dall'analisi contestuale e situazionale di una definizione alla prova dell'azione pratica. Tuttavia in questo modo non si capisce *che cosa* sia la bontà delle virtù e delle loro implementazioni. In effetti, Broadie non fornisce mai una spiegazione del significato di 'buono/Bene', se non forse (p. 49) che esso è «nothing more complicated than 'desirable, to be pursued'»[21]. Un'altra possibile risposta a quale sia la base per stabilire che qualcosa sia buono è riassunta a p. 211:

> For the interrogative form of the good to be truly sun-like, in other words an actual source for accurately tracking the natures of other intelligibles – the ones that have to be recognized in order for the good to be empirically brought about and possessed – the interrogation must be conducted by an intelligence in the right condition, and this condition is a love of truth trained to be at home in the world of intellect and undergirded by unshakeable moral excellence.

Questo approccio però è a rischio di circolarità: l'unica garanzia della verità di un'apprensione corretta e la conseguente implementazione di una virtù è dato dalla G-question, eppure qui si dice che la capacità della G-question di illuminare la verità delle virtù si basa sul fatto che chi pone la domanda possegga quelle virtù. Risulta quindi molto difficile che la base che *garantisca* la conoscenza delle risposte alla G-question sia l'essere virtuoso dell'interrogante, visto che uno può conoscere veramente se si è virtuosi solo sulla base della 'luce' derivante dalla G-question.

2.2 Ontologia

Molto velocemente vorrei ora considerare le due proposte che Broadie fornisce per interpretare il ruolo ontologico del Bene nei confronti degli altri intelligibili, le quali risultano strettamente dipendenti dalla lunga discussione del ruolo epistemologico del Bene. Broadie rifiuta le due letture principali secondo cui il Bene fornisce i caratteri comuni alle altre idee (immutabilità, etc.)

[20] Cfr. G. Giannantoni, *La dialettica platonica*, «Elenchos», 15, 1994, pp. 105-115. Broadie, *apertis verbis*, nega che questo approccio possa essere applicato al Bene, cfr. 152.
[21] Correttamente Broadie (p. 99) riconosce la corrispondenza tra valore assiologico e valore conoscitivo, su cui però cfr. M. Vegetti, *Dialettica*, cit., p. 431, il quale evidenzia la caratura morale intrinseca alla dialettica, senza però intenderla come un *practical reasoning*. Scopo del buono è la *valorizzazione* cioè trasferimento di utilità, desiderabilità, quindi intenzionalità conoscitiva al campo delle idee in quanto tali, della verità e della scienza.

o che il Bene costituisca il sistema di tutte le idee nei loro rapporti e nel loro ordinamento. Delle due proposte che offre, Broadie predilige la seconda. Per sommi capi sono le seguenti:

I) L'aggiunta dell'essere agli intelligibili (virtù) starebbe nel fatto che essi da entità inerentemente e isolatamente buone siano impiegati come misura della virtù di una certa azione o norma. In altre parole, l'essere sarebbe l'utilizzo nell'implementazione della loro natura in un contesto empirico-pratico, attraverso la G-question.

II) Broadie parte col definire intelligibili alcuni *types* di azioni (ad es. mantenere la parola data, etc.) di per sé moralmente neutre. Quando si considerano questi tipi di azioni come miranti a rappresentare un comportamento virtuoso, il Bene causerebbe il loro *essere* effettivamente virtuosi, se la risposta alla G-question è positivo.

Broadie ritiene non sia ovvio stabilire quale sia più adatta e sebbene, a sua detta, la prima si adatti meglio all'analogia del sole, ella afferma che la seconda rende meglio conto della connessione tra i testi che abbiamo visto in 1.1 e in 1.2 assieme a passi dalla fine del libro V. Non è possibile qui entrare nel dettaglio, mi pare tuttavia che entrambe le opzioni non siano scevre da difficoltà. Per quanto riguarda (I), come abbiamo già visto, si interpreta l'essere come il divenire nell'ambito dell'empirico, sebbene nella veste della 'implementazione', e quindi della partecipazione, il che non può essere ciò che Platone intende con 'essere' in quel contesto. Per quanto riguarda (II), che effettivamente sembra più plausibile tra le due, ritenere che tipi di azioni siano intelligibili nel senso platonico non appare legittimo o quanto meno questi ultimi non sono ciò che degli intellegibili coglie l'essenza, ma al massimo una loro qualità (il riferimento è qui ovviamente al *Menone*, 71b-d).

2.3 Questioni aperte

Sebbene la valutazione complessiva di questo testo sia critica, deve essere detto che l'interpretazione di Broadie è molto sofisticata e sottile, oltre che molto scrupolosa nella discussione della letteratura critica e nella valutazione di diverse prospettive interpretative. Inoltre, pone diverse questioni che ritengo essere di grande importanza e che richiedono ulteriore discussione e riflessione. Molto brevemente vorrei riportare almeno tre punti.

1) L'interpretazione di Broadie prende molto seriamente la misteriosa asserzione di Socrate per cui l'idea del Bene è al di là dell'*ousia*. La natura del Bene è eccentrica per la sua preminenza e questo si riflette sulle procedure conoscitive messe in atto dal dialettico. L'interpretazione ha il pregio di coniugare questo aspetto del Bene riconoscendogli un ruolo di fondazione della verità e

conoscibilità degli altri intelligibili senza abbandonare la sua natura di Bene (non offrendo la tentazione, probabilmente scorretta, di ridurre l'idea di Bene a una idea di realtà o di essere *tout court*). In altre parole, Broadie sviluppa una teoria complessa che tenta, giustamente, di tenere assieme molti aspetti apparentemente contrastanti: l'appartenere il Bene all'essere, la sua eccentrica non definibilità, il suo essere causa di verità, conoscibilità ed essere, la sua natura di Bene che rende buone le altre cose. In aggiunta, Broadie coniuga questi aspetti con la natura pragmatica dell'operazione della *Repubblica*, indubbiamente presente: in fin dei conti, Platone sta illustrando una Callipoli che può esistere e a cui si deve tendere; sebbene l'autrice enfatizzi troppo questo aspetto, usandolo come chiave interpretativa univoca. Per il lettore che non fosse convinto della soluzione di Broadie, rimane comunque il fatto che tentativi complessi di tenere assieme questioni solo apparentemente distanti o incompatibili è giusto siano compiuti.

2) Uno dei punti più convincenti dell'interpretazione dell'epistemologia dell'immagine della linea è la spiegazione del fatto che la modalità di procedere della dialettica sia non deduttiva (pp. 65-71). Qui Broadie offre una serie di argomentazioni che sono ovviamente funzionali alla sua interpretazione interrogativa del Bene, ma che possono essere anche valorizzate da prospettive differenti. L'idea di fondo è che il Bene non costituisce un principio da cui dedurre la totalità dell'esistente e che in generale la scelta terminologica dell'ascesa al principio sia da collocarsi nella prospettiva di maturazione epistemica e allo stesso tempo morale che è rappresentata dall'immagine della linea nella sua interezza. Il Bene, quindi, è correlato a una forma di razionalità né matematica né meta-matematica, che *è* la dialettica e che, agli occhi di Platone, è superiore alla matematica.

3) Infine, c'è un punto di potenziale interesse (pp. 111-123) che non ha ricevuto, mi pare, sufficiente attenzione dagli altri studiosi, sebbene non sia stato del tutto ignorato: si può assumere che conoscere un'idea in quanto natura perfettamente determinata e unica sia in grado di fornire il fondamento a un'infinità di ragionamenti e pensieri che mirano a descriverla e definirla, sebbene sia ragionevole pensare che un solo discorso ne cattura univocamente l'essenza (*logos tes ousias*). L'autrice è interessata ai contesti empirici di applicazione di una virtù, per cui trova inaccettabile che conoscere ad esempio l'idea di Giustizia implichi avere a priori un registro di tutte le situazioni possibili in cui si può essere giusti. Certamente non sembra che Platone sostenga qualcosa del genere, ma dato che Broadie ritiene che conoscere sia sapere come implementare nell'orizzonte pratico una virtù, è costretta a schiacciare i due livelli (empirico e ideale), mentre possiamo benissimo pensare che il dialettico conosca la natura di una virtù e sulla base di questa *episteme* sappia come agire empiricamente di volta in volta, senza far collassare il primo livello sul secondo. Ad ogni modo, ritengo che questo ordine di problemi sia di grande interesse rispetto agli sviluppi del tardo Platone (questione non menzionata da Broadie) in cui la natura *relazionale*

delle idee o generi diviene di importanza capitale. Conoscere un'idea significa conoscerne, allo stesso tempo, *tutte* le relazioni? Se sì, c'è il rischio che conoscere un'idea implichi conoscerle tutte allo stesso tempo, il che sembra un esito indesiderato per il filosofo platonico.

Conclusione

Per concludere vorrei ribadire che il libro di Broadie è uno studio molto sofisticato che esibisce una capacità analitica fuori dal comune, sia rispetto al senso complessivo dell'operazione, sia rispetto ad alcuni dettagli, sulla maggior parte dei quali non c'è stata la possibilità di soffermarsi. Questo fatto emerge anche per chi scrive, che, come sarà apparso chiaramente, non trova in ultima analisi ineludibili le conclusioni a cui Broadie giunge, né è del tutto persuaso dei suoi punti di partenza. L'operazione esegetica di Broadie rientra nel legittimo confronto di interpretazioni e scambio di idee, rappresentando in maniera chiara un'opzione interpretativa molto definita, audace e come l'autrice stessa riconosce (p. 206) «disconcertingly unorthodox». Sebbene non paia il modo migliore per rendere conto del testo platonico, penso che chiunque abbia motivo di sostenere un'interpretazione in cui la dialettica sia essenzialmente un *practical reasoning* possa farlo proficuamente sulla scia di questo dettagliato studio. Detto questo, è necessario evidenziare un ultimo aspetto.

Come molti studiosi di storia della filosofia antica sanno, una parte degli studi interpreta il compito dello storico come quello di 'smascherare' le deficienze del testo del pensatore antico, fenomeno che ha investito soprattutto Platone. Sarebbe impossibile qui riportare l'enorme numero di esempi di questo atteggiamento, basti pensare a tutte le volte in cui si conclude che il pensatore in questione 'è confuso' oppure 'non ha capito che…', o ancora 'non aveva ancora a disposizione questa terminologia o concetto filosofico'. Il mio timore è che questo atteggiamento risulti anacronistico e nella maggior parte dei casi anche fuorviante per una corretta esegesi, anche perché probabilmente la sua origine è di natura ideologica e/o sociologica e non corrisponde a un approccio pienamente scientifico. Duole dire che questo atteggiamento, contrario a un ragionevole 'principio di carità', sebbene non sia preponderante e sia spesso accompagnato da apprezzamenti e attenzione al testo, non è assente dal testo di Broadie. Ad esempio, a p. 97 si dice che Platone non abbia capito che matematica e dialettica appartengono a due generi diversi di conoscenza e non a gradi diversi della stessa 'virtù intellettuale' (sic), questo perché per Broadie la dialettica è un *practical reasoning* e la matematica no. Ma è chiaro che la cosa possa essere letta molto diversamente: forse Platone non compie questa distinzione perché questa distinzione, per lui, non c'è (o magari la dialettica non è un *practical reasoning*).

Si veda anche ad esempio p. 107, dove si parla della «almost pitiful poverty of his [*scil*. Plato's] logical vocabulary». In questo caso lo sguardo nei confronti di Platone è di commiserazione, sulla base di requisiti impostati anacronisticamente, come è chiaro dalla battuta poco oltre nel libro: «It is not impossible that when

he wrote the *Republic* he even lacked a reflective grasp of the difference between a fact or state of affairs as represented by a proposition, and a complex object». Eppure, potrebbe essere che questa distinzione non sia nell'orizzonte di pensiero e di interesse del filosofo e quindi non abbia senso pretendere che sia una scoperta 'oggettiva' da fare per condurre correttamente le proprie riflessioni filosofiche.

Si consideri infine il seguente passo, dove il giudizio si sarebbe potuto esprimere in maniera più appropriata (p. 135): «For all his dazzling acuity and glorious eloquence, he [*scil*. Plato] was terminologically handicapped to a degree we can hardly begin to imagine what it would be like to live with». Soprassediamo sul fatto che 'handicapped' sia un termine problematico[22], e che il riferimento alla disabilità come paradigma della manchevolezza sia ormai messo in discussione. Platone è senza dubbio uno dei più grandi pensatori dell'umanità, che tra le altre cose ha esplorato il proprio genio nel conio di decine di termini dando vita alla filosofia occidentale (non ultimo il parlare stesso di 'filosofia'). Che proprio Platone sia definito terminologicamente deficitario, tanto da non poter immaginare come potesse vivere in una tale condizione – perché il suo pensiero non corrisponderebbe alle categorie concettuali degli interpreti di più di duemila anni dopo – non rientra nel legittimo conflitto delle interpretazioni, ma risulta, francamente, arbitrario.

Lorenzo Giovannetti
Istituto per il Lessico Intellettuale Europeo
e Storia delle Idee - CNR
✉ lorenzo.giovannetti@cnr.it

Bibliografia

Adams, R. et al. 2015. *Disability*, in Iid. (eds.), *Keywords for Disability Studies*, New York, pp. 30-44.

Baltes, M. 1997. *Is the Idea of Good in Plato's Republic Beyond Being?*, in M Joyal (ed.), *Studies in Plato and the Platonic Tradition*, Aldershot, pp. 3-23.

Cambiano, G. 1991. *Platone e le teniche*, Roma-Bari.

Dixsaut, M. 2000. *L'analogie intenable: le Soleil et le Bien*, in Ead., *Platon et la question de la pensée. Études platoniciennes*, Paris, pp. 121-151.

Dixaut, M. 2001. *Métamorphoses de la dialectique dans le dialogues de Platon*, Paris, pp. 94-102.

Ferber, R. 1989. *Platos Idee des Guten*, Sankt Augustin.

[22] Cfr. ad esempio Adams, R. et al., *Disability*, in Id. (eds.), *Keywords for Disability Studies*, New York 2015, pp. 30-44; Kirchner, C., L. Ben-Moshe, *Language and terminology*, in S. Burch, (ed.), *Encyclopedia of American Disability History*, New York 2009, pp. 546-550.

Ferrari, F. 2001. *La causalità del bene nella* Repubblica *di Platone*, «Elenchos», 22, 2001, pp. 5-38.

Ferrari, 2003, *L'idea del bene: collocazione ontologica e funzione causale*, in M. Vegetti (ed.), *Platone. La Repubblica*, Vol. V, Napoli 2003, pp. 287-325.

Franco Repellini, F. 2003. *La linea e la caverna*, in M. Vegetti (ed.), *Platone. La Repubblica*, Vol. V, Napoli, pp. 355-403.

Fronterotta, F. 2001. *ΜΕΘΕΞΙΣ La teoria platonica delle idee e la partecipazione delle cose empiriche. Dai dialoghi giovanili al* Parmenide, Pisa.

Giannantoni, G. 1994. *La dialettica platonica*, «Elenchos», 15, pp. 105-115.

Kirchner, C., Ben-Moshe, L. 2009. *Language and terminology*, in S. Burch, (ed.), *Encyclopedia of American Disability History*, New York, pp. 546-550.

Lafrance, Y. 1987. *Pour interpréter Platon. La ligne en* République *VI 509d-511e. Bilan analytique des études (1804-1984)*, Montréal-Paris.

Moss, J. 2021. *Plato's Epistemology. Being and Seeming*, Oxford.

Natorp, P. 1999. *Dottrina platonica delle Idee*, Milano.

Santas, G. 2000. *The Form of the Good in Plato's Republic*, in G. Fine (ed.), *Plato*, Oxford, pp. 249-276.

Sillitti, G. 1980. *Al di là della sostanza: ancora su* Resp. *VI 509b*, «Elenchos», 1, pp. 225-244.

Smith, N. D. 1996. *Plato's Divided Line*, «Ancient Philosophy», 16, pp. 25-46.

Vegetti, M. 2003. *Dialettica*, in Id. *Platone. La Repubblica*, Vol. V, Napoli, pp. 405-433.

Vegetti, N. 2003. Megiston mathema. *L'idea del "buono" e le sue funzioni*, in Id. (ed.), *Platone. Platone. La Repubblica*, Vol. V, Napoli, pp. 253-286.

Vegetti, M. 2018. To agathon: *buono a che cosa? Il conflitto delle interpretazioni sull'idea del buono nella* Repubblica, in Id., *Il potere della verità. Saggi platonici*, Roma, pp. 113-133.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771612

Discussione

Schelling e Pareyson: sulla rilevanza filosofica del cristianesimo come pensiero tragico

Un commento a F. Tomatis, *Il Dio vivente. Libertà, male, Trinità in Schelling e Pareyson*

Morcelliana 2022

Stefano Piazzese

Schelling and Pareyson: on the philosophical relevance of Christianity as tragic thought. A commentary on F. Tomatis, *Il Dio vivente. Libertà, male, Trinità in Schelling e Pareyson*

The purpose of this discussion is to highlight one aspect of Tomatis' work *Il Dio vivente. Libertà, male, Trinità in Schelling e Pareyson*, or rather the relevance of Christianity for philosophy understood as tragic thought. To do this we will try to follow the stages of the study path traced by the author where emerge the relationship between Christian revelation and philosophy and the terms of this problematic relationship. Schelling and Pareyson did found in the Christian revelation a theoretically fertile ground, in order to answer some fundamental questions of existence and to build their own systems of thought.

