S&F_n. 27_2022

scienzaefilosofia.it

S&F_n. 27_2022

ISSN 2036_2927

S&F_Scienzaefilosofia.it

Comitato Scientifico

Paolo Amodio	Docente di Filosofia Morale _ Università degli Studi di Napoli Federico II
Guido Barbujani	Docente di Genetica _ Università degli Studi di Ferrara
Edoardo Boncinelli	Docente di Biologia e Genetica _ Università "Vita-Salute San Raffaele" di Milano
Rossella Bonito Oliva	Docente di Filosofia Morale _ Università degli Studi di Napoli – L'Orientale
Barbara Continenza	Docente di Storia della scienza e delle tecniche _ Università degli Studi di Roma "Tor Vergata"
Massimiliano Fraldi	Docente di Scienza delle costruzioni _ Università degli Studi di Napoli Federico II
Orlando Franceschelli	Docente di Teoria dell'evoluzione e Politica _ Università degli Studi di Roma "La Sapienza"
Elena Gagliasso	Docente di Filosofia e Scienze del vivente _ Università degli Studi di Roma "La Sapienza"
Gianluca Giannini	Docente di Filosofia Morale _ Università degli Studi di Napoli Federico II
Pietro Greco	Giornalista scientifico e scrittore, Direttore del Master in Comunicazione Scientifica della Scuola Internazionale Superiore di Studi Avanzati (SISSA) di Trieste
Giuseppe Lissa	Professore Emerito di Filosofia Morale _ Università degli Studi di Napoli Federico II
Giuseppe O. Longo	Docente di Teoria dell'informazione _ Università degli Studi di Trieste
Roberto Marchesini	Centro Studi Filosofia Postumanista_Direttore della Scuola di Interazione Uomo Animale (SIUA)_Docente di Scienze Comportamentali Applicate
Davide Marocco	Docente di Psicometria _ Università degli Studi di Napoli Federico II _ Già Lecturer in Cognitive Robotics and Intelligent Systems, Centre of Robotics and Neural Systems, School of Computing and Mathematics, University of Plymouth, UK
Alessandro Minelli	già Docente di Zoologia _ Università degli Studi di Padova
Maurizio Mori	Docente di Bioetica _ Università degli Studi di Torino
Telmo Pievani	Docente di Filosofia della Scienza _ Università degli Studi di Milano-Bicocca
Vallori Rasini	Docente di Filosofia Morale _ Università degli Studi di Modena e Reggio Emilia
† Stefano Rodotà	Docente di Diritto Civile _ Università degli Studi di Roma "La Sapienza"
Giacomo Scarpelli	Sceneggiatore cinematografico e storico della filosofia e delle idee, insegna all'Università di Modena e Reggio Emilia. È Fellow della Linnean Society of London e della Royal Geographical Society
Settimo Termini	già Docente di Cibernetica _ Università degli Studi di Palermo
Nicla Vassallo	Docente di Filosofia Teoretica _ Università degli Studi di Genova

International Advisory Board

David Banon	Professeur au Département d'études hébraïques et juives, Université de Strasbourg; Membre de l'Institut Universitaire de France; Prof. invité au départment de pensée juive, Université hébraïque de Jérusalem
Renaud Barbaras	Professeur à l'Université Paris 1 Panthéon-Sorbonne; Membre de l'institut universitaire de France; Grand prix de philosophie 2014 de l'Académie Française
Mylene Botbol-Baum	Professeure de philosophie et bioéthique, facultés de Médecine et des Sciences Philosophiques; Professeure à l'Unité d'éthique biomédicale – UCL (Université Catholique de Louvain), Bruxelles, Belgique
Colas Duflo	Professeur de Littérature française et d'histoire des idées du dix-huitième siècle, Université Paris-Nanterre, France
Edward K. Kaplan	Kevy and Hortense Kaiserman Professor in the Humanities, Brandeis University, Waltham, Massachusetts
Neil Levy	Deputy Director (Research) of the Oxford Centre for Neuroethics; Head of Neuroethics at the Florey Neuroscience Institutes, University of Melbourne
Anna Lissa	Maitre de Conférence Littérature juive et hébraïque Département d'Etudes hebraïques Universite Paris-8, France
Diego Lucci	Professor of History and Philosophy, American University in Bulgaria
Max Stadler	Professur für Wissenschaftsforchung, Eidgenössische Technische Hochschule, Zürich

Redazione

Paolo Amodio (Direttore)	Università degli Studi di Napoli Federico II_ Dipartimento di Studi Umanistici_ Via Porta di Massa, 1 80133 Napoli tel. +390812535582 fax +390812535583 email: paamodio@unina.it
Luca Lo sapio (Coordinatore di Redazione)	Università degli Studi di Torino
Viola Carofalo	Università degli Studi di Napoli_L'Orientale
Didier Alessio Contadini	Università degli Studi di Milano_ Bicocca
Mario Cosenza	Università degli Studi di Napoli_Federico II
Cristian Fuschetto	Università degli Studi di Napoli_Federico II
Fabiana Gambardella	Università degli Studi di Napoli_Federico II
Delio Salottolo	Università degli Studi di Napoli_L'Orientale
Alessandra Scotti	Università degli Studi di Napoli_Federico II
Rosa aka Vittoria Spagnuolo Vigorita	Università degli Studi di Napoli_Federico II
Aldo Trucchio	Université du Québec à Montreal

INTRODUZIONE

Scienza**&F**ilosofia 27_2022, con un Dossier sull'Intelligenza artificiale. S&F_ sempre online per scelta, in ordine al dinamismo e all'immediata disponibilità della ricezione, adattandosi volentieri ai tempi e agli spazi che la rete in genere istituisce: vorrebbe essere agile e facilmente fruibile per chi è interessato a prender parte alle nostre discussioni. La sua *mission* non può dunque che essere diretta e senza scolastici orpelli:

Preoccupata di istituzionalizzarsi come depositaria della coscienza etica del progresso scientifico, a quasi trent'anni dalla sua nascita la bioetica sembra essere a un bivio: rinnovare il suo statuto o rischiare di smarrire definitivamente la sua mission di disciplina di incrocio tra sapere umanistico e sapere scientifico. È nostra convinzione che la bioetica possa continuare a svolgere un ruolo solo se, piuttosto che salvaguardare principi assiologici di una realtà data, sia intenzionata a ripensare criticamente i valori alla luce dei cambiamenti, epistemologici prima ancora che ontologici, dettati dall'età della tecnica. Il nostro obiettivo è quello di individuare ed evidenziare il potenziale d'innovazione filosofica tracciato dalla ricerca scientifica e, al contempo, il potenziale d'innovazione scientifica prospettato dalla riflessione filosofica.

Da questa *mission* la rivista trova l'articolazione che ci è parsa più efficace. Anche questo numero conterrà perciò le tipiche sezioni:

DOSSIER Il vero e proprio focus tematico scelto intorno al quale andranno a orbitare

STORIA Esposizione e ricostruzione di questioni di storia della scienza e di storia di filosofia della scienza con intenzione sostanzialmente divulgativa;

ANTROPOLOGIE Temi e incroci tra scienze, antropologia filosofica e antropologia culturale;

ETICHE Riflessioni su temi di "attualità" bioetica;

LINGUAGGI Questioni di epistemologia;

ALTERAZIONI Dalla biologia evoluzionistica alla cibernetica, temi non direttamente "antropocentrati";

COMUNICAZIONE La comunicazione della scienza come problema filosofico, non meramente storico o sociologico. In altri termini:

quanto la comunicazione della scienza ha trasformato la scienza e la sua percezione?;

ARTE Intersezioni tra scienze e mondo dell'arte;

RECENSIONI&REPORTS Le recensioni saranno: *tematiche*, cioè relative al dossier scelto e quindi comprensive di testi anche non recentissimi purché attinenti e importanti; *di attualità*, cioè relative a testi recenti. *Reports* di convegni e congressi.

Per favorire la fruibilità telematica della rivista, i contributi si aggireranno tra le 15.000 – 20.000 battute, tranne rare eccezioni, e gli articoli saranno sempre divisi per paragrafi. Anche le note saranno essenziali e limitate all'indicazione dei riferimenti della citazione e/o del riferimento bibliografico e tenderanno a non contenere argomentazioni o ulteriori approfondimenti critici rispetto al testo.

A esclusione delle figure connesse e parti integranti di un articolo, le immagini che accompagnano i singoli articoli saranno selezionate secondo il gusto (e il capriccio) della Redazione e non pretenderanno, almeno nell'intenzione – per l'inconscio ci stiamo attrezzando – alcun rinvio didascalico.

Le immagini d'apertura ai singoli articoli – coperte da copyright – che appaiono in questo numero, sono di MAC EPARWA, giovane e talentuoso artista filippino, cui va il nostro ringraziamento (https://www.instagram.com/m.eparwa/?igshid=YmMyMTA2M2Y=)

In rete, giugno 2022

La Redazione di S&F_

S&F_n. 27_2022

INDICE

1 **Introduzione**
4 **Indice**

Dossier
7 **Intelligenza artificiale e nuove piattaforme per un umanesimo a venire**
12 **Edmondo Grassi** *Di automi e persone*
28 **Gianluca Giannini – Antonio Pescapè [Luca Lo Sapio]** *AI e futuro di* sapiens *tra nuovi orizzonti e antichi timori*
43 **Andrea Calandrelli – Alessandra Nicolini** *Dono e dolo: dal digital turn al tecnostress*
54 **Simona Tiribelli** *Artificial Agency and Moral Agency: Conceptualizing the Relationship and its Ethical Implications on Moral Identity Formation*
69 **Cristiano Calì** *Algoritmi e processo decisionale. Alle origini della riflessione etico-pratica per le IA*
88 **Settimo Termini** *L'IA tra Katzuo Ishiguro e Federico Faggin*
109 **Maria Teresa Catena** *L'intelligenza delle mani*
126 **Pierluigi Ametrano** *AI e archivio: un'analisi foucaultiana*
137 **Pierpaolo Marrone** *Spazi hobbesiani?*
160 **Valerio Specchio** *Scomparire nell'alterità. L'intelligenza artificiale attraverso Jean Baudrillard*
179 **Alfonso di Prospero** *Intelligenza artificiale e inferenza non-monotona: modelli culturali e questioni epistemologiche*
197 **Edoardo Datteri** *Robotica e filosofia della scienza*

Storia
218 **Alfonso Lanzieri** *Da Matrix a Don't look up : il realismo come questione etico-politica*

Antropologie
237 **Mauro Mandrioli** *From genome editing to human genetic enhancement: a new time for discussing eugenics?*

Etiche
251 **Delio Salottolo** *È davvero necessaria un'etica delle piante?*

Linguaggi
278 **Giacomo Scarpelli** *Freud e Conan Doyle. Vite parallele*

Alterazioni
294 **Giovanni Altadonna** *La critica all'antropologia razziale in Stephen Jay Gould*

Comunicazione
314 **David Benatar** *Perché è meglio non venire mai al mondo*

Arte
337 **Armando Minopoli** *L'essenziale non è invisibile agli occhi. Divagazioni sul concetto di essenza in architettura*

RECENSIONI&REPORTS

report

357 **Tessuto e sospensione. Il corpo inaggirabile tra io e mondo**
Merleau-Ponty: Ritornare alla percezione, Napoli 2022
(Valerio Specchio)

recensioni

368 **Yuk Hui, *La questione della tecnologia in Cina,*** Neri, Roma 2021
(Francesca De Simone)

375 **Luciano Floridi, *Etica dell'intelligenza artificiale. Sviluppi. Opportunità, sfide,*** Raffaello Cortina, Milano 2022
(Stefano Palumbo)

383 **Angela Balzano, Elisa Bosisio, Ilaria Santoemma (a cura di), *Conchiglie, pinguini, staminali. Verso futuri transpecie,*** DeriveApprodi, Roma 2022
(Melania Moltelo)

386 **Telmo Pievani, *Serendipità. L'inatteso nella scienza,*** Raffello Cortina, Milano 2021
(Giovanni Altadonna)

397 **Margaret A. Boden, *L'intelligenza artificiale,*** Il Mulino, Bologna 2019
(Stefano Palumbo)

404 **Maria Teresa Catena, *Breve storia del corpo,*** Mimesis, Milano-Udine 2020
(Alberto Giovanni Biuso)

408 **Federica Porcheddu, *Ripensare il terzo a partire da Lévinas,*** Mimesis, Milano-Udine 2021
(Sergio Labate)

414 **Carlo Rovelli, *Helgoland,*** Adelphi, Milano 2020; **Telmo Pievani, *Finitudine. Romanzo filosofico su fragilità e libertà,*** Raffaello Cortina, Milano 2020
(Giordana Colicchio)

419 **Adriano Pessina (a cura di), *Vulnus. Persone nella pandemia,*** Mimesis, Milano-Udine 2022
(Viola Carofalo)

S&F_n. 27_2022

DOSSIER

Intelligenza artificiale e nuove piattaforme per un umanesimo a venire

ABSTRACT: ***ARTIFICIAL INTELLIGENCE AND NEW PLATFORMS FOR AN INCOMING HUMANISM***

Artificial intelligence is a research field of computer science that since 1956, when the expression was coined, has produced countless devices capable of performing complex tasks, in some cases, comparable to human performativity. In 1983, Elaine Rich defined artificial intelligence as «the branch of computer science concerned with the possibility of building computers capable of performing tasks that humans currently perform better». Rich's definition differs from McCarthy's more challenging definition of AI as «a science and engineering branch aimed at building intelligent machines», because it avoids the hard problem of giving a definition of "intelligence".

On the other hand, Roger Penrose, pointed out that «true intelligence cannot be presented without consciousness, and for this reason intelligence can never be produced by some algorithm implemented on a computer».

This consideration draws a new idea of artificial intelligence: no longer a mere transposition of human intelligence into a machine, but «a new form of capacity to act» (Luciano Floridi). Faced with the possibilities opened up by this new form of ability to act, «it is up to each one of us to decide how to use our latest invention and choose which future we want to contribute to realizing for the Homo faber in us» (Federico Cabitza).

This Dossier explores some lines of reflection raised by the developments of artificial intelligence that, more and more present in our daily lives and in a future that is just around the corner, announces unprecedented possibilities and new potential risks for sapiens. On the one hand, the moral issues raised by the possible emergence of a Superintelligence capable of self-assigning goals, not necessarily compatible with those of sapiens, on the other hand, the many applications related to social robotics that lead to a critical reconsideration of categories such as care and assistance, work, leisure, etc..

Will the digital era be dominated by robots? Is the utopia of a world freed from drudgery, built around an industry (5.0) that is able to realize the potential inherent in advanced digitization, Big data and artificial intelligence finally upon us? Are the once central distinctions between strong and weak AI still valid? What possibilities are opened up by sectors such as Agrifood, in which artificial intelligence is used in the food manufacturing process? Does the encounter between artificial intelligence and quantum mechanics open up new pathways capable of accelerating the emergence of a Superintelligence? Are we ready to govern a new model of intelligence that is not anthropocentric? What repercussions do the new lines of research of deep learning and machine learning have on the self-representation of sapiens? What new narrative and self-narrative scenarios are unfolding in front of sapiens?

Per parafrasare il vangelo di Giovanni, il Verbo si è fatto macchina, lo spirito soffia anche nell'inorganico e la ragione e il linguaggio, oggettivati in forma di algoritmo, abitano in corpi non umani, creando una «umanità aumentata».

Remo Bodei

Nessuno ne parla in questo modo, ma penso che l'intelligenza artificiale sia quasi una disciplina umanistica. È davvero un tentativo di comprendere l'intelligenza umana e la cognizione umana.

Sebastian Thrun

La cosa triste dell'intelligenza artificiale è che manca di artificio e quindi di intelligenza.

Jean Baudrillard

Prima di lavorare sull'intelligenza artificiale, perché non facciamo qualcosa per la stupidità naturale?

Steve Polyak

Di qualunque cosa si tratti, io sono contro.

Groucho Marx

L'intelligenza artificiale (AI) è un settore dell'informatica che si propone di realizzare sistemi hardware e software in grado di fornire all'elaboratore elettronico prestazioni che *tenderemmo* ad attribuire *solo* agli esseri umani. In realtà, *già* questa definizione *appare fortemente problematica*. Innanzitutto, pur essendo *associata* all'informatica, l'AI è «un campo molto vasto», dal momento che la disciplina comprende, al contempo, «la logica, la probabilità e la matematica del continuo; la percezione, il ragionamento, l'apprendimento e l'azione; l'equità, la fiducia, il bene sociale e la sicurezza, oltre ad applicazioni che spaziano dai dispositivi microelettronici ai robot per l'esplorazione planetaria, fino ai servizi online con miliardi di utenti»[1]. Inoltre, con l'acronimo AI (*Artificial intelligence*) indichiamo *oggetti* molto differenti, dal *Machine learning,* al *Deep learning*, dalla *Computer Vision,* alla *Pattern Recognition,* dall'*Automated Reasoning,* al *Game theory,* e poi ancora *Logics, Multi-Agents, Fuzzy systems, Knowledge representation, Speech*

[1] S. Russell, P. Norvig, *Intelligenza artificiale. Un approccio moderno* (2020), tr. it. Pearson, Milano-Torino 2022, p. XVII.

Recognition, Natural Language Processing, Cognitive *Robotics* e tanto altro[2].

In aggiunta a ciò, va evidenziato che anche il riferimento al comportamento umano è *questione difficile da districare*. Nel 1980 il filosofo inglese John Searle propose la nota distinzione tra una *Weak AI* e una *Strong AI*[3], la prima capace di agire *come se* fosse intelligente, la seconda capace di agire *in maniera realmente intelligente*. Servendosi, poi, in uno scritto di alcuni anni dopo, dell'argomento della stanza cinese[4], mise in evidenza che nessuna AI avrebbe potuto *realmente* eguagliare un essere umano, dal momento che quest'ultimo opera a un livello non solo *sintattico* ma anche *semantico*, non limitandosi a manipolare simboli, sulla base di un *manuale di istruzioni*, ma *accedendo* anche al loro *significato*. Nell'ottica di Searle, dunque, nessuna intelligenza artificiale potrebbe *essere umana* essendo nelle cosiddette macchine intelligenti assenti i *qualia*, cioè una visione della realtà *in prima persona*, un'interiorità.

Che si accettino o meno le conclusioni di Searle, va sottolineato che, dopo una prima fase in cui la *Strong AI* sembrava essere, per gli addetti aI lavori, l'obiettivo primario della ricerca, si è passati a una *diversa tendenza*, in cui l'AI *task oriented*, di carattere ingegneristico-riproduttivo si è imposta rispetto all'AI cognitivo-produttiva[5].

Tale apparente novità può essere letta nei termini, esplicitati da Luciano Floridi, di un divorzio tra *intelligere* e *agere*. L'AI sarebbe, in questa prospettiva, una nuova forma dell'agire *più che* una riproduzione delle *capacità cognitive umane*[6].

---

[2] Cfr. *infra* G. Giannini, A. Pescapè, L. Lo Sapio, *AI e futuro di sapiens tra nuovi orizzonti e antichi timori*, in «S&F_scienzaefilosofia.it», 27, 2022.

[3] Cfr. S. Russell, P. Norvig, *op. cit.*, p. 341.

[4] J. Searle, *Consciousness, explanatory inversion and cognitive science*, in «Behavioral and Brain Sciences», 13, 1, 1990, pp. 585-642.

[5] Cfr. L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina, Milano 2022.

[6] L. Floridi, F. Cabitza, *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Milano 2021.

Partendo da questa idea, in apparenza paradossale, di una AI *intelligence-free*, si avvia la strutturazione di una piattaforma di comprensione attraverso la quale l'AI non viene né sovradimensionata né sottodimensionata nelle sue prerogative.
Sono molti gli ambiti di applicazione dell'AI, così tanti che ormai *si farebbe* prima a elencare dove *non* sia presente.
Essendo *ubiqua* l'AI solleva numerose questioni, via via che si interseca con ambiti più o meno articolati dell'*agere* umano, costringendo *sapiens* a fare i conti con un prodotto del proprio *ingenium* che retroagendo su di lui, lo ridetermina nelle proprie prerogative e nella costruzione del proprio *ubi consistam*.
Veicoli a guida autonoma, smart-fridges, armi letali autonome, Big Data e Data mining, nuovi scenari per la privacy, la sorveglianza di massa e la sicurezza, il futuro delle democrazie liberali e le nuove possibilità offerte ai regimi autoritari, i diritti dei robot, la sicurezza dell'AI e il suo impiego in contesti di cura e assistenza, il futuro del lavoro, sono soltanto alcuni degli *items* intorno ai quali si sviluppano i *pluristratificati* sentieri dell'Intelligenza artificiale.
In gioco tante questioni, come già profeticamente Alan Turing aveva intravisto nel suo articolo del 1950 *Computing machinery and intelligence*[7].
Tuttavia, a noi sembra siano due i temi-caldi intorno ai quali le altre questioni si addensano, fino a formare un nuovo "sistema planetario": "quale etica per l'intelligenza artificiale?" e "quale ruolo *anthropos* riuscirà a ritagliarsi all'interno del *loop* che lo vede *co-protagonista* insieme ai frutti del proprio ingegno?"[8].
In merito al primo punto, si tratta di andare oltre la *semplice* (e per certi versi *semplicistica*) dicotomia "consequenzialismo/deontologismo" (con una timida possibilità di

[7] A. Turing, *Computing machinery and intelligence*, in «Mind», LIX, 236, 1950.
[8] Il tema risuona anche nel testo di G. Lissa, *Morte e/o trasfigurazione dell'umano*, Giannini Editore, Napoli 2019.

apertura all'etica della cura), che pur essendo efficace a fornire coordinate per comprendere alcuni aspetti della Rivoluzione in atto, ci impedisce di coglierne altri. Si tratta, dunque, di recuperare il senso più autentico di *Etica*, non come applicazione astratta di principi e regole alle fattispecie particolari ma come *spazio dell'abitare* in cui, sempre più, le alterità non umane, e ora anche le *agentività artificiali* (come qualcuno ama definirle) sono co-presenti.

A partire da qui, la seconda questione. Al di là delle distopie di autori come Nick Bostrom[9] o Stephen Hawking[10], il ruolo che *sapiens* saprà ritagliarsi *entro* il nuovo spazio dell'abitare, risulterà *cruciale*. L'allineamento dei valori delle nuove macchine intelligenti agli scopi di *sapiens* sarà, dunque, *la* questione decisiva intorno alla quale, nei prossimi anni, verosimilmente si struttureranno molti dei dibattiti sull'intelligenza artificiale, sia nella forma *task oriented* sia nella forma, di cui alcuni autori preconizzano la diffusione *urbi et orbi*, dell'AGI (*Artificial general intelligence*), ossia una intelligenza artificiale non solo in grado di *apprendere* dall'esperienza (come avviene oggi con i sistemi di *machine learning*) ma radicalmente *task-free*.