***

L'opera di Tomatis *Il Dio vivente. Libertà, male, Trinità in Schelling e Pareyson* è un rigoroso e profondo percorso teoretico volto a scandagliare alcune questioni fondamentali della filosofia, rinvenibili a partire dal titolo stesso, attraverso Schelling e Pareyson. Due nomi e due situatività storico-filosofiche diverse che hanno saputo accogliere nei rispettivi pensieri le sfide principali – domande fondamentali – che al mutare delle epoche storiche rimangono incessantemente presenti nell'istante del domandare, mai risolte del tutto, irrimediabilmente aperte a nuove risposte, sempre prospicienti su nuovi itinerari.

L'aspetto dell'opera che qui s'intende sviluppare riguarda la rilevanza del cristianesimo per la filosofia e i termini di tale rilevanza che Tomatis fa emergere indagando il pensiero dei due filosofi: una questione che caratterizza la cultura europea e la sua storia ponendosi come fondamento della stessa, eppure, spesso, posta in secondo piano poiché, stando allo spirito del tempo e ad alcune mode filosofiche del presente, non risulterebbe connessa al principio di realtà, alla concretezza della vita. Il richiamo al trascendente o, ancor di più, all'*Uno* trascendente, significherebbe aprire i confini del ragionamento a un inquinamento mistico del pensiero (in tal caso, sarebbe lecito parlare di 'inquinamento' visto che vi è anche un *mistico* tutto filosofico?[1]) e, stante lo stesso pregiudizio, attuare una svilente flessione dello sguardo speculativo verso l'*altrove* che impedisce di cogliere il *qui e ora*: un tradimento nei confronti della terra, insomma. Schelling e Pareyson, invece, hanno trovato nella rivelazione cristiana un terreno teoreticamente fecondo per rispondere ad alcune questioni fondamentali dell'esistenza e per edificare i propri sistemi di pensiero.

Partendo dal concetto schellinghiano di *stupore della ragione* in Pareyson, l'autore pone immediatamente in evidenzia che per il filosofo italiano l'esperienza religiosa è rilevante per comprendere l'esistenza, laddove quest'ultima è paradossale coincidenza di relazione con sé e relazione con l'altro, ovvero «quell'esperienza di trascendenza che nella stessa limitata situazione dell'uomo intende il darsi di Dio, comprende l'approssimarsi di Dio all'uomo che crede in Lui, per instaurare liberamente una relazione positiva e significativa, salvifica per la persona mortale» (p. 17). Va da sé che tale esperienza religiosa per Pareyson non corrisponde, per forza, all'esperienza di trascendenza costitutiva dell'esistenza umana. Solo la seconda è legittimamente estendibile a tutti gli uomini, poiché mantiene la questione su un piano "laico" fondato sul rigore filosofico, dice Tomatis, del trascendentalismo kantiano e fichtiano. Solo l'esperienza di trascendenza, dunque, rende possibile un'ontologia della libertà, la quale presuppone, a sua volta, una libertà originaria, abissale, ancipite. Per tale ragione, *ontologia della libertà* ed esperienza religiosa non risultano necessariamente implicantisi, anche quando si tratta di quella particolare esperienza «che inoltre Pareyson interpreterà nella sua ermeneutica filosofica dell'esperienza religiosa cristiana, e che pur può fornire elementi, significati, simboli con cui la filosofia puramente esistenziale e trascendentale può pensare e dialogare» (*ibid*.).

Tale dialogo ha avuto luogo nella storia della filosofia determinando la nascita e i successivi sviluppi teoretici delle idee che hanno plasmato la storia, e che dalle contingenze storiche sono sorte. La concezione cusaniana di Dio come *coincidentia oppositorum* testimonia quanto sia Cusano che, successivamente, Schelling abbiano attinto, in grande misura, alla fonte della tradizione biblico-cristiana e al neoplatonismo reinterpretando queste due tradizioni in modo originale. Anche se il nome di Cusano non è presente negli scritti schellinghiani pervenutici, l'autore ricostruisce rigorosamente l'influenza di alcuni concetti

[1] Cfr. M. Cacciari, *Paradiso e naufragio*, Einaudi, Torino 2022, pp. 10, 87.

cusaniani nella teoresi del filosofo tedesco. Schelling, «indirettamente, in particolare attraverso Giordano Bruno e Johannes Reuchlin, recepì alcune sue problematiche specifiche, *in primis* quella di Dio come *coincidentia oppositorum*» (p. 25).

Il Cristo, definito da San Paolo θεοῦ δύναμιν (*1Cor* 1, 24), è ri-definito da Cusano *virtus actu infinita*, *infinita potentia*, *omnipotens*: potere e divinità in Dio sono congiunti nel concetto di *possest* che dice la connessione di potenza e atto, via apofatica e catafatica insieme. Siamo al cospetto di una *theologia eminentiae* che permette al pensante di partire da ciò che si può nominare per ascendere all'innominabile. Al di là della ὁμοίωσις θεῷ e della *coincidentia oppositorum sive contradictoriorum*, in Schelling vi è anche il momento dello *stupor* scaturente dall'incontro personale con Dio, ovvero lo stupore o vertigine «di fronte al nudo *quod* della coincidenza dei contraddittori» (p. 49). L'imprepensabile, *das Unvordenkliche*, anipotetico inizio, si può parzialmente conoscere solo riconoscendo di non sapere, «stupendo di fronte al suo mero "che", *actu* meramente essente» (p. 24). L'uomo colto da stupore scorge una 'barriera' di fronte a sé, nonostante la sua apertura alla rivelazione. Egli rimane, epperò, sempre posto dinanzi a *quell'* invalicabile che tuttavia lo reclama.

Tomatis precisa che dopo il 1460

> Cusano riconoscerà nettamente lo scarto fra ascesi mistica e rivelazione divina. Quell'abissale fessura inscalabile e insaltabile fra stupore perché il muro della coincidenza dei contraddittori non risulta ancora il Paradiso e meraviglia per la grazia teofanica dell'energia divina, quando donata direttamente dal Dio libero. […] Lo stesso libero automanifestarsi e ri-velarsi di Dio è colto da Schelling nell'interpretare il passo biblico. È il Dio vivente, il Dio in divenire, nascosto e rivelantesi, imperscrutabile eppure personale (p. 49).

Tale passaggio risulta fondamentale se si considera che il concetto di *stupore della ragione* – mai adoperato direttamente dal filosofo tedesco, ricordiamo, ma dedotto dalle sue pagine – permette a Schelling il passaggio dalla filosofia negativa alla filosofia positiva, e che Pareyson lo assume negli ultimi anni della sua attività filosofica per indicare che la seconda è, a tutti gli effetti, una filosofia della rivelazione che «aprendosi alla fede come fonte veritativa storica, cioè accesso libero e interpretativo a eventi sovrannaturali eppure liberamente fattisi storia, arricchisce la ragione di contenuti e verità che la ragione stessa, nella propria autonomia, non avrebbe potuto raggiungere da sola» (p. 15).

Il passaggio dalla filosofia negativa alla filosofia positiva sancisce il coabitare di filosofia e rivelazione cristiana:

> con l'emergere della distinzione dei due volti della filosofia o dell'affiancarsi di due differenti filosofie, unite nell'unica ricerca del sapere, filosofia negativa e filosofia, parallelamente all'esaurirsi della formulazione di un empirismo filosofico regressivo e progressivo. Schelling non fa che approfondire il rapporto fra filosofia e rivelazione, cioè fra la imprescindibile trascendentalità della filosofia e la rivelazione cristiana come

realtà storica superiore a ogni sensibile empirismo, ulteriore a ogni pur fattuale e reale rappresentazione mitologica, essendo le stesse raffigurazioni di dèi ancora presenti nella coscienza umana stessa, seppur come necessariamente, naturalmente subite (p. 83).

In merito a ciò, non si può non udire un forte riverbero agostiniano. Per comprendere in cosa consiste la dimensione della rivelazione che Tomatis pone in primo piano nello studio del filosofo tedesco, e quale può essere il suo rapporto con il pensiero critico, la filosofia, non si può prescindere dal pensiero e dalla vicenda esistenziale di Agostino. Pertanto, poniamo la domanda: cosa vuol dire *credere* e, nello specifico, credere nel modo della rivelazione cristiana, ovvero quel credere che indica l'evento Καὶ ὁ λόγος σὰρξ ἐγένετο?[2] Evento che attraversa, scuote e anima la filosofia di Schelling e di Pareyson.

Luigi Alici, nella sua introduzione a *La città di Dio,* afferma che permane in tutte le opere di Agostino, e quindi 'nonostante' la conversione al cristianesimo, un «atteggiamento di ricerca; l'incontro con la verità cristiana non coincide con una smobilitazione o un ridimensionamento del suo impegno intellettuale, che, al contrario, viene ad essere ricostituito dentro una prospettiva più ampia e più completa»[3]. Oltre alla considerazione della vicenda esistentiva di Agostino, la rivelazione cristiana, in sé, è e rimane un punto di congiuntura tra evento storico e comprensione del dramma umano in molte delle sue tragiche sfaccettature. Ecco perché, «a partire dalle opere giovanili fino a quelle più note della maturità, Agostino ricorda continuamente il carattere vivo e dinamico della fede; questa infatti indica una fiducia originaria, un'apertura di credito che precede e fonda ogni esplicitazione razionale»[4].

La prospettiva agostiniana appena esposta, sussumibile nelle celebri formule *credo ut intelligam* e *intelligo ut credam*, è a fondamento di tutto il sapere che si è sviluppato in Occidente. Per tale ragione, la rivoluzione culturale di matrice teologico-filosofica poggiante su questo evento storico indica un nuovo significato di *Dio* che spinse pensatori di epoche successive come Kierkegaard ad affermare che «il concetto centrale nel cristianesimo è la "pienezza dei tempi" (*Gal* 4, 4), nella quale tutto si rinnova; ma la pienezza dei tempi è il momento inteso come l'eternità, eppure questa eternità è tanto il futuro quanto il passato»[5]. Va precisato che, per quanto concerne il pensiero di Schelling, il riferimento ad Agostino è presente più volte nell'apparato delle note della monografia introduttiva di Tomatis *Kenosis del logos. Ragione e rivelazione nell'ultimo Schelling*[6], in particolare quando l'autore affronta gli argomenti del *male*, della *Trinità* e della *libertà* in Dio.

[2] *Gv* 1, 14.
[3] L. Alici in Agostino, *La città di Dio*, trad. di L. Alici, Bompiani, Milano 2001, p. 10.
[4] *Ibid*.
[5] S. Kierkegaard, *Il concetto dell'angoscia* (*Begrebet Angest*, 1844), in *Le grandi opere filosofiche e teologiche*, trad. di C. Fabro, Bompiani, Milano 2013, Parte II – *L'intermezzo filosofico*, p. 483.
[6] Cfr. F. Tomatis, *Kenosis del logos. Ragione e rivelazione nell'ultimo Schelling,* in Schelling, *Sui principi sommi. Filosofia della rivelazione* (*Über die Höchsten principien philosophie der offenbarung* 1841/42), trad. di F. Tomatis, Bompiani, Milano2016.

Ora, alla luce di quanto detto sin qui, e dunque in forza del valore della rivelazione cristiana per la filosofia degli autori di cui parliamo, è opportuno porre una domanda fondamentale: chi è il vero filosofo? Seguendo il sentiero teoretico schellinghiano,

> il vero filosofo dev'essere iniziato anche a questa sorpresa paradossale, al rivolgimento d'ogni credenza e previsione, anche allo scoprire l'opposto di quanto creduto e ricercato con tanto studio, immane fatica. Come un combattente, non deve fuggire innanzi al nemico, bensì proprio nel contrasto, nella lotta contro quanto di più oscuro e ciecamente distruggente ogni altra potenzialità, cercare la via della vittoria (pp. 88-89).

Ma è proprio qui, in questa linea di confine, oltrepassate le Colonne d'Ercole della ragione, che piomba l'angoscia caratterizzante la ricerca filosofica volta a ciò che oltrepassa l'essere indagando, in tal modo, la propria possibilità fondativa, originaria. Ecco il paradosso ove la filosofia confligge con l'essere stesso, «su cui non ha potere alcuno, non avendone nemmeno la potenza dell'essere, il potente-essere, che viene invece inghiottito inesorabilmente dall'essere di cui dovrebbe essere inesauribile fonte e potenzialità» (p. 88).

Angst autentica. Tremenda nella sua radicalità. Vorace nel suo continuo inghiottire il filosofo – "immediato venir fagocitato dall'essere". Eppure, l'angoscia è l'*orizzonte* esistenziale a cui il filosofo si volge consapevolmente dopo ogni rivolgimento, alla fine di ogni contrasto con l'essere, al termine dell'ultimo sconvolgimento, per riconoscere in esso la «realtà inaggirabile ma non intransitabile e invincibile del trapassare immediato della potenza nel cieco essere, del precipitare della possibilità di esistere da una punta inconcepibile, all'istante, nella nascita di un'immemore molteplicità» (p. 89). L'apertura al *paradossale* comporta *Schreck*, spavento, terrore; il terrore che caratterizza la scelta di non fuggire la contrapposizione, ma di abbracciarla teoreticamente ed esistentivamente nelle pieghe della storia in cui essa dispiega il dramma dell'esser gettati di fronte al nulla.

È nell'ultimo Schelling che lo *stupore della ragione* permette l'aprirsi della filosofia alla comprensione della Trinità cristiana (*Mysterium Trinitatis*) e all'ermeneutica che vuole rendere ragione del suo valore storico-ontologico. Si tratta di «una radicale critica dell'onto-teo-logia che conduce il movimento del pensiero schellinghiano, volto ad accogliere al fondo spoglio della soggettività l'assoluto, che si mostra libero dell'essere e da ogni sua entificazione» (p. 131). Non vi è una denotazione ontologica o teologica che possa cogliere appieno questa *in-differenza* di essere e non essere. Appare chiaro, allora, il motivo per cui il passo biblico preso in esame da Schelling, come Tomatis giustamente sottolinea, non può che essere tratto dal prologo del vangelo giovanneo, dove la parola di Dio, nel senso soggettivo del genitivo, testimonia il darsi insieme di necessità e possibilità. Difatti, il celebre inno al λόγος del *Vangelo di Giovanni* è «pro-logo teologico nel senso logico di teo-logico, dicendoci quanto a priori è necessario e possibile dire su ὁ λόγος pre-concetto, *prius* della Divinità, preliminare rispetto

alla sua stessa divinità, quindi ὁ λόγος prima e a prescindere ancora» (p. 159). L'*actus purissimus*, inizio imprepensabile o ἀρχή intemporalmente nell'eternità, si è fatto carne, è entrato nella storia e «nello stesso imprepensabile *prius* logico era ὁ λόγος, ed eternamente, anche quando si penserà di attribuirgli una divinità o percorrerà *usque ad mortem* le umili strade della nostra storicità» (*Ibid.*).

Il dramma di questa condizione del λόγος che precede la stessa incarnazione è enunciato anche nella *Lettera di Paolo ai Filippesi.* Ivi leggiamo chiaramente che

> "ὃς ἐν μορφῇ θεοῦ ὑπάρχων" (*Fil* 2, 6a), "trovandosi in forma di Dio", indica il suo stato transitorio: la forma di Dio, nella quale non è quindi propriamente Dio. Né uomo né Dio, il λόγος, preesistente si trova in forma di Dio, ha cioè la stessa signoria propriamente divina sull'essere. […] La signoria sull'essere resta la forma della sua divinità, ma l'essere è extradivino, quindi il λόγος ἐν μορφῇ θεοῦ è soltanto extradivinamente divino, non divinamente divino come seconda Persona della compiuta Trinità (pp. 160-161).

Né Dio né uomo, chi è allora questo λόγος che precede la stessa κένωσις? Egli è «il Signore dell'essere extradivino, che è tale solo alla fine del processo mitologico aperto dalla caduta dell'uomo» (p. 161). La caduta dell'uomo, quindi, reclama quel λόγος preesistente che nell'incarnazione rivela la propria interiore e vera volontà sempre in costante comunione con il Padre. In quanto libero di scegliere anche la rinuncia nei confronti della stessa signoria, è il Figlio, pertanto, a essere il signore dell'essere extradivino. La κένωσις stessa non conduce al ristabilimento immediato della comunione trinitaria e creaturale; essa è piena assunzione da parte dell'essere extradivino della propria umanità, dello stato mortale. Apertura, questa, che dice l'immediato farsi uomo del dio cristiano; quello *svuotamento* che definisce e realizza storicamente l'*umanità di Dio*, come ebbe a dire Karl Barth[7].

A fondamento di questa esegesi teologico-filosofica vi è una concezione del monoteismo cristiano – «L'unicità del Dio unico monoteista» – che Schelling espone nelle lezioni consacrate al suddetto tema, *Der Monotheismus*, e, ancor prima, nella *Historisch-kritische Einleitung in die Philosophie der Mythologie*. In primo luogo, bisogna comprendere la «distinzione fra unicità di Dio assoluta e in quanto tale, ossia fra monoteismo astratto e negativo, sedicente assoluto ma di fatto relativo, e monoteismo realmente assoluto, reale e vero, trinitario» (p. 165). Certamente, il discorso che apre il capitolo nono assurge ad *argumentum* necessario per comprendere «l'unicità del dio monoteista» sostenuta da Schelling, in quanto tale *unicità* risulta imprescindibile affinché si dispieghi nella storia la molteplicità di significati e interpretazioni dello stesso monoteismo per le altre religioni.

[7] Cfr. K. Barth, *L'umanità di Dio. L'attualità del messaggio cristiano* (*Die Menschlichkeit Gottes*, 1956), trad. di S. Merlo e S. Rostagno, Claudiana Editrice, Torino 1997, pp. 89-115.

Quando ci si riferisce schellinghianamente al Dio cristiano è opportuno parlare del *Dio plurale*. La profonda teoresi del filosofo tedesco sembra, sin da subito, delegittimare ogni ermeneutica sbrigativa, e dunque superficiale, che nel monoteismo cristiano vede l'assoluto contrapposto alla pluralità di cui invece il politeismo sarebbe alfiere e geloso custode. Cosa significa, innanzitutto, *monoteismo*? Esso è un «"concetto storico-mondiale", che "appartiene all'umanità", "già presente e universalmente concesso" – benché Schelling ammetta che esso «domina certo la parte migliore e più educata, ma pur sempre solo ancora minore, del genere umano» (*ibid.*). L'affermarsi del monoteismo va interpretato come un processo storico nel senso di un procedere critico-filosofico che, progressivamente, rigetta ciò che si mostra errato, e dunque non più accettabile, a causa di una profonda crisi della coscienza umana.

Si rimane, qui, sempre nell'ambito della mitologia, ovvero entro i confini di «una religione immemorabile al pensiero, imprepensabile, indipendente tanto dalla ragione filosofica quanto dall'autorità della rivelazione» (p. 175). Dallo studio del primo monoteismo preistorico e del politeismo (come si legge in *Filosofia della mitologia*) emerge che

> il politeismo successivo, storico, autentico, cioè il vero politeismo, *Vielgötterei*, mostra più dèi, i quali possono regnare solo successivamente, ciascuno come sovrano nella coscienza; i dominanti le varie epoche mitologiche sono: Urano, Crono, Zeus, essendo la mitologia greca la coscienza suprema e l'essenza di tutte le mitologie. Il politeismo successivo include quindi nel suo insieme anche quello simultaneo (XI 122, 257). Se il politeismo successivo è stato un avvenimento reale nella storia dell'umanità, allora dev'esser esistito precedentemente anche un primo dio che si presentasse come un dio assolutamente unico, dominante incondizionatamente la Terra intera ed esclusivamente la coscienza di tutti gli uomini come su di essi il cielo, conferente la pace e quiete d'una signoria incontrastata, priva tuttavia di coscienza e libertà (XI 104, 136) (p. 177).

Dunque, solo quando il dio relativamente unico (come il dio assolutamente unico) non avrà più signoria incontrastata verrà la conoscenza del vero Dio inteso come «Spirito perfetto, assoluto, positivamente differenziato da quello non-vero» (p. 179).

Teologia assurge a scienza che in sé accoglie sapere filosofico e sapere storico (Schelling, *Lezioni sul metodo dello studio accademico – Sulla costruzione storica del cristianesimo*, 1802). Ma tale capacità della teologia di accogliere la sistematicità filosofica e la fattualità storica propria del teologare non è una determinazione appartenente anche alle altre religioni, agli altri discorsi intorno a Dio, bensì è quanto di più proprio vi è nel cristianesimo; è stato quest'ultimo a far sorgere nell'orizzonte del sapere umano l'interpretazione dell'universo come storia. E ciò avvenne «sia attraverso la concezione del peccato originale come distacco dal mondo naturale, sia più in generale intendendo come libertà l'essenza di Dio e anche quella dell'uomo» (p. 181). A testimoniare che la storicità sia un portato esclusivo del cristianesimo è il concetto di incarnazione, secondo quanto afferma

la stessa cristologia con la formula Λόγος – σὰρξ. Dio si fa uomo. Cristo «viene elevato in Dio attraverso il farsi storicamente sofferente di Dio» (*Ibid*.).

In questa storicità muta anche il modo in cui la vita umana concepisce se stessa. Difatti, prendendo le mosse dalla costruzione storica del cristianesimo, Pareyson afferma che la persona è *iniziativa*, *esigenz*a. Ne consegue che «l'esperienza morale della persona, al culmine delle sue possibilità […] si fa problema a se stessa» (p. 200). Sorge una delle domande fondamentali – *chi è l'uomo?* – e Tomatis, seguendo il sentiero battuto da Schelling e Pareyson, risponde:

> nella sua finitezza, l'uomo è in quanto tale esperienza di trascendenza: della legge morale, della natura, del passato, dell'inconscio, del male, della stessa propria libertà. Sebbene le esperienze di trascendenza, costitutive dell'uomo a partire dalla sua mera autorelazionalità, siano possibili sedi di esperienza religiosa (OL 97-99), cioè del rapporto fra il divino e l'uomo tale che Dio istituisca il rapporto senza esserne fagocitato, risultandone così sua costruzione o proiezione, relazione quindi libera fra irrelativo che pone la relazione e relativo, tuttavia non è necessaria l'esperienza religiosa effettiva, né soltanto ermeneuticamente interpretata e comunicata, perché l'uomo comprenda sé come persona, cioè come libera autorelazione che è tale solamente in quanto eterorelazionata, non solo nel senso di essere eticamente e socialmente orientata, ma innanzitutto poiché posta da qualcosa di trascendente ogni limite umano (p. 203).

Uno degli elementi che rende originale l'*ontologia della libertà* elaborata da Pareyson consiste nell'aver saputo cogliere il baratro, l'errore delle ideologie politiche otto-novecentesche che si palesa nell'incapacità di saper strutturare un pensiero che renda ragione di verità e libertà, trascendenza e incarnazione, espressione e rivelazione. Ovvero un pensiero che, nel suo volgersi alla trascendenza, riesca a penetrare, in modo profondo, gli aspetti più pregnanti dell'esperienza umana, e viceversa. *Trascendenza*, qui, è il modo più autentico per rimanere fedeli alla terra e per esplorare gli anfratti degli aspetti più problematici dell'esistenza.

Nell'*ontologia ermeneutica della libertà originaria* viene posto il problema della lotta contro il male e la possibile vittoria che può prendere forma nella storia solo a partire dalla libertà della persona umana fondata, a sua volta, in Dio «perché proprio da un'originaria libertà è giunta la nostra libertà, cosicché per i puri di cuore non c'è – schellinghianamente – che un dire: libertà cerca libertà» (pp. 205-206). *Bene* è vittoria sul male, non mera innocenza. Il centro della questione della libertà ruota attorno a una comprensione fondamentale: la *mia* libertà è possibile solo se viene compresa come libertà donata da *un'altra* libertà. La libertà donata dal Dio che ha conosciuto la sofferenza nella persona del Figlio, sofferenza rivelativa, espiativa, redentiva, «*copula mundi* che tiene insieme Dio e uomo e mondo in una immane tragedia cosmoteandrica» (p. 213).

Donde la nostra libertà?

> Dio è essere e bene anche nell'eterno rischio di questa implosione che è l'eone aperto dall'uomo come risposta tragica alla sua libertà originaria profusasi altruisticamente in liberalità e generosità per la propria libera creatura? Forse proprio

con questi pulsanti interrogativi in cuore, ancora maggiormente sapremo, lungo le nostre insonni vie, nei passi più umili o celestiali, rasentando l'abisso del nulla o il più inattaccabile e irriconoscibile male, decidere ogni istante l'essere e il bene, iniziando originalmente non solo la nostra vita, ma di più, nel balzo stesso e intemporale della libertà (p. 216).

La suddetta domanda posta da Tomatis, in conclusione della parte seconda dell'opera, *Male in Dio e libertà in Pareyson*, pone, a sua volta, la questione principale di questa discussione: a che una filosofia tragica ispirata al cristianesimo? Stando a Pareyson, la risposta a questa domanda afferma, innanzitutto, che solo una simile filosofia è in grado di rispondere alle questioni poste dall'ateismo e dal nihilismo contemporanei. Ecco perché *Ontologia della libertà* (1995) «consiste principalmente in una ermeneutica dell'esperienza religiosa avente il proprio nucleo rivelativo di verità in Gesù Cristo» (p. 228).

Se Gesù Cristo è la risposta al problema del male – Tomatis precisa che non si tratta di avere una risposta o di enunciare una risposta, ma di *essere* risposta –, allora è possibile un pensiero tragico che sia all'altezza di interpretare filosoficamente la verità che è incarnata nella vicenda di Gesù, nella sua esistenza, nella sua sofferenza, nella sua morte innocente e nella sua resurrezione. Dostoevskij, tanto caro a Pareyson, ha colto ciò riflettendo sul mistero della sofferenza nel mondo: solo la cristologia è il discorso ermeneutico capace di affrontare in termini filosoficamente rilevanti la questione del male. Gesù Cristo è la risposta al male perché lo attraversa subendolo ingiustamente. Da questa prospettiva emerge la ragione per cui solo una cristologia che si pone filosoficamente come "pensiero tragico" (o, per così dire, laica), «non per tradurla in ribaltamenti speculativi, ma per mostrare l'esserci interrogante e significante che Gesù Cristo crocifisso è, possa affrontare il problema del male nella sua disarmante realtà» (p. 231).

Sorge, così, un pensiero tragico al cui centro vi è la verità rivelata in Gesù Cristo e nella sua morte da innocente a fondamento di un *esistenzialismo ermeneutico veritativo* – interpretare significa trascendere, e dunque non vi è trascendimento senza verità – «pluralista e veritiero, dialogico e singolare perché interpretativamente aperto alla trascendenza della verità infinita» (p. 251).

Tenuto conto di ciò,

> non si dovrà allora infine dire, seppur sempre disdicentemente, che ricerca infinita, verità inesauribile, libertà originaria abissalmente imperscrutabile, eppure esistenzialmente esperibile in dotta ignoranza o trascendenza esperienziale, non sono che il nostro portato, la situazione europea, la tradizione nostra storico-esistenziale? Non sono che il presupposto socratico, ma poi anche cristiano, storicamente-ermeneuticamente presupponibile a tutto il libero aprirsi circolare del pensiero rivelativo? Presupposto abissale, nel quale si può ancora (perché senza il quale non si potrebbe) ricercare interpretativamente ogni rivelazione nuova della verità originariamente inesauribile e libera (p. 254).

Tale prospettiva non permette alcuna riduzione dell'esistenza a negatività, come invece accade per l'esistenzialismo di matrice kierkegaardiana. E Pareyson, difatti, contrappone alla negatività del finito, nella sua triplice matrice novecentesca

(Jaspers, Heidegger e Barth), l'affermazione filosofica della trascendenza in forza di un'ontologia cristico-tragica. Ovvero, una filosofia che nella sua matrice jaspersiana guarda anche al mito come "plesso originario e sorgivo": la filosofia, pertanto, «è chiarificazione filosofica interpretativa della situazione esistenziale» (p. 280). Ma cosa accade quando vi è eccellenza e profondità nell'*indagandum* al punto tale che si verifica una lancinante frattura tra speculazione ed esperienza vivente?

Ci si volge al μύθος per ri-scoprire, ancora una volta, il Dio che parla all'uomo. Quel particolare *parlare* che è *Escatologia*. Ecco l'affermazione radicale di Pareyson nella quarta delle *Lezioni di Napoli*: «tutti i maggiori problemi si trovano nell'escatologia» (p. 289). Nell'ἔσχατον abbiamo la rivelazione di Dio fondata nell'autonominazione divina, espressione di abissale libertà (p. 295) che tuona nella forma di una tragica domanda: perché l'amore del Cristo si è spinto sino al sacrificio sulla croce? Sino a patire l'abbandono nella sofferenza? Perché l'ἀγών di Dio contro Dio in Dio? «Sulla croce il Cristo ripete kierkegaardianamente, rinnova abissalmente, deve e vuole liberamente rifare da capo la teogonia, l'autooriginazione divina come autoposizione di Dio nello scegliersi come essere e bene, lottando e vincendo contro il nulla» (p. 297). Nell'abisso della fede l'escatologia indica la salvezza dell'uomo nella consapevolezza della possibile vittoria sul male: viene Dio, il non-ancora; viene il futuro non solo tumultuoso, ma scosso anche dai bagliori della speranza. L'orizzonte filosofico del pensiero di Pareyson è caratterizzato da un *primato del positivo* che, va da sé, non va confuso banalmente con il cieco ottimismo.

Se senza l'unicità della verità non si dà alcun prospettivismo interpretativo, bisogna riconoscere, allora, la rilevanza filosofica del cristianesimo. La condizionalità storica della filosofia definisce e reclama, da sempre, il carattere rivelativo della stessa. Tale rilevanza, emergente dallo studio di Schelling e Pareyson, supera la caratterizzazione negativa del finito di matrice kierkegaardiana per aprirsi al primato del positivo, e afferma l'attualità del cristianesimo

> come questione filosofica. Infatti è l'esperienza religiosa cristiana ad aver contemplato in sé, almeno in quanto possibilità, come proprio problema interno, l'anticristianesimo, l'ateismo, il nihilismo: problemi caratteristici dell'attuale crisi del pensiero e della civiltà. Un recupero o ritrovamento del Cristianesimo nell'attuale situazione storica, fra le rovine della cultura moderna, oppure il suo rigetto, è una questione che interessa tutti, credenti e non credenti, sollecita la coscienza critica contemporanea assediata da nulla e avvolta dal nihilismo (p. 315).

Il venerdì santo costituisce la possibilità dell'ateismo interna allo stesso cristianesimo; possibilità che ha la sua risposta in Cristo, colui che ha affrontato e superato il negativo, che ha vinto il male e la morte. Possibilità del nulla «contemplata da Dio in se stesso e messa a tacere per sempre proprio in quanto tenuta presente e serbata eternamente nelle sue più immemorabili profondità» (p. 315). Kierkegaard e Dostoevskij sono tra coloro che hanno tracciato questa strada. Come non dar ragione all'osservazione di Tomatis secondo cui il pensiero

di Pareyson è attualissimo «per la capacità della sua filosofia dell'interpretazione di affrontare in particolare la questione del dialogo fra differenti persone, prospettive, culture, religioni, fra laici e credenti al tempo stesso» (p. 316)? *Attualità* che risiede, certamente, anche nel perspicuo richiamo alla trascendenza che definisce i limiti dell'esistenza «oltre i quali la pur onniespansiva virtù umana non può procedere» (p. 317).

In conclusione, va ribadito che l'*ontologia della liberà originaria*, dal punto di vista di Pareyson, è pensabile *etsi Deus non daretur*, a prescindere dall'esperienza religiosa e dall'esperienza religiosa cristiana. Motivo per cui, il criterio veritativo di questo pensiero è e rimane filosofico, non religioso. Posto ciò, se l'ermeneutica veritativa non esclude, anzi assume, l'esperienza religiosa può nascere una filosofia che prende in seria considerazione i significati di tale esperienza, una cristologia tragica in grado di oltrepassare i confini della mera appartenenza a una religione, a una determinata fede. Una filosofia che, come la grandezza di Giuseppe declamata da Giacobbe, si estende oltre i confini[8] plasmanti un modo particolare dell'esistenza umana per giungere a colui che indaga non per convertirlo (non è questo il senso, anche perché la fede cristiana nasce esclusivamente da un *incontro/evento*), ma per ampliare lo sguardo filosofico che vuole elaborare risposte concrete alle domande fondamentali. Una filosofia che, in quanto animata dal principio speranza, afferma e mantiene il primato del positivo

> poiché solo la libertà originaria ha autooriginato se stessa *ab aeterno*, vincendo per sempre il nulla come unità irrevocabile di inizio e scelta, è possibile la libertà esistenziale, dalla originaria originata; libertà, benché derivata, anch'essa inizio e scelta, ma, in quanto posta, come paradossale coincidenza (non irrevocabile unità originaria) e quindi sempre ancora revocabile, sospendibile, per quanto responsabilmente realizzabile: nel rischio della libertà (p. 322).

Stefano Piazzese
✉ stefano.piazzese.s@gmail.com

Bibliografia

Alici, L. 2001, in Agostino, *La città di Dio*, trad. di L. Alici, Milano, Bompiani

Barth, K. 1956, *L'umanità di Dio. L'attualità del messaggio cristiano* (*Die Menschlichkeit Gottes*), trad. di S. Merlo e S. Rostagno, Torino 1997, Claudiana Editrice

Cacciari, M. 2022, *Paradiso e naufragio*, Torino, Einaudi

Kierkegaard, S. 1844, *Il concetto dell'angoscia* (*Begrebet Angest*), in *Le grandi opere filosofiche e teologiche*, trad. di C. Fabro, Milano 2013, Bompiani

[8] *Gen* 49, 22.