P.A., L.L.S.

[9] N. Bostrom, *Superintelligenza. Tendenze, pericoli, strategie* (2014), Bollati Boringhieri, Torino 2018.

[10] Autore quest'ultimo, nel 2015, insieme ad altri studiosi di una *Open letter on artificial intelligence*.

EDMONDO GRASSI

DI AUTOMI E PERSONE

1. *Annotazioni sull'artificio "umano"* 2. *L'utopia androide nella post-realtà*
3. *Ramificazioni e proiezioni*

ABSTRACT: ***ON ROBOTS AND PERSONS***
The immanence of digital technologies and their function as agents acting in the social context has allowed a new debate on the figure of the person, his value and his relationship with technologies. The intent of this contribution is to theoretically outline the ontological relationship between a person in his relationship with the automaton, whether in the form of a machine or an algorithm, which defines its natural artificiality. This relationship has crossed history and, to date, it is no longer conceivable to debate scientifically through dichotomous classes but it becomes essential to recognize the same generative matrix within the social context, since the person is such in the manifestation of his relationship with the collective environment. through the multiple communicative grammars expressed through the technological apparatus.

1. *Annotazioni sull'artificio "umano"*

Le funzioni politiche, le strutture istituzionali, la grammatica urbanistica, le declinazioni dei rapporti collettivi e la narrazione dei valori morali della persona attraverso i quali una società organizza il proprio corpo sociale e la coesistenza degli individui sono elementi scaturiti da scelte condotte attraverso l'osservazione del progresso scientifico-tecnologico in specifici spazi che sono indirizzati ed edificati attorno al grado di sviluppo della cultura materiale e intellettuale. Anche l'essere umano risulta essere un meccanismo finalizzato a una precisa configurazione per la sussistenza della sua specie e della società

di appartenenza secondo le necessità degli interessi dominanti: l'afflato della perpetuazione della vita diviene razionalità tecnologica che muta in controllo di ogni azione del soggetto.

Dalla prima rivoluzione industriale alla comparsa dell'intelligenza artificiale, l'affermazione della tecnologia come elemento immanente nella biologia della persona è divenuta piena dimensione politica e ambiente vitale nel quale articolare la realizzazione di specifiche mete culturali e di osservazione della natura, umana e non, come strutture complesse per l'attuazione di esperienze, mutamenti e organizzazione della nuda vita come campo di potere, manifestazione del sé e della sua potenziale coercizione. Ciò si riflette anche nella visione dell'automazione totalizzante che va ricercata nella patologia linguistico-cognitiva sociale della ragione priva di plasticità, subalterna all'immediato e incapace di dialettica nell'ipotesi di un reale perduto.

I mutamenti dell'essere sono già narrati da Ovidio attraverso i miti che, secondo una prospettiva sociale contemporanea, possono divenire strumenti per comprendere l'individuo, le interrelazioni sociali, i sentimenti intimi e collettivi, il rapporto con il sé e con le istituzioni attraverso l'espediente metamorfico finalizzato a osservare, analizzare, destrutturare o ricomporre il soggetto delle storie, donando un quadro macrosociologico a un ordine sociale di carattere mitologico/universalistico. All'eroe unico che racchiudeva i valori di un tempo e di una cultura classica – si pensi a Odisseo o a Enea – si contrappone il coro di voci delle *Metamorfosi* quali manifestazione del mutamento e del superamento dei confini dei singoli esseri[11].

L'epidermide rappresenta l'ultimo confine del corpo umano, un limite che permette, però, di comunicare costantemente con l'ambiente circostante grazie alla sua capacità di assorbire,

[11] Cfr. E. Pianezzola, *Ovidio, dalla cosmogonia alla metamorfosi: per la ricomposizione di un ordine universale*, in «Materiali e discussioni per l'analisi dei testi classici», 65, 2010, pp. 59-68.

cernere ed espellere tutto ciò con cui viene a contatto, rappresentando una membrana contigua tra natura sociale e corpo biologico, venendo declinato in nuove apparizioni rispetto a ciò che esperisce costantemente[12]: una forma di compenetrazione costante che definisce non più l'individuo, nel suo essere *dividuo*, separato, a sé stante, ma il suo divenire *con-dividuo*, olobionte di popolazioni di specie diverse che turba la rappresentazione dell'individuo che si autocompleta e che vive un proprio senso del "sé"[13], una struttura comunitaria costantemente mutevole e plasmata da elementi divisibili nella loro temporalità e spazialità. Inoltre, la filosofia dei valori e la distinzione tra *sein* (essere) e *sollen* (dover essere) sono elementi centrali nella disamina sociale[14], avendo l'interesse di dimostrare come il mutamento sia una costante per il riconoscimento della pluralità dei valori del singolo e della collettività. Se nel secolo scorso, la persona ha esperito la frattura tra mito e quotidianità rivelando una struttura sociale fondata sulla ragione, sul calcolo e sull'agire meccanico, nella post-realtà[15], è la persona stessa a divenire dato e sistema di controllo privato e pubblico[16].

---

[12] Cfr. E. Grassi, *The forgery of Deepfake and the "advent" of Artificial Intelligence*, in «Intercultural Relations», 10, 2021, pp. 76–91.

[13] Cfr. M. Monti, C.A. Redi, *Noi siamo l'intestino*, in «La Lettura/Corriere della sera», 8 aprile 2018.

[14] Cfr. M. Weber, *La scienza come professione. La politica come professione* (1919), tr. it. Einaudi, Torino 2004.

[15] Nel contesto contemporaneo, la tensione dell'individuo tende ad allontanarsi sempre più dal contesto reale, in modo tale da poter vivere nella post-realtà. Con tale termine, si vuole indicare un contesto dove sottostare a una finzione tutelare, una vita che abdica alla progettualità costitutiva dell'essere a favore di una costante distrazione attraverso la quale confutare ciò che accade. La rappresentazione diviene sostitutiva della realtà e giunge a ridurre la propria partecipazione sociale a simulacro, in ragione di una verità auto-affermata. L'evento collettivo non è più accolto come rottura radicale della monotonia quotidiana, da indagare e analizzare secondo competenze scientifiche ma è contro-spazio nel quale far emergere teorie, ipotesi, fantasie prive di qualsiasi fondamento razionale, frutto solo di una esasperazione del reale e della sua sempre più complessa analisi. Dato che la realtà, a oggi, si manifesta attraverso una molteplicità di media – una ricerca accademica, una registrazione, una foto o un video – si troveranno passaggi per invalidarne la verità, favorendo la contemplazione di molteplici micro-universi nei quale edificarne una propria.

[16] Cfr. J. Van Dijck, *Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology*, in «Surveillance & Society», 12, 2014, pp.

Nella società in cui digitale, biologico, artificiale, naturale, sintetico e organico sono divenuti sinonimi della costruzione sociale dell'essere, la definizione di persona vive il restauro delle fondamenta ontologiche e ontogenetiche della sua rilevanza nelle scienze sociali e umane: l'individuo non è più il fine ultimo, delimitato e circoscritto, degli studi scientifici, poiché «l'uomo non è più il metro di misura della società»[17]. La conoscenza del mondo è dissoluzione della sua compattezza poiché ogni essere nella sua unicità si trasforma in altro, vive «come distinto e come unico essendo fra uguali»[18], non vi sono gerarchie cristallizzate, di un potere naturale cementificato, stabilendo così una parità, se non un'uguaglianza tra naturale e artificiale: l'essere è per sua natura artificiale sin dalla nascita.

La mutazione dell'essere è un aspetto insito nella natura umana: dalla sua forma cellulare, al divenire embrione, sino al processo di senescenza che decreta il superamento di uno stadio verso un altro. Le innovazioni tecnologiche per essere definite tali devono disporre di un contesto sociale che sia consono al loro funzionamento, alla loro diffusione e a un consequenziale sviluppo che possa favorire la loro acquisizione di ruolo di agente attivo nella dinamica dei fatti sociali. Rispetto all'evoluzione del settore tecnologico degli ultimi secoli, avvenuta in tempi dilatati secondo la processazione del passato al fine di migliorare il presente, nella contemporaneità, si assiste a innovazioni ultra accelerate che valicano i confini strutturali dell'essere nella sua comparazione con la società, manifestando in modo assiomatico che il centro della vita collettiva è fondato sull'interazione, sulla compenetrazione di tutti i fenomeni e

197-208; D. Cardon, *Che cosa sognano gli algoritmi?* (2015), Mondadori, Milano 2016; D. Lupton, *The Quantified Self: a Sociology of Self-tracking*, Polity press, Cambridge 2016.

[17] Cfr. N. Luhmann, *Teoria della società* (1974), tr. it. FrancoAngeli, Milano 2013.

[18] Cfr. H. Arendt, *Vita activa. La condizione umana* (1958), tr. it. Bompiani, Milano 2017.

sulla *Wechselwirkung*[19]: è solo attraverso la collettivizzazione che il capitale sociale di ogni soggetto può produrre la propria manifestazione che verrà riconosciuta nel continuo scavalcamento di forme e nella reciproca conoscenza[20].

Se la prospettiva filosofica e quella sociologica sono state dominate, per anni, dal dualismo cartesiano tra *rex exstensa* e *rex cogitans*, dall'opposizione tra cultura e natura, dallo scontro tra individuo e società, quindi, da prospettive fondate sui dualismi, a oggi, le visione interdisciplinari, necessarie all'evoluzione del sapere, hanno permesso di adottare un orizzonte che riconosce l'individuo come una relazione con l'ambiente, con la peculiarità del soggetto sociale che si definisce in funzione del proprio progetto esistenziale[21], caratterizzandolo in base alla possibilità di esprimere coralmente quanto vissuto. Tale approccio conferma la persona come elemento strutturale della biologia sociale, in cui i rapporti sono stabiliti tra la parte e il tutto e in cui il singolo non è ascrivibile nel rappresentare un'ipotetica universalità astratta come affermato da Kant e Hegel[22]. La tecnica è il modello organizzativo della società dal momento in cui si è assisto al disincantamento del mondo[23] e all'allontanamento dalle arti introspettive, sfruttando il progresso tecnologico non come semplice applicazione di un congegno o di una macchina ma come costituzione intellettuale che permette alla persona di essere un vivente in movimento, un automa – inteso come ciò che si muove, non solo fisicamente ma anche grazie a una potenza di azione/pensiero – alimentato dalla

[19] Cfr. G. Simmel, *Sociologia* (1908), tr. it. Meltemi, Sesto San Giovanni 2018.
[20] Cfr. P. Jedlowski, *La conversazione socievole. Simmel, Chakrabarty e l'orientamento fatico della comunicazione*, in «Studi culturali», IV, 1, 2007, pp. 3-26.
[21] Cfr. J.P. Sartre, *Critica della ragione dialettica. Teoria degli insiemi pratici* (1960), tr. it. Il Saggiatore, Milano 1982.
[22] Cfr. F.A. Trendelenburg, *Per la storia del termine persona* (1908), Morcelliana, Brescia 2015.
[23] Cfr. M. Weber, *La scienza come professione*, cit.

cultura[24]. Biologia sintetica, intelligenza artificiale, nanotecnologia hanno abbandonato le strutture di ricerca per riversarsi nell'ambiente sociale su scala globale, richiedendo a tutte le scienze di adottare riflessioni prospettiche e speculative che siano ponte di superamento dell'antropocene[25].

In questo contributo non si ha la pretesa di sgranare la definizione di persona ma di porla in comunicazione con la sua vicinanza ai concetti di automa, come già accennato, e di macchina nella prospettiva in cui l'artificialità dell'essere non venga considerata come tratto aggiuntivo, bensì costitutivo[26], poiché il corpo è lo strumento per eccellenza dell'essere umano, in quanto costituito di organi e di sensi che permettono di trasformare l'ambiente circostante costruendo ulteriori protesi e apparati[27]. La tecnologia è l'intersezione che unisce essere e società: il processo attraverso il quale l'esistenza dell'essere si modella in funzione degli obiettivi e degli ambienti circostanti che a loro volta sono modellati e generano nuovi ambiti di interazione e commistione in cui la tecnologia dona consapevolezza dei propri limiti e ne colma le carenze[28].

Nella contemporaneità, la tecnologia dell'intelligenza artificiale richiama l'essere umano a riflettere sui cambiamenti dovuti all'era dell'antropocene e della rottura dell'antropocentrismo, contemplando una forza naturalmente artificiale che si innesta pervasivamente e immanentemente nella propria vita nell'ottica dell'etica dei principi e dell'etica delle responsabilità. L'essere umano è definibile quale automa dal momento in cui crea il congegno – parola che indica il potenziamento dell'ingegno, il

[24] Cfr. C. Sini, *L'uomo, la macchina, l'automa*, Bollati Boringhieri, Torino 2009.
[25] Cfr. J. Lovelock, *Novacene. L'età dell'iperintelligenza* (2019), tr. it. Bollati Boringhieri, Torino 2020.
[26] Cfr. S. Natoli, *Manipolazione. Promesse di salute e umanità artificiale*, Festival della Filosofia, Modena, 19 settembre 2020.
[27] Cfr. G.O. Longo, *Il nuovo Golem. Come il computer cambia la nostra cultura*, Laterza, Roma-Bari 2000.
[28] Cfr. A. Gehlen, *L'uomo. La sua natura e il suo posto nel mondo* (1940), Mimesis, Milano 2010.

suo utilizzo in modo attivo, fattuale, risolutivo, il co-ingegnare – definendo la macchina come estrinsecazione delle proprie progettualità future per il compimento di un fine ma, simultaneamente, intrinsecazione della stessa che muta la quotidianità del soggetto e, soprattutto con l'immanenza dell'intelligenza artificiale e delle sue declinazioni, retroagisce sull'individuo, indirizzandolo culturalmente, ridefinendo le sue strutture comunitarie e lasciando che il proprio destino sia iscritto negli strumenti che produce e condivide.

2. *L'utopia androide nella post-realtà*

Charlot (*Tempi Moderni*) è il manifesto dell'operaio che trascorre le sue giornate nella ripetitività dei suoi gesti, delle sue mansioni, introiettando ritmi alienanti e compiti spersonalizzanti, così come Maria (*Metropolis*) è l'insegnante che racchiude in sé la veste di profeta che guida i figli degli operai e conosce il mondo delle macchine: entrambi sono modelli rispetto al quale l'immaginario ricalcherà il prototipo di automa che la collettività sociale assorbirà per decenni, non solo nella proiezione dei due lungometraggi ma anche nel riflesso dei miti e dei modelli di comportamento moderni. Si tratta dell'esplicitazione del concetto che considera lo sviluppo dello strumento come uno sforzo umano concentrato e della strutturazione di un sistema operativo comportamentale che è incastonato già ai prodromi dell'umanità[29]. In che modo conciliare la cultura umanistica e la tecnologia? Come colmare lo iato tra etica e tecnica? Sembrerebbero domande di difficile approccio ma la filosofia e, in seguito, la sociologia sono di per sé elementi fondamentali nel dispiegamento della tecnica, in quanto sono lo spirito e la struttura della sua progettualità esistenziale. Nel

[29] Cfr. G. Simondon, *Sulla tecnica* (2014), tr. it. Orthotes, Napoli-Salerno 2017.

rapporto con l'essere, è l'etica che, attraverso i mutamenti culturali, struttura valori che siano strettamente dipendenti, nella loro proiezione futura, al contesto socioculturale nel quale si svilupperanno[30] in funzione di una maggiore comprensione dei principi morali in un orizzonte integrativo e complementare, nella sua capacità trasduttiva, ovvero di sostenere, delineare, rafforzare contenutisticamente lo sviluppo esponenziale dei congegni.

Nella dimensione simbolica, inoltre, viene rappresenta la coesistenza del mondo interiore dell'individuo che si connette con lo spirito collettivo, divenendo una delle condizioni vitali per il fatto sociale, nel quale i sistemi simbolici acquisiscono una propria semantica fondata sull'interdipendenza dei segni e dei significati. L'immaginario rappresenta, quindi, la dimensione costitutiva delle prassi attraverso le quali l'individuo si confronta con sé, con l'ambiente e con l'altro, rappresentando quel fantastico conscio e inconscio di cui l'essere non può privarsi per la sua esistenza[31].

In *Ghost in the Shell*, manga cyberpunk del 1989, vengono declinate le preoccupazioni di una società in cui il soggetto è sempre più isolato ma coniuga tale critica con una vocazione risolutiva esistenzialista, psicoanalitica e introspettiva. La persona e l'automa vengono posti in un ambiente post-informatico, ricercando una contro-finalità, una praxis che sia immersa in una natura artificiale in-sé e per-sé in bilico tra differenti stadi dell'evoluzione sociale. Come Turing si interrogò sulla possibilità di creare computer pensanti, così Kusanagi, eroina androide, si interroga sulla natura di un'intelligenza artificiale che viene posta nel corpo di un robot, di quel fantasma che viene immerso in un guscio, e che vive su una Terra completamente

[30] Cfr. E. Grassi, *Etica e intelligenza artificiale. Questioni aperte,* Aracne, Roma 2020.
[31] Cfr. G. Durand, *Les structures anthropologiques de l'imaginaire*, Dunod, Paris 2006.

digitalizzata, con una popolazione postumana, transumana, robotizzata o con chip sottopelle per potenziare abilità fisiche e cerebrali. Si tratta di un'umanità che deve riscoprire il suo essere nomade[32], quale luogo di transizioni e trattazioni nei corpi che sono enti dinamici, mobili, che possiedono memoria incarnata in cui avvengono trasformazioni e mutamenti di differente intensità per la sua costante riscoperta di coordinate per nuove manifestazioni corporali di natura tecnologica, insettologa, postumana che già sono presenti nel tessuto del vivere collettivo[33].

Nuovamente, l'immaginario diviene lo specchio e le fondamenta di una riflessione etica per l'identità collettiva che può dibattere sulle proprie scelte morali attraverso la condivisione di finzioni di umanità, metafore, narrazioni e leggende che definiscono l'essere e le sue peculiarità. In questi frangenti, l'individuo si trova a dover dialogare con una doppia coscienza: da una parte, quella reale, quotidiana, condivisa con le norme sociali, dall'altra, la proiezione in una finzione che suscita dubbi, incertezze, quesiti, ma che ha proprie strutture e norme che anticipano, a volte, quelle della realtà. Secondo tali prospettive, emerge che l'artificializzazione del sé e dell'ambiente sono necessarie al fine di sviluppare prestazioni che siano contestuali al mutevole corpo del sociale in cui la macchina sviluppa due aspetti cardine per l'essere:

- la macchina quale ente «difensivo contro» il tempo, nella sua possibilità di arginare aspetti lesivi, corrosivi, *sgualcenti* dell'esistenza della persona; che contiene la sua medicalizzazione; che permette la sua stabilizzazione e ne organizza la società, poiché senza di essa non vi sarebbe l'individuo;

[32] Cfr. R. Braidotti, *Soggetto nomade. Femminismo e crisi della modernità* (1994), tr. it. Donzelli, Roma 1995.
[33] Cfr. Id., *Materialismo radicale. Itinerari etici per cyborg e cattive ragazze*, Meltemi, Milano 2019.

- la macchina quale ente «estensivo oltre» il tempo, nella sua formulazione di preveggente che scardina la temporalità e la spazialità; che crea la storia dell'umanità e ne delinea le vie di fuga prospettiche future; è il corpo che si rigetta nel mondo e prosegue in altri altrove attraverso organi esosomatici del sé e, quindi, della società.

Rispondere, quindi, alla domanda «che cos'è persona?» o quali siano le sue caratteristiche in rapporto alla tecnologia e al suo essere automa potrebbe trovare una prima e approssimativa declinazione nell'affermazione che l'essere non è altro che la storia del suo progresso, il riverbero delle sue creazioni, delle sue macchine, dei suoi congegni e degli apparati, di quegli organi che permettono al biologico di essere sociale e al sociale di accrescere il proprio portato nodale nella vita degli individui: il corpo è una costruzione sociale incardinata nell'utilizzo della tecnologia.

La misura dell'essere non è quella del singolo in quanto unità discreta ma diviene cenosi tra una molteplicità di specie, di manifestazioni e di ambienti in cui la persona dovrà contemplare la necessità di allearsi con molte altre creature al fine di costituire luoghi salvifici, rigenerativi e ricreativi[34] e, tra queste creature, vi sono anche le proiezioni artificiali, dalla biotecnologia sintetica, all'informatica, alla robotica sino all'intelligenza artificiale, rappresentando un'ulteriore proposizione di modernità multiple[35]. Nella loro declinazione fisica – attraverso la robotica – e non, le intelligenze artificiali[36] acquisiscono capacità di azione in ambienti

[34] Cfr. D. Haraway, *Chthulucene. Sopravvivere su un pianeta infetto* (2016), tr. it. Nero, Roma 2020.

[35] Cfr. S.N. Eisenstadt, *Multiple Modernities*, in «Daedalus», 129, 1, 2000, pp. 1-29.

[36] L'uso del sostantivo al plurale vuole indicare la differenziazione e la personalizzazione che tale tecnologia subisce entrando in contatto con la molteplicità di soggetti sociali. Per ogni società, cultura, grado di istruzione, specificità d'uso e altri elementi caratterizzanti, ogni intelligenza artificiale si adatta, si plasma, evolve e muta a seconda dei dati che riceve. I big data di cui fruisce sono lo specchio dei nostri sensi, sono

collettivi sino a ora di pertinenza umana: robot impiegati nelle fabbriche, robot attivi nel settore dell'assistenza domiciliare, algoritmi intelligenti operativi nella finanza, sistemi ibridi utilizzati per rafforzare gli approcci educativi o per agire in scenari di guerra, robot e algoritmi che recitano o cantano, sino alla loro manifestazione di agenti sessuali. La società odierna è generata e posta in essere da un movimento continuo e risolutivo operato da intelligenze artificiali e robotica. Presa coscienza della loro immanenza, emerge la riflessione etica per cui i sistemi tecnologici devono essere accompagnati da analisi che permettano una interconnessione tra saperi per l'interno ciclo vitale delle innovazioni proposte e rilasciate nel corpo del sociale, in cui percezione umana, produzione digitale, manifestazione biologica e tecnologica proseguono in una costante integrazione. Dalla visione di Hishiguro[37] in cui umani e umanoidi condivideranno gli stessi spazi sociali, agli studi di Young[38] che denotano sempre meno umani nelle strutture cardine delle società, alle proiezioni di Tegmark[39] in cui le intelligenze artificiali diverranno parte integrante della vita futura dell'essere, sino alle visioni contemporanee di nazioni in cui esistono reti sociali che mettono in stretta dipendenza essere e algoritmo[40] in una politica degli artefatti[41], nella contemporaneità, si può assistere alla crescita costante di agenti intelligenti in modo aurorale nei corpi delle persone e oltre questi, superando il concetto di protesi estensive e acquisendo quello di tecnologie

l'elaborazione di uno specifico bagaglio esperienziale che permette alla tecnologia di interfacciarsi in modo sempre più performativo con il soggetto che la interpella.