Tomatis, F. 2016, *Kenosis del logos. Ragione e rivelazione nell'ultimo Schelling,* in Schelling, *Sui principi sommi. Filosofia della rivelazione* (*Über die Höchsten principien philosophie der offenbarung* 1841/42), trad. di F. Tomatis, Milano Bompiani

Tomatis, F. 2022, *Il Dio vivente. Libertà, male, Trinità in Schelling e Pareyson,* Brescia, Morcelliana

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771614

Recensione

M. Pietropaoli, *La società dell'estinzione. Liberazione dell'essere umano da se stesso*

Orthotes 2022

Nicolò Galasso

La scomparsa del genere umano da questa terra è sempre stata collegata, nelle varie culture, con la fine del mondo, l'apocalisse che travolge ogni cosa per volontà di un dio. Nell'ultimo secolo l'idea dell'estinzione dell'umanità è diventata, da figura mitica, possibilità reale: piuttosto che la fine dei tempi decretata da una divinità lo sviluppo inarrestabile della società contemporanea e gli effetti sulla tenuta ambientale in termini di inquinamento e di esaurimento delle risorse, insieme con la spaventosa potenza distruttrice degli armamenti nucleari, hanno dato consistenza all'ipotesi di una fine non troppo lontana dell'*homo sapiens* su questa terra e forse della terra medesima. In *La società dell'estinzione. Liberazione dell'essere umano da se stesso* Matteo Pietropaoli ribalta provocatoriamente la narrazione dominante: l'estinzione non sarebbe un effetto indesiderato dello sviluppo delle nostre società, bensì l'ultimo stadio del processo di individualizzazione dell'uomo a cui non rimane che, una volta abbattuti tutti i limiti sociali, culturali e morali che ne reprimevano l'autonomia, distruggere se stesso e tutto ciò che lo circonda per essere finalmente e completamente libero. Per capire meglio il senso della tesi dell'Autore bisogna prendere le mosse dall'articolata analisi dell'individuo contemporaneo che viene fornita nel libro.

L'individuo odierno è orientato al consumo. Lo strumento che la società iper-moderna o post-moderna ha messo a punto per far raggiungere la felicità ai suoi membri consiste nel consumare. Il processo di individualizzazione – sostiene Pietropaoli – ossia di progressiva espressione e affermazione del sé è un processo *eudemonico*, che pone in secondo piano la cura del mondo naturale e di quello sociale di cui ogni individuo fa parte.

Una caratteristica paradossale della società attuale consiste nella presenza di un malessere generalizzato, nonostante la disponibilità di beni materiali e

le incredibili potenzialità tecniche che distinguono l'età contemporanea da qualsiasi altro periodo storico precedente. Una delle cause di questo disagio – argomenta Pietropaoli – consiste nell'eccessiva aspettativa generata in essi dalla società dei consumi. Da una parte una tale aspettativa rappresenta la molla dell'agire dell'individuo, il quale si industria per raggiungere i propri obiettivi di affermazione e riconoscimento. Dall'altra parte, trattandosi di obiettivi spesso irrealistici o almeno di ardua realizzazione, l'individuo si scontra con la propria imperfezione e con i limiti intrinseci all'esistenza umana. A differenza di quanto accadeva in altri contesti culturali, l'individuo odierno non riesce a dare senso al dolore e ad accettare la persistenza dell'infelicità, stante il vertiginoso progresso tecnologico.

Rifacendosi agli studi di Gilles Lipovetsky, Pietropaoli divide lo sviluppo della società dei consumi in tre fasi. I primi passi di essa si possono riscontrare già nella seconda metà dell'Ottocento, quando si verificò un impetuoso sviluppo tecnico e produttivo che determinò la nascita di una prima forma di mercato di massa. Tra gli anni '50 e '60 del Novecento – continua Pietropaoli sulla falsariga di Lipovetsky – si colloca il passaggio a una seconda fase, quella dei 'consumi di massa'. La società del dovere lascia il passo a quella del desiderio. Si assiste a una democratizzazione dell'acquisto che permette la progressiva integrazione delle masse nel sistema consumistico. Queste profonde trasformazioni socioeconomiche comportano un incisivo mutamento della cultura e dei valori di riferimento. Dagli anni '80 in poi si entra, infine, nella società iper-moderna segnata dall'iper-consumo, dove la realizzazione materiale ed emotiva dell'individuo si libera da ogni codificazione sociale, costituendo il fine ultimo di ogni acquisto.

I tratti che distinguono la società iper-moderna vengono esemplificati dall'Autore, sempre sulla scorta di Lipovetsky, in sette figure metaforiche. Innanzitutto, la *Penia*, ossia la mancanza, rappresenta un aspetto caratteristico della società contemporanea. Benché quella attuale si possa ragionevolmente indicare come l'era dell'abbondanza, persiste la mancanza, giacché la felicità dell'iper-consumatore «risiede non in quello che egli ha ma in quello che ricerca, a partire da quello che ha» (p. 43), fungendo la mancanza ancora da stimolo fondamentale dell'agire sociale. Tuttavia, mentre nelle società precedenti la mancanza aveva un carattere prettamente materiale, in quella iper-moderna è trasmigrata piuttosto nelle relazioni, nelle esperienze e nella comunicazione tra le persone. La ragione di questo sviluppo 'immateriale' di *Penia* può essere in buona parte ricondotto alla crescita esponenziale delle aspettative e all'articolazione di bisogni sempre più sofisticati, una volta che i problemi materiali sono stati risolti dall'incremento della produzione.

Un'altra figura affrontata in queste pagine dall'Autore è quella del dio greco *Dioniso*, il quale secondo alcuni studiosi rappresenterebbe in modo adeguato la liberazione degli istinti e delle pulsioni che ha segnato gli ultimi decenni. Tuttavia, Pietropaoli mostra i limiti di questa lettura. Al contrario della rottura del *principium individuationis* e della conseguente liberazione delle pulsioni

rappresentate dalla figura di *Dioniso*, infatti, «l'edonismo dell'iperconsumatore si mostra invece controllato, moderato e al fondo spaventato» (p. 49). Si tratta, quindi, di un individuo piuttosto fragile e, come *Narciso*, focalizzato su di sé e il proprio benessere corporeo e mentale, ossessionato da *Crono*, il passare inesorabile del tempo con le conseguenti ripercussioni fisiche e psichiche.

Un'ulteriore figura che metaforicamente indica alcuni tratti dell'iper-consumatore è quella di *Superman*. Mentre però per alcuni critici, *in primis* Alexander Lowen, l'individuo contemporaneo sarebbe totalmente invischiato in una cultura del superamento dei propri limiti e della perenne ottimizzazione di sé, Pietropaoli distingue, insieme a Lipovetsky, le 'mitologie sociali' dagli effettivi comportamenti dei singoli, oggigiorno molto più attenti al proprio benessere anche emotivo che non al mero rendimento lavorativo. Pertanto, più che *Superman* la figura che meglio esprime l'ambivalenza dell'individuo contemporaneo è *Giano Bifronte* ossia il tentativo di coniugare l'efficienza lavorativa e sociale con una certa dose di felicità personale ed equilibrio emotivo.

Il processo di individualizzazione e liberazione, arrivato quasi al culmine nella società contemporanea, lascia, quindi, l'individuo isolato, privo di codici di comportamento e valori condivisi, responsabile del senso della propria vita ed estremamente impaurito da tutto ciò che potrebbe diminuire la quantità di esistenza o la quantità di piacere (non solamente fisico, ma anche emotivo e relazionale) in cui, in ultima analisi, egli ritiene che consista la felicità. Inoltre, il superamento della società della vergogna e della colpa, dovuto alla percezione positiva dell'ostentazione della ricchezza e dei consumi, se da un lato libera la società dall'invidia in favore dell'emulazione dei 'vincenti', dall'altro sposta l'inquietudine, dall'esterno del soggetto, al suo interno, ponendo in essere meccanismi di autovalutazione (*Nemesi*). L'individuo iper-moderno risulta, pertanto, fragile e a rischio sempre di crollare, in virtù del fatto che, per essere autenticamente se stesso come gli viene richiesto dalla società, non può che fondare su di sé il senso del mondo, rimanendo in uno stato di perenne precarietà.

L'analisi portata avanti dall'Autore si sposta, nel secondo capitolo, dalla economia del consumo alla politica della soggettività. La società contemporanea viene definita da Zygmunt Bauman 'società degli individui' di contro a quella di 'massa' che l'ha preceduta. Tuttavia, questi individui sono – è questo il paradosso sottolineato da Pietropaoli – in cerca di identità. In termini politici si assiste, pertanto, dagli anni '80 in poi, a un abbandono dell'impegno per la collettività e a una crescente diffidenza nei confronti dell''altro'. Nello stesso torno di anni nasce quella che Christopher Lasch ha definito efficacemente la 'cultura del narcisismo', con la riduzione della dimensione pubblica a rappresentazione del sé e la progressiva abolizione di codici e vincoli anche famigliari per dare a ogni individuo l'opportunità di perseguire la *propria* felicità.

Con questi cambiamenti culturali e antropologici di vasta portata, anche le élite al governo subiscono mutamenti significativi. Riprendendo il Lasch de *La rivolta delle élite*, Pietropaoli mostra come la classe dirigente dei paesi più sviluppati persegua obiettivi non sempre collimanti con l'ideale democratico. In

primo luogo, tra i membri della classe dirigente è presente la segreta convinzione che i problemi più profondi (come la crisi climatica) siano in realtà insolubili. In secondo luogo, le élite sono sempre più lontane dal cittadino medio in termini di reddito, educazione, conoscenze e opportunità. Ora, è evidente che i membri della classe dirigente, in ogni epoca storica, abbiano avuto stili di vita e redditi diversi da quelli della maggioranza della popolazione. Tuttavia, mentre in altre fasi storiche le élite, giacché fortemente radicate in un territorio, si sentivano responsabili sia del luogo in cui vivevano sia della società di cui erano in un certo senso garanti (*noblesse oblige*), nella contemporanea società globalizzata queste non sono più espressione di un luogo preciso, «divenendo quella tribù transnazionale di esuli volontari» (p. 82). Non appartenendo più a un territorio e finanche a uno Stato determinato, queste élite non si sentono responsabili verso alcuna comunità. Per disinteresse o convenienza si assumono i vantaggi del potere senza le corrispettive responsabilità. Gli effetti di questa 'rivolta' delle élite sul cittadino medio sono, come giustamente sostiene l'Autore, la rabbia sociale, la disillusione e la fuga dallo spazio pubblico con le evidenti ripercussioni politiche che questi atteggiamenti comportano.

Dopo avere analizzato le dinamiche politiche generate dalla iper-modernità, Pietropaoli focalizza la sua attenzione sulla formazione del soggetto e al rapporto con l'altro nel mondo contemporaneo. La società moderna si è contraddistinta per la contrapposizione nei confronti dell'estraneo. L'altro culturale, politico, sociale o pulsionale che fosse, veniva riconosciuto – e permetteva anche a chi lo riconosceva di strutturarsi in quanto tale – in forza del conflitto, il quale assumeva un senso centrale nella definizione del singolo e della comunità. Il negativo implicito nel conflitto ha ceduto il passo – qui Pietropaoli si rifà alle analisi di Byung-Chul Han – al positivo. Nella società attuale non vi è più *alterità* con cui confrontarsi dialetticamente, bensì rimangono solo *differenze*. Mentre l'alterità comporta conflitto e, per il suo tramite, un ampliamento dell'orizzonte di comprensione, le differenze si stagliano su un piano neutro, pacificato e che non produce resistenze, funzionale quindi alla nuova forma di capitalismo globale. Alain Ehrenberg collega la *deconflittualizzazione* dell'epoca odierna, che elimina il negativo dai meccanismi di riconoscimento sociale e di formazione del soggetto individuale, con la dinamica psicologica ed esistenziale propria dell'individuo iper-moderno, che senza più punti di riferimento e narrazioni condivise si ritrova a doversi assumere la completa responsabilità per le proprie azioni e le proprie scelte, dando luogo a un'identità fragile e sempre in bilico tra depressione e dipendenza.

Nell'ultimo capitolo del libro, l'individuo della iper- o post-modernità viene qualificato da Pietropaoli come 'surrogato' di contro a quello 'mitico'. L'utilizzo sempre più massiccio e onnipervasivo dei mezzi tecnologici rappresenta un aspetto tipico della quotidianità contemporanea. Il mondo materiale, composto sia da cose che da persone, viene sempre più allontanato attraverso il *medium* tecnologico producendo un cambiamento antropologico di notevole portata. Come spiega accuratamente l'Autore, l'individuo iper-moderno o surrogato si

distingue sia da quello mitico sia da quello moderno. Mentre quest'ultimo aveva dominato gli istinti e le pulsioni dell'uomo primitivo sublimandole in vista della convivenza sociale e dello sviluppo culturale, l'individuo odierno dà invece sfogo alle sue più recondite perversioni in forma mediata e controllata. Come acutamente nota Pietropaoli, egli «è un *portatore sano di inciviltà*, invece di un febbricitante barbaro (il selvaggio) o di un nevrotico civilizzatore (il borghese)» (pp. 132-133).

Il mondo dell'individuo surrogato non costituisce più un'unità di senso, forte e condivisa come invece era per l'individuo mitico e, sebbene in modo diverso ed estremamente 'culturalizzato', per il borghese moderno. Tale diversità si riscontra, secondo Pietropaoli, nel modo di affrontare la morte che caratterizza i due prototipi di individui. Mentre per quello mitico la morte è inclusa nel senso del mondo, costituendone una parte ineludibile e giustificata, la sua controparte surrogata comprende la morte come ciò che non può essere mediato e, dunque, ciò da cui si deve con ogni mezzo fuggire.

Pertanto, l'utilizzo massivo dei mezzi tecnologici non rappresenta, di per sé considerato, il discrimine tra l'uomo mitico e l'uomo surrogato. Il processo di individualizzazione e di liberazione, che trova nel tipo di individuo contemporaneo il suo apogeo, comporta una progressiva perdita di senso del mondo che, benché meglio conosciuto e dominato rispetto al passato, si svuota di significato e di mistero. L'ossessiva mediazione a cui l'individuo surrogato sottopone la vita serve – argomenta Pietropaoli – a distrarlo da questa insensatezza che avrebbe fatto inorridire il suo antenato. L'individuo che abita le attuali società sviluppate e iper-tecnologiche non è più guidato dal mito né da principi indiscutibili, ma non è neanche in grado di dare senso alla vita se non appellandosi alla ricerca, sempre più ossessiva e sempre più frustrata, di piccoli piaceri, sicuri e mediati. Questa fuga dalla politica, dal collettivo e dalla stessa materialità del mondo e degli altri esseri umani, può essere letta, secondo l'Autore, come un fastidio per la stessa società, dovuto al raggiungimento dell'apice del processo di individualizzazione che, in termini kierkegaardiani, pone «per la prima volta nella storia l'*individuo* dinanzi alla *specie* in termini d'importanza» (p. 142). Fastidio che, parafrasando il sottotitolo del libro, può forse indicare l'ultimo stadio del secolare processo di liberazione e individualizzazione, ossia la distruzione non solo dei vincoli sociali, bensì della civiltà stessa per arrivare, infine, alla liberazione dalla propria materialità e dall'imperativo di essere felici.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771616

Recensione

M. Mustè, S. Trinchese e G. Vacca (a cura di) *Bertrando Spaventa tra unificazione nazionale e filosofia europea*

Viella 2018

Dario Pizzi

Il volume raccoglie i contributi presentati al convegno su *Bertrando Spaventa tra coscienza nazionale e filosofia europea* che si è tenuto a Chieti il 23-24 febbraio 2017 in occasione del bicentenario della nascita del filosofo abruzzese. Non si tratta di una pubblicazione con intenti meramente celebrativi, ma del primo tassello di un progetto più ampio ed ambizioso, volto a dare all'opera di Spaventa quell'attenzione che merita, pienamente giustificata dalla dimensione europea del suo pensiero e dall'influenza che esso ha esercitato sulla filosofia. Il progetto comprende innanzi tutto l'edizione critica integrale dell'epistolario, che è stato recentemente pubblicato (B. Spaventa, *Epistolario (1847-1883)*, a cura di M. Diamanti, M. Mustè e M. Rascaglia, Roma, 2020) e copre l'intero periodo dal 1847 (anno della prima lettera al fratello Silvio) alla morte del filosofo (1883). Si tratta di un prezioso strumento di ricerca per gli studiosi, che consente di gettare la viva luce dello scambio epistolare non solo sul pensiero di Spaventa, ma altresì sul suo impegno politico, come testimonia il contributo di Maria Rascaglia, *Filosofia e politica nell'epistolario di Bertrando Spaventa*, che illumina le posizioni assunte da Spaventa in merito alla questione nazionale, dalle speranze iniziali alla delusione al cospetto dei risultati raggiunti. Obiettivo finale del progetto, che merita di essere realizzato, è quello, apertamente dichiarato dai curatori del volume, di «una edizione completa degli scritti di Bertrando Spaventa» (p. 9). Un'edizione critica è ormai un'esigenza ineludibile, come sottolinea Giuseppe Landolfi Petrone, *Le edizioni di Spaventa. Vicende e prospettive*, che propone «un'edizione completa basata su criteri strettamente cronologici» (p. 361), e ne stigmatizza il mancato rispetto nelle edizioni sinora realizzate: da quella più

nota di Cubeddu-Giannantoni (Firenze, Sansoni, 1972), in cui l'ordinamento cronologico risulta «privo di qualsiasi spiegazione filologica plausibile» (p. 372), in quanto basato «sull'uscita delle edizioni e non sulla pubblicazione da parte di Spaventa» (p. 371), a quella curata per Bompiani nel 2009 da Francesco Valagussa che, mantenendone inalterato l'assetto, ha finito con l'essere una riproduzione della precedente.