[37] Cfr. H. Hishiguro, K.F. MacDorman, *The uncanny advantage of using androids in cognitive and social science research*, in «Interaction Studies», 7, 3, 2006, pp. 297–337.

[38] Cfr. L. Young, *Machine Landscapes. Architecture of the Post-Anthropocene*, Wiley, Hoboken 2019.

[39] Cfr. M. Tegmark, *Vita 3.0. Essere umani nell'era dell'intelligenza artificiale* (2017), tr. it. Raffaello Cortina, Milano 2018.

[40] Cfr. E. Grassi, *Etica e intelligenza artificiale. Questioni aperte*, Aracne, Roma 2020.

[41] Cfr. L. Winner, *Do artifacts have politics?*, in «Dedalus», 109, 1, 1908, pp. 121-136.

«astensive» rispetto all'essere umano. Le I.A. nascono dall'agire dell'individuo, delle sue scelte, pulsioni e desideri ma, sincronicamente, diventando sempre più autonome, attivando nella persona la possibilità di agire per una dislocazione del sé, intesa come protesi intangibile e mentale che esternalizza quanto pensato dal soggetto, trasformandosi in istituzioni, codici[42], e sistemi riorganizzativi di spazi collettivi agiti da *data doubles*[43]. Risulta ancora difficoltoso rintracciare ciò che del soggetto viene registrato, conservato ed elaborato. Tuttavia, tali architetture digitali rappresentano una parte centrale della vita postmoderna. L'artificio intelligente è costantemente attivo nel suo essere collaborativo per la risoluzione dei problemi, nel suo essere essenziale per comunicare con l'ambiente contemporaneo e stabilirne nuovi piani di interazione ed evoluzione, racchiudendo in sé rapporti di potere e di condotta[44]: diviene quel congegno che è in grado di guidarci verso la scoperta del nostro costruirci singolarmente e costituirci socialmente, nella contemplazione della macchina quale speciazione dell'ente che si disegna nell'osservazione dell'essere, della persona che attinge dall'automa – in questo caso ampliandone la visione non solo a una struttura hardware ma alla controparte software – per comprendere le proprie origini e il superamento dei confini e, infine, di una futura convivenza che allo stato attuale potrebbe divenire, immaginificamente, emancipazione dell'umano.

L'esplorazione dell'universo delle innovazioni artificiali reclama la fondante visione di concepire la tecnologia e i suoi organi quali enti retroagenti – dal pensiero di Parsons, che concepiva l'oggetto come manchevole di ragionamento introspettivo, o di Goffman, il quale vedeva negli strumenti meri oggetti di scena

[42] Cfr. L. Lessig, *Code: and other laws of Cyberspace*, Basic Books, New York City 1999.
[43] Cfr. K.D. Haggerty, R.V. Ericson, *The Surveillant Assemblage*, in «The British Journal of Sociology», 51, 4, 2000, pp. 605-622.
[44] Cfr. B. Latour, *Non siamo mai stati moderni* (1991), tr. it. Elèuthera, Milano 2018.

necessari agli umani per migliorare la loro performance – capaci di modificare l'essere e le sue azioni nel momento in cui entrano in contatto anche se indirettamente – riflessione che si colloca nella teoria dell'attore-rete di Michel Callon, Bruno Latour e John Law, indicando la valenza di un modello costruttivista per la realtà sociale, nella quale l'attore-attante è qualsiasi tipo di entità in grado di compiere un'azione con ripercussioni sociali, anche se tale gesto è privo di intenzionalità, contemplando in tal modo anche ciò che è non umano – momento in cui, crolla il fissismo della persona e del suo corpo cedendo il passo a una mutevolezza costante e permeabile dall'agente artificiale esterno. Cade la concezione culturale delle antinomie, in particolar modo quella tra naturale e tecnologico, dato che la tecnologia non è più concepita come esterna al corpo dell'individuo, nella misura in cui i suoi artefatti sono parte di esso e del suo esperire.
La storia dell'umanità è costituita di artifici attraverso i quali ripensare la società, per la disamina dell'esistenza e per l'esternalizzazione delle pulsioni che la costituiscono in cui la protesi definita intelligenza artificiale diviene specchio e lente della persona in grado di donarle una nuova valenza escatologica.

3. *Ramificazioni e proiezioni*

Nelle *Metamorfosi*, non vi è barriera tra divinità e mortali, tra coloro che osservano, seguono, prevedono e manipolano le narrazioni degli altri, provando a immedesimarsi nella pelle degli umani, dei loro sentimenti e del loro quotidiano. Così come, oggi, non vi è più confine tra le intelligenze artificiali, che animano i futuri automi, e le persone che si legano a esse, inconsapevolmente ma immanentemente in un continuo movimento di dati, di informazioni, di storie che saranno elaborate secondo prospettive altre rispetto alla propria.
Nello studio del rapporto tra individuo e macchina, le dimensioni della sovrapposizione e della fusione sono divenute sempre più

preminenti, sin da quando McLuhan[45] definì la tecnologia come prolungamento naturale del sistema nervoso dell'essere umano e rete che ingloba la società e i suoi rapporti o da quando Donna Haraway descrisse, con l'ausilio della figura simbolico-reale del ragno Primoa Cthulhu, la società artificiale come l'era delle connessioni sempiterne, della comunicazione invisibile, delle strutture digitali tentacolari, sistema olistico che ha bisogno della persona come parte non centrale di un ingranaggio molto più complesso, assieme a una pervasività tecnologica che possa sostenere l'essere umano nella tutela del pianeta, così come indicato da Edward Tenner[46] nella sua storia antropologica della scienza che ha esteso a tutte le creazioni tecnologiche e artificiali il potere di plasmare e modellare la struttura dell'individuo sociale e delle sue istituzioni, sino a radicarsi nel suo modo di leggere il presente e di immaginare il futuro. Bisogna abitare la catastrofe come condizione umana: il baratro è un punto di vista della fragilità che può agire sulla coscienza e sulla politica nell'architettura di un'etica delle soglie: la crisi del soggetto dovrebbe essere percepita come esperienza di attraversamento verso un controspazio inesplorato e portatore di meraviglia.

La società artificiale può essere considerata come l'ambiente di questa crisi della persona e come lo specchio di una sintesi umana verso la costruzione di un proprio alter-ego automatizzato, la realizzazione di quei robot intelligenti che, grazie all'implementazione degli algoritmi di apprendimento su organismi meccanici, siano in grado di sostituirsi alla vita biologica nella responsabilizzazione e nella gestione dei problemi più complessi e che richiedono celerità nelle decisioni, divenendo i collaboratori perfetti verso un futuro incerto, in cui potrebbe manifestarsi

[45] Cfr. M. McLuhan, *Gli strumenti del comunicare* (1964), tr. it. Il Saggiatore, Milano 2015.
[46] Cfr. E. Tenner, *Our Own Devices: The Past and Future of Body Technology*, Vintage Books, New York City 2004.

quella singolarità che potrebbe renderli autonomi e semicoscienti. L'essere umano viene rappresentato quale risultato complesso delle sue invenzioni, dei suoi strumenti, delle sue tecnologie che si sono susseguite lungo la linea storica, la quale è essa stessa la manifestazione della tecnologia come scelta dell'umano di vivere la propria esistenza ed è per questo che l'essere umano è il primo automa, è esso stesso la sua utopia ma reale:

- l'essere è di per sé tecnologico vivendo costantemente di espedienti, di elaborazioni, di congegni, sin dalla sua comparsa, dalla sua nascita con l'uso della parola, il veicolo comunicativo per eccellenza che, nel suo evolversi, è divenuto segno, scrittura, dati, algoritmo;
- per perpetuare la sua esistenza sulla Terra, l'individuo ha costante necessità di protesi, elementi che siano in grado non solo di potenziare le sue prestazioni ma di esternalizzare l'intangibilità del suo pensiero;
- nella sua dimensione biologica, l'epidermide è la manifesta rappresentazione della metamorfosi, del mutamento che permette al corpo di essere un soggetto nomade in cui esistono molteplicità di organismi e ambienti plurimi;
- la persona è tale nel suo fare, nel suo produrre, quindi, nel suo pensare, azione costitutiva cardine del suo vissuto mnesico che è storia.

Considerando tali aspetti, emerge la centralità della narrazione, del dialogo, del dato che, a differenza di quanto comunemente immaginato non è liquido, non rappresenta una società della fluidità, bensì è elemento grezzo, solido, tellurico che si incide nella parabola del cammino umano. Le macchine apprendono a ogni passo della persona, la persona lascia tracce, impronte, a ogni suo gesto e fa in modo che si imprimano nella storia del progresso, stratificandole, solidificandole e costituendole quale scheletro di una iperidentità estensiva del sé interiore, innervandole nel continuum delle scosse polimorfiche della società

contemporanea. In questa visione, assurgono intelligenze artificiali che sono mimesi della persona nel loro stadio infantile ma che, grazie alla relazione con la persona, crescono e costituiscono un altro dal quale apprendere.

La persona è la storia delle sue creature. È narrazione di metamorfosi.

EDMONDO GRASSI è Ricercatore di Sociologia presso l'Università Telematica San Raffaele di Roma

edmondo.grassi@uniroma3.it

GIANLUCA GIANNINI – ANTONIO PESCAPÈ
[LUCA LO SAPIO]

AI E FUTURO DI *SAPIENS* TRA NUOVI ORIZZONTI E ANTICHI TIMORI

ABSTRACT: ***AI AND FUTURE OF SAPIENS BETWEEN NEW HORIZONS AND ANCIENT FEARS***
AI in its various forms and articulations represents a range of new opportunities and strategies that our species can use to improve its condition. Often the scenarios presented in association with AI speak of a humanity threatened by machines or, on the contrary, of a utopian future in which man and machine will be completely hybridised in a harmonious manner.
In this interview, a moral philosopher, Professor Gianluca Giannini, and Antonio Pescapè, professor of computer engineering, will try, each from their own perspective, to answer some of the key questions emerging in the field of artificial intelligence, on the one hand avoiding sensationalist journalistic exaggerations, and on the other hand highlighting that the real challenge posed by intelligent machines lies in the new ways in which humans will narrate themselves and their possibilities of existence.

S&F: Ci sono numerose definizioni di Intelligenza Artificiale. Se rivolgiamo la nostra attenzione a "intelligenza", ad esempio, troviamo almeno una sessantina di definizioni[1]. D'altro canto, anche la definizione di artificiale presenta delle difficoltà, dal momento che spesso viene strutturata in opposizione a naturale. Dunque, ci sono evidenti difficoltà sia a definire che cosa sia l'intelligenza, sia nel definire che cosa sia l'artificiale. Ritenete sia possibile trovare un comun denominatore in grado di mettere d'accordo le diverse prospettive?

A.P.: Sicuramente è più che complesso, se non impossibile, trovare un comune denominatore in grado di "accordare" le diverse prospettive, soprattutto perché risulta piuttosto difficile trovare una definizione unica di Intelligenza Artificiale. Volendo

[1] S. legg, M. Hutter, *Universal intelligence: a definition of machine intelligence*, in «ArXiv: 0712.3329 [cs.AI]», https://doi.org/10.48550/arXiv.0712.3329.

partire da un punto di vista tecnologico, quando si parla di Intelligenza Artificiale ci si riferisce a un ampio insieme di tecnologie e, di conseguenza, a uno scenario altrettanto ampio nel quale trovano spazio applicazioni molto diverse tra loro. Oggi quando discutiamo *lato sensu* di Intelligenza Artificiale parliamo di aree tematiche quali *Computer Vision, Pattern Recognition, Automated Reasoning, Game theory, Logics, Multi-Agents, Fuzzy systems, Knowledge representation, Speech Recognition, Natural Language Processing, Machine Learning, Deep Learning, Cognitive Robotics*. Tutte queste aree tematiche trovano la loro collocazione (fonte Scimago) in più di 165 riviste internazionali e sono state alla base di più di 900 conferenze internazionali di settore; si può facilmente intuire come il termine Intelligenza Artificiale sia, quindi, un termine *"umbrella"* sotto il quale si trovano cose molto diverse tra loro. A quanto finora detto va aggiunta un'altra importante distinzione, quella tra *Strong AI* e *Weak AI*. Con *Strong AI* ci si riferisce a scenari nei quali le macchine sono effettivamente in grado di pensare ed eseguire compiti da sole, essere veri e propri *alias* di un essere umano. In questa categoria sono ricompresi il *Turing Test* (Turing), il *Coffee Test* (Wozniak), il *Robot College Student Test* (Goertzel), l'*Employment Test* (Nilsson), il *Flat Pack Furniture test* (Tony Severyns), il *Mirror Test* (Tanvir Zawad). Con *Weak AI* ci si riferisce invece a scenari in cui ci si concentra su un compito ristretto, riferendosi al fenomeno per il quale le macchine che non sono troppo intelligenti per svolgere il proprio lavoro possono essere costruite in modo tale da *sembrare* intelligenti, simulando solo la funzione cognitiva umana. Questo tipo di AI "debole" agisce in maniera "semplice" ed è vincolata dalle regole che le vengono imposte senza mai poter agire oltre tali regole. Al di là di queste due definizioni, *Strong AI* e *Weak AI*, la tassonomia dell'Intelligenza Artificiale si specializza ulteriormente in altri quattro tipi:

· AI di Tipo 1: si riferisce a una AI reattiva e specializzata in una singola area. Ad esempio, la redazione e la revisione di contratti di finanziamento di natura commerciale.

· AI di Tipo 2: si riferisce a una AI che ha memoria o "esperienza" appena sufficiente per prendere decisioni adeguate ed eseguire azioni appropriate in situazioni o contesti specifici.

· AI di Tipo 3: si riferisce a una AI che ha la capacità di comprendere pensieri ed emozioni che influenzano il comportamento umano.

· AI di Tipo 4: si riferisce a una AI come quella tipicamente rappresentata in TV (nelle rappresentazioni cinematografiche/nella fantascienza). Le macchine che utilizzano questo tipo di intelligenza artificiale sono autocoscienti, super-intelligenti, senzienti e coscienti.

G.G.: Anzitutto consentitemi di fare due considerazioni di "contesto". Una apparentemente più complessiva e che parte da una

domanda di fondo: "perché tanto interesse per l'AI?". La risposta è sotto gli occhi di tutti: come ci è stato appena ricordato da Antonio, il termine *umbrella* cela letteralmente un pluriverso. Lo si ritrova, il senso complessivo di questo pluriverso, persino in approcci meramente illustrativi e/o divulgativi *a buon mercato*. È ben descritto, ad esempio, dalla docuserie *The Age of A.I.* andata in onda su YouTube, che nel concentrarsi su tutta una progressione di dettagliate ricostruzioni di situazioni e contesti (dall'occasione domestica all'ambito delle fabbriche di automobili, passando per le sale operatorie, arrivando alla *driverless car*), mostra chiaramente come la tecnologia *smart* abbia già plasmato il mondo in cui ci muoviamo attimo per attimo. Va infatti detto che l'AI ha una potenzialità di interazione, trasformazione e rideterminazione del contesto che può essere, potrà essere, unico nella storia di *sapiens*. I programmi di AI stanno infatti già mutando e trasformando ampi campi dell'agire umano: dall'economia a quelli più *banalmente* del quotidiano. Si pensi agli assistenti vocali (sempre più diffusi nei dispositivi elettronici); alle ANN (*Artificial Neural Network*) ovvero le reti neurali artificiali che setacciano i Big Data in cerca di modelli che serviranno a prevedere tendenze, gusti, desideri… E così via. Anzi forse si può dire, con buona dose di certezza, che l'ombrello omnicomprensivo che sottende l'acronimo AI, al momento e in linea di principio, concerne forme applicative d'*intelligenza* automatica limitate a compiti ben definiti in ambiti ben precisi che procedono, però, in direzioni plurime al punto da investire tutta la sfera dell'agire umano.

La seconda considerazione è legata proprio ai termini in questione. AI, *Intelligenza Artificiale*, infatti, è una sorta di ossimoro: intelligenza, cioè un qualcosa di naturale coniugato ad artificiale, che letteralmente è: non-naturale. Che intendiamo allora con AI? In realtà se ci riferiamo agli albori[2] «far sì che una macchina agisca con modalità che sarebbero definite intelligenti se un essere umano si comportasse allo stesso modo», ci troviamo già in un incavo problematico. È infatti un'espressione *controfattuale*, perché anzitutto è del tutto evidente che quello cui si mira non ha nulla a che vedere con il pensiero bensì con il comportamento. *Se l'essere umano si comportasse… sarebbe definito intelligente*: non significa che la macchina sia intelligente o che addirittura stia pensando. In questo allora, forse, e me ne rendo conto in via del tutto esemplificativa, soccorrono le definizioni per restringere il perimetro. In linea di massima, io mi riferisco a due direttrici definitorie per poi arrivare a una terza che potrebbe fungere in qualche modo da "collante". E quindi una definizione *teorico-disciplinare*, per cui l'AI è quella «disciplina, appartenente all'informatica che studia i fondamenti teorici, le metodologie e

[2] A. Turing, *Computing machinery and intelligence*, in «Mind», LIX, 236, pp. 433-460.

le tecniche che permettono di progettare sistemi hardware e sistemi di programmi software capaci di fornire all'elaboratore elettronico delle *prestazioni* che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana», da cui il termine chiave in cui si risolve il "comportamento" è il rassicurante (?) "prestazione". Poi, in seconda battuta, una definizione *applicativa*, per cui «l'obiettivo non è quello di replicare o simulare l'intelligenza umana» bensì «quello di *riprodurre* o *emulare* l'intelligenza umana, in quanto non vi è alcun motivo a priori che impedisca che *talune prestazioni* dell'intelligenza umana possano essere fornite anche da una macchina». È del tutto evidente che nel caso dell'emulazione le prestazioni 'intelligenti' sono ottenute utilizzando dei meccanismi propri della macchina, eventualmente differenti da quelli ipotizzabili per l'uomo, ma appunto tali da fornire *funzioni/risultati* qualitativamente equivalenti e quantitativamente superiori a quelli umani. Qui nell'intreccio riproduzione-prestazione, emulazione-funzioni-risultati si gioca la ricaduta di ordine applicativo giacché l'obiettivo dell'AI è costruire entità intelligenti *capaci-di*. Cioè macchine in grado di calcolare come agire in modo efficace e sicuro in un'ampia varietà di situazioni. Ovviamente, in maniera diretta e/o indiretta, avendo l'uomo come fine. Tuttavia qui, mi permetto, è possibile inserire quella che considero una definizione *filosofica* (e spero che Antonio non se ne abbia a male) di AI. Se quello che possiamo circoscrivere come modello standard di AI riguarda l'agire razionale, in altri termini il fine è quello di metter su un'entità, un soggetto agente intelligente, cioè un soggetto agente intelligente *ideale* che intraprenda in ogni situazione la migliore (l'aggettivo è meramente descrittivo) delle azioni possibili, abbiamo già posto le basi per un qualcosa di assolutamente inedito.
Ovvero di un peculiare e diverso *artificium* che fondi la sua singolarità, in qualche modo emulativa di quella dell'uomo, in una sorta di autonoma creatività sorretta da un paradigma costitutivo *ingaggio-esperienza-apprendimento-fare* tipica del vivente. Questo peculiare *artificium* sarebbe in tutto e per tutto un'alterità assoluta che, sfuggendo alle dinamiche interpretative passate rispetto ai canoni tradizionali di 'macchina', si staglierà per l'uomo come un vero e proprio interlocutore-altro. E questo lo dico senza cedere alle tentazioni dei tecnofobici e/o, al contrario, dei tecnoentusiasti.

S&F: Una delle dicotomie più influenti nella storia dell'intelligenza artificiale è, come ci ha appena ricordato il Prof. Pescapè, quella tra una Strong AI e una Weak AI o, in altri termini, un'intelligenza artificiale cognitivo-produttiva e un'intelligenza artificiale ingegneristico-riproduttiva. Le linee

di ricerca più recenti sembrano aver puntato più su quest'ultima che sull'intelligenza artificiale forte. Ha forse ragione, dunque, Floridi[3] quando sostiene che per comprendere l'intelligenza artificiale bisogna partire dalla constatazione del divorzio tra intelligenza e capacità di agire? In tal senso l'AI sarebbe più una nuova capacità di agire che una riproduzione dell'intelligenza umana.

A.P.: Concordo con le linee di ricerca recenti: anche io nel mio laboratorio sto portando avanti approcci tipici della *Weak AI*, applicati ai dati di rete così come ad altri domini applicativi come quelli della salute, ambito nel quale l'AI può fare molto. E provo a dire qui, in parole semplici, il principio-base che guida gli approcci di *Weak AI* e il perché oggi gli sforzi maggiori, in termini sia di ricerca sia di innovazione, si profondono qui e non nella *Strong AI*. L'intuizione di fondo è quella che, come nel caso di un bambino, si può fare (o provare a fare) qualcosa non perché la si sappia fare bensì perché si è osservato qualcuno (un adulto, ad esempio) fare quella o una cosa simile (che è il darwininsmo della macchina?). Questa intuizione negli approcci di *Weak AI* si sostanzia nella fase di apprendimento (training) – durante la quale si costruisce il modello addestrando una rete – e in quella di uso (testing) – durante la quale si valuta la capacità del sistema addestrato di operare su dati diversi da quelli utilizzati nella fase di addestramento. Il *Machine Learning* (ML) – e il più recente *Deep Learning* (DL) – fa parte di questa categoria. Sebbene approcci di *machine learning* fossero stati già proposti alla fine degli anni '50, l'*hype* a cui abbiamo assistito e a cui stiamo assistendo è legato alla combinazione di tre importanti condizioni: disponibilità di considerevole potenza di calcolo a costi contenuti, disponibilità di considerevole spazio di memorizzazione a costi contenuti (entrambi abilitati anche dal paradigma cloud), infine disponibilità di quantità di dati mai viste prime. Grazie a queste tre condizioni, gli algoritmi di ML e di DL oggi cominciano ad avere performance tali da poter prevedere un loro utilizzo massivo in ogni attività quotidianamente.

G.G.: Mi sembra di aver cominciato a dire qualcosa in tal senso, laddove ho ricostruito il perimetro entro cui calare questo macro-fenomeno che porta le insegne "AI" e ho poi accennato al paradigma interpretativo cui in qualche modo riferirsi: quello di *alterità assoluta*. La capacità di agire è indefettibilmente la prerogativa dell'*artificium* che è progettato e realizzato quale punto di caduta di un insieme di tecniche computazionali ispirate dal modo in cui gli esseri umani utilizzano il proprio sistema nervoso e il

[3] Cfr. L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità sfide*, Raffaello Cortina Editore, Milano 2022. Dello stesso autore cfr. anche il meno recente *Infosfera. Etica e filosofia nell'età dell'informazione*, Giappichelli Editore, Torino 2009.

proprio corpo per sentire, imparare, ragionare e agire. È del tutto evidente, e proprio attraverso l'uso di modelli computazionali, che l'*artificium*, a partire dal proprio "linguaggio costitutivo" e dalla propria materiale difformità, opererà – e non a caso utilizzo questo verbo per sottolineare ancora una volta che trattasi di questione legata all'*agere* e non all'*intelligere* – in maniera del tutto diversa dall'uomo. È questo, in maniera del tutto avalutativa, che mi fa dire che si apre lo spazio di un altro correlazionale. Il richiamo alla *Strong AI*, che prevede un automa cosciente della sua intelligenza e capace quindi di poter agire in modo indipendente, e la *Weak AI*, che prevede invece un automa limitato a un'attività specifica per cui è progettato senza alcuna possibilità di espansione autonoma – e sappiamo che al giorno d'oggi per la difficoltà di costruire delle vere *Strong AI*, con totale autonomia e capacità di adattamento, si è dato maggior credito all'approccio *Weak*, che rappresenta la maggior parte delle AI con cui abbiamo a che fare nel quotidiano – non mi sembra dirimente segnatamente a questo aspetto che pongo in rilievo. Del resto, mi sembra chiaramente emerso già dalla prima risposta.

S&F: ***Alcuni studiosi (ma non solo) impegnati nel campo dell'AI – da Bostrom a Stephen Hawking passando per Elon Musk e Max Tegmark – mettono in guardia dai suoi potenziali pericoli. Credete che si tratti di meri scenari distopici o immaginate il futuro della nostra specie realmente minacciato dalla diffusione dell'AI?***

A.P.: L'utilizzo massivo di algoritmi di *Weak AI* apre un tema importante legato a una serie di questioni cruciali quando si utilizzano approcci di *Machine/Deep Learning*; questioni che oggi rappresentano la frontiera della ricerca in questo settore. In particolare, l'AI deve (quanto più possibile) essere etica (provando a capire cosa significa), trasparente, *trusted* (affidabile), antropocentrica, inclusiva, responsabile e neutrale (nel senso che non operi distinzioni, ad esempio di sesso o di razza o di religione). A tal fine vi sono due questioni di estrema importanza: da un lato, abbiamo il *bias* introdotto dai dati di addestramento nella decisione presa dall'approccio di AI, dall'altro l'interpretabilità del comportamento, del risultato e della predicibilità del comportamento. Per comprenderci meglio possiamo prendere come esempio il caso della corte americana che ha utilizzato un algoritmo di AI – addestrato su dati relativi a reati di una particolare area di Los Angeles – e nel fare screening si è scoperto che l'algoritmo soffriva di un "bias di razza". Oppure possiamo richiamare il caso della AI che si rivolge in maniera discriminatoria nei confronti delle donne perché "addestrata" con dati (quelli prodotti dagli uomini) affetti da bias di genere; è chiaro, quindi, che ci potremmo trovare davanti

a un complesso di decisioni che vengono prese in una certa maniera perché condizionate da dati polarizzati dal punto di vista, ad esempio, etnico o sociale. Mentre per quanto concerne la questione dell'interpretabilità fino a quando non saremo in grado di comprendere i meccanismi e le motivazioni per le quali un algoritmo di AI prende una decisione non potremo mai utilizzarlo in contesti dove il risultato (responso, verdetto) richiede una "certificazione" frutto di una interpretabilità del processo di decisione.

Molte sono le preoccupazioni nei confronti dell'AI, tra cui quelle di Elon Musk, Bill Gates, Steve Wozniak, e del compianto Stephen Hawking (in alcuni casi le posizioni arrivano a sfiorare forme di tecno-luddismo). Ma sono molti anche i tecnoentusiasti. Io non mi iscrivo a nessuna delle due categorie. Il CEO di Google Sundar Pichai – sollecitato a rispondere riguardo l'impatto dell'Intelligenza Artificiale sul lavoro e sulla paura che la società nutre nei confronti di questa tecnologia – l'ha paragonata in termini di portata e importanza a scoperte come il fuoco; ed è proprio così. La scoperta del fuoco ha cambiato la storia dell'umanità: la paura, l'ingegno, la fantasia e l'incoscienza hanno reso possibile l'utilizzo del fuoco, che da iniziale pericolo si è rivelato il più prezioso strumento per il progresso dell'umanità. Dal fuoco per illuminare, per riscaldare, per difenderci. L'evoluzione e il progresso portano all'oggi e alla gestione del fuoco con le sue regole, le prescrizioni, e le misure di sicurezza, con professionalità specializzate al controllo e alla gestione del fuoco. E, come è stato per il fuoco, così oggi dobbiamo imparare a gestire l'Intelligenza Artificiale Generale. A tal proposito, il Regolamento Europeo sull'AI dell'aprile 2021 (a mio avviso tardivo) propone, tra le varie, una classificazione delle applicazioni di AI: Vietate, Alto Rischio, Rischio Limitato, Rischio Minimo. E, per capirci, nella classe delle applicazioni Vietate troviamo:

- l'uso di sistemi di IA che distorcono il comportamento di una persona attraverso tecniche subliminali;
- l'uso di sistemi di IA che sfruttano qualsiasi vulnerabilità in modo da causare o essere suscettibili di causare danni fisici o psicologici;
- l'uso di sistemi di IA che consentono la valutazione/classificazione dell'affidabilità di persone fisiche mediante l'attribuzione di un punteggio sociale (*social score*)

L'economia predittiva abilitata dall'IA che influenza gli esseri umani cambiando il loro processo decisionale e, di conseguenza, il loro comportamento è accettabile? (seti è piaciuto questo vino potrebbe piacerti anche quest'altro, e lo si acquista). Ma l'economia predittiva diviene poi informazione predittiva e ciò permette di moltiplicare la capacità di fare propaganda (abbiamo visto cosa è successo durante il COVID-19, stiamo vedendo che cosa

sta accadendo nella discussione sulla guerra Russa/Ucraina): la novità è che si può mirare agli individui, si genera una informazione personalizzata: questa è probabilmente la più grande minaccia per la stabilità della società e della democrazia, per come la conosciamo oggi.

G.G.: Per ora, almeno, preferirei non entrare nel merito di formule ammalianti come "l'AI deve essere etica… etc.". Al di là dalle seduzioni di cui dicevo prima tra tecnofobia e tecnoentusiasmo che animano solitamente i dibattiti gazettieri, mi limito a fare una considerazione che ritengo preliminare se si ragiona nell'alveo del rapporto tra "scienza e filosofia" anche segnatamente al macro-fenomeno AI: ogni nuova tecnologia, sin dai tempi del Cannocchiale di Galileo Galilei ha sollevato e solleva dubbi, preoccupazioni, resistenze, incredulità e proiezioni d'ogni tipo non tanto per l'*ente* in sé ma, ovviamente, segnatamente alle ricadute che può avere. Sovente, invece di mettere in forma e affrontare le questioni che potrebbero/dovrebbero individuarsi, ci siamo lasciati (e ci lasciamo ancora) influenzare da una sorta di congenita diffidenza, specie per cambiamenti che potrebbero apparire troppo rapidi ma di cui, in fondo, siamo pur sempre gli attori sintagmatici, anche se tendiamo a dimenticarcene. Non v'è dubbio che, in particolar modo quelle in atto da oltre un quarto di secolo (ma in realtà potremmo pre-datare di qualche decennio, se si pensa alle tecnologie del Bios), le rivoluzioni tecniche e tecnologiche presentino risvolti complessi e controversi, prefigurando scenari prossimi e non troppo a venire di radicali modificazioni dell'antropo-sfera. Tuttavia, una cosa va anzitutto rilevata e sottolineata, altrimenti ogni discorso, già dalle sue premesse, ricade in una chiacchiera di senso comune: le rivoluzioni tecniche e tecnologiche non sono mai estranee al contesto antropico in cui si determinano e si sviluppano. Non solo ne sono il portato ma, finanche, vi interagiscono in maniera profonda, al punto da trasfigurare e rideterminare il contesto stesso e, quindi, come sempre e come da sempre, l'umano stesso che non solo non è mai stato un qualcosa di dato una volta e per tutte ma, anzi, ha rinvenuto la possibilità persistentiva più propria nella sua costante trasfigurazione. Di fatto, si può dire che le rivoluzioni/evoluzioni tecniche e tecnologiche non sono mai neutre, indifferenti e imparziali. Finanche nei termini di accessibilità. Per cui, rispetto al taglio della sua domanda, mi preoccuperei più della detenzione/disponibilità e, quindi, dell'uso potenziale, di determinate "risorse". Mi sembra, in quest'ottica, che imprenditori come Musk, Zucherberg etc. vadano ascoltati alla stessa stregua di tutti gli altri "addetti ai lavori" ricordando, però, che non sono estranei al contesto ma, anzi, per certi versi *sono* essi stessi il contesto.

S&F: ***L'AI rappresenta una manipolazione di terzo livello, vale a dire una tecnologia che riesce, in maniera autonoma, a mettere in comunicazione artefatti prodotti da sapiens, senza che quest'ultimo debba essere coinvolto, in qualche misura, nel loro funzionamento e nella loro supervisione. Ritenete che questo possa determinare una marginalizzazione di sapiens e un rifluire di quest'ultimo all'interno di una dimensione costituita da agenti artificiali che non hanno più sapiens stesso come terminus ad quem?***

A.P.: Siamo ancora lontani da scenari di questo tipo e dobbiamo anche evitare che i tecnoentusiati di turno o gli appassionati di film ambientati nel futuro tecnologico confondano la realtà con una plausibile finzione, sebbene le finzioni sembrano sempre più vicine alla realtà che stiamo vivendo. Sicuramente c'è però anche l'idea dell'umano intrappolato nella macchina e tutta la lunga narrazione sul preoccupante ruolo del progresso (spesso solo sviluppo o confuso come tale), demonizzato in funzione di una centralità dell'uomo. C'è poi il rapporto tra *homo sapiens* e *homo technologicus*, un rapporto che sembra sostanzialmente destinato a essere sempre più simbiotico. Nel 1973 un'azienda tedesca, la Kuka, entrò nel mercato della robotica con una macchina pionieristica e la chiamò Famulus. La scelta del nome è quantomai interessante perché all'epoca, l'idea dominante era di una macchina a completo servizio dell'uomo; oggi, invece, i software di AI – in particolare quelli delle applicazioni più avanzate del *Machine Learning* – richiedono una "mediazione" dell'umano, ossia della conoscenza e dell'esperienza dei *sapiens*, in base al "principio" dell'uomo nel loop, dove l'esperienza umana è centrale nei processi di apprendimento dell'AI. La guerra tra Russia e Ucraina ci ha mostrato come la tecnologia e in particolare l'intelligenza artificiale sia centrale sia nella guerra "materiale" sia in quella immateriale della "cyber guerra". E a oggi in campo militare le intelligenze artificiali prevedono che sia sempre l'uomo a prendere l'ultima decisione: non solo quindi "in the loop", ma "at the end of the loop".

G.G.: Visto che in qualche modo insiste sul punto…
Non v'è dubbio che la macro-questione etica che sottende ogni tipo di riflessione (anche e soprattutto filosofica) ruota attorno all'interrogativo: "AI così potenti potrebbero a loro volta progettare macchine ancor più intelligenti di loro stesse e di *sapiens* ragion per cui *sapiens* si troverebbe a condividere il pianeta con una nuova *specie* (artificiale) che a tal punto sarebbe dominante?".
Ora, al di là di retro-pensieri più o meno fantascientifici, la questione non è banale tant'è che numerose personalità del mondo

della scienza, della tecnologia e della ricerca a tutto tondo (Hawking, Wilczek, Bill Gates, Wozniak, lo stesso Musk come veniva ricordato prima) hanno espresso cautele e si sono fatti promotori di una lettera aperta (*Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter*) in cui, tra le altre cose, è sostenuta da un lato e con forza la ricerca sull'AI avvertendo, però e contestualmente, sulla necessità che l'AI *faccia quel che le chiediamo di fare*. Ovvero: «a causa del grande potenziale dell'IA, è importante ricercare come sfruttarne i vantaggi evitando potenziali insidie». Sostanzialmente, e come rilanciato in sede di UE, la ricerca e lo sviluppo dell'AI dev'essere comunque "umanocentrica". Posta così la questione sembra essere risolta, e in fondo la "pezza a colori" di stampo etico sembra fungere da buon palliativo. In realtà a me la questione, che tra le altre cose ruota attorno al posto dell'uomo *nel Loop*, qui e ora ma anche e soprattutto domani, mi pare si presenti come epocale, al punto da far deflagrare gli approcci etici tradizionali. E la formula "umano-centrismo" sembra risolvere tutto ma, in realtà, è assolutamente vuota.
Ripartiamo allora dal problema che ci si pone in maniera ritornante e che, appunto, il presunto fine eticamente assoluto dell'umanocentrismo non risolve in premessa, se non è ovviamente riempito di contenuti. La questione che, in quanto tale è capitale, è quella di raggiungere un accordo tra le nostre reali preferenze e l'obiettivo posto alla/nella macchina. Cioè, e in maniera stringente, il problema che ci si pone non è banalmente mezzi-fine bensì, e più radicalmente, dell'allineamento dei valori. I valori e dunque gli obiettivi affidati alla macchina devono essere allineati a quelli dell'uomo. E questa è, fondamentalmente, questione filosofica, perché in gioco vi sono valori e obiettivi di ànthropos. Presente e prossimo-venturo. Il *piccolo* problema è che rispetto a questo, ancor prima di fantasticare sulle *potenzialità* dell'*artificium*, dovremmo esser consapevoli che i valori sono storici e non assoluti e metafisici. Dov'è il punto allora… Trovare una soluzione al cospetto di questioni non lineari, ovvero autentici dilemmi – che è letteralmente un problema decisionale tra due imperativi morali, nessuno dei quali è del tutto preferibile – segnatamente alla coniugazione tra pratica e valore. La questione dell'allineamento dei valori cui accennavo. L'espressione che mi piace molto, *Human in the Loop*, fa riferimento a questo snodo qui e non tanto, e semplicisticamente, a che tipo di postazione operativa l'uomo ha e avrà nella sua relazione-coniugazione con la macchina. Rispetto a questo, approcci tradizionali – ovviamente parlo dell'etica e della filosofia morale – non reggono più. Sono, a mio modo di vedere, vetusti e assolutamente non all'altezza della sfida.

S&F: ***Solitamente le problematiche etiche dell'AI sono esaminate a partire da due modelli normativi: l'etica consequenzialista (in***

particolare in una delle versioni dell'utilitarismo) o l'etica deontologica, nella versione kantiana o nella versione pluralista, ad esempio il principialismo di Ross, Beauchamp e Childress (vale a dire più principi da usare in situazione e attraverso l'applicazione di criteri di bilanciamento). Ritenete che questi approcci siano sufficienti? O credete piuttosto che di fronte ai rapidi avanzamenti delle tecnologie digitali e dell'AI siamo costretti ad adottare forme, più o meno sofisticate, di una qualche '"etica della situazione" non sussumibile in principi pre-stabiliti?

A.P.: *The Moral Machine Survey* (pubblicato su Nature nel 2018)[4] ha mostrato – nel caso delle auto a guida autonoma – come sia praticamente impossibile pre-programmare con un approccio *Top-Down* una AI di un veicolo autonomo con una disciplina morale. Qualunque essa sia. Sia essa una disciplina di natura consequenzialista/utilitarista, fare cioè ciò che provoca la minore quantità di danni aggregati (e la maggiore quantità di benefici aggregati); sia essa una disciplina costruita intorno a una qualche forma di imperativo categorico, dove si fissa *ex ante* una regola pratica, indipendentemente dagli scopi e dalle circostanze specifiche. 40 milioni di decisioni in dieci lingue raccolte da milioni di persone in 233 paesi hanno mostrato come la morale non è un concetto universale: ad es., i Paesi orientali considerano meno accettabile di noi Europei – in un incidente causato da un'auto a guida autonoma – "sacrificare" un anziano per salvare un bambino, mentre quelli di cultura ispanica e francese sono più orientati a salvare le donne. Ne viene comunque fuori un paradosso: nella maggior parte dei casi emerge che gli utenti preferirebbero che un'auto a guida autonoma salvasse i pedoni a scapito del suo equipaggio, ma anche che non comprerebbero mai un'auto a guida autonoma programmata in questo modo. E quindi mi viene da concludere citando Megatron Transformer di Nvidia, una intelligenza artificiale di Oxford a cui è stato posto il quesito: "L'intelligenza artificiale sarà mai etica?" La risposta della "macchina", di Megatron Transformer, è stata: "l'AI è uno strumento e, come ogni strumento, viene utilizzato nel bene e nel male. Non esiste una buona intelligenza artificiale. Solo umani buoni e cattivi."

G.G.: È quello a cui accennavo in precedenza: gli approcci – diciamo così – tradizionali a cui allude anche nella domanda, sono letteralmente evaporati e perciò decisamente inutilizzabili e forse è arrivato il momento di sbarazzarcene e di liquidare una volta e per tutte le raccapriccianti pretese di ancorare la morale al naturale, facendo finanche della morale qualcosa di naturale…

[4] E. Award, S. Dsouza, R. Kim *et al.*, *The moral machine experiment*, in «Nature», 563, 7729, 2018, pp. 59–64.

Vediamo di capirci: siamo abituati – immagino che ad Antonio sia già capitato di imbattersi in filosofi morali in tal senso orientati decine e decine di volte – a tipologie d'approccio al dilemma etico di stampo *verticale* e/o *orizzontale*. Lo sperimentiamo ancora nel solco di quel sapere senza statuto epistemologico qual è la Bioetica. Ci si riempie la bocca, si monopolizzano dibattiti pubblici d'ogni tipo e sorta, riproponendo questo schema bipolare verticale-orizzontale di fatto non sortendo effetto alcuno nei termini di incidenza e incisività. Se da un lato, infatti, l'approccio verticale – fuori spazio-tempo da circa cinque secoli – si riferisce a un codice di valori metafisicamente fondato, quindi prescrittivo e indiscutibile, in cui il valore trova la sua legittimazione in un'istanza extra-fisica (Dio, mondo delle idee etc.), dall'altro, quello orizzontale, sembra presentare aspetti di maggiore efficacia ed efficienza. Non mira infatti a imporre valori eterni e si dimostra solitamente attento alle esigenze umane che tengano conto delle condizioni e delle trasformazioni storiche. In quest'ottica rientra, ad esempio, la cosiddetta etica ingegneristica, oppure l'etica *delle* nuove tecnologie e, comunque, ogni piattaforma riflessiva che accompagna un complemento di specificazione al sostantivo "etica". In questo solco, che potremmo definire a-religioso e che, indefettibilmente, è quello in cui si gioca il dibattito etico più serio e problematico come pure si accennava, Bentham e Kant sono i due *campioni* di riferimento. Ovvero come sintetizzato dal Prof. Pescapè, da un lato, il *conseguenzialismo/utilitarismo* in ragione del quale è utile ciò che ha come conseguenza la più grande felicità del maggior numero di persone e che tende a fare dell'etica persino una scienza esatta alla stregua della matematica. Dall'altro, appunto, il *de-ontologismo* in ordine al quale un atto morale è un atto che sarebbe giusto per qualsiasi tipo di persona, in circostanze simili a quelle nelle quali un soggetto si trova nel momento di eseguirlo. Se poi tutto questo è addirittura sotteso dalle formulazioni dell'imperativo categorico, e cioè dall'*operare in modo che la massima della volontà possa sempre valere in ogni tempo come principio di una legislazione universale* fino all'*agisci in modo da trattare l'umanità sia nella tua persona sia in quella di ogni altro sempre come fine e mai come mezzo*, sembra proprio che abbiamo dato un contenuto all'umano-centrismo di cui prima. Questo approccio sembra infatti mettere l'uomo al centro del mondo morale perché lo considera come noumeno e lo esorta a non contemplare mai soltanto la parte fenomenica di se stesso ma, anche, quella razionale. Solo così l'uomo può evitare di far danno a se stesso e agli altri.
Questo è l'orizzonte etico-filosofico nel quale siamo abituati a muoverci. Primo problema: nessuna filosofia, o teoria etica, è assumibile in tutte le circostanze, al massimo possiamo parlare di criteri orientativi. Secondo problema: come ho già detto, i valori sono storici (e storicizzabili). Il valore è l'attività della

valorazione che è sempre in una collocazione spazio-temporale. È una tensione a far-valori, ma il valore è ciò che qui e ora vale attraverso l'azione. Quello che oggi per noi è valore, domani potrebbe essere disvalore e, soprattutto, disgiunto dall'azione è niente più che una dichiarazione d'intenti. Terzo problema: da quanto ci siamo detti sino a ora, e come accennava nella domanda, l'etica è sempre un'etica della situazione. Qualcuno in tal senso ha parlato di *relativismo etico*… va bene. Purché non nel senso di abbandono del valore e dell'agire etico. Bensì nel senso del valore e dell'agire etico sempre e indefettibilmente relativamente-*a* quale, dunque, permanente etica situazionale.
Perciò, se relativamente alla macro-questione AI, si cerca una risposta "chiusa"… beh è sbagliata l'aspettativa riposta nella domanda. La domanda di per sé è sempre aperta, e non potrebbe essere altrimenti. La risposta "chiusa"… che è quello che si cerca in vista di una pre-programmazione dell'*artificium*, cade miseramente al cospetto dell'imponderabile.
Si è detto: è l'aspettativa riposta nella domanda a esser fuori fuoco; alla fine tutto rifluisce solo in un problema di *Manuale d'istruzioni etiche*, in un'etica da etichetta, che giocherà e danzerà sempre, in maniera del tutto inconcludente, sugli assetti variabili dei due criteri di cui prima in seno all'approccio orizzontale. Ovvero: o l'etica precede oppure corriamo ai ripari dopo nei termini del dispositivo costi-benefici. O, ancora: il computabile dispositivo costi-benifici è addirittura prodromo in maniera tale da dare all'etica una fondazione sincretistica tra *conseguenzialismo/utilitarismo* e *de-ontologismo*.
Sotto questo profilo, scienziati, ingegneri informatici, biotecnologi e filosofi ed eticisti parleranno sempre lingue estremamente distanti, avendo finalità distinte che sembreranno trovare soluzioni a un problema solo quando, per mero caso, gli interessi s'incrociassero… Possiamo permettercelo?
Direi di no. Direi proprio di no.
Per me la questione è questa e si gioca al di qua del bene e del male: oggi, nell'era dell'AI, l'uomo è sottoposto a una pressione decisiva che lo proietta verso un proprio, ulteriore (e forse definitivo) trascendimento. Come conseguenza delle rivoluzioni biotecnologiche e informatiche, si impone che l'uomo si proietti oltre se stesso, verso qualcosa d'altro. Una trasfigurazione complessiva che ingiunge di ripensare qual è la condizione umana passata, presente e futura.
Il problema della scelta, della scelta etica, se c'è, è purtroppo dopo…. Propongo per questo, in ultima battuta, un approccio circolare. La domanda sull'AI è anzitutto domanda sull'uomo. Sull'uomo che ha messo mano all'AI; sull'uomo che la utilizza e utilizzerà e, soprattutto, sull'uomo del domani.
Come s'è provato a dire sin dalle prime battute, non v'è dubbio che l'Intelligenza Artificiale sia una delle più grandi promesse dell'umanità; grazie ai suoi sviluppi, attuali e dietro l'angolo,