Una parte cospicua dei contributi proposti dal volume affronta i temi della riforma della dialettica hegeliana e della «circolazione», che rappresentano i due capisaldi del progetto culturale di Spaventa. Ma la sua biografia intellettuale offre una ricchezza (a volte inesplorata) ed una straordinaria attualità che il volume riesce in vari modi a far emergere, gettando ampie luci sui rapporti fra il pensatore abruzzese ed il suo tempo, sui suoi debiti nei confronti di altri autori, sulla peculiarità del suo hegelismo, sugli scritti 'giovanili' ed infine sulla sua presenza nella filosofia successiva.

A *L'influenza di Jacobi nel pensiero di Spaventa* è dedicato il lavoro di Federica Pitillo. L'autrice inquadra la ricezione di Jacobi nel pensiero di Spaventa nel contesto dello studio da questi condotto sui rapporti fra la filosofia classica tedesca e la filosofia italiana del Rinascimento di cui Bruno, precursore di Spinoza, rappresentava per Spaventa la vetta più elevata. Proprio a questo parallelo fra Bruno e Spinoza è strettamente legata la ricezione di Jacobi nella filosofia di Spaventa. Pitillo ricorda infatti che già Jacobi, «inaugurando un motivo filosofico destinato ad avere ampie ripercussioni nel corso di tutto l'Ottocento», aveva individuato negli *Spinoza Briefe* una corrispondenza fra i due pensatori volta a «mostrare come la filosofia del Nolano non fosse altro che una diversa declinazione di quel panteismo caratteristico, a suo modo di vedere, dello spinozismo» (p. 224). Pitillo mostra che questa idea di un parallelo fra Bruno e Spinoza derivò a Spaventa proprio da Jacobi: chiara testimonianza di ciò gli appunti spaventiani degli anni 1853-1854, pubblicati per la prima volta nel 2000 da Maria Rascaglia e Alessandro Savorelli con il titolo *Lettera sulla dottrina di Bruno* (Napoli, Bibliopolis, 2000), contenenti una molteplicità di riferimenti a Jacobi ed una esplicita ripresa di «alcune espressioni jacobiane» (p. 225). L'autrice non si limita ad evidenziare il debito di Spaventa nei confronti di Jacobi, ma sottolinea giustamente l'enorme portata delle conseguenze che il riconoscimento del parallelo fra Bruno e Spinoza ebbe sugli sviluppi del pensiero di Spaventa. Innanzi tutto, esso «influì sulla sua interpretazione di Spinoza, che si distingueva sensibilmente dalla lettura di Hegel»: laddove Hegel aveva interpretato il sistema di Spinoza in termini di «acosmismo», Spaventa invece riteneva che lo spinozismo «non fosse altro che un 'sistema della *causalità*', ovvero un sistema fondato sulla mediazione». In secondo luogo, «contribuì a rafforzare la tesi della circolazione del pensiero, introducendo anche gli autori italiani contemporanei» (p. 228): quel riconoscimento infatti, spingendo Spaventa verso un approfondimento della filosofia spinoziana, si rivelò fondamentale «anche in relazione alla successiva lettura della filosofia italiana contemporanea e, in particolare, della filosofia di Gioberti» (p. 229).

Proprio alla lettura spaventiana di Gioberti è dedicato il contributo di Marcello Musté, *Spaventa e Gioberti*, volto ad evidenziare la centralità della presenza di Gioberti nella filosofia di Spaventa. Una presenza inaggirabile, della cui importanza forse lo stesso Spaventa non fu sempre pienamente consapevole. L'autore sottolinea che la filosofia di Gioberti, malgrado la crisi in cui il giobertismo era caduto dopo il 1848, influenzò in modo significativo la nuova cultura napoletana di matrice idealistica, interagendo proficuamente con la crescente intelligenza dell'hegelismo: se da una parte la filosofia di Gioberti fu letta «con le categorie distintive della fenomenologia e della logica hegeliane», dall'altra molti temi hegeliani «vennero filtrati e a volte rielaborati sulla base di posizioni che provenivano dalla lettura delle edizioni giobertiane» (p. 125). Tale interazione emerge dal rapporto fra Spaventa e Gioberti, che Mustè intende far emergere in tutta la sua complessità, andando al di là delle stesse dichiarazioni di Spaventa. È il caso della nota 'confessione' del 1867, dove «per marcare in una forma più netta il mutamento intervenuto nel suo giudizio sui filosofi italiani contemporanei, Spaventa accentuò, e in definitiva ingrandì, la prima valutazione negativa di Gioberti» (p. 126). Lo studio di Musté mostra che già nel 1850, in particolare nei *Pensieri sull' insegnamento della filosofia*, «l'influenza delle categorie giobertiane appariva significativa», e che «il confronto con Gioberti risultò dominante» (p. 127) nella visione speculativa di Spaventa. Fino alla metà del decennio la filosofia di Gioberti restò sempre al centro della riflessione spaventiana, «intrecciandosi (per analogia o per distinzione) con il continuato studio delle opere hegeliane» (p. 130), il che portò Spaventa «a leggere i testi dell'abate torinese con le categorie hegeliane, ma anche, al contrario, a innestare motivi giobertiani nella genesi del suo peculiare idealismo (pp. 130-131).

L'autore sottolinea giustamente che sulla genesi dell'interpretazione spaventiana di Gioberti, e quindi sugli sviluppi della filosofia di Spaventa, influì un'altra linea di ricerca, destinata a rivelarsi feconda. Si tratta degli studi su Bruno, subito considerato l'instauratore della filosofia moderna e destinato poi a diventare uno dei principali protagonisti della teoria della «circolazione». Fu infatti l'indicazione di Bruno quale precursore di Spinoza ad innescare la ben nota lettura spaventiana della filosofia italiana ed a condizionare altresì la sua interpretazione dello spinozismo «non secondo la chiave ermeneutica dell'eleatismo e dell'immobilità della sostanza, ma, al contrario, come 'sistema della *causalità*'» (p. 132). Interpretazione, questa, che non solo allontanava Spaventa da Hegel, ma era destinata a fecondare in tutta la ricostruzione della filosofia italiana, e soprattutto nella successiva lettura dell'abate torinese, come l'autore spiega nel prosieguo del suo lavoro, in cui ampio spazio è dato alla complessa genesi della *Filosofia di Gioberti*.

Musté scandisce puntualmente i passaggi che portano Spaventa alla stesura finale del libro su Gioberti: innanzi tutto, la lunga lettera al fratello Silvio dell'11 ottobre 1857 che, se da una parte testimonia la delusione al cospetto della ripresa della lettura di Gioberti, dall'altra rappresenta «il concreto inizio» (p. 135) della monografia sul pensatore torinese; in secondo luogo, pochi mesi dopo (8 febbraio

1858), l'individuazione di «un nuovo e imprevedibile nesso, quello di Gioberti con Spinoza» (p. 136), fondato sull'interpretazione causalistica dello spinozismo elaborata diversi anni prima attraverso Bruno; infine, il 2 settembre 1858, la dichiarazione di aver sostanzialmente concluso il volume, ma soprattutto, e per la prima volta, di essersi «un po' riconciliato con quest'uomo» (B. Spaventa, *Epistolario (1847-1883)*, cit., p. 158). La composizione dell'opera, insomma, «aveva attraversato, in un breve volgere di tempo, fasi e momenti diversi», il che «spiega in larga parte la struttura disarmonica dell'intera opera, che in certo modo seguì l'evoluzione di pensieri dell'autore» (p. 139). Com'è noto, la *Filosofia di Gioberti* è un'opera incompiuta. Il lavoro di revisione del manoscritto proseguì dalla fine del 1858 alla pubblicazione nel 1863, ma «senza trovare un vero e proprio punto di approdo» (p. 139). La meditazione di Gioberti, tuttavia, era intanto proseguita ed aveva trovato la sua espressione più compiuta nelle prolusioni del biennio 1860-1861: «I nuovi pensieri che attraverso la lettura delle postume erano maturati nella sua riflessione» (p. 141) emersero negli scritti cui Spaventa consegnò la sua più matura versione della teoria della «circolazione», dove campeggiava una lettura di Gioberti ben diversa da quella consegnata al libro non ancora edito ma comunque già concluso in precedenza. Nelle due prolusioni, dunque, si concludeva il lungo, tormentato e fecondo dialogo con l'abate torinese, nel quale Spaventa individuava «l'ultimo grado al quale è salita la speculazione italiana», ma soprattutto un punto di coincidenza «coll'ultimo risultato della speculazione alemanna».

Legato al tema dei rapporti tra Spaventa e il suo tempo è anche il contributo di Alessandro Savorelli, *Il decennio di preparazione e l'*ideologia italiana, espressamente dedicato al duro confronto tra Spaventa e l'«ideologia italiana», come l'autore definisce lo spiritualismo, corrente di cui Terenzio Mamiani con la sua rivista, la «Filosofia delle scuole italiane» sorta nel 1870, intendeva essere il riorganizzatore. Savorelli fa emergere con chiarezza che «l'opposizione all'*ideologia italiana* è una costante che non si esaurisce nel decennio di preparazione» (p. 151), ma sopravvive alle svolte più radicali del pensiero di Spaventa. Egli evidenzia con cura i momenti più significativi della reazione spaventiana a questa filosofia: la celebre lettera sul *Paolottismo* (1868), la prefazione agli *Studii sull'etica di Hegel* (1869), la stroncatura (1872) della *Vita di Bruno* di Domenico Berti. E sottolinea giustamente che Spaventa, se negli anni Cinquanta poteva anche minimizzarne la forza, a distanza di anni è costretto a prendere atto della sua pericolosità ed a cercare di assestare ad essa un colpo definitivo. Le ragioni della polemica infatti non erano venute meno. Spaventa non solo «non riteneva del tutto superato lo scontro con quella cultura», ma «lo considerava momento fondante della sfida per un rinnovamento culturale che accompagnasse il percorso politico interrotto nel Quarantotto», ed era ben consapevole che essa era «negata e ostile – anche per condizionamenti e pregiudizi confessionali – alla comprensione della modernità», inadeguata insomma «a contribuire alle basi intellettuali di un nuovo Stato» (p. 152).

Savorelli intende far emergere «il motivo portante della critica all'*ideologia italiana*, e cioè che al fondo di quella costellazione teorica, metodologica, teoretica e politica c'è un'unità astratta, un pensiero dell'*identità* che ne occulta la strutturale *contraddittorietà*» (p. 155). Ai suoi avversari, in altri termini, Spaventa rinfaccia una sostanziale superficialità, che si rivela patentemente in un uso spregiudicato dei concetti; un atteggiamento antistorico, che non solo non permette di comprendere lo spirito di una data epoca, ossia di «valutare svolte, interruzioni, mutamenti di paradigmi e prospettive filosofiche», ma neanche «di comprendere le contraddizioni interne agli attori della storia» (p. 155). In particolare, Spaventa comprese «la necessità di contestare il mito dell'*unità* intellettuale italiana presente nelle varie versioni, illustri e non, del *primato*, che costruiva il quadro storico (o piuttosto, pseudostorico) di quell'identità astratta. [...]. E comprese perfettamente, e se ne assunse l'impegno, che tra i compiti dell'intellettualità di un nuovo Stato c'era la denuncia di quella leggenda, elaborata, diffusa e resa centrale dall'*ideologia italiana*». Spaventa si assunse già nel decennio di preparazione quell'impegno, teoretico e civile, che lo accompagnerà per tutta la vita, e il cui adempimento ancora oggi rimane, come scrive Savorelli, «l'acquisto più stabile e incontestabile consegnato da Spaventa alla sua generazione, insieme a una *traduzione* di Hegel in forma autonoma anche se discutibile e a lungo discussa, e alla costruzione di un determinato modello di storia della filosofia» (p. 158).

A quella «leggenda», com'è noto, Spaventa oppose la sua teoria della «circolazione», che ancora oggi, secondo una certa storiografia (Savorelli cita, fra gli altri, Paolo Casini, *L'antica sapienza italica. Cronistoria di un mito*, Bologna, 1998), proprio del «primato» riprodurrebbe la mentalità mitografica e la vocazione antistorica. A tal proposito, Savorelli giustamente afferma che ciò su cui Spaventa «insiste ripetutamente, dal 1851 al 1861, non è un primato alternativo: ma, viceversa, un *rovesciamento* del primato, che ne smentisce intenzionalmente le tecniche argomentative» (p. 159). L'idea che la tradizione filosofica italiana possa essere definita «attraverso *caratteri* identitari stabili, una presunta *indole* diacronica (spinta fino alle nebbie delle sterminate antichità pelasgiche e pitagoriche) e una posizione autosufficiente e privilegiata» (p. 159) rappresentano un bagaglio antistorico che non fa parte della prospettiva spaventiana. Rappresentano tutto ciò che Spaventa contrastò risolutamente e contro cui combatté dal Quarantotto sino alle «Lezioni» del 1861 e oltre: «La sua concezione cosmopolita della nazionalità suona smentita definitiva della gretta chiusura municipale 'italica'» (p. 159). Individuato con precisione il significato dell'operazione spaventiana, Savorelli può qualificare come insostenibile non solo l'idea che il «primato» e la «circolazione» fossero «un contributo ideologicamente contrapposto ma in qualche modo, al di là delle reciproche scomuniche, convergente nel progetto di costruzione di una storiografia filosofica adeguata ai tempi» (p. 160) ma anche qualunque accostamento tra la storiografia spaventiana e le opere prodotte nel quadro dello spiritualismo. Laddove queste appartengono al genere dossografico, «legato a moduli settecenteschi e costituzionalmente incapace – e indifferente ad

essa – di una lettura critica dei testi» (p. 160), gli studi spaventiani sul pensiero rinascimentale e moderno sono «letture analitiche, all'altezza della storiografia europea dell'epoca o addirittura l'anticipazione di nodi poi dibattuti in Europa» (p. 161), come ad esempio il confronto di Bruno con Cusano. Rappresentano ciò che ha consentito a Spaventa di rinnovare la storiografia filosofica italiana, «gettando le premesse di una scuola, da Fiorentino a Tocco, modernamente attrezzata, influente fino in pieno Novecento e ai cui lavori ricorre talora persino lo storico di oggi» (p. 162).

La teoria spaventiana della «circolazione» conteneva una rivalutazione del Rinascimento che non solo era diretta a scalzare la visione che di quel periodo era emersa nel contesto dello spiritualismo, ma si opponeva anche alla nota tesi hegeliana che individuava la modernità nella Riforma. Lo sottolinea Fernanda Gallo, *La costruzione dell'identità nazionale e le interpretazioni del Rinascimento nello hegelismo napoletano*, che ricorda il grande merito spaventiano di aver riscoperto il significato filosofico del Rinascimento e di aver individuato in Bruno «una concezione della libertà più profonda di quella proposta dalla Riforma» che, lungi dal limitarsi all'ambito religioso, si estendeva anche a quello filosofico e politico. «L'immanenza della natura divina nella natura umana, la dignità e sacralità della persona, l'autonomia della coscienza e della libertà morale» (p. 111) rappresentavano agli occhi di Spaventa la «via italiana alla modernità» consentendogli di combattere efficacemente «la dominante tesi storiografica della 'superiorità protestante'» (p. 121).

Emblema del Rinascimento, «instauratore» della filosofia moderna e protagonista della teoria della «circolazione», Giordano Bruno rappresenta una presenza decisiva nell'opera spaventiana, come mostra il lavoro di Biagio de Giovanni, *Il Giordano Bruno di Spaventa*, che affronta il rapporto tra la lettura spaventiana di Bruno e il principio della «mediazione». È proprio «la mediazione, come mediazione di contrarii, punto di convergenza dell'infinito e del finito, dell'eterno e del contingente, dell'essere e del non-essere» (p. 81), il tema che da una parte permise a Spaventa di valorizzare il pensiero di Bruno, dall'altra lo spinse ad impegnarsi sulla *Fenomenologia dello spirito*, fondata proprio «sulla necessità della mediazione» (p. 88). De Giovanni sottolinea che «le riflessioni critiche sulla *Fenomenologia* hegeliana sono scritte negli anni medesimi della riscoperta di Bruno, e la riscoperta di Bruno è modulata anch'essa sulla centralità della *mediazione*» (p. 90). Non è dunque peregrino ritenere il nolano una chiave essenziale per la lettura della *Fenomenologia*, forse il «protagonista decisivo» (p. 101) dell'apertura di Spaventa sul tema della mediazione. «Senza passare attraverso la sua lettura di Bruno – scrive infatti de Giovanni – alcuni passaggi essenziali della interpretazione spaventiana di Hegel resterebbero più oscuri, forse incompresi nella loro genesi» (p. 99). Il che ovviamente non esclude una reciproca interazione fra le due letture.