saremo probabilmente in grado di fare cose che oggi sarebbero impensabili. Vivremo meglio, e magari più a lungo e più felici. E tuttavia non è possibile non cogliere sino in fondo anche quelli che potrebbero essere, ma che *in nuce* già sono proprio come segnalato, i risvolti connessi a questo tipo di tecnologia che «raggiungerà e poi supererà enormemente la finezza e la duttilità di quelli che consideriamo i migliori fra i tratti umani».
Ho detto qual è la questione: parlare di AI implica parlare di tecnologia e filosofia, di macchine e di uomini, *naturale* e *artificiale* in termini del tutto nuovi.
Significa sviluppare sempre più la capacità di conoscere l'essere umano coi suoi bisogni, e sviluppare la capacità di integrarsi con l'innovazione tecnologica senza perdersi in essa. E in questo la coappartenenza in un univoco progetto di filosofi e scienziati è decisivo.
Motivo per il quale dobbiamo metter mano a un Lògos flessibile e polifunzionale e strutturare una lingua comune come base di confronto che esca sia dai gerghi del linguaggio scientifico sia da alcune complessità del linguaggio filosofico.
Questo costituisce lo scaffolding euristico per l'edificazione di un nuovo umanesimo, una nuova condizione dell'umano che riparta da una ritrovata intimità scientifico-filosofica che è ormai ineludibile e vitale proprio per l'umano a venire.
In fondo dobbiamo ripartire dalla consapevolezza che, ogni volta, le nostre storie hanno inizio da una fine. Dalla fine di un'idea di noi stessi su noi stessi che ci ha accompagnati produttivamente anche per secoli.
Nietzsche le chiamava *metafore vitali*, anche nei termini delle nostre auto-narrazioni che ci re-insediano nel nostro rapporto con il mondo esterno, anche per dare un senso. Da quando abbiamo preso a costruirle e raccontarle, però, il tema è sempre stato il desiderio di emanciparci da noi stessi per divenire qualcosa di diverso e d'altro per continuare a essere.
In quanto esseri umani, viviamo tra le rovine di uno splendore immaginato, mai raggiunto e mai raggiungibile. Abbiamo sempre avuto un'idea ben più alta del nostro destino. Siamo e continuiamo a essere, né più né meno che alla stessa maniera degli altri viventi, a partire dai connotati della nostra strategia persistentiva di specie.
Certo complessa, complessissima…
Ma questo non è che fa di noi un vivente speciale e/o che gode di uno statuto ontologico privilegiato.
Ne dice semplicemente del fatto che, in una circolarità che non conosce soluzione di continuità, nel farci in relazione con l'ente, anche l'ente artificiale creato da noi, troviamo il modo di continuare a essere e, perciò, ridirci, nei termini anche di ipotesi auto-narrative che si superano costantemente.

E alla fine, e giusto per chiudere – mi accorgo di essere andato un po' troppo oltre –, rispetto all'AI la vera domanda è: "ne vale la pena?".
A me sembra che abbiamo già risposto, e da tempo.

GIANLUCA GIANNINI insegna Filosofia Morale presso l'Università Federico II di Napoli e Coordinatore del Corso di Laurea Triennale in Filosofia
gianning@unina.it

ANTONIO PESCAPÈ insegna Sistemi Informatici presso l'Università Federico II di Napoli e Delegato del Rettore all'Innovazione e alla Terza Missione. È inoltre direttore della DIGITA Academy
antonio.pescape@unina.it

LUCA LO SAPIO è Ricercatore di Filosofia morale presso l'Università degli Studi di Torino e Docente esterno incaricato di Percezione ed Etica delle Biotecnologie Industriali presso l'Università degli Studi di Napoli Federico II
luca.losapio@unina.it

ANDREA CALANDRELLI - ALESSANDRA NICOLINI

DONO E DOLO: DAL DIGITAL TURN AL TECNOSTRESS

1. *La digitalizzazione dell'esperienza tra risorse e fattori di rischio*
2. *Il tecnostress*
3. *Attivazione della risposta infiammatoria indotta dal tecnostress*

ABSTRACT: ***GIFT AND "DOLUS": FROM DIGITAL TURN TO TECHNOSTRESS***
The digital revolution, by reconfiguring our way of being in the world as a stay-always-on, has altered the principle of separation between online and offline. Continuous and unrestricted connection can have negative effects on quality of life and productivity, compromising individual psychosomatic equilibrium. The health effects associated with the pervasive use of digital devices and their applications are often summarised under the term "technostress". Protracted exposure to an excessive information load determines a response involving the activation of psycho-neuro-endocrine-immunological processes whose alteration plays a crucial role in the pathogenesis of the symptoms most frequently associated with technostress. We call this response "e-inflammation".

1. *La digitalizzazione dell'esperienza tra risorse e fattori di rischio*

L'introduzione delle tecnologie dell'informazione e della comunicazione (ICT), e la loro ampia diffusione in tutti i campi della vita, personale, sociale, lavorativa, hanno cambiato il nostro modo di stare al mondo. Non solo noi agiamo attraverso di esse, ma esse agiscono su di noi, non soltanto per mezzo delle abitudini che introducono nella nostra quotidianità, ma modificando il nostro modo di fare e accedere all'esperienza di noi stessi, degli altri e del mondo.

Le nuove tecnologie non offrono semplicemente un supporto operativo, un prolungamento, una "estensione" alle nostre possibilità di pensare, agire e sentire, ma hanno mutato in modo radicale le coordinate del nostro abitare, fungendo da veri e propri organizzatori del dominio esperienziale, riconfigurando/trasfigurando il nostro modo di percepire lo spazio, il tempo e il nostro corpo.

La discontinuità e frammentarietà dell'esperienza prodotta dal "multiliving"[1] genera nuovi modi del sentire che si connotano per la flessibilità e velocità di adattamento con cui essi mutano al variare, allo "switch", dei contesti in rapido cambiamento.

Le nuove tecnologie non rappresentano più solo una tra le diverse dimensioni del reale, ma sono vere e proprie «forme di vita», come suggerisce il titolo del saggio *"Technology as Forms of Life"*[2], ovvero prospettive da cui osservare il mondo, gli altri e noi stessi.

Sono, al contempo, «un modo del disvelamento»[3] e un campo di manifestatività del nostro esserci. Esse cioè si costituiscono come un vero e proprio ambiente, luogo del commercio quotidiano con il mondo.

I nuovi dispositivi di comunicazione mobile, ad esempio, come gli smartphone o i tablet, non sono più soltanto degli strumenti per telefonare, ma inaugurano un nuovo mo(n)do della comunicazione, che cambia il rapporto tra presenza e assenza, tra concentrazione e distrazione, tra pubblico e privato.

Lo smartphone, se da un lato abolisce la distanza, dall'altro abolisce la vicinanza poiché genera una condizione in cui tutti sono, allo stesso tempo, vicini e lontani, indipendentemente dal loro grado di prossimità fisica. Si può, cioè, essere seduti al

[1] T. Mazali, *Not too close not too far, just right. Sherry contro Turkle*, in «Mediascapes Journal», 2, 2013, pp. 155-158.
[2] L. Winner, *Technologies as Forms of Life*, in «Epistemology, Methodology and the Social Sciences», eds. Cohen and Wartofsky (Kluwer Academic Publishers), 1983.
[3] M. Heidegger, *La questione della tecnica* (1953), tr. it. goWare, Firenze 2017.

bar, di fronte a una persona, ma "chiacchierare" con chi non è presente. Questo fenomeno è conosciuto in letteratura con il termine "phubbing", un neologismo derivato dall'unione delle parole "phone" e "snubbing" che rimanda, per l'appunto, all'atto di ignorare qualcuno in un contesto sociale[4].

Nel riconfigurare il nostro modo di stare al mondo, in rapporto con gli altri e con noi stessi, come uno stare-sempre-connessi, "always on", lo smartphone ha dunque alterato il principio di separazione tra l'online e l'offline. Viviamo, cioè, in una "zona grigia", termine con il quale vogliamo indicare quella realtà ambigua, ibrida, dai contorni mal definiti, che insieme separa e congiunge l'online dall'offline.

Il carattere "onlife"[5] dell'esperienza, la costante reperibilità e connettività rendono, perciò, sempre più difficile la distinzione tra vita privata e vita professionale, tra tempo libero e tempo di lavoro. Il confine si è talmente assottigliato al punto che non sembra avere più senso parlare di questa separazione; nell'era dei device mobili, di Internet e dei Social network, essere online sembrerebbe, infatti, essere la norma.

Ma questa con-fusione tra i domini dell'esperienza che il massiccio innesto delle nuove tecnologie nella vita odierna ha generato, introduce nuove opportunità, o arreca danni? Rappresenta una risorsa, o un fattore di rischio?

È indubbio che gli individui, le organizzazioni e la società in generale abbiano ottenuto benefici significativi dall'implementazione e dall'utilizzo delle nuove tecnologie, attraverso l'accesso maggiore e rapido alle informazioni, l'ampliamento delle possibilità di comunicazione, i miglioramenti nelle prestazioni e nella produttività.

[4] E. Karadağ, S.B. Betül Tosuntaş, E. Erzen, P. Duru, N. Bostan, B.M. Şahin, I. Çulha, B. Babadağ, *Determinants of phubbing, which is the sum of many virtual addictions: A structural equation model*, in «Journal of Behavioral Addictions», 4, 2, 2015, pp. 60–74.

[5] L. Floridi, *La quarta rivoluzione*, Raffaello Cortina, Milano 2017. Con questo termine Floridi identifica l'ibridazione tra vita ed esperienza online.

Ma ogni nuova tecnologia, anche la più promettente, ha sempre una contropartita. Esse, infatti, rivelano una originaria e ineliminabile ambivalenza poiché, se da un lato offrono nuove opportunità, dall'altro nascondono molteplici insidie. Come ci ricorda il mito di Prometeo, il dono è sempre anche un "dolos", un inganno. Ogni progresso di carattere tecnologico, se da un lato pone condizioni che possono, ad esempio, portare alla emancipazione dal lavoro, dall'altro impone nuove condizioni di dipendenza da esso. La libertà permessa dalla connessione senza fili, se da un lato ci consente di essere sempre da un'altra parte, d'altro canto ci rende sempre raggiungibili, ovunque ci si trovi, e in qualsiasi momento. Possiamo allontanarci dal luogo di lavoro, eppure continuare a essere lì, così come possiamo trovarci sul posto di lavoro ma essere affacciati sulla soglia di un altrove.

L'accessibilità resa dai device senza fili crea, inoltre, aspettative relative al comportamento di risposta. Anche quando non siamo presenti, siamo comunque disponibili. Distanti, ma raggiungibili, "reperibili". Ogni chiamata non risposta, ogni email in giacenza, o messaggio WhatsApp spuntato ma non evaso, rimane registrato, e questa registrazione implica una responsabilità che è quella del rispondere. Essa può generare sentimenti di sovraccarico a causa dell'elevato numero di notifiche più o meno rilevanti, più o meno desiderate, che ciascuno di noi riceve giornalmente da parte dei contatti social, un fenomeno questo che troviamo descritto in letteratura con l'espressione "social network overload".

Potenzialmente, quindi, non c'è fuga dal lavoro, dalla famiglia, dagli amici, o da chiunque altro, ed è proprio il fatto di non essere cablati, ma wireless, a renderci legati!

Il contatto perpetuo e la co-presenza di reti sociali dovuti alla portabilità possono creare enormi richieste e generare importanti ricadute sulla salute e il benessere individuale. Il carattere di

disponibilità, infatti, se da una parte facilita la comunicazione superando i vincoli fisici del faccia-a-faccia, può essere percepito come un fattore di stress. Una connessione continua e senza vincoli di tempo e spazio può avere effetti negativi sulla salute, la qualità della vita e sulla produttività, compromettendo il nostro equilibrio psicosomatico.

2. *Il tecnostress*

Le continue sollecitazioni provenienti dalle nuove tecnologie e dagli online social network possono determinare notevoli percezioni di stress negli utenti.

Uno studio svedese ha riscontrato un aumento dello stress percepito in relazione al carico informativo e sociale. L'uso combinato di computer e telefono cellulare, così come il numero di messaggi scambiati giornalmente, l'email e le chat online, sono stati associati a stress prolungato, depressione e disturbi del sonno[6].

Lo stress associato all'uso pervasivo delle nuove tecnologie dell'informazione e della comunicazione nella società contemporanea è stato denominato "tecnostress". Il primo a osservare e descrivere questo fenomeno fu lo psicologo americano Craig Brod che lo definì *«un disturbo causato dall'incapacità di gestire le richieste provenienti dalle moderne tecnologie informatiche in maniera sana»*[7]. Le nuove tecnologie richiedono, infatti, l'acquisizione di alcune skills, tra cui la gestione di un maggior flusso di informazioni, una maggiore velocità di elaborazione, l'aggiornamento continuo di applicazioni e software, così come la gestione contemporanea di più funzioni (multitasking).

[6] S. Thomée, F.M. Eklöf, E. Gustafsson, R. Nilsson, M. Hagberg, *Prevalence of perceived stress, symptoms of depression and sleep disturbances in relation to information and communication technology (ICT) use among young adults: An explorative prospective study*, in «Computers in Human Behavior», 23, 3, 2007.

[7] C. Brod, *Technostress: the human cost of the computer revolution*, AddisonWesley, Reading 1984.

Questa definizione è stata successivamente ampliata fino a includere *«qualsiasi impatto negativo su atteggiamenti, pensieri, comportamenti o fisiologia del corpo causato direttamente o indirettamente dalla tecnologia»*[8].

L'avvento della tecnologia mobile, potremmo dire "prêt-à-porter", che, come abbiamo argomentato nel paragrafo precedente, ha permesso un uso costante dei flussi informativi, senza vincoli di spazio e di tempo, ha introdotto una sostanziale differenza con la prima, embrionale definizione di Brod sviluppata a partire dall'osservazione delle reazioni psicofisiche di soggetti che lavoravano seduti alla propria scrivania.

Le forme di questo disagio possono delinearsi in due modi differenti; possiamo, infatti, considerare il tecnostress un termine ombrello che include due fenomenologie distinte, ma correlate: il technostrain o tecnofobia, e la technoaddiction o tecnomania. Mentre il technostrain è caratterizzato dalla difficoltà ad accettare le nuove tecnologie e si manifesta con sentimenti e comportamenti negativi associati al loro uso, la technoaddiction, invece, consiste in una iper-identificazione con lo strumento, caratterizzata da un utilizzo compulsivo delle nuove tecnologie e da stati di ansia associati ai periodi di disconnessione. Tra i fenomeni descritti in letteratura citiamo, ad esempio, i casi di "Nomo Phobia", un termine derivato da "nomo", abbreviazione di no-mobile, e "phobia", con il quale si definisce la paura incontrollata di rimanere senza connessione alla rete mobile, a cui si associano sintomi panicoformi, o la FOMO, acronimo di Fear of Missing Out, una condizione caratterizzata dal timore di rimanere tagliati fuori dal contatto con gli eventi, le esperienze e le "stories" che avvengono all'interno dei social networks, e che può portare al controllo compulsivo delle pagine social e della messaggistica istantanea

[8] M.M. Weil, L.D. Rosen, *Technostress: coping with technology @work @home @play*, Wiley, New York 1997.

per mantenere la connessione sociale ed evitare la perdita del contatto con quelle esperienze.

La diffusione del lavoro a distanza, fortemente incentivato dalla pandemia di Covid-19, ha portato, inoltre, a osservare sempre nuove forme di disagio derivanti, ad esempio, dall'esposizione prolungata a sessioni di videoconferenze, denominata "affaticamento da riunione virtuale" (Zoom fatigue). Tale condizione è stata identificata come un fattore di rischio per il benessere e la produttività dei lavoratori e risulta da un insieme di fattori tra cui l'aumento del carico cognitivo dovuto allo sguardo prolungato da parte degli altri; l'apparente vicinanza degli altri e la riduzione della mobilità; le aspettative non soddisfatte relativamente alla sincronia e ai segnali non verbali; la ridotta distribuzione del lavoro tra i colleghi; l'aumento dell'attenzione sulla propria immagine visualizzata sullo schermo. Rispetto a quest'ultimo fattore è stato osservato che nelle persone che presentano una maggiore insoddisfazione per il proprio aspetto, in particolare il volto, la visualizzazione della propria immagine sullo schermo determina un'attenzione su di sé negativa che le espone a un maggior rischio di sviluppare disturbi quali "ansia da specchio" ("mirror anxiety")[9]. A questa aumentata preoccupazione per il proprio aspetto è, inoltre, corrisposto un incremento dell'interesse verso l'utilizzo della cosmesi e della chirurgia estetica del volto, come dimostrato da uno studio che ha analizzato i trend di ricerca online sui principali browser da parte degli utenti durante il lockdown[10].

Possiamo, dunque, considerare il tecnostress come un fenomeno "psicosomatico" che emerge dall'incontro con le nuove tecnologie,

[9] R. Ratan, D.B. Miller, J.N. Bailenson, *Facial Appearance Dissatisfaction Explains Differences in Zoom Fatigue,* in «Cyberpsychology, behavior and social networking», 25, 2, 2022, pp. 124-129.

[10] K. Thawanyarat, S. Francis, T. Kim, C. Arquette, S. Morrison, R. Nazerali, *The Zoom Effect: A Google Trends Analysis*, in «Aesthet Surg J.», 42, 1, 2022, NP76-NP82.

sia in ambito individuale, sia in contesti organizzativi come disturbo lavoro-correlato, interferendo con la quotidianità sociale e relazionale e con la produttività lavorativa. Come abbiamo visto, esso è associato all'insorgenza di disturbi di diversa natura, che coinvolgono la sfera emotiva e comportamentale, con conseguenze sul piano della concentrazione e del rendimento, o sulla qualità del sonno[11], ma l'uso prolungato dei dispositivi tecnologici può comportare anche disturbi a carico dell'apparato muscolo-scheletrico, come dolori cervicali dovuti alla posizione della testa perennemente chinata in direzione dello smartphone/tablet ("text neck"), sindrome del tunnel carpale, tendinite da "texting", gomito da mouse ("mouse elbow"), o disturbi a carico dell'apparato visivo come affaticamento oculare, bruciore, arrossamento, lacrimazione, secchezza oculare, fotofobia, annebbiamenti visivi transitori (sindrome da visione del computer).

Nonostante la varietà di forme attraverso cui il tecnostress trova espressione a livello individuale, questi sintomi presentano un comune denominatore: l'infiammazione, la quale gioca un ruolo cruciale nella patogenesi dei disturbi a esso più frequentemente associati.

3. *Attivazione della risposta infiammatoria indotta dal tecnostress*

L'esposizione protratta a un carico informativo e social(e) eccessivo determina una attivazione neuro-endocrina e immunitaria che ha come effetto la produzione di sostanze responsabili dell'infiammazione. I sistemi neuro-endocrino e immunitario formano un network di comunicazione bidirezionale attraverso le citochine, una classe di molecole messaggere prodotte da diversi

[11] Gli schermi retroilluminati di pc, tablet e cellulari inibiscono la produzione di melatonina che serve ad avere un sonno di qualità ed entrare nel cosiddetto *"default mode network"*, ossia lo stato di semi-veglia che precede il sonno.

tipi di cellule, immunitarie e non, in risposta a stimoli di diversa natura.

In circostanze normali, la maggior parte delle citochine non sono espresse, o sono espresse a concentrazioni molto basse, mentre la loro produzione è tipicamente dis-regolata durante l'esposizione a eventi di natura stressante, quali infezioni, lesioni, tossine, ma anche stati emozionali. Diversi studi, infatti, dimostrano che livelli più elevati di citochine pro-infiammatorie circolanti nel sangue sono associati a emozioni negative quali rabbia, paura, vergogna, tristezza, ma non a quelle positive[12].

Lo stress è noto per attivare il sistema nervoso simpatico e l'asse ipotalamo-ipofisi-surrene i quali regolano la produzione di citochine attraverso il rilascio corrispettivo di adrenalina e noradrenalina negli organi linfoidi, e di cortisolo. Mentre la noradrenalina stimola il sistema immunitario a produrre citochine pro-infiammatorie, il cortisolo, a dosi fisiologiche, esercita una potente azione anti-infiammatoria. Tuttavia, in condizioni di stress eccessivo e prolungato, aumenti persistenti del cortisolo si traducono in una ridotta risposta recettoriale. Di conseguenza, la capacità del sistema immunitario di rispondere ai segnali antinfiammatori si riduce e i processi infiammatori si sviluppano.

La persistenza degli stressors, o l'incapacità adattativa della persona alle richieste provenienti dalle nuove tecnologie, può determinare, nel medio e lungo periodo, una condizione di infiammazione cronica di basso grado ("low grade inflammation"), derivante dalla mancata risoluzione dell'infiammazione e/o dalla stimolazione prolungata e ripetuta del sistema immunitario. Proponiamo di definire tale condizione *"e-inflammation"*.

[12] D.C. Slavish, D.R. Jones. J.M. Smyth, C.G. Engeland, S. Song, N.M McCormick, J.E. Graham-Engeland, *Positive and Negative Affect and Salivary Markers of Inflammation Among Young Adults*, in «Int J Behav Med.», 27, 3, 2020, pp. 282-293.

L'esposizione protratta a un carico informativo eccessivo determina, infatti, una risposta che coinvolge l'attivazione di processi psico-neuro-endocrino-immunologici la cui alterazione gioca un ruolo cruciale nella patogenesi e nel mantenimento dei sintomi più frequentemente associati al tecnostress.

Diversi studi dimostrano, infatti, come l'implementazione delle nuove tecnologie, così come l'esposizione a problemi specifici derivanti dal loro utilizzo (ad esempio, sovraccarico informativo e sociale, multitasking, lunghi tempi di risposta, guasti al sistema, interruzioni del lavoro), possano causare effetti dannosi sulla salute attraverso il rilascio di sostanze che, in maniera diretta o indiretta, sono coinvolte nella regolazione dei processi infiammatori. Questi studi, ad esempio, hanno rilevato aumenti dei livelli di noradrenalina[13] e cortisolo[14], e diminuzione delle concentrazioni di melatonina[15] che sappiamo svolgere una importante azione antiossidante e possedere proprietà immunomodulanti e anti-infiammatorie.

Uno studio, in particolare, ha riscontrato la presenza di livelli elevati di citochine pro-infiammatorie, in particolare interleuchina-6 (IL-6), nei figli adolescenti rispetto ai loro genitori, in relazione al maggiore tempo dedicato da parte dei primi, se confrontati con gli adulti, all'utilizzo delle nuove tecnologie e dei social network[16].

Questi risultati ci portano a osservare che il modo in cui siamo coinvolti nel mondo e con gli altri attraverso la tecnologia ha

[13] C. Korunka, K.H. Huemer, B. Litschauer, B. Karetta, A. Kafka-Lützow, *Working with new technologies: hormone excretion as an indicator for sustained arousal: a pilot study*, in «Biological Psychology», 42, 3, 1996, pp. 439-452.

[14] R. Riedl, H. Kindermann, A. Auinger, A. Javor, *Technostress From a Neurobiological Perspective: System Breakdown Increases the Stress Hormone Cortisol in Computer Users*, in «Business & Information Systems Engineering», 4, 2, 2012, pp. 61-69.

[15] B.B. Arnetz, M. Berg, *Melatonin and adrenocorticotropic hormone levels in video display unit workers during work and leisure*, in «Journal of Occupational and Environmental Medicine», 38, 11, 1996, pp. 1108-1110.

[16] T.D. Afifi, N. Zamanzadeh, K. Harrison, M.A. Callejas, *WIRED: The impact of media and technology use on stress (cortisol) and inflammation (interleukin IL-6) in fast paced families*, «Computers in Human Behavior Volume», 81, 2018, pp. 265-273.

profonde ricadute sulla funzionalità immunitaria, inducendo, ad esempio, una modificazione nei geni di trascrizione pro-infiammatoria. In questo senso possiamo allora capire come lo stress derivante dall'uso pervasivo delle nuove tecnologie possa radicarsi biologicamente, "incarnarsi" di volta in volta nell'incontro con esse, modellando i fenotipi e la suscettibilità alle malattie[17].

Rimane da chiedersi, tuttavia, se le nuove tecnologie siano la "causa", o semplicemente il campo di manifestazione di un disagio che, attraverso di esse, trova la sua attuazione.

Andrea Calandrelli svolge attività didattica in Clinica medica ed è Direttore scientifico dei Master di "Psico-Neuro-Endocrino-Immunologia" e "Integrazione tra medicine convenzionali e non convenzionali" presso l'Università degli studi di Roma - La Sapienza

andreacalandrelli@virgilio.it

Alessandra Nicolini è Psicologa clinica e specialista in Psicoterapia a indirizzo fenomenologico-ermeneutico

alessandra.nicolini1@virgilio.it

[17] A. Calandrelli, A. Nicolini, *Sintomi funzionali. Dalla scissione mente-corpo al paradigma corpo-mondo*, in «InCircolo. Rivista di filosofia e culture», 11, 2021, pp. 168-188.

SIMONA TIRIBELLI

ARTIFICIAL AGENCY AND MORAL AGENCY: CONCEPTUALIZING THE RELATIONSHIP AND ITS ETHICAL IMPLICATIONS ON MORAL IDENTITY FORMATION

1. *Introduction* 2. *Artificial Agency and Moral Agency*
3. *AI over Epistemic and Moral Autonomy: who decides what?*
4. *Conclusive Reflections: AI over Moral Identity Formation*

ABSTRACT: *ARTIFICIAL AGENCY AND MORAL AGENCY: CONCEPTUALIZING THE RELATIONSHIP AND ITS ETHICAL IMPLICATIONS ON MORAL IDENTITY FORMATION*
The paper aims to conceptualize the relationship between artificial agency, that is, the kind of agency connoting artificial intelligence (AI) systems, especially machine learning (ML) and deep learning (DL) algorithms, and moral agency, that is, the agency characterizing individuals as moral agents (persons), by highlighting its main ethical implications for moral identity formation. To this aim, after having unpacked the different forms of agency into question and clarified their main features, an ethical inquiry is carried out in order to show how artificial agency, as currently designed, not only interacts with, but it might also endanger the genuine development and expression of moral agency, by undermining individuals' epistemic and moral autonomy. The paper concludes by showing the main consequences that might arise from this impact, in particular in relation to those processes of individual and collective moral self-representation that nurture and steer the genuine formation and the flourishing of moral identity.

1. *Introduction*

The increasing advances in the design and use of artificial intelligence (AI), and especially of machine learning (ML) and deep learning (DL) algorithms, have prompted a robust corpus of literature in the field of philosophy and applied ethics over the last decade, assessing their opportunities and risks for individuals and societies. On the one hand, AI-based systems such as ML- and DL-based technologies have shown their huge capacity to perform more and more of our tasks and decisions efficiently

(i.e., quickly and at low costs) in many social domains[1], from communication and advertising to education, justice, and healthcare, by enabling their widespread conceptualization as a new «growing resource of interactive, autonomous, and self-learning agency»[2], along with their increasing delegation of – previously just human – choices, actions, and activities[3]. On the other hand, ML and DL algorithms, and thus, the new forms of "artificial agency"[4] they constitute, have been also widely unveiled to be very often flawed and/or biased[5] and to produce controversial effects, ranging from fostering social discrimination and inequalities[6] up to facilitating individuals' manipulation[7] and self-deceit[8]. In this scenario, whilst great attention has been paid to understanding and critically assessing whether ML and DL algorithms may be entitled to moral status[9], their interaction *with* and/or their impact *on* our moral agency have not been sufficiently explored thus far[10]. Nevertheless, this

[1] B.D. Mittelstadt *et al.*, *The Ethics of Algorithms: Mapping the Debate*, in «Big Data & Society», 3 (2), 2016.
[2] L. Floridi, *What the Near Future of Artificial Intelligence Could Be*, in «Philosophy & Technology», 32, 2019, p. 2.
[3] B.D. Mittelstadt *et al.*, *The Ethics of Algorithms*, cit.
[4] The concept of "artificial agency" is widely in use in the debate in ethics of AI to mean the capacity of AI systems (e.g., ML- and DL-based systems) to perform (also human) tasks and decisions (algorithmic decision-making) in a way that does not necessarily entail their further attribution of moral agency.
[5] A. Tsamados *et al.*, *The ethics of algorithms: key problems and solutions*, in «AI & Society» 37, 2002, pp. 215-230.
[6] C. O'Neil, Weapons of Math Destruction, Penguin, London 2016; V. Eubanks, *Automating Inequality*, St Martin's Publishing, New York 2018; S. Noble, *Algorithms of Oppression*, NYU Press, New York 2018; R. Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code*, Polity, Medford 2019.
[7] See the European "Declaration on the Manipulative Capabilities of Algorithmic Processes" (EU Decl [13/02/2019]).
[8] S. Natale, *Deceitful Media. Artificial Intelligence and the Social Life after the Turing Test*, Oxford University Press, Oxford 2021.
[9] S.M. Liao, *The Moral Status and Rights of Artificial Intelligence*, in «Ethics of Artificial Intelligence», Oxford University Press, Oxford (UK), 2020; L. Floridi, J. Sanders, *On the Morality of Artificial Agents*, in «Minds and Machines», 14, 2004, pp. 349-379.
[10] The majority of the discussion in the field of ethics and AI concerning moral agency is devoted to understanding artificial agency as morally connoted. AIs' interaction with human moral agency is less investigated and almost exclusively framed in relation to phenomena of social manipulation, surveillance and persuasive technology, and technological paternalism. See B.

inquiry sounds today needed considering how these systems, by deploying subtle and fine-grained techniques, have shown a huge capacity to capture, steer, and influence our preferences and beliefs[11] up to deeply reshape our decision-making processes and social and political[12] (including moral[13]) choices, by threatening our control and autonomy over them[14].

This paper aims to contribute to filling this gap and conceptualizing the relationship between *artificial agency* and *moral agency*. By artificial agency is meant the capacity of ML and DL that rule the functioning of today's majority of digital information and communication technology (ICTs) to "act" or "behave" (according to technical vocabulary)[15] by making decisions and accomplishing tasks in order to achieve certain pre-set goals. By moral agency is meant our capacity to develop genuine values, reasons, commitments, goals and moral ground-projects and to reflectively endorse them as motives for our choices, actions, and behavior. Specifically, the goal of the paper is to unpack this relationship and show how artificial agency can reshape but also hinder the genuine development of our moral agency and moral identity, by affecting the processes through which a) we self-

---

Frischmann, E. Selinger, *Re-Engineering Humanity*, Cambridge University Press, Cambridge 2018.

[11] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford 2014, p. 58; L. Royakkers *et al.*, *Societal and Ethical Issues of Digitalization*, in «Ethics and Information Technology», 20, p. 2.

[12] Consider the Facebook emotion experiment (2004) or the case of Cambridge Analytica (2016).

[13] On the impact of algorithms-based systems on our moral choices (and, specifically, on our freedom of choice), see S. Tiribelli. *Predeterminazione algoritmica e libertà di scelta*, in L. Alici e F. Miano (a cura di), *Etica nel Futuro*, Orthotes, Napoli 2020.

[14] BD. Mittelstadt *et al.*, *The Ethics of Algorithms*, cit., p. 9; M. Hildebrand, *Smart technologies and the end(s) of law: Novel entanglements of law and technology*, Cheltenham: Edward Elgar.

[15] This paper does not assess whether AI systems can be considered as moral agents, as the scope of this inquiry is limited to their interaction with individuals as moral agents. Thus, the use of the term "agency" (or "behavior") in relation to AI is that adopted in the technical debate: it refers to AI capacity to operationalize instructions (as in the case of deterministic algorithms) and/or to develop new paths to achieve a certain goal (as in the case of indeterministic algorithms).

represent and act as morally-connoted agents and b) we self-represent and act as part or members of a common-good-driven collectivity (a society sharing common substantial values). To this aim, a first section is devoted to introducing the specific features characterizing artificial agency and human moral agency. A second section clarifies how they mutually relate and interact (i.e., modality of inter-action) and how artificial agency, as currently designed, might endanger our moral agency by undermining our autonomy at the epistemic level (*epistemic autonomy*), thus, weakening us as *knowers*, and at the very stringent moral level (*moral autonomy*), weakening us as *moral agents*. A final section sheds light on further implications of artificial agency's impact on moral agency, by focusing especially on the role of the former in reshaping and/or hampering individual and collective moral self-representation processes which in turn shape the genuine formation and expression of our moral identity.

2. *Artificial Agency and Moral Agency*

The understanding and conceptualization of the relation between artificial agency and (human) moral agency cannot prescind from first clarifying the features mainly connoting each form of agency into question. Indeed, even if with the well-known "fourth revolution"[16], leading scholars have argued the blurring of the distinction between artificial agents and human agents on the basis of an unprecedented common capacity to process (acquire and act over) information, and being processed as information, when it comes to considering moral agency, this common broad definition of agency might result too partial and turn out to be misleading. Beyond the pretense of exhaustiveness of the wide critical debate on moral agency developed in moral philosophy, there is a widespread Kantian-inspired agreement on the fact that moral agency requires certain basic features of rationality, and

[16] L. Floridi, *The Fourth Revolution*, cit.

specifically, a certain degree of *moral reasoning* (that is, the capacity to form and reason on moral concepts, ideas of good, as well as moral reasons, values, and beliefs), especially when it comes to consider deliberative processes on alternative options that we might choose as reasons for our agency. Moral reasoning indeed allows the genuine formation and the exercise of the reflective endorsement of what can drive and motivate our agency (*moral motives*), from values and beliefs to shared goals, joint commitments, and moral ground projects, which normatively denote a mere agency as a morally-connoted one. Put it differently: moral reasoning allows the formation of our specific *moral knowledge*, that is, the formation of values, moral reasons, and ideas of good, and thus, broadly, of the moral motives that normatively steer our choosing and agency by developing a specific *ought to*. Therefore, moral reasoning enables *moral responsiveness* (both as *accountability* and *answerability*), that is, to offer others reasons (i.e., account) for our actions and be responsive to reasons provided by others. It sounds noteworthy underlying that in order to form a *genuine* moral knowledge, we need to be exposed to different ways of thinking, acting, and behaving (different moral practices), insofar as moral heterogeneity is crucial to assess whether our values, reasons, beliefs (and so forth) we endorse as moral rules for our choosing and agency are the optimal ones, or instead there are reasons to revise them. According this standard view, the genuine development and reflective endorsement of values, reasons, beliefs (…and so forth) as moral motives for our choices express the exercise of our autonomy and enable the genuine expression and flourishing of our moral agency. To sum up: moral reasoning is a basic feature of moral agency and it requires a heterogeneous moral exposure to enable truly the exercise of our *epistemic autonomy*, that is, to form our moral knowledge in a way that is genuine, as what we will endorse as moral motives (normative moral rules) for our behavior. Once we endorse those

moral motives and choose and act (or choose to do not act) on the basis of them, we exercise our *moral autonomy*, that lies in our reflective endorsement, and thus give shape to and express our moral agency. While the standard view on moral agency stresses the import of rationality features and moral reasoning, as it has been pointed out by theories of bounded rationality and studies in cognitive moral phycology[17], our deliberative processes and choices are not only a matter of rationality, but are also influenced and driven by emotional facts[18]. Hence, our moral agency and moral identity can be the result of choices also motivated by morally-loaded emotions. The latter, in turn, play a bivalent role, potentially having both a negative impact on moral agency, namely, counteract moral reasoning and endorsement, or instead motivating and informing it[19], and this mainly depend on what emotion is at stake and how that is triggered, thus, depend on particular forms that emotions can take in particular individuals. If our moral agency can be steered by motivations that involve the interplay of rationality and emotions, the role of the latter should be acknowledge to understand what characterizes our moral agency, especially when – as we will see later – the latter is considered in a new scenario of inter-action with artificial agency. When it comes to understanding artificial agency, the debate in ethics of AI very often uses a similar vocabulary to that pertaining to moral agency; however, substantial differences in meaning need to be highlighted. As anticipated before, by artificial agency is meant the capacity of

[17] On theories of bounded rationality, see D. Kahneman (2011), *Thinking, Fast and Slow*, Farrar, Straus & Giroux. For a contribution at the intersection of cognitive sciences and philosophy of mind, see M. De Caro, M. Marraffa, *Debunking the Pyramidal Mind: A Plea for Synergy Between Reason and Emotion*, in «Journal of Comparative Neurology», 8, 524, 2016.

[18] See A.E. Galeotti, *Political Self-Deception*, Cambridge University Press, Cambridge 2018; A. Mele, *Self-Deception Unmasked*, Princeton University Press 2000.

[19] On the role of positive and negative emotions for moral reasoning, see the contributions of M. Nussbaum, *Upheavels of Thought: The Intelligence of Emotions*, Cambridge University Press, 2003; *The Monarchy of Fear: A Philosopher Looks at Our Political Crisis*, Simon & Schuster, 2018.

ML and DL algorithms to make (also human) tasks and decisions to achieve certain preset goals. The latter are usually humanly decided, that is, pre-determined by designers according to technology providers' vision and, as widely documented[20], in line with third-party interests. To date, these goals, especially in ICTs, have been mainly driven to maximize utility and economic interest (e.g., revenue from ads); therefore, they have little to do with human moral goals or ground-projects. However, the recent rise of AI ethics as a discipline and movement is working to counteract this "hetero-determination"[21], by encouraging AI providers and designers to embrace a value-oriented approach to design AI systems in a way that respect ethical principles and foster the social good. Nonetheless, this virtuous scenario still requires some time to be soundly realized, and current goals of algorithms ruling ICTs are mainly preset to capture individuals' attention and maximize revenues resulting from clicks, purchases, and advertising and, therefore, they do not usually align with specific human (moral) goals, as well as no form of "autonomy" of these systems is exercised for the definition of such goals. In this regard, it is important to clarify that when we refer to artificial agency we mainly refer to probabilistic ML and DL algorithms, rather than deterministic (statistical) models, which capacity to perform tasks consists of a mere operationalization (automation) of instructions or rules humanly defined. Probabilistic or indeterministic algorithms show instead a heuristic and predictive power (i.e., *knowledge discovery method*) that expresses itself in the capacity to find, given a certain goal, existing or new paths to achieve it, by inferring predictive patterns or correlations. To sum up: they are not trained to achieve a certain goal but, given that, they can behave (deploy techniques) in order to *learn how* to achieve it. In this learning

[20] B. Frischmann, E. Selinger, *Re-Engineering Humanity*, cit.
[21] S. Tiribelli, *Predeterminazione algoritmica e libertà di scelta*, cit.

or "smart" feature lies what many scholars describe, with much criticism, as the algorithmic capacity of "reasoning" (or the "intelligent" side of these systems). Also, it is specifically in the consideration of this algorithmic capacity to self-learn (especially in unsupervised models) how to achieve certain goals that some scholars attribute a certain kind of "autonomy" to these systems – here too with not a few criticisms. Indeed, the more the algorithms prove the ability to find patterns that help to accurately predict how things will go even without instructions to follow, the more they are perceived as smart, autonomous, and thus powerful artificial agents. If the different features of the two kinds of agency into inquiry are now clearer, in the next section, we explore how artificial agents relate and interact with us as moral agents, by showing how the former, as currently behave, allow to conceptualize this specific new relation, more than as an inter-action, as a sort of over-action: how artificial agents more than *inter-act with* us seem to *over-act on* us as moral agents.

3. *AI over Epistemic and Moral Autonomy: who decides what?*

In the previous section, we have underlined how today we – as agents – are increasingly sharing our space – whose informational understanding (informational space: everything can be understood as information)[22] is now widely acknowledged – with artificial agents, with whom we share the capacity to unceasingly process and be fed by information, while, as previously clarified, we perform different kinds of agency. In this informational space, artificial agents, that is, ML and DL algorithms, sit in a privileged position both in knowledge and action in respect to ours both as knowers and agents. Firstly, they have the capacity to elaborate huge amounts of data, and also thanks to our continuous interaction with them (*unconscious human training*), semanticize them, by transforming vast streams of disaggregated and

[22] L. Floridi, *The Fourth Revolution*, cit.

heterogeneous data in valuable information, about us and the world, in the form of predictive patterns or correlations. In this sense, they have an *epistemic advantage*: they see what is invisible to us (including information about us) without being seen by us. This advantage is also a *practical* one, that is, it is a power asymmetry in knowledge and action: insofar they are informational gatekeepers, they can decide what option to show to us in informational terms; in other words: what *can* or *cannot* inform and motivate our agency. In this regard, that profiling and persuasive techniques are the means through which algorithms are mainly designed to achieve preset third-party (economic) goals is something today well documented[23]. Less explored instead is how they *inter*-act or *over*-act with/on our moral agency while they work to achieve their goals. Let us see it.

In mature informational societies, where our daily experience is almost ubiquitously mediated by digital ICTs and by the algorithms ruling them, the latter more than mediate are deeply reshaping our choice-contexts, the context where we form and make our choices and actions, that is, where we develop and express our moral agency. To clarify this claim, it suffices to think about how algorithms through filtering and classifying techniques determine which information (or informational options) we can get access to. These informational options do not only include trivial contents. Indeed, since algorithms today mediate and therefore embeds almost each aspect of our life (datification process), from affiliations, professional relationships, and political and religious orientation to values, reasons, and beliefs, informational options – algorithmically chosen – can embed almost everything that might become a motive for our action, including what we have defined as moral knowledge. In this sense, it sounds reasonable claiming that ML and DL algorithms increasingly decide what (also morally-connoted) informational option we can get access to and process –

[23] BD. Mittelstadt *et al.*, *The Ethics of Algorithms*, cit.

and adopt as a moral motive – for our choices and actions. More the algorithmic presence increases in our lives, more their agency on what can inform and drive our choosing and agency can become invasive. It follows that whether we experience the above described algorithmic action, we as moral agents do not only just interact with artificial agents, by providing them information on how understanding us and the world that they in turn process and re-use to feed us and influence our agency. This algorithmic agency that is designed toward the achievement of specific preset goals expresses itself as a conditioning third-party oriented pre-determination of our choice-contexts, limiting by pre-selecting what can or cannot become part of our moral knowledge (values, reasons, ideas of good, etc.). If we experience this pre-determining artificial action on our choice-contexts and moral knowledge we can fairly claim that our epistemic autonomy is endangered[24]. Indeed, to be threatened is our capacity to genuinely form and adopt – via critically assessment of – our moral motives (that we choose as normative rules for our agency) as the optimal ones. Indeed, algorithms tend to predetermine our informational context and exposure on the basis of subtle profiling and data mining techniques which process huge streams of data in order to create profiles as labels to attach to us, exploited in turn to categorize us in specific groups and to test on us (*fine-tuning*), as part of these groups, what preselected informational options (and contexts) work better in order to achieve their preset goals (from capturing human attention to increasing click and advertising or purchasing revenues). As documented in the debate[25], these ML techniques tend to foster filter bubbles or echo-chambers of profiled like-minded people, where phenomena like social polarization, social cascades, as well

[24] S. Tiribelli, *Predeterminazione algoritmica e libertà di scelta*, cit.
[25] C. Sunstein, *Democracy and the Internet*, In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy*, Cambridge University Press, pp. 93-110; E. Pariser, *The Filter Bubble*, Penguin, 2011.

as thought radicalization are fostered due to a lack of diverse points of view, including morally heterogeneous reasons, values, beliefs (and so forth). In this sense, to achieve third-party preset goals, algorithms have a considerable impact on our moral agency by over-acting on our capacity to form *genuine moral motives* that we can critically assess and endorse for our agency. However, our epistemic autonomy is not the sole to be endangered. Indeed, one might say that, even if our epistemic autonomy is affected, we can exercise our autonomy via the exercise of the reflective endorsement on available options, even if this means exercising our moral agency by embracing motives that are not developed in a truly genuine way (thus, critically assessed via a morally heterogeneous exposure). However, we claim that also our reflective endorsement, thus, our moral autonomy, might be algorithmically undermined.