Il tema dell'hegelismo di Spaventa è affrontato nel volume da prospettive molto diverse. Alla rivendicazione della necessità di sottrarre Spaventa alla lettura gentiliana si ispirano i contributi di Vincenzo Vitiello, *I due 'cominciamenti'.*

Spaventa interprete di Hegel, e di Jonathan Salina, *Fenomenologia, logica e sistema in Spaventa*. Vitiello sottolinea che Gentile, «trascurando affatto la *Fenomenologia*, ha concentrato l'attenzione sulla parte più debole degli studi spaventiani su Hegel – quelli dedicati alle prime categorie della Logica» (p. 260) e mette in rilievo «l'importanza dell'interpretazione che Bertrando Spaventa ha dato del rapporto Fenomenologia-Logica» (p. 275). Salina, al fine di liberare il pensiero spaventiano da gravose ipoteche attualistiche, evidenzia l'aderenza dell'hegelismo di Spaventa all'architettura sistematica hegeliana. Laddove la riforma dell'hegelismo condotta da Gentile si basa sulla critica della ripartizione sistematica hegeliana (la critica, cioè della tripartizione dell'Idea in Logica, Natura e Spirito e della distinzione tra fenomenologia e logica), Spaventa opera la famosa 'revisione' del *Frammento inedito* «non tanto nei termini di una risistemazione globale del pensiero hegeliano, quanto in quelli di una 'critica di dettaglio'». In altri termini, il fatto che Spaventa abbia ritenuto inadeguata la spiegazione hegeliana delle prime categorie della logica, in ciò influenzando la riforma gentiliana, «non significa che egli abbia rigettato il modo hegeliano di porre la Logica in quanto tale: la prima esplicazione dell'Idea nel contesto di un'articolazione sistematica triadica» (p. 239). Neanche è possibile affermare che «Spaventa, mediante le lezioni napoletane, operi una revisione del sistema hegeliano fondata su un differente statuto da attribuirsi alla fenomenologia»; anzi va rilevato «come l'esigenza di una fenomenologia del tipo di quella hegeliana fosse asserita con vigore dal nostro autore ancora nel 1867» (p. 247). Ciò pone Spaventa ben al di qua della riforma gentiliana, che «abolirà non soltanto la valenza pre-sistematica della fenomenologia, ma la stessa differenza tra fenomenologia e logica» (p. 250). «Alla verità – così Salina riassume la posizione spaventiana – si accede in un modo determinato e dal quale non si può prescindere (la fenomenologia) – ma la verità resta unica, ed è quella data dalla filosofia in quanto tale – dall'impalcatura sistematica» (p. 251). Il dettato hegeliano insomma, malgrado le difficoltà da lui stesso evidenziate, poteva per Spaventa, «qualora venisse 'interpretato' in un certo modo, godere ancora di una piena validità teoretica» (p. 256).

Lo scritto di Francesco Valagussa, *Bertrando Spaventa. Raffigurare la logica di Hegel*, intende mettere in luce un aspetto della lettura spaventiana di Hegel generalmente poco curato dalla critica. L'autore, ricordando l'enorme compito che Spaventa si era assunto incaricandosi «di diffondere l'ostica dialettica hegeliana, di penetrare i segreti di questa nuova concezione del mondo», sottolinea che «l'operazione venne condotta con rigorosa accuratezza teoretica, ma insieme attingendo al patrimonio peculiare della tradizione italiana, ossia al 'pensiero figurale'. Spaventa – scrive Valagussa – analizza con acribia infinita gli snodi della logica hegeliana, ma il suo contributo inestimabile raggiunge l'apice quando arriva a sintetizzare le complesse strutture metafisiche tramite una metafora, una prosopopea, una figura emblematica, un paragone. Sono *immagini pensanti* quelle che spesso si trovano all'interno degli scritti di Spaventa: sono 'figure che pensano'» (p. 280). Valagussa evidenzia lo «sforzo costante di spiegare Hegel

associando la profondità della struttura logica alla figura emblematica» (281) e rintraccia nelle pagine spaventiane diversi esempi di quel «linguaggio figurato» con cui Spaventa traduce i concetti hegeliani e che rende un'opera come *Logica e metafisica* non un semplice compendio della *Scienza della logica* di Hegel, ma «una sorta di racconto illustrato, una *Scienza della logica* sviluppata per figure» (p. 284).

Il lavoro di Valagussa ha senz'altro il merito di affiancare all'immagine tradizionale di un pensatore dallo stile scarno e arido quella di un efficace disegnatore di movimenti concettuali. Un punto del suo lavoro, tuttavia, ci lascia più di una perplessità: l'affermazione, affatto ingiustificabile, che «la lettura spaventiana di Vico venne influenzata da C. Cantoni, G. B. Vico. *Studii critici e comparativi*, Torino, Civelli, 1867» (p. 279 *n*). Innanzi tutto, sull'interpretazione spaventiana di Vico, maturata lungo gli anni Cinquanta e culminata nelle «Lezioni» napoletane del 1861, gli *Studi* di Cantoni (pubblicati nel 1867) non potevano certo aver influito. In secondo luogo, pur volendo riferirci ad altri scritti, sarebbe davvero arduo individuare la benché minima traccia di una «influenza» sul Vico di Spaventa da parte di un pensatore come Cantoni che proveniva da ben altri orizzonti speculativi e fu animato nella sua lettura di Vico da presupposti ed interessi ben lontani da quelli spaventiani (al riguardo ci permettiamo di rinviare al nostro *Carlo Cantoni interprete di Vico*, «Giornale critico della filosofia italiana», settembre-dicembre 2012, pp. 615-644).

In *Spaventa, De Sanctis e la filosofia di Hegel in Italia* Marco Diamanti, attraverso un costante parallelismo fra i due più cospicui rappresentanti dell'hegelismo napoletano, ricostruisce la presenza di Hegel lungo il complesso *iter* speculativo di Spaventa. Partendo dagli esordi giovanili, in cui la filosofia di Hegel «rappresentava un 'culto', una 'religione ideale', un credo irrefrenabile capace di effondere un messaggio di speranza nel miglioramento e nel progresso» (p. 186), l'autore evidenzia come nel corso degli anni Spaventa, malgrado la «distanza progressivamente assunta dal pensiero hegeliano» (p. 202), non abbia cessato di vedere in Hegel un punto di riferimento costante di fronte ai pericoli del positivismo, del materialismo e di ogni filosofia che implicasse una rinuncia alla metafisica. La sua scelta di continuare a professarsi idealista (ma di quell'idealismo che, lungi dal negarli, includeva il positivismo, il naturalismo e il materialismo) dimostrava «lo sforzo di mantenersi dentro l'orbita del pensiero filosofico, e quindi metafisico», su cui soltanto «poteva costituirsi l'orizzonte della verità e quindi dello sguardo autentico sulla realtà» (p. 202).

Proprio alla metafisica, alla complessità del ruolo che essa gioca nel pensiero spaventiano e, in particolare, ai ripensamenti cui fu sottoposto negli anni da parte del pensatore abruzzese è dedicato il lavoro di Stefania Achella, *La metafisica di Bertrando Spaventa tra logica e psicologia*. La tesi hegeliana della coincidenza tra metafisica e logica rappresenta infatti una prospettiva che Spaventa, «alla luce delle questioni che emergeranno nel dibattito di fine Ottocento tra filosofia e scienze», supererà in direzione «di un riavvicinamento alla metafisica di Kant» (p. 291). L'autrice individua le ragioni del mutamento nel confronto sempre

più intenso, ed ineludibile («alla luce delle questioni poste prima da Herbart, poi da Spencer e dal darwinismo»), da parte di Spaventa «con le questioni della sensibilità, del dato, dell'esperienza». Fu dunque nel tentativo di rispondere alle «critiche mosse a Hegel da parte delle istanze realiste» che Spaventa passò «a una 'nuova' metafisica», intesa non più quale «luogo di incontro tra essere e pensiero», ma come «la legge della relazione tra soggetto ed esperienza» (p. 302). Il contributo di Achella evidenzia come il pensiero di Spaventa, ben lungi dal porsi «come ripetizione di quello hegeliano», vada inteso invece «come uno sforzo all'interno del dibattito europeo di fine Ottocento per uscire dai limiti di una ragione astratta» (p. 308). Uno sforzo tuttavia che cerca di tutelare le istanze del pensiero metafisico, «di salvare la conoscenza filosofica dagli attacchi delle scienze positive» (p. 308). Consapevole della necessità per la filosofia di «difendersi dalla sua riduzione a conoscenza ancillare delle scienze», Spaventa vede nella «metafisica critica» lo strumento più adatto «per opporsi a una riduzione a sua volta 'metafisica' dell'uomo alla dimensione empirica» (p. 309).

Alle interpretazioni del pensiero spaventiano è dedicato il lavoro di Ambrogio Garofano, *Spaventa e la dialettica hegeliana. Una riforma da riformare?*. L'autore prende in considerazione lo Spaventa di Gentile, gli studi elaborati nell'ambito del marxismo, infine i più recenti lavori di Vincenzo Vitiello. Ciascuna di queste interpretazioni ha puntato ad offrire un'immagine globale del pensatore abruzzese, «che implica, anzi tutto, la determinazione della precisa posizione di Spaventa rispetto al pensatore di Stoccarda» (p. 312). In tutti e tre i casi è emersa l'immagine di uno Spaventa proiettato oltre Hegel, sia pure in direzioni diverse: ora in direzione dell'attualismo, ora della filosofia della *praxis*, «ora in direzione di un essere infondato e sottratto ad ogni pensare» (p. 313). Spaventa interprete di Hegel è tema insidioso ma ineludibile. Garofano lo affronta cercando di far emergere quei nodi che giustificano il disaccordo fra le diverse interpretazioni da lui considerate, ciascuna delle quali «non si presenta come estrinseca rispetto alla speculazione del filosofo di Bomba, ma piuttosto di quest'ultima mette in evidenza uno solo dei termini che ne costituiscono il difficile e instabile intreccio» (p. 314). Un intreccio che emerge sin dagli *Studii sopra la filosofia di Hegel* (1850), dove «compaiono già quei termini – e il problema del loro difficile e instabile rapporto – che saranno al centro della più matura riflessione su Hegel, vale a dire l'essere, il pensare (o il conoscere) e la volontà (il tema della prassi)» (p. 318). È dunque nell'intrinseca complessità del testo spaventiano che trovano la loro giustificazione le tre diverse linee esegetiche, ciascuna delle quali ha potuto far leva «su uno dei tre termini che compongono la trama più essenziale della speculazione del filosofo di Bomba» finendo «col presentarsi come una riforma della riforma spaventiana della dialettica» (p. 326).

L'influenza di Spaventa sulla filosofia italiana è un tema che il volume non poteva certo eludere. Alla presenza di Spaventa nell'interpretazione di Marx elaborata dal giovane Croce è dedicato il lavoro di Claudio Tuozzolo, *Spaventa nel 'Marx possibile' di Croce e la riforma italiana di Hegel*. L'autore, ricordando la diffidenza più volte manifestata da Croce nei confronti della filosofia di Spaventa,

sottolinea tuttavia che alcuni aspetti del pensiero spaventiano sono penetrati in profondità anche nell'opera crociana, segnatamente nella sua metodologia storiografica. Di ciò è chiara testimonianza il noto saggio *Marxismo ed economia pura*, dove emerge «un vero, accorato, *elogio dello Spaventa storico del pensiero*», motivato dal fatto che il «modo» in cui Spaventa conduce l'indagine dei pensatori italiani è agli occhi di Croce «lo stesso con cui egli ha affrontato il pensiero di Marx» (p. 334). In altri termini, il metodo storiografico che Croce ritrova in Spaventa è il corretto modo di affrontare la storia della filosofia, in quanto basato sulla «distinzione tra pensiero reale ed elementi estranei» (B. Croce, *Marxismo ed economia pura*, in *Materialismo storico ed economia marxistica*, Bari, 1951, p. 167). Una distinzione, questa, che costituisce per Croce un principio irrinunciabile della storiografia: lo storico deve infatti cercare «non ciò che lo scrittore ha creduto di pensare, ma ciò che ha realmente pensato per la forza delle premesse da cui moveva» (ivi, p. 166), laddove trascurabili sono invece gli «elementi *estranei*», quelli cioè che non appartengono «*allo spirito profondo* degli scritti dell'autore, in quanto *non sono coerenti con le* "*premesse* da cui muoveva" lo scrittore indagato» (p. 335).

Tuozzolo individua nell'adesione crociana al metodo storiografico spaventiano l'acquisizione fondamentale che sta alla base della nota distinzione tra il vero e il falso, tra il vivo e il morto, che Croce applicherà nel celebre saggio del 1906 *Ciò che è vivo e ciò che è morto della filosofia di Hegel*. Opportunamente sottolinea le differenze fra il modo in cui Croce opera quella distinzione in Marx e in Hegel. Laddove nel caso di Hegel, colpevole di non aver scorto la distinzione tra la teoria degli opposti e la teoria dei distinti, Croce guarda il pensatore indagato dall'esterno, ovvero dal punto di vista di chi è in grado di chiarire tale distinzione, e quindi il falso (o il morto) «è individuato a partire da una prospettiva, assunta come vera, *estranea* a quella dell'autore indagato»; nel caso di Marx, invece, «l'interprete si pone *dal punto di vista del pensiero reale dello scrittore indagato*» (p. 340), e «il 'vero' che consente di giudicare 'falsa' una parte delle affermazioni dello scrittore indagato è, in questo caso, qualcosa che si trova nel 'fondo' del suo stesso pensiero, e che anzi costituisce *il* 'fondo' di tale pensiero». Ciò rende il Marx del giovane Croce «il frutto del *metodo interno hegeliano-spaventiano* consistente nel far emergere 'dalla cosa stessa' (dall'indagine del pensiero del filosofo indagato) lo sviluppo della verità» (p. 341).

A *Spaventa, Labriola e il problema della mediazione* è dedicato uno dei due lavori che *ex professo* affrontano il tema dei rapporti tra Spaventa e il marxismo italiano. Luca Basile vi sottolinea il debito di Antonio Labriola nei confronti di Spaventa e dell'hegelismo quale componente fondamentale della formazione del cassinate: al pensiero della mediazione, alla concezione dialettica della storia, infatti, essenziale per intendere il socialismo scientifico, Labriola era stato già avviato proprio da Spaventa, che ai suoi occhi ben rappresentava la «*coscienza della storicità*, dell'immanenza e del dialettismo innervante il reale» (p. 433). Una significativa influenza di Spaventa sul pensiero di Antonio Gramsci, in parte mediata da Labriola, emerge dal contributo di Giuseppe Vacca, *Spaventa*

nel marxismo italiano. L'autore individua innanzi tutto una evidente analogia fra il programma di ricostruzione della cultura italiana sviluppato da Spaventa dopo il 1848 e quello gramsciano di rifondazione del marxismo nei *Quaderni*: «il primo si proponeva di innestare la filosofia classica tedesca nella cultura della nuova Italia per portarla all'altezza del pensiero europeo più avanzato; il secondo si proponeva di sviluppare la filosofia di Marx per sollevare il socialismo europeo dalla crisi in cui era precipitato con la Grande guerra» (p. 415). Debitore altresì della visione spaventiana dei filosofi e della rivoluzione è il ruolo che Gramsci attribuiva alle élites intellettuali e la conseguente visione dello Stato italiano, che «non sarebbero concepibili prescindendo dalla lezione spaventiana» (p. 416). Queste ed altre testimonianze della presenza spaventiana nell'opera di Gramsci, puntualmente individuate da Vacca nella sua ricostruzione, consentono di concludere «che il pensiero di Spaventa sia stato una fonte essenziale – diretta o mediata da Labriola – nell'emancipazione di Gramsci dalle influenze crociane e gentiliane della giovinezza, e abbia contribuito significativamente a ispirarne la rielaborazione del materialismo storico in chiave di filosofia della praxis» (p. 418).

Polemista ostinato, impenitente e pertinace, Spaventa si confrontò duramente nella sua attività giornalistica con «La Civiltà cattolica». Il contributo di Luciano Malusa, *Bertrando Spaventa e padre Luigi Taparelli d'Azeglio*, intende ricordare il pensatore abruzzese attraverso l'attenzione nei confronti della figura del gesuita Luigi Taparelli d'Azeglio, con cui si scontrò in alcuni articoli scritti sulle riviste «Il Cimento» e «Il Piemonte» negli anni 1854-1856. Malusa ricostruisce la polemica, che ruotava in particolare attorno al tema dell'origine dell'autorità, nei termini di «un confronto a livello intellettuale tra la visione cattolica della politica e delle istituzioni nello Stato e la visione laico-liberale che si stava affermando in Piemonte» (p. 45), con l'esplicito e dichiarato intento di valorizzare al cospetto di Spaventa la posizione, poco nota, del suo avversario.