In the previous section, we highlighted that emotions play an important role in motivating our agency and that their enabling and disabling role in the light of moral agency depends on the particular forms that emotions can take in particular individuals. Indeed, we as particular individuals are more than mere instances of rational agency (abstract selves): we are deeply connoted by specific attitudes, matured via affiliations, social relations, and experiences, as well as shaped by vulnerabilities, fears, and physical or psychological weaknesses. These traits connote us intimately, by shaping the way in which we live our life and rationally or emotionally respond to life events and other people. In other words, these personal traits shape how an event or information is perceived or interpreted and which kind of emotion might be generated in response. In this regard, that fine-grained algorithmic profiling techniques are able to capture these traits have been extensively shown.[26] However, algorithms, as currently designed, in order to achieve certain preset goal, do not only

[26] EU Decl [13/02/2019].

infer these traits and use them to sub-categorize us in groups to which show through filtering and classifying techniques preselected informational options rather than others, predetermining the epistemic conditions of our choosing and agency. Even more impactful techniques such as micro-targeting algorithmic recommender systems (RSs) are increasingly deployed to meet third-party goals. These work by exploiting inferred intimate traits (such as weakness or vulnerabilities) in order to trigger specific disruptive emotions, especially primary emotions (like fear, disgust, anger), which – even if functional to achieve preset goals – are more likely to disenable, rather than enabling, our moral reasoning and the exercise of our moral autonomy. In these cases, we might assist to a phenomenon we can define as the *suspension of the reflective endorsement*. Primary emotions indeed are defined in moral psychology as those we have in common with children, as they are instinctive and innate, and differ from secondary emotions that instead require a certain degree of awareness and socialization, along with the formation of an idea of good. RSs, by exploiting specific informational contents to target these specific individual traits, can elicit primary emotions, which in turn can trigger instinctive behavior-responses. The latter can suspend the exercise of our reflective endorsement, thus leading an algorithmically recommended emotionally-loaded informational option to determine our choices and actions at our own place. According to this standpoint, we claim that artificial agency, as currently designed, might undermine also our moral autonomy, specifically by deploying vulnerabilities-exploiting and primary-emotions-triggering techniques that can convert an – algorithmically recommended – informational option from being a *motive* (that inform and we can embrace) for our choices and actions to be the main *cause* of them, triggering instinctive or emotional choice-behavior responses that suspend our reflective endorsement. In this sense, we might not

only result as weakened as knowers, that is, in our capacity to reason and develop genuine moral motives (values, reasons, beliefs, etc.) to endorse in our agency, but also as moral agents, that is, in our deep normative capacity to endorse reflectively what can become a moral motive for our agency. It follows that our capacity to develop, express, and exercise our moral agency, more than just algorithmically informed, shaped, or distorted, might result deeply compromised.

4. *Conclusive Reflections: AI over Moral Identity Formation*

In the previous section, we have unpacked the relation between artificial agency, as currently resulting from the dense network of multiple algorithmic techniques that govern our pervasive ICTs, and our moral agency, showing how the former can undermine our epistemic and moral autonomy, and thus, more than *inter*-ac, can *over*-act on or *counter*-act the genuine expression and exercise of our moral agency. This final section sketches a few main implications of this impact for the formation of moral identity. Indeed, that digital ICTs reshape how we self-understand and constitute as individuals, hence, how we form our personal identity, has been already discussed in the debate[27]; less covered instead are the specific implications of algorithms for our moral identity's formation. We aim to shed light on a few of them in order to pave the way for further research on the topic. That artificial agency is currently shaping the social practices through which we develop our values, reasons, beliefs up to shared commitments and moral ground-projects should be now clearer. Indeed, it has been shown how algorithms affect our informational (and relational) exposure and pre-determine our choice-contexts in a way that tend to categorize individuals in *moral echo-chambers* where moral heterogeneity is diminished and people are incline to radicalize their beliefs and thoughts in order to be accepted by

[27] L. Floridi, *The Fourth Revolution*, cit.

other members of their group[28]. Also, we have argued how less social and moral heterogeneity (in thoughts, values, and reasons) tend to disenable moral reasoning and hamper the possibility to develop genuine moral motives which connote our agency as specific moral persons, that is, our moral agency. It follows reasonable claiming that artificial agency, as currently operating, is affecting our possibility to form over time genuine moral identities. Indeed, when we unconsciously end up in less heterogeneous and thus increasingly enclosed groups, formed through the exploitation of primary emotions (especially negative emotions), the capacity to critically assess if the reasons, beliefs, values (moral motives) we embrace are the optimal ones is eroded, due to the lack of a critical and heterogeneously informed dialogue, favoring a more emotional adherence to the claims made by members belonging to the same group (confirmation bias)[29]. This lack of heterogeneity and moral reasoning in turn undermine also our capacity to offer to others reasons (i.e., moral responsiveness) for our choosing and agency. It follows that algorithms not only reshape but might also hamper our processes of moral self-representation both at the individual and at the collective level: that is, how we self-represent and over time constitute as genuine moral agents (moral identity formation) both as individual persons and as member of a social collectivity that shares common substantial values and that is oriented toward the common good. It is likely indeed that the more our capacity to form genuine reasons and be responsive to different reasons is eroded, the more it is difficult to self-understand and self-represent as moral agents, by questioning if we are acting according to moral reasons, values, and projects that are aligned with the moral persons we would like to become. Conversely, we might find out – over time – that we have emotionally followed or

[28] Cass R. Sunstein, *Democracy and the Internet*, cit.
[29] C. O'Neil, *Weapons of Math Destruction*, cit.

approved practices that are disaligned to the kind of moral progress we would like to see in our society, experiencing a *moral disorientation* between who we are and who we wanted to be from a moral standpoint. This phenomenon can have also broader or collective implications. In fact, in like-minded and increasingly self-enclosed groups, along with thoughts' and beliefs' radicalization and the gradual erosion of the capacity to account for our actions, we might also become less responsive to others' diverse reasons, once we will encounter them, that is, less open and capable to understand and embrace other ways of thinking and acting (i.e., different moral practices). This phenomenon might lead us, as members belonging to different groups, to be less open and capable to achieve a potential agreement on topics of societal matter, thus, it might hinder our capacity to form collective joint commitments and share common values and goals – that is, to self-represent and act as part of a common-good-driven collectivity – which in turn might result in a diminished capacity of joint action and planning, which instead are crucial in order to create the conditions for a stable and morally flourishing society. It sounds the present as the right time to tackle or prevent these phenomena: revising algorithms' design in a way that considers the above-mentioned implications and allows the [re]alignment between moral agency and artificial agency is only the first step of a novel *ethics of inter-action* between *moral agents* and *artificial agents* that has just begun to take its first steps and imperatively demands to be further developed.

Simona Tiribelli è Ricercatrice in Filosofia Morale presso l'Università di Macerata e Research Affiliate presso il MIT Media Lab del Massachusetts Institute of Technology (USA)

simona.tiribelli@unimc.it

CRISTIANO CALÌ

ALGORITMI E PROCESSO DECISIONALE.
ALLE ORIGINI DELLA RIFLESSIONE ETICO-PRATICA PER LE IA

1. *Introduzione* 2. *La decisione demandata alle macchine* 3. *Le decisioni delle macchine*

ABSTRACT: *ALGORYTHMS AND DECISIONAL PROCESS. AT THE ORIGINS OF ETHICAL-PRACTICAL REFLECTION FOR AI*
This contribution aims to investigate not so much the ethical implications of utilizing intelligent machines in specific contexts, (human resources, self-driving cars, robotic hospital assistants, et cetera), but the premises of their widespread use. In particular, it aims to analyze the complex concept of decision making - the cornerstone of any ethical argument - from a dual point of view: decision making assigned to machines and decision making enacted by machines. Analyzing the role of algorithms in decision making, we suggest a redefinition of the ethical vocabulary for AI, specifically regarding the concept of agency.

1. *Introduzione*

Nel 2018 usciva la prima monografia in lingua italiana che trattava, in modo quasi esclusivo, del problema, abbastanza recente, dell'etica delle macchine intelligenti[1], preceduto di diversi anni dal suo corrispettivo in lingua inglese[2]. Non erano di certo mancati articoli che avevano avviato la discussione, ma soltanto negli ultimi cinque anni, a seguito della rinnovata "estate" che l'intelligenza artificiale (IA) sta vivendo, vi è stata un'innumerevole produzione di testi legati

[1] Cfr. P. Benanti, *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Marietti, Bologna 2018.
[2] Cfr. M. Anderson, S. Leigh Anderson, *Machine ethics*, Cambridge University Press, New York 2011.

all'impatto etico delle nuove tecnologie[3]. Il catalizzarsi dell'attenzione verso quest'ultime non è stato operato esclusivamente dal mondo accademico, ma anche da quello politico. Dal 2016, infatti, «quando i governi nazionali, le organizzazioni non governative e le aziende private hanno iniziato ad assumere un ruolo di primo piano nel dibattito sull'IA e sugli algoritmi "equi" ed "etici"», i lavori sull'etica degli algoritmi sono aumentati in modo significativo[4] e a partire dal 2017, quando furono definiti alcuni principi nel documento di Asilomar, le pubblicazioni intorno all'etica delle IA e i documenti da parte di governi e commissioni internazionali si sono moltiplicati a dismisura[5].

Se si dà un rapido sguardo alle prime, le pubblicazioni, gli argomenti trattati sono quasi sempre gli stessi: macchine a guida autonoma e problemi correlati, armi autonome e politiche che le normino, IA e mondo del lavoro, cybersicurezza e privacy, impatto ambientale dell'IA e sex robot, diritto *delle* IA e IA *nel* diritto[6]. Indubbiamente questi sono i temi più cogenti e anche i

[3] L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, tr. it. Raffaello Cortina, Milano 2022, p. 147.

[4] 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*, in «Internationl Journal of Communication», 10, 2016, pp. 4972-4990; R. Binns, *Fairness in machine learning. Lessons from political philosophy*, ArXiv 2018, 1712.03586 [cs]; A.D. Selbst, D. Boyd, S.A. Friedler, S. Venkatasubramanian, J. Vertesi, *Fairness and abstraction in sociotechnical systems*, in «Proceedings of the Conference on Fairness, Accountability, and Transparency-fat*'19», ACM Press, Atlanta/GA, pp. 59-68; R. Ochigame, Rodrigo, The Invention of "Ethical AI", in https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ (13/03/2022).

[5] Oltre ai 23 principi frutto del convegno ad Asilomar in California, si pensi anche alla *Partnership on AI*, finalizzata allo sviluppo di *best practices*, e al progetto l'*AI for Good*, con il coinvolgimento dell'ONU. Per una rassegna commentata cfr. Floridi, *op. cit.*, pp. 94-105.

[6] Cfr. N. Turi, *Guida per umani all'intelligenza artificiale. Noi al centro di un mondo nuovo*, Giunti, Milano 2019; G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Carocci, Roma 2020; A. Longo, G. Scorza, *Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà*, Mondadori università, Milano 2020; M. Chiriatti, *Incoscienza artificiale. Come fanno le macchine a prevedere per noi*, Luiss University Press, Roma 2021; F. Fossa, V. Schiaffonati, G. Tamburrini (a cura di), *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Carocci, Roma 2021; F. Pasquale, *Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'intelligenza artificiale*, Luiss

più attraenti. Pensare a un veicolo a guida autonoma (e a tutti i possibili scenari in cui questo potrebbe ritrovarsi coinvolto) rimanda ai classici dilemmi etici come quello del *trolley problem*, e a concetti tipici della riflessione morale, primo fra tutti quello di responsabilità morale. Prendere le mosse da questo tipo di IA è altresì giustificato dal fatto che, al giorno d'oggi, le macchine a guida autonoma, sperimentate per prime da Google e poi da Waymo (ai quali sono seguito Tesla, Uber, Toyota e Nvidia), sono «un vero concentrato di IA»[7] perché esse devono avere

> la capacità di "vedere" la strada, i segnali, gli eventuali pedoni e gli ostacoli, perciò hanno bisogno di tecniche di interpretazione delle immagini. In più, devono prevedere le azioni dei pedoni e delle altre auto per evitare gli incidenti. Per questo hanno bisogno di ragionare statisticamente e gestire l'incertezza delle possibili situazioni che possono capitare tramite tecniche di IA come la pianificazione e la ricerca ottimizzata[8].

Se ci si rivolge quindi ai documenti, si notano invece i principi ispiratori dell'etica delle IA e quelli che debbono guidare la realizzazione degli algoritmi: trasparenza, inclusione, responsabilità, imparzialità, sicurezza e privacy, affidabilità[9]. Da questa velocissima disamina si nota come la riflessione si sia focalizzata in modo preponderante sul cosiddetto campo delle etiche applicate eludendo – o semplicemente dando per presupposta – una riflessione che si collochi piuttosto alla base del discorso etico, che ne indaghi la sua possibilità prima ancora della sua applicabilità. Senza volere (né potere) entrare qui nel dibattito tra IA debole (o leggera) o IA forte (o sarebbe meglio dire AGI), bisogna notare, infatti, che nel momento in cui parliamo di etica delle macchine si assume tacitamente (anche da parte di quegli autori che a buon diritto sono scettici se non del tutto

---

University Press, Roma 2021; P. Severino (a cura di), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*, Luiss University Press, Roma 2022.

[7] F. Rossi, *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale*, Feltrinelli 2019, p. 22.

[8] *Ibid*.

[9] Cfr., *Rome Call for AI Ethics*, in https://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febbraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_con%20firme_.pdf (ultimo accesso 12/3/2022).

negazionisti della IA forte[10]) che le macchine siano come gli esseri umani. Questo dato traspare già dal vocabolario al quale si ricorre per parlare delle macchine e delle loro operazioni.

Si parla di «*autonomia* delle macchine»[11], la quale deve «essere limitata e resa intrinsecamente reversibile, qualora l'autonomia umana debba essere protetta o ristabilita»[12]; si fa riferimento alle «*capacità* di *operare* in maniera intelligente»[13]; si parla di vero e proprio *agire*, non di un suo surrogato, al punto tale che per Floridi l'IA deve essere intesa primariamente come forma di un «agire artificiale»[14] più che di un pensare artificiale.

Anche guardando a una delle possibili (seppur abbastanza generica) definizioni di robot, si nota come essi vengano definiti in quanto «dispositivi hardware supportati da un software che permette loro di funzionare e *agire* secondo determinati *scopi* nel mondo fisico reale»[15]. L'attribuzione di elementi tipici dell'agire umano alle macchine è pertanto prassi comune e non si limita al puro ambito lessicale. Piuttosto, quest'operazione lessicale è susseguente alle applicazioni in ambito pratico in cui prodotti sofisticati come, ad esempio, Shakey, uno dei primi robot mobili, è «in grado di *pianificare* da solo le sue azioni»[16].

Se è vero, allora, che ci troviamo dinanzi a un uso metaforico del linguaggio, è altrettanto vero – e non è un fattore da sottovalutare – che ci troviamo sprovvisti di un dizionario adeguato ad affrontare questi argomenti. Questo dato richiede quindi di essere assunto come sintomo di una patologia da indagare: quando attribuiamo alle IA nozioni come responsabilità,

[10] Penso sia opportuno rammentare che, sebbene il progetto dell'IA forte non sia l'orientamento maggioritario, a esso lavorano le big dell'automazione come DeepMind.
[11] L. Floridi, *op. cit.*, p. 99, corsivo mio.
[12] *Ibid.*
[13] M. Ienca, *Intelligenza. Per un'unione di intelligenza naturale e artificiale*[2], prefazione di A. Scoccia Pappagallo, Rosenberg & Sellier, Torino 2019, p. 13, corsivo mio.
[14] L. Floridi, *op. cit.*, p. 12.
[15] F. Rossi, *op. cit.*, p. 40, corsivo mio.
[16] *Ibid.*, p. 39.

libertà, autonomia, capacità (e volutamente per questo discorso non chiamo in causa coscienza e intelligenza), sinora appannaggio esclusivo degli agenti umani, dobbiamo chiederci se esse mantengano ancora il loro classico significato, ovvero siano una semplice trasposizione dal piano naturale a quello artificiale dei medesimi concetti, o se in questo *slittamento contestuale* vengano risignificate e, conseguentemente, aiutino a ricomprendere l'umano.

Questo contributo vuole dedicarsi all'analisi di una di queste nozioni, anzi di quella nozione che può essere individuata come minimo comune multiplo di tutte le altre: la nozione di decisione e/o scelta (in questo contesto la distinzione tra le due perde di significato in quanto la seconda sarebbe l'atto conclusivo, il momento volitivo, al quale conduce la prima). Quando, infatti, intavoliamo un qualsiasi discorso etico, non facciamo altro che indicare i principi normativi dai quali una scelta dovrebbe essere diretta o forniamo un giudizio di valore sulla scelta compiuta. Cercherò quindi di circoscrivere alcuni caratteri peculiari della scelta per notare *come* e *se* le macchine abbiano influito dall'esterno sul nostro processo interno di deliberazione, per analizzare poi se quegli stessi elementi possano essere attribuiti anche alla "scelte" operate dalle IA. In altre parole, non andrò a indagare le applicazioni pratiche (e quindi etiche) della scelta nei suoi vari ambiti, ovvero gli ambiti nei quali le macchine possono dirsi responsabili e libere, tenterò piuttosto di individuare *se* esse possano essere "agenti capaci di compiere scelte".

Un tale discorso mi appare tutto fuorché inutile dal momento che, se è vero che molti pensano che gli algoritmi non potranno mai prendere delle decisioni importanti perché queste richiederebbero una dimensione etica (che gli algoritmi non possono né potranno mai comprendere), è altrettanto vero che

> non c'è ragione di presumere che gli algoritmi non saranno in grado di superare l'uomo medio anche nell'etica. Già oggi, mentre apparecchi

> come gli smartphone o i veicoli a guida autonoma prendono decisioni che un tempo erano monopolio dell'uomo, altri simili apparecchi iniziano ad affrontare problemi etici analoghi a quelli che tormentano gli uomini da millenni[17].

2. *La decisione demandata alle macchine*

Se parlare di *self driving cars* può sembrare ancora qualcosa di distopico – benché ormai non lo sia affatto – l'IA ritenuta oggigiorno in grado di prendere decisioni non si limita a quell'ambito di applicazione e, ben lungi dall'essere aliena dalle nostre vite, le permea ormai completamente. Oggi l'IA, infatti, non è perennemente presente soltanto nella vita della generazione Z e di quelle successive ma anche in quella delle generazioni precedenti. GPS, assistenti vocali, *software* per ogni tipo di lavoro, sistemi di identità digitale per accedere ai più basilari servizi al cittadino, sono ormai imprescindibili e fruibili da chi, anche superata la soglia degli ottant'anni, ha acquisito un minimo di dimestichezza digitale. L'IA è «entrata a far parte delle nostre abitudini e dei nostri gesti quotidiani, talvolta modificandoli profondamente»[18].

Questo ingresso, massiccio e non graduale, non ha trasformato soltanto le nostre abitudini ma ha cambiato anche il nostro cervello, o per essere più precisi, ha modificato le nostre capacità cognitive. Si pensi a come l'utilizzo costante del GPS (anche per muoversi nella propria città) abbia comportato – dati alla mano – una perdita notevole del senso di orientamento, inteso come la capacità di avere punti di riferimento con i quali orientarsi nello spazio[19]. Viviamo ormai in uno spazio *smart*, interattivo, ma l'interazione non è soltanto unidirezionale (da noi al *social*, o ad Alexa o a Siri, ecc.) ma bidirezionale: noi

[17] Y.N. Harari, *21 lezioni per il XXI secolo* (2005), tr. it. Bompiani, Milano 2018, p. 68.
[18] F. Rossi, *op. cit.*, p. 26.
[19] Per un'introduzione all'argomento, cfr. *Il GPS ha cambiato il nostro mondo e sta cambiando la nostra testa!*, in https://www.solotablet.it/lifestyle/il-gps-ha-cambiato-il-nostro-mondo-e-sta-cambiando-la-nostra-testa.

usiamo i *social* e loro "usano" noi; noi adoperiamo Siri e lei (o lui, a seconda della voce che abbiamo definito sui nostri dispositivi Apple) "adopera" noi, per apprendere ad esempio, ma non solo. Un primo dato che è necessario far emergere, allora, risiede nel cogliere come la nostra decisione di agenti razionali sia stata modificata dai sistemi di IA. Quando adottiamo l'IA e il suo agire *smart*, infatti,

> cediamo volontariamente parte del nostro potere decisionale ad artefatti tecnologici. Per questo, affermare il principio di autonomia nel contesto dell'IA significa trovare un equilibrio tra il potere decisionale che ci riserviamo e quello che deleghiamo agli agenti artificiali. Il rischio è che la crescita dell'autonomia artificiale possa minare il fiorire dell'autonomia umana[20].

L'impatto dell'intelligenza artificiale per le nostre vite è stato diffusamente e magistralmente descritto da Luciano Floridi, che non si limita a vedere nei progressi della *computer science* una nuova rivoluzione industriale né una rivoluzione epistemologica (sulla base della rivoluzione copernicana di Kant) ma una vera e propria rivoluzione ontologica dal momento che «il digitale ri-ontologizza il reale»[21]. Assunta questa prospettiva, allora, è di cruciale importanza comprendere cosa questa riontoligizzazione implichi per l'agire in generale e per l'agire morale in particolare. L'intelligenza artificiale, infatti, ha iscritto nel nostro universo «un nuovo agire non naturale che ha sempre maggior successo nelle sue interazioni con la realtà»[22]. L'impatto dell'IA sulle *nostre* decisioni, infatti, è già in atto.

A chi non è capitato di essere totalmente confuso quando si deve scegliere quale film guardare tra le migliaia di titoli offerti da Netflix, o quando bisogna selezionare una tra le centinaia di offerte di Glovo? È proprio in questi casi che qualcuno avrà gradito quella sezione delle nostre applicazioni che ha il titolo di *scelti per te*, una selezione *ragionata* in base ai contenuti

[20] L. Floridi, *op. cit.*, p. 98.
[21] Id., *Agere sine intelligere. L'intelligenza artificiale come nuova forma di agire e i suoi problemi etici*, in L. Floridi, F. Cabitza, *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Milano 2021, p. 66.
[22] *Ibid.*

precedentemente scelti (ma anche a quelli non scelti e che pertanto potrebbero destare il nostro interesse) volti a facilitare la nostra decisione. Se qualcuno non si è ancora trovato in queste situazioni paradigmatiche, basti pensare allora che quotidianamente i nostri *browser* di ricerca per il web decodificano i nostri interessi per selezionare i siti internet di maggior interesse per noi, cosicché se io studio filosofia in Italia alla digitazione delle lettere *cin* apparirà subito *cineca* e non *cinema*. Questo procedimento – si noti bene – non avviene nel momento in cui noi digitiamo nella barra di ricerca una qualunque parola (che il browser andrà poi a rintracciare tra miliardi di pagine web) ma avviene ancor prima che noi digitiamo le parole. Per spiegarlo meglio: la profilazione dei nostri interessi e la possibile nostra volontà di cucinare un tiramisù è calcolata dall'algoritmo molti mesi prima di quando ci vorremmo industriare nell'impresa culinaria, per esempio quando – mentre facevamo un giro per Roma – abbiamo digitato sul motore di ricerca del nostro smartphone 'migliori ristoranti a Trastevere'. Questa dinamica, che appare tanto consueta quanto comoda, modifica, anche se non ce ne rendiamo immediatamente conto, la nostra capacità di scelta, o sarebbe meglio dire di scelta *libera*.

È necessario allora fare una breve precisazione. Non sarebbe possibile neppure accennare al problema se una scelta possa dirsi libera o no, e in quale senso si debba intendere questo predicato[23]. Al fine della mia argomentazione si darà per scontato che la scelta sia libera nel senso comune che conferiamo a questo attributo, ovvero che appartiene all'agente, senza ulteriori specificazioni. Di fatto, l'oggetto di ogni riflessione etica sono sempre le scelte ritenute *libere*, e non quelle coartate o impedite.

[23] Per un primo approccio alla questione, cfr. S.F. Magni, *Teorie della libertà. La discussione contemporanea*, Carocci, Roma 2017, cap. 2.

Ora, perché si possa parlare di scelta libera è indispensabile che si abbia la possibilità – oltre che la capacità (altro distinguo necessario ma che qui può essere eluso senza che l'argomentazione ne abbia detrimento) – di scegliere tra due o più possibili opzioni. Perché possa operare una scelta tra i diversi contenuti offerti da Netflix è necessario che quei contenuti siano due o più, perché qualora avessi uno e un solo film a disposizione potrei sicuramente scegliere, ma senza che quella scelta possa ritenersi davvero genuina e libera. Nel momento in cui mi accomodo sul divano, allora, e Netflix seleziona per me 30 tra le migliaia di titoli, quale ruolo sta giocando l'algoritmo delle preferenze del colosso americano sulla mia scelta? sta forse riducendo la mia capacità di scelta, dal momento che limita le altre possibili opzioni? Stessa cosa si potrebbe dire del GPS: quando imposto un tragitto – nella mia stessa città – e il navigatore satellitare mi fornisce due o più percorsi possibili, ma io so che ve ne sono altri senza che il navigatore li indichi (a volte solo perché più lunghi di poche decine di metri anche se più lineari o più usuali per l'utente autoctono) sta forse riducendo la mia capacità di discernimento e di valutazione dei possibili corsi d'azione? In definitiva, si dovrebbe riconoscere che «la tecnocrazia invita a non scegliere, introducendoci in un orizzonte nel quale convivono "troppe" possibilità, che non siamo in grado di vagliare nemmeno volendolo, e in cui soltanto il sistema può stabilire cosa è meglio per noi»?

Reputo che la situazione non sia così drammatica e rammento una mia studentessa che – seguendo un ragionamento del tutto controintuitivo – mi disse "più scelte si hanno e meno si è liberi". Seguendo questa suggestione potremmo quindi dire che gli algoritmi designati a selezionare e discernere alcuni tra i molteplici corsi d'azione al posto nostro, piuttosto che essere identificati come un ostacolo alla nostra libertà di scelta si configurano come un *auxilium*, e non è forse questo che ricerchiamo

quando impostiamo il navigatore dal nostro smartphone? Questa dinamica, tuttavia, necessita di un ulteriore *focus*.

Nel momento in cui permettiamo all'algoritmo di *scegliere per noi*, seppur circoscrivendo la vasta gamma di opzioni percorribili, è utile chiederci cosa stiamo demandando, all'interno del processo deliberativo, alle IA. La deliberazione, infatti, caratterizzata dal soppesare diversi motivi per percorrere A o B, e quindi i *pro* e i *contro* sia di A che di B, è sempre stata un procedimento tipico del nostro apparato cerebrale, al punto tale che Thomas Hobbes riconosceva in questo procedimento la sede della libertà[24]. Oggi quel processo tipicamente umano viene demandato in parte alle IA, le quali, attraverso i loro algoritmi, selezionano per noi soltanto un *range* limitato di opzioni. Il vero problema è che le altre opzioni non vengono neanche proposte e quindi, di fatto, sono a noi del tutto precluse. Come si è avuto modo di notare, questo procedimento ci appare un aiuto all'esercizio della nostra libertà, in quanto ci aiuta a discernere solo su *concreti* e *possibili* corsi d'azione perché confacenti ai nostri interessi. Un aspetto, tuttavia, rimane però da essere rilevato; e per far ciò viene in aiuto uno dei *social* più nuovi che sta spopolando tra i *teenager*: TikTok. Il social, nato soltanto nel 2016, che si basa né su contenuti testuali (com'era in origine Facebook) né su foto (come il più recente Instagram) ma su video. A differenza dei sue social sopracitati, in cui sono gli utenti a individuare gli account o le pagine di proprio interesse che desiderano seguire (su Instagram o Facebook, ad esempio all'utente appaiono soltanto le *stories* e i contenuti dei profili che l'utente *ha scelto* di seguire), TikTok è pensato e progettato perché sia l'algoritmo a stabilire cosa tu debba vedere, in base ai tuoi interessi e ai video sui quali ti sei soffermato per più secondi. Anche in questo caso risulta particolarmente comoda una selezione dei contenuti

[24] Cfr. T. Hobbes, *Il corpo* (1655), in T. Hobbes, *Elementi di filosofia. Il corpo e l'uomo*, tr. it. Utet, Torino 1972, IV, 25, 13, p. 393.

secondo i nostri gusti, ma a cosa condurrebbe un mondo virtuale in cui siamo sottoposti solo ed esclusivamente ai contenuti in linea col nostro pensiero? Se siamo di destra vedremo solo contenuti del nostro orientamento politico, se amiamo i cani vedremo sempre e solo video con questi animali a quattro zampe e mai con felini, se siamo annoverabili tra i *novax* non ci verranno mai proposti contenuti dell'orientamento opposto. Osservato da questa prospettiva l'*auxilium artificialis* smette di avere un ruolo di *adiutorium*, per divenire un vero e proprio *impedimentum*. Se è vero che una scelta è sempre nostra perché discende dal soggetto, è altrettanto vero – benché alcune posizioni compatibiliste in merito alla libertà non concorderebbero – che quando siamo orientati, o sarebbe meglio dire influenzati solo e soltanto in una direzione, quella scelta da noi compiuta perde di efficacia e, sebbene non arriviamo a dire che essa sia stata costretta, potremmo dire che essa è stata sicuramente indotta.

Senza voler demonizzare, tuttavia, questa tipologia di algoritmi preposti alla "scelta per noi", (atteggiamento spesso portato avanti soprattutto da una certa saggistica che si concentra sugli scenari distopici) e tenendo conto che non ne potremmo fare a meno (sarebbe impossibile navigare sul web senza un algoritmo che ci orienti), rimane allora da capire *come* la decisione di selezione delle opzioni percorribili venga esercitata dall'algoritmo. Questa disamina si rende necessaria perché, mentre per gli agenti umani la valutazione dei possibili corsi d'azione è un momento previo alla scelta, per l'algoritmo la valutazione diventa vera e propria scelta, e quindi vero e proprio atto. Rimane quindi da capire in che modo l'IA prenda le sue decisioni.

3. *Le decisioni delle macchine*

È ormai famoso il caso di Eric Loomis, un cittadino statunitense che nel 2013 fu condannato dalla Corte suprema del Wisconsin, la quale – per emettere il verdetto – si basò su una valutazione del

rischio di recidiva formulata da un algoritmo: COMPAS, *Correctional Offender Management Profiling for Alternative Sanctions*. Quel caso divenne paradigmatico per le riflessioni sull'utilizzo in ambito giurisprudenziale delle IA, ma in questa sede assume rilevanza per un duplice ordine di fattori (che peraltro furono già fatti emergere dagli avvocati della difesa): l'algoritmo di COMPAS è a tutt'oggi segreto, in quanto prodotto da una società privata e tutelato dalle leggi sul diritto d'impresa, e quindi, benché si sappia che esso giunge a un verdetto combinando alcuni dati standard, non si ha conoscenza del procedimento esatto attraverso il quale viene emessa la valutazione dell'imputato[25]. Il secondo fattore rilevante è emerso a seguito di un'inchiesta condotta da *ProPublica* nel 2016, la quale, avendo condotto diverse ricerche statistiche, mostrò come l'algoritmo COMPAS presentasse forti pregiudizi nell'attribuire agli afroamericani un tasso di rischio sempre elevato[26].

[25] Per un'analisi particolareggiata del caso di Loomis, mi sia permesso rimandare a C. Calì, *L'imparzialità del giudicante. Alcune implicazioni etiche dell'utilizzo dell'intelligenza artificiale in giurisprudenza*, in *La pubblica amministrazione del futuro. Tra sfide e opportunità per l'innovazione del settore pubblico*, a cura di A. Alù e A. Ciccarello, Editoriale Scientifica, Napoli 2021, pp. 121-134.

[26] «Was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants» J. Angwin et al., *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, ProPublica, 23 maggio 2016, in www.propublica.org (8 gennaio 2021). All'inchiesta è seguito un acceso dibattito, al quale hanno partecipato sia la società produttrice di COMPAS con uno studio [cfr. W. Dieterich, C. Mendoza, T. Brennan, *COMPAS Risks Scales: Demonstrating Accuracy Equity and Predictive Parity*, in go.volarisgroup.com (6 gennaio 2021)] sia un *team* di ricercatori (cfr. A. Flores, K. Bechtel, C. Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks"*. Federal probation, 80 (2016) 2, pp. 38-46; J. Jung, et *al.*, *Simple rules for complex decisions*, in «arXiv», 1702 (2017), 04690]. A conclusioni simili a quelle edite da *ProPublica* nel 2016 sono giunti invece, nel 2018, due ricercatori del Dartmouth college, Julia Dressel e Hany Farid, i quali hanno condotto uno studio per mostrare che nel valutare la potenziale recidività di un individuo, COMPAS non è più affidabile di un gruppo di volontari scelti a caso su internet. Cfr. J. Dressel, H. Farid, *The accuracy, fairness, and limits of predicting recidivism*, Science Advances, 4 (2018), eaao5580.

Dinanzi all'assenza di trasparenza del procedimento algoritmico che conduce all'azione sembrerebbe configurarsi la prima grande differenza col processo decisionale degli animali razionali. Il *libero accesso* al processo con cui si è addivenuti a un *output*, infatti, se è facile da ottenere con un'IA basata su un approccio simbolico, è pressoché impossibile da ottenere quando si adoperano IA basate sul *machine learning*. Questi sistemi, infatti, «non descrivono i passi necessari per risolvere un certo problema, ma sanno solo apprendere come risolverlo a partire dall'osservazione di una grande quantità di esempi»[27]. Si deve notare il fatto per cui mentre questo tipo di IA non sono in grado di riconsegnare il procedimento che ha condotto a un determinato esito, un agente è sempre capace – salvo casi di lesioni cerebrali o di ipnosi, o altri stati di coscienza attenuata – di dare conto della propria decisione e del perché sia addivenuto a tale verdetto, fosse anche che quest'ultimo suoni come "ho voluto perché ho voluto", "sono stato spinto dalla foga del momento" o altre argomentazioni che dicono molto poco dei reali motivi per i quali si è agito. Il nostro cervello, inoltre, non soltanto ha la capacità di ricostruire i processi decisionali ma, nientemeno, anche di inventarli *ex post*. Per dirla con un gergo più tecnico, il cervello riesce a *confabulare*, ovvero a razionalizzare *a posteriori* un'azione intrapresa non consapevolmente o, quantomeno, un'azione che abbiamo preso in modo leggermente differente da quella di cui poi ci viene richiesto di dare conto[28]. Come si è accennato, questo processo con l'IA è al momento impossibile, e questo fattore, definito come argomento della *black box*, è stato considerato un problema talmente inaggirabile da essere utilizzato sovente per gettare discredito sull'IA, tant'è che è stato

[27] F. Rossi, *op. cit.*, p. 52.

[28] Gli esperimenti in merito sono innumerevoli, ne riporto due particolarmente significativi: R.E. Nisbett, T.D. Wilson, *Telling More Than We Can Know*. Verbal Reports on Mental Processes, in «Psychological Review», 84, 1977, pp. 231-259. P. Johansson, L. Hall, S. Sikström, A. Olsson, *Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task*, in «Science», 310, 2005, pp. 116-119.

fortemente contrastato dalla normativa europea secondo la quale un soggetto deve poter accedere al meccanismo col quale l'IA ha preso una decisione, almeno in quelle circostanze in cui quest'ultima ha un impatto sulla vita del soggetto[29]. È altresì vero che questo problema si sta avviando, seppur lentamente, a una risoluzione dal momento che diversi ricercatori stanno lavorando a fornire all'IA la capacità di spiegare perché abbia preso una decisione (es. ha interpretato un'immagine come un possibile tumore o no), ovvero della capacità di *spiegare* le proprie decisioni. Non entro qui nella questione, perché lo sviluppo di questa capacità algoritmica è appena incominciato e io stesso sto conducendo alcune ricerche in merito: solo un appunto – neanche troppo originale – ritengo però possa essere mosso[30]. Mancando all'IA delle preferenze, delle inclinazioni, in quanto le IA sono prive di carattere, in base a cosa la scelta sarebbe operata? Sarebbe soltanto un risultato ottimale a seguito di un calcolo? Anche ammettendo quest'ultima possibilità, viene da chiedersi quante volte noi calcoliamo approfonditamente e soppesiamo accuratamente i pro e i contro prima di deciderci ad agire[31]. Sembra ancora una volta come l'IA, lungi dal fare qualcosa di anche vagamenti simile all'attività umana, crei qualcosa di profondamente diverso, una nuova forma di *agere*, un *agere* che Floridi definisce come «agere sine intelligere»[32]. Rimane quindi da analizzare il secondo elemento che ho fatto emergere portando l'esempio di COMPAS e di Eric Loomis: i *bias* impliciti presenti anche negli algoritmi.

[29] Per i casi giuridici e i riferimenti normative dell'Unione Europea, cfr. T. Cassauwers, *Opening the "black box" of artificial intelligence*, in https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence (12/3/2022).

[30] Ho già suggerito una simile riflessione nel momento in cui si valuta l'utilizzo delle IA per scelte e previsioni in ambito giuridico, cfr. C. Calì, *op. cit.*, pp. 132-134.

[31] Si noti per altro che questa critica è spesso mossa contro coloro che difendono posizioni libertarie in ordine alla libertà metafisica. Si veda a tal proposito, R. Kane, *The Significance of Free Will*, Oxford University Press, New York 1996.

[32] L. Floridi, *Etica*, cit., p. 13.

Un anno dopo la condanna di Loomis con l'ausilio di COMPAS, nel 2014, si riscontrò un errore nel *software* utilizzato da Amazon per lo screening dei *curricula vitae* dei candidati da assumere. Si notò come l'algoritmo penalizzasse i candidati di sesso femminile soprattutto quando si trattava di posizione nel settore tecnologico dell'azienda[33]. In quel caso «il modello ha riconosciuto in modo automatico un pattern che delineasse i migliori candidati, inglobando tra le caratteristiche ideali il genere maschile, e incorrendo così in un bias»[34]. L'errore – viene detto in molti dei commenti sulla vicenda – «era dovuto ai dati con cui il modello è stato addestrato: dati reali, contenenti i curricula ricevuti dalla società nei 10 anni precedenti; CV prettamente maschili, data la maggioranza di uomini nel settore tecnologico»[35].

Nel 2018, una giovane ricercatrice per Microsoft, poi assunta da Google, insieme con un collega del MIT, notò come certi algoritmi modellassero le elaborazioni linguistiche e il riconoscimento facciale su basi razziste, sessiste e offensive[36] (una precedente versione del software di Google, ad esempio, etichettava come *gorilla*, i cittadini di pelle scura[37]). L'indagine si estendeva ad alcuni algoritmi adoperati da altri fra i GAFAM. Gli algoritmi che stanno alla base del funzionamento di Google Cloud Vision, Microsoft Azure Computer Vision e Amazon Rekognition, ad esempio,

[33] Cfr. J. Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, in https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, (ultimo accesso 22/4/2022).

[34] Cfr. *Amazon: intelligenza artificiale discriminava curriculum donne*, in https://www.ansa.it/sito/notizie/tecnologia/tlc/2018/10/10/amazon-stampa-intelligenza-artificiale-discriminava-donne_1b1ebaca-f000-48a6-89b4-ff9854ef75e7.html (ultimo accesso 22/4/2022).

[35] G. Rizzi - M.T. Cimino, *Bias negli algoritmi: come le macchine apprendono i pregiudizi dagli esseri umani*, in https://ibicocca.unimib.it/bias-negli-algoritmi-come-le-macchine-apprendono-i-pregiudizi-dagli-esseri-umani/ (ultimo accesso 21/4/2022).

[36] Cfr. J. Buolamwini - T. Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in «Proceedings of Machine Learning Research», 81(2018), pp. 1-15.

[37] Cfr. C. Dougherty, *Google Photos Mistakenly Labels Black People 'Gorillas'*, in https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/ (ultimo accesso 22/4/2022).

osservavano e valutavano donne e uomini in modo molto differente, stimando le prime su criterii legati all'aspetto estetico e sessuale[38]. Noto *en passant*, come nel 2020 la ricercatrice – dedicatasi all'interno dell'azienda alla *machine ethics* – sia stata licenziata dal colosso della Silicon Valley a causa –stando alla versione ufficiosa – di un articolo scientifico in cui esponeva i numerosi rischi legati allo sviluppo delle IA.

Nel 2021, infine, il tribunale del lavoro di Bologna, con una sentenza storica, ha rintracciato atteggiamenti discriminatori nell'algoritmo di nome Frank, utilizzato da Deliveroo, nota piattaforma di *food delivery*. L'algoritmo discriminava i riders autonomi sulla base delle loro performance senza distinguere tra quelli che si assentavano per motivi futili e quelli che invece ricevevano un annullamento dell'ordine. Anche in questo caso è stato detto che «non è colpa dell'algoritmo: semplicemente Frank ripete e applica, in modo più o meno continuo, le direttive che gli sono state impiantate»[39].

È evidente come se la situazione in ordine alla decisione umana demandata alle IA destava qualche perplessità, la situazione in riferimento alle decisioni *delle* IA non versa in condizioni migliori. Benché in molti ripetano che l'IA sia neutra, in quanto non esiste un'IA buona e una cattiva, diviene sempre più difficile definire chi è il soggetto responsabile per una determinata decisione. Il tribunale di Bologna ha condannato e obbligato al risarcimento Deliveroo, ma è davvero l'azienda londinese di Will Shu a essere responsabile o, piuttosto, i programmatori assunti dalla stessa o, ancora, coloro che 'tra gli umani' procedono troppo spesso a valutazioni sommarie e poco accurate, sulla base delle quali l'IA apprende il proprio *modum operandi*? Anche su

[38] C. Schwemmer, C. Knight, E.D. Bello-Pardo, S. Oklobdzija, M. Schoonvelde, J.W. Lockhart, *Diagnosing Gender Bias in Image Recognition Systems*, in «Socius: Sociological Research for a Dynamic World», 6, 2020, pp. 1-17.

[39] E. Verga, *Intelligenza artificiale, la discriminazione da parte degli algoritmi è un pericolo reale*, in https://tech4future.info/discriminazione-degli-algoritmi-di-intelligenza-artificiale/ (ultimo accesso 22/4/2022).

questo secondo versante, nondimeno, si sta correndo ai ripari, grazie al campo di ricerca definito *algorithmic fairness*,

> volto a mitigare gli effetti di pregiudizi e discriminazioni ingiustificate sugli individui nell'apprendimento automatico, principalmente incentrato sul formalismo matematico e sulla ricerca di soluzioni per questi formalismi. È un ambito di ricerca interdisciplinare che ha l'obiettivo di creare modelli di apprendimento in grado di effettuare previsioni corrette dal punto di vista di equità e giustizia[40].

Anche in questo caso però il compito è particolarmente arduo dal momento che una prima difficoltà si riscontra nell'incapacità di ottenere una definizione esaustiva e universale di correttezza (*fairness*)[41], poiché quest'ultima risulta enormemente variabile in base ai contesti sociali, politici e religiosi di riferimento. È proprio a partire da questo tentativo col quale i programmatori vogliono correre ai ripari che la riflessione etica in riferimento alle IA deve interrogarsi. Ammesso che si possa dare responsabilità senza coscienza e, altresì, responsabilità senza libertà, è ancora opportuno continuare a utilizzare un termine che sin dalla notte dei tempi è stato, è, e sembra ancora essere – nonostante qualsiasi descrizione del mondo la fisica ci fornisca sulla nostra libertà – un termine imprescindibile per l'agire umano?

Floridi, come pocanzi accennato, si è mosso da tempo in questa direzione, comprendendo l'IA come una «nuova forma di agire intelligente, determinata dal disallineamento digitale tra azione e intelligenza»[42], un *agere sine intelligere*, al punto tale che l'ormai consueta sigla di AI (*artificial intelligence*) potrebbe essere anche riformulata in AA (*artificial agency*). Questa riformulazione dell'usuale binomio tra intelligenza e artificio in termini di *agency* penso sia cruciale dal momento che viene separata l'*attività* dalla *razionalità*, un binomio canonico tanto

[40] G. Rizzi, M.T. Cimino, *Bias negli algoritmi*, cit.
[41] A tal proposito, cfr. N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, *A Survey on Bias and Fairness in Machine Learning*, in «ACM Computing Surveys», 54, 2021, pp. 1-35.
[42] L. Floridi, *Etica*, cit., p. 21.

nell'antichità quanto nel periodo della Scolastica. Questo scollamento ha notevoli ripercussioni e fornisce innumerevoli prospettive d'indagine. Qui mi limito a percorrerne soltanto una. Nel momento in cui la decisione artificiale non è informata dall'intelligenza sembra configurarsi, ancora una volta, una situazione ibrida in cui la macchina viene in aiuto all'umano e l'umano corregge o sistema l'operato della macchina. Alla luce di ciò non si parlerà più di *autonomia* ma di meta-autonomia, o modello di decisione di delega:

> Gli esseri umani dovrebbero mantenere il potere di decidere quali decisioni prendere, esercitando la libertà di scelta dove necessario e cedendola nei casi in cui ragioni di primaria importanza, come l'efficacia, possano prevalere sulla perdita di controllo sul processo decisionale. Ma qualsiasi delega dovrebbe anche rimanere in linea di principio rivedibile, adottando come ultima garanzia il potere di decidere di decidere di nuovo[43].

Anche per la scelta, allora, si avrà una decisione composita in cui interagiscono tanto gli esseri umani quanto le macchine. Oggi, dice Marcello Ienca, la nostra specie si trova a un crocevia epocale:

> Davanti a noi, infatti, si profila una transizione storica senza precedenti noti nella storia