All'attività giornalistica di Spaventa si lega anche il lavoro di Nicola D'Antuono, *Discorrendo di De Sanctis e Bertrando Spaventa*, che intende far emergere, tra gli interessi scientifici spaventiani non attentamente valutati, il testo *Dell'importanza civile del teatro drammatico*, edito nel «Cimento» del 30 giugno 1855. L'obiettivo si scontra però con i dubbi sull'attribuzione del testo espressi da Italo Cubeddu nella sua *Bibliografia* (in B. Spaventa, *Opere*, a cura di G. Gentile, 3 voll., Firenze, 1972, III, p. 905). D'Antuono non ritiene pertinenti tali dubbi, ma non raccoglie prove sufficienti per legittimare la sua tesi. «Utili indizi – scrive – sono offerti dalle lettere del 18 novembre 1854, del 16 giugno 1856, del 5 gennaio 1857, del 3 ottobre 1859 e del 24 dicembre 1859, tutte al fratello Silvio» (p. 174 *n*). In realtà, le lettere indicate (segnaliamo qui che quella del 3 ottobre 1859 non risulta nell'*Epistolario*: dovrebbe trattarsi pertanto della missiva del 30 ottobre 1859) contengono generiche affermazioni di Spaventa sulla sua attività giornalistica a Torino, un invito al fratello a scrivere sulla «Nazione», brevi commenti sugli articoli di Silvio ed infine l'ammissione, di per sé non certo decisiva, di essere lettore dell'*appendice teatrale* del giornale

fiorentino (cfr. B. Spaventa, *Epistolario (1847-1883)*, cit., pp. 85, 122, 129, 185, 195-196).

Tuttavia l'aspetto più rilevante che emerge dall'attività giornalistica di Spaventa, segnatamente dagli scritti apparsi sul «Progresso», è senza dubbio la sua straordinaria modernità e attualità, che alcuni contributi opportunamente mettono in luce. È il caso di Luigi Mastrangelo, *Giuseppe Capograssi e i fratelli Spaventa*, che evidenzia come Spaventa riesca a «definire la natura di uno Stato, nel quale si svolga un processo di progressiva estensione di competenze sociali, anticipando scenari avanzati, ancora in corso di definizione» (p. 459). Sul terreno della rivendicazione di una 'democraticità' della concezione spaventiana dello Stato si muove anche Aldo Marroni, *L'idea di progresso: da Melchiorre Delfico a Spaventa*, per il quale gli scritti spaventiani del «Progresso» consentono di delineare il profilo di un autentico 'riformatore', animato dalla «volontà di trasformare la società e lo Stato» e dall'idea secondo cui l'umanità è avviata verso il riconoscimento che l'uguaglianza e la libertà, da intendersi in senso sostanziale e non meramente formale, «sono beni etici irrinunciabili» (p. 206).

Agli scritti spaventiani del «Progresso» è dedicato *ex professo* il contributo di Nicola Capone, *Spaventa e la nuova immagine del diritto. Gli scritti del «Progresso» del 1851*, che in essi individua «un'immagine del diritto del tutto inedita per l'epoca» ed una brillante anticipazione di «quella che sarà la funzione del diritto costituzionale propria dell'esperienza giuridica del secondo Novecento» (p. 33). Egli sottolinea, innanzi tutto, l'idea spaventiana della necessità delle rivoluzioni «nel momento in cui il diritto entra in conflitto con i bisogni reali» (p. 34), una necessità che tuttavia non implica alcun determinismo «perché le contraddizioni storiche, anche le più drammatiche, non danno automaticamente luogo ad una fase rivoluzionaria. Perché questo avvenga è necessario che si abbia chiara coscienza della fase storica e delle sue antinomie» (p. 34). Ed è qui che entra in gioco la funzione per Spaventa rivoluzionaria dei filosofi, i quali «fanno maturare nelle popolazioni in rivolta la *coscienza del diritto*, di un *novello diritto*. Essi sono chiamati a pensare una nuova forma di diritto che corrisponda all'idea della nuova epoca, alla nuova e concreta immagine dell'essere umano». Il profilo del filosofo che Spaventa delinea è dunque quello di un attento e vigile interprete dei travagli della sua epoca, che lavora «per favorire l'emersione di un ordine nuovo» (p. 35) il cui principio ispiratore sia l'uguaglianza, che però non può risolversi nella mera uguaglianza dinnanzi alla legge ma deve essere sostanziale. Perché l'uomo sia veramente libero, è necessario per Spaventa, come ricorda Capone, «che gli vengano assicurate tutte quelle condizioni che promuovono lo svolgimento della libertà» (p. 39). Parole, queste, dinnanzi alle quali non si può non notare un'«impressionante assonanza» con il secondo comma dell'articolo 3 della Costituzione italiana, che diviene ancora più forte quando Spaventa pone il lavoro «come condizione di vita», quale «manifestazione umanitaria». In tale prospettiva tutti coloro che possono e vogliono lavorare, dal contadino all'operaio, dallo scienziato al filosofo, hanno il diritto di chiedere che lo Stato non faccia mancare loro quegli strumenti (dall'aratro alle scuole, dalle biblioteche

alle accademie) che, favorendo di fatto il lavoro, contribuiscono alla piena realizzazione dell'individuo. In questo modo lo Stato, lungi dal rappresentare una forma di dominio, diventa «strumento di libertà e di liberazione» (pp. 39-40), una realtà attorno a cui tutti possano riconoscersi.

La straordinaria attualità di queste idee ci porta infine al contributo di Stefano Trinchese, *Unione o Unità d'Italia. Per una rilettura della storia patria*. L'autore, partendo dalla constatazione della mancanza nel nostro Paese, tradizionalmente avvezzo alle divisioni (a volte laceranti), di «un sentimento radicato in una storia comune» (p. 13), fornisce un'interessante rivisitazione della nostra vita nazionale, finalizzata alla ricerca di una storia condivisa che possa garantire una vera «unione» degli italiani. Egli la individua negli anni drammatici di quella che ancora oggi molti faticano a definire una guerra civile (1943-1945), nella caduta della dittatura e nella liberazione, e più ancora nella scelta repubblicana del 1946 e nella Costituzione del 1948. Forse è in questo periodo che «rinacque l'unione di tutti gli italiani, quell'unione democratica che gli italiani avevano cercato nel 1848 e non avevano vista realizzata nel 1860», e che è dunque possibile rintracciare «in qualche modo un filo rosso della storia condivisa dagli italiani»: «l'orditura difficile di una trama democratica, invocata dal Risorgimento» e «realizzata un secolo dopo dall'incontro degli antichi esclusi dall'Unità: i cattolici, i socialisti, i laici, i radicali, i repubblicani, i democratici» (p. 25). Di quella «trama democratica» proprio Spaventa, in ciò simbolo non solo di «unità» ma altresì di autentica «unione», contribuì certamente a tracciare le linee essenziali.

Dal volume emerge in definitiva il quadro di un autentico classico, di un pensatore ancora vivo, il cui profilo è lontano anni luce da quello, delineato da Armando Plebe, di un filosofo «tutto legato a certi provincialismi della cultura filosofica italiana» (A. Plebe, *Spaventa e Vera*, Torino, Edizioni di «Filosofia», 1954, p. VI). Altro che provincialismi! Dall'opera di Spaventa si staglia l'immagine di un pensatore pronto a confrontarsi autorevolmente con tutte le correnti della cultura europea contemporanea, disposto ad accettarne le sfide e ad accoglierne criticamente le istanze. Di un autentico protagonista in campo filosofico, storiografico e politico. Di un protagonista non solo del suo, ma anche del nostro tempo.

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771620

Recensione

A. Cerutti, G. Dettori (a cura di), *La rivoluzione in esilio. Scritti su Mario Tronti*

Quodlibet 2021

Andrea Prizia

La rivoluzione in esilio. Scritti su Mario Tronti, a cura di Giulia Dettori e Andrea Cerutti, è un volume collettaneo edito da Quodlibet che raccoglie contributi di studiose e studiosi, spesso giovani, che riflettono sul lascito del pensiero di Mario Tronti. Non è senza motivo che si sottolinea la giovane età di molti autori e autrici, come è anche importante notare l'ambito in cui si muovono, a volte fuori dell'accademia: negli interventi – così come nel pensiero di Tronti – le riflessioni politiche e teoriche sono affrontate e percepite come problemi attuali, rivolte alle questioni del tempo presente, affidate anche ad una forma stilistica lontana dallo stile scientifico – come nel caso di alcuni agenti scismatici del partito immaginario o l'intervento di Andrea Cerutti.

La parola rivoluzione che appare nel titolo è da considerare come il nome del pensiero più radicale cui Tronti ha dedicato i propri scritti e che oggi è del tutto lontano e forse inaccessibile. Uno degli scopi di questo volume è proprio quello di mostrare come oggi tale pensiero radicale che si intendeva rivoluzionario sia costretto a ripensarsi, nel doppio senso di ripensare agli elementi che lo strutturavano e a ripensarsi da capo, a rinnovarsi. Questa operazione, che è solo una delle molte possibili a partire dal volume, si muove essenzialmente in tre movimenti: la postura politica del soggetto militante; gli strumenti teorici mobilitati da Tronti e i suoi interpreti; la tradizione di pensiero e gli autori di riferimento da recuperare.

Per quanto riguarda il soggetto politico, il militante, ossia chi agisce il pensiero della rivoluzione, cosa si intende per «postura»? A tutti gli effetti, la posizione che esso assume rispetto all'ordinamento nemico, capitalista. Delle

suggestioni trontiane riguardanti il «dentro e contro», il «punto di vista», due interventi nel volume spiegano con chiarezza la centralità della questione. Daria Saburova dedica un paragrafo del suo contributo al «principio di parzialità del punto di vista operaio» (pag. 58) che costituisce la «soggettività antagonista» come fulcro della «soggettività operaia»: in effetti, l'antagonismo, il lato negativo del soggetto operaio militante, riguarda non solo il sistema capitalista, ma la stessa classe operaia, come colei che deve paradossalmente rifiutare se stessa in quanto parte dello stesso sistema. A questa vocazione di parte della classe operaia, Andrea Cerutti associa la capacità «di dare corpo allo spirito distruttivo della filosofia nietzscheana» (pag. 17), individuando in Nietzsche un'ispirazione fondamentale del pensiero di Tronti: «pensare la rivoluzione» significa assumere un atteggiamento «anarchico e istituzionale, negatore e organizzato» (pag. 23), entro il quale Tronti è dunque, «come Nietzsche, un pensatore del conflitto» (pag. 28). Tale conflitto, con Soburava, prende forma nella pratica dell'organizzazione e ha una determinazione del tutto politica, tale per cui il soggetto rivoluzionario, nietzscheano e antagonista, si muove non attraverso un percorso economicista di sviluppo del capitale, ma nello spazio autonomo, di lotta e imprevedibile del politico.

Il tema dell'organizzazione, in quanto momento segnatamente politico dell'antagonismo, ci conduce al secondo aspetto che emerge chiaramente nel volume, ossia il dibattito attorno alle nozioni trontiane più pregnanti e, principalmente, il concetto di autonomia del politico, verso il quale convergono in definitiva gli altri (punto di vista, lotta di classe, la teologia politica). L'elaborazione nel 1972 della teoria dell'autonomia del politico è tradizionalmente considerata una svolta – o almeno l'inizio di una nuova fase – del pensiero di Tronti, ma almeno due interventi raccolti nel libro tentano di smentire o mitigare questa interpretazione. Il «tradimento» che Tronti avrebbe compiuto con l'intervento del '72, è definito «presunto» da Andrew Anastasi e Matteo Mandarini (pag. 117), che considerano l'idea di autonomia del politico «né frattura epistemologica né conseguenza di premesse già viziate» (pag. 134), bensì un'intuizione trontiana presente già nella riflessione su Lenin, esemplarmente nell'interpretazione della NEP come tentativo «di anticipare il movimento del capitale verso il terreno politico» (pag. 129): già dagli scritti precedenti al '72 – e, da notare, parallelamente ad Althusser - Tronti avrebbe quindi sempre avuto l'intenzione teorica di uno «sviluppo del potere immediatamente *politico* della classe» (pag. 122). Passando per il più cauto intervento di Michele Filippini, che descrive un percorso più lineare della genesi del concetto (pag. 103), si arriva al saggio del gruppo Epimeteo, in cui si propone una lettura «tendenziosa» (pag. 89) di *Operai e capitale* a partire dalla questione del politico, rileggendo le varie nozioni trontiane come fossero già *in nuce* nell'opera più celebre (politico, linea di condotta, *kairòs,* punto di vista) e passando in rassegna i vari autori di riferimento e di ispirazione per Tronti.

Il saggio di Epimeteo suggerisce così la terza tendenza degli interventi raccolti, ossia il bilancio, in parte ancora da sviluppare, della tradizione di

pensiero che Tronti utilizza e mobilita. Oltre ai riferimenti classici della tradizione marxista (Lenin, Gramsci, ovviamente Marx), sono ricorrenti i nomi di Carl Schmitt, Walter Benjamin, Jacob Taubes, Max Weber, Thomas Hobbes. Certo questi nomi non ricorrono solo per un'anagrafe delle fonti, ma in un dialogo che tenta di rimettere in questione le interpretazioni che di esso Tronti ha suggerito negli anni, per confermarle o muovere critiche. In particolare, Marten Björk e Giulia Dettori puntano l'attenzione su una delle questioni più recenti aperte da Mario Tronti, ossia ala rinnovata attenzione alla teologia politica e alla tradizione conservatrice. L'«indivisibile costellazione» (Dettori, pag. 249) composta da Benjamin, Taubes e Schmitt, consente a Tronti di pensare, ancora una volta, «dentro e contro», anzi fuori, la pesante immanenza della politica contemporanea, in una politica ormai «al tramonto». Nella rivalutazione del ruolo del teologico nel politico si cerca la via di fuga da un sistema che sembra inscalfibile e che ha sconfitto, forse definitivamente, l'idea di rivoluzione. La «tensione apocalittica» di questi autori serve a pensare di «realizzare, qui e ora, le condizioni di una possibile rivoluzione» (pag. 269). Il politico va dunque «riteologizzato» (Björk, pag. 231) nella misura in cui «l'anima della politica sarà in grado di fronteggiare il corpo della storia» (pag. 247): in altre parole, di fronte a ciò che la storia oggi rappresenta, ovvero nient'altro che il fallimento di un *eschaton* ormai dissoltosi nella «catastrofe della nostra epoca iperindustrializzata» (*Ibidem*), è di nuovo lo spazio organizzato a autonomo del politico a individuare una trascendenza, un atteggiamento teologico che «sottometta il ritmo della macchina» (pag. 232) e crei una contraddizione tra politica e storia del capitale. Infine, su un'ulteriore relazione - anch'essa in parte contradditoria - tra politica e vita, scrive Elettra Stimilli in un intervento che vuole far dialogare Max Weber e Tronti in vista del comune tentativo di «definire l'autonomia dei mezzi politici rispetto al processo» di neutralizzazione della politica (pag. 175). A questa intenzione fa da *pendant* l'idea di una politica come 'vocazione' e anzi come vocazione di parte, antagonista. La strategia trontiana di operare la rivoluzione con uso politico della macchina statale si è dimostrata «strategicamente fallimentare», ma la questione centrale rimane invariata: lo strumento politico trova, in Tronti, senso ed efficacia fintanto che è organizzato e direzionato da una postura militante e di parte. Come lo stesso Tronti sostiene nel saggio-intervista che chiude il volume, «il punto di vista è la verità del rapporto sociale» (pag. 331) e, in questa prospettiva, «la lotta per la rivoluzione è il *Beruf*, la vocazione-professione» (pag. 344.), cui non può esimersi chi entra nello spazio autonomo del politico.

In conclusione, si può sottolineare ancora come il tentativo di pensare la rivoluzione comporti necessariamente tornare a interrogarsi tanto sull'idea di soggettività politica che sui mezzi del politico. Il suggerimento di Tronti verso la necessità di una «rivendicazione della differenza come nuova frontiera per la rivolta del soggetto» (Stimilli, pag. 186) diventa l'indicazione per un punto di rottura di un sistema che non è altro che un «capitalismo che ha integrato tutto il politico» marchiato dal «segno totalitario delle democrazie liberali». (Tronti, pag. 331). La rivoluzione in esilio alla fine del mondo non è forse un pensiero

così immediatamente traducibile in azione politica né facilmente operabile: deve, con Benjamin, rammemorare un passato in cui è stato decisivo e mostrarsi rinnovata. Questo volume, in questo senso, riprende le traiettorie del pensiero trontiano inteso come strumento per i problemi contemporanei, mappa le possibili contraddizioni, indicando in fondo anche la speranza di ritrovare quella invenzione penetrante del pensiero e della pratica politica rappresentata dalla nozione di rivoluzione, cercando di ridefinirla per il «lungo periodo» contemporaneo. Rappresenta, per così dire, un certo tipo di vita di questo pensiero, che cerca di nuovo di riposizionarsi nonostante l'avvenuta sconfitta: la vita in esilio che è, con Pavese, «uguale alla vita di prima, solo più sporca».

© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7771624

Recensione

E. Laclau, *Dibattiti e scontri: per un nuovo orizzonte della politica*, a cura di F. M. Cacciatore e S. Mazzolini

Mimesis 2020

Alessandro Volpi

La traduzione italiana di *Debates y Combates* di Ernesto Laclau (*Dibattiti e scontri: per un nuovo orizzonte della politica*, a cura di Fortunato M. Cacciatore e Samuele Mazzolini, Milano-Udine: Mimesis, 2020) riunisce una serie di saggi dell'autore che si trovavano, tanto nell'originale inglese come in alcune traduzioni nel nostro paese, sparse su riviste e altre raccolte. Ma la raccolta non è inedita, si tratta infatti della traduzione di un libro già esistente in questa forma, solo in lingua spagnola, pubblicato nel 2008 in Argentina.

Paese, data di pubblicazione e scelta dei saggi da parte dell'autore sono rilevanti per capire il senso della costruzione di questa opera, e la brevissima premessa all'edizione originale firmata dallo stesso Laclau ci rende conto del senso di questa raccolta. Laclau infatti ci dice che si tratta – come del resto si desume dal titolo – di scontri teorici diretti, o battaglie (*combates* è scontri ma può essere anche combattimenti, battaglie, lotte) con alcuni fra i principali teorici della sinistra contemporanea (prevalentemente *radical thinkers*). Gli 'avversari' (p.45) (così si riferisce agli autori trattati) con cui si confronta sono nello specifico: Slavoj Žižek, Giorgio Agamben, Alain Badiou, Toni Negri e Michael Hardt. L'obbiettivo dei saggi – ci dice Laclau – è «riprendere l'iniziativa politica» (p. 46) e fin da subito si percepisce, in un autore che lavora spesso ad un livello teorico che potrebbe essere percepito come astratto e lontano dai conflitti ideologici, la concretezza politica degli scontri. In questi lavori infatti – più che nei suoi classici, probabilmente – il teorico argentino mostra apertamente le scelte strategiche che conseguono direttamente dalle differenze teoriche, passando agilmente da questioni di ontologia politica, alle conseguenze nella pratica dei soggetti politici. In questo senso questa pubblicazione si colloca nell'orizzonte storico della «lotta per l'emancipazione» che «i nostri popoli latinoamericani»

(p. 46) stanno portando avanti in quegli anni (i saggi sono stati scritti negli otto anni che precedono la pubblicazione nel 2008, che coincidono con l'affermarsi di diversi governi populisti e socialisti nel continente). Il fatto allora che il libro sia in *castellano,* pubblicato originariamente in Argentina è già una presa di posizione, parla *desde el sur* – come del resto *La ragione populista*, ma in realtà sottotraccia tutta la sua produzione teorica – al mondo della sinistra euro-(nord) americana proponendo, da una posizione non esterna a questo dibattito (del resto Laclau è considerato un esponente del post-marxismo britannico) delle scelte teoriche che danno battaglia ad altre, per imporsi e guidare la riscossa delle forze emancipatrici.

Questo 'Laclau contro tutti' ci mostra, andando più nello specifico, la sua specifica posizione rispetto alle ipotesi politiche che nascono dalla fine del socialismo reale e dalla decomposizione – iniziata prima – del monopolio marxista come ideologia della trasformazione sociale nella modernità capitalistica (in conseguenza della crisi della fase fordista del capitalismo). Gli autori con cui si confronta Laclau sono tutti esponenti che vengono da un pensiero post- (moderno, strutturalista, marxista, operaista) interpretato secondo diverse direttrici teoriche. Tutti frammenti dell'esplosione di un mondo che cercano di ricomporre una teoria radicale della propria contemporaneità. Laclau mostra in questi saggi il suo essere interno a questo mondo post- e la sua specifica risposta per superarlo, per ricostruire una teoria per la prassi che possa finalmente – come stava avvenendo in America Latina – guidare una riscossa delle classi popolari orfane di un'ideologia politica mobilitante. Nel teorico argentino non c'è né la nostalgia di un ritorno all'ortodossia comunista (non fosse altro perché ha passato la sua vita a combattere tutte le incrostazioni residue di quella visione del mondo, come del resto faceva anche in questi saggi) né la decisione per l'abbandono di una prospettiva di costruzione di una nuova unità politica popolare, in favore del trionfo della differenza o per l'esodo e il rifiuto della prassi, ma c'è la convinta battaglia – contro tutti, nella sua prospettiva – per una lotta per l'egemonia. E all'interno del suo campo la battaglia per l'affermazione (insieme come comprensione e come prassi) di un'ontologia politica che è a nostro avviso quanto di più interessante e importante si presenti oggi al pensiero politico che vive la lunga crisi delle visioni del mondo moderne e che rende la proposta teorica laclauiana un punto di partenza decisivo.

Seguiamo quindi Laclau nei suoi scontri, cercando di delineare per ogni posizione che critica, la dimensione più strettamente teorica (e ontologica) politica e le conseguenze strategiche.

L'autore a cui è dedicato il primo saggio (*Perché costruire un "popolo" è il principale compito della politica radicale*), quello nei confronti del quale mostra più acredine – alla luce di un dibattito di più lungo corso che li aveva portati da un confronto amicale allo scontro aperto – è il filosofo sloveno Slavoj Žižek. L'intervento di Laclau si concentra sulla risposta ad un articolo (*Against the populist temptation*, «Critical Inquiry», 32 (3), 2006, pp. 551-574) in cui Žižek aveva critica il suo libro *La ragione populista* definendo il consenso populista

una forma di reificazione laddove l'idea, l'universale, nel suo incarnarsi in una persona, si sottometterebbe alla logica di questo particolare. Questo significa per Laclau che il filosofo sloveno accetta solo una mediazione impossibile che è quella della piena presenza dell'universalità, pensabile solo idealmente nella prospettiva di una totalità trasparente con un soggetto che ne sia pieno inveramento: il proletariato del marxismo, laddove «la vera coscienza [è] totalmente esente dalla mediazione simbolica». (p. 57)

Qui Laclau imputa a Žižek fondamentalmente di aver frainteso Lacan (il comune riferimento, vero oggetto della disputa teorica) e di averlo sovrainterpretato per farlo rientrare all'interno di una filosofia storico-sociale di matrice hegelo-marxiana, pensando così il Reale non come un'eterogeneità costitutiva ma come logica sottostante una totalità compiuta. Questo significa di conseguenza non riconoscere la negatività costitutiva che attraversa e determina (indeterminando) il politico. L'ontologia politica Žižekiana ricadrebbe in fin dei conti in un positivismo storico-dialettico, al contrario della logica della mancanza come interruzione della relazione fra struttura ed evento che assume l'ontologia politica laclauiana sulla base della teoria psicoanalitica lacaniana (e della linguistica). In questo senso la critica a Žižek alla fine si dipana in una serie di critiche alla sua ortodossia hegelo-marxista solo mascherata da un gergo lacaniano.

Le conseguenze politiche di questa ontologia sono la ricerca di un soggetto inesistente della rivoluzione (*in attesa dei Marziani* dice ironicamente Laclau), rifiutando la dinamica reale (che è populista ed egemonica) di creazione della soggettività rivoluzionaria. Nel suo rifiuto del populismo Žižek sta rifiutando ogni ipotesi di articolazione egemonica, ed in ciò, secondo Laclau, si avvicinerebbe ad un 'rifiuto marxista' della politica come prassi contingente.

Il secondo autore trattato è Badiou nel saggio *Un'etica del compromesso* militante: qui ricompare in apertura di nuovo Žižek ma a sugellare con una citazione (dal suo punto di vista criticamente) un'affinità fra il filosofo francese e Laclau, rispetto alla quale questi si riconosce. La teoria politica di Badiou infatti ha degli elementi in comune con quella di Laclau, nel suo riconoscere la centralità della decisione (quindi di un momento etico) nella cesura fra situazione ed evento. La filosofia politica di Badiou condividerebbe con quella di Laclau il pensare una cesura nell'oggettività della struttura alla base della decisione politica (che rimane puntuale), anche se fondata su un'ontologia matematica basata sulla teoria degli insiemi che si distingue dall'ontologia linguistico-discorsiva di matrice sausarriana di Laclau (cfr. pp. 120-1). Lungo tutto il saggio il discorso del teorico argentino procede riconoscendo la validità dei problemi posti da Badiou, e la corretta direzione intrapresa, salvo però mostrare delle aporie che possono essere risolte con le proprie proposte teoriche.

L'errore teorico fondamentale di Badiou sarebbe l'aver posto una distinzione fra l'adesione all'universalità vuota dell'evento e la sostituzione con quello che viene definito simulacro. Per Laclau, se non si tradisce la purezza dell'evento, non si innesca quel gioco fra pieno e vuoto, fra universalità e particolarità, che

permette di pensare l'azione politica in senso non esclusivamente destituente. Nella logica egemonica la «distinzione tra evento vero e simulacro collassa» (p. 108) ma ciò permette la nominazione del vuoto attraverso un particolare che dà nome all'evento secondo la logica del significante vuoto. In questa logica si può pensare la «situazionalità» (p.103) come funzione strutturante di un ordine che è già compreso nel momento di destrutturazione del precedente. Quello che manca in Badiou affinché si possa uscire dall'impossibilità della scelta etica – perché basata su una distinzione puramente formale – è che l'evento invece che essere pura destituzione venga pensato contemporaneamente anche come articolazione contingente di ordine. Scrive Laclau: «quello che in questi casi è l'oggetto di un investimento etico non è il contenuto ontico di un certo ordine, ma il principio dell'ordinare in quanto tale» (p.115)

Questa coazione all'ordinare, che permane all'interno dell'evento destituente viene efficacemente descritta da Laclau con il caso del gramsciano «divenire Stato della classe operaia» (p.112), mostrando così le conseguenze politiche di questa differenza di ontologia politica. Badiou cioè nella sua logica del puro evento, e della pura molteplicità non riesce a concepire la costruzione dell'egemonia e dello Stato, cosa invece che Laclau nella sua ontologia dell' «unità fallita» (p. 124) può pensare.

Nuda vita o indeterminazione sociale è sostanzialmente una lettura critica di *Homo Sacer* (Einaudi, Torino, 2005) di Giorgio Agamben, rispetto al quale il giudizio di Laclau è politicamente molto liquidatorio, pur riconoscendone l'erudizione (che però lo porta all'ipostatizzazione dell'etimologia – e quindi dell'origine – come momento disvelante del concetto) (cfr. p. 129). Anche nel caso di Agamben Laclau critica il rifiuto radicale della logica dell'unità, che in questo caso si specifica nell'unità del popolo nella sovranità che secondo il filosofo italiano darebbe vita ad un governo totalitario. Sussiste quindi una radicale opposizione fra un potere onnipervasivo e un'alterità assoluta. Troviamo cioè la netta distinzione fra un di fuori bandito che è alla mercé del potere sovrano e il sovrano stesso che può disporre della nuda vita, creando attraverso la violenza unità fittizie come quella del popolo (che sutura la pluralità delle identità sociali). Al contrario per Laclau la funzione della sovranità e la dialettica fra popolo e plebe seguono una logica egemonica, sono cioè momenti di contesa. Il bandito non è un corpo ridotto a mera vita, ma è anzi il punto di partenza per la costruzione di un blocco contro-egemonico, che nella misura in cui può arrivare a produrre una legge rivoluzionaria che si oppone alla legge sovrana, un bando contro il bando, entra in una relazione antagonistica prettamente politica con il potere vigente. In questo però è l'alterità stessa che non deve mantenersi tale, ma farsi unità, «divenire stato» (per seguire ancora Gramsci), e pensare la possibilità di sfruttare il momento dell'eccezione come occasione per la produzione di una nuova sovranità. Qui Laclau sostanzialmente critica le acquisizioni della biopolitica foucaultiana che operano alla base del discorso di Agamben e che ne determinano la specifica ontologia politica. Ma la conseguenza politica della sua particolare rilettura di Foucault è per Laclau proprio la spoliticizzazione,

l'assenza della possibilità dell'azione che in ogni caso rimarrebbe presa nell'ambito della sovranità come subalternizzazione dell'alterità. Il «nichilismo politico» (p.143) e quindi l'esodo dalla politica è la sola possibilità quando l'unica società alternativa immaginabile è quella di un'escatologia anarco-comunista: la società pienamente riconciliata, senza potere. Con Agamben ci troviamo quindi, secondo Laclau, di nuovo in una logica meccanica, che non riesce a leggere la natura discorsiva ed egemonica dell'ontologia politica, e dividendo il mondo in secche e rigide alternative dalle quali non si può rifuggire, non può che arrendersi all'impossibilità di un'unità politica altra rispetto a quella di un potere sovrano totale sui corpi che ne dispone secondo il modello paradigmatico del campo di concentramento.

Il confronto con l'opera di Toni Negri e Michael Hardt (*L'immanenza può spiegare le lotte sociali?*) che chiude il libro, è forse il testo più noto di questa raccolta, anche perché è qui che si scontra con una proposta ontologico politica che ha avuto una messa in atto strategica alternativa a quella populista, è stata cioè messa alla prova della storia. Qui lo scontro si articola intorno all'alternativa – strettamente connessa con quella fra unità e pluralità o differenza che abbiamo visto all'opera negli altri saggi – fra trascendenza e immanenza. Scrive Laclau a tal proposito:

> "Il popolo" è il soggetto specifico della politica, un termine questo che presuppone una netta divisione nel corpo sociale non riconducibile ad alcun tipo di unità immanente. *Impero*, al contrario, fa dell'immanenza la sua categoria centrale e il fondamento ultimo dell'unità della moltitudine. (p.145)

Nella principale opera di Negri e Hardt, *Impero* (Milano: Rizzoli, 2003) troviamo quindi una qualche forma di unificazione della soggettività trasformatrice, però questa unità si dà come immanente e spontanea, non come il frutto di articolazione egemonica che implica – come abbiamo visto nella sezione dedicata a Badiou – un'intenzionalità e una trascendenza politica – la logica dell'evento, della cesura ecc… – rispetto al piano del sociale.

Per Laclau la storia dell'immanentismo risale fino a Scoto Eriugena e sostanzialmente serve a conciliare l'onnipotenza divina con la presenza del male: dalla teologia alla filosofia spinoziana e poi ad Hegel e Marx ci troviamo di fronte all'idea di una logica soggiacente (una divinità immanentizzata) come spiegazione della presenza del negativo, la cui esistenza è così giustificata nell'ordine positivo (cfr. p. 126-7). Laclau al contrario pensa il male (o, in senso secolarizzato, l'antagonismo sociale) come qualcosa di irriducibile: una negatività ontologica, e si contrappone alla tradizione immanentistica, che riassorbe, positivizzandolo in un automatismo puramente logico, il negativo. Da un lato una logica dell'immanenza, dell'automatismo, della totalità dialettica e infine, nella versione di Negri e Hardt dell'affermazione in senso nietzscheano, dall'altro, le filosofie della negatività radicale, dell'unità fallita e dell'impossibilità della conciliazione finale: è qui che si gioca il problema ontologico politico fondamentale. Laclau lo dice chiaramente:

Troviamo qui il vero spartiacque teorico delle discussioni contemporanee: *o* si ritiene che vi sia la possibilità di un'universalità non politicamente costituita e mediata, *o* si ritiene che ogni universalità sia precaria e dipenda dalla combinazione storica di elementi eterogenei. (p.149-50)

La conseguenza politica di questa concezione è di nuovo la rinuncia alla possibilità dell'articolazione politico-egemonica e la negazione dei principi della sovranità (cfr. p. 150) e della rappresentanza/rappresentazione (cfr. p. 151), visti come strumenti che vogliono imporre una trascendenza neutralizzante alla potenza immanente della moltitudine. Sul piano dell'immanenza la volontà della moltitudine si identifica immediatamente con la propria potenza e con l'universalità, e non necessita quindi di nessuna mediazione articolatoria fra la particolarità degli agenti e l'universalità (fallita) della comunità. Nello scontro tra la soggettività politica pensata da Negri ed Hardt – la moltitudine – e quella pensata da Laclau – il popolo – possiamo cogliere tutta la distanza fra i sostenitori di una logica tutta movimentista e antagonistica e quella del «divenire stato» che il populismo latinoamericano sostenuto dal teorico argentino stava sperimentando.

Nella piena politicità dello scontro con Negri e Hardt cogliamo quindi, in conclusione, il senso di questi *dibattiti e scontri* fra alternative ontologico politiche della contemporaneità. Unità e pluralità, positivo e negativo, trascendenza e immanenza, concetti che appartengono alla tradizione filosofica moderna qui scendono nell'agone dello scontro politico e mostrano la loro rilevanza teorico-pratica, oltre le letture «volgarmente empiriche» che troppo spesso non ne colgono le logiche profonde che li strutturano. Con questa raccolta, meritoriamente proposta ora in Italia da Samuele Mazzolini e Fortunato M. Cacciatore, autori anche dei due saggi introduttivi molto dettagliati che possono essere di grande aiuto per il lettore, auspichiamo che a partire da Laclau si possa allargare sempre di più il campo di questi studi e allo stesso tempo che la politica possa trarne insegnamento per dare un nuovo impulso per una politica emancipativa, da troppo tempo impantanata nella ripetizione delle stesse stanche e inefficaci proposte strategiche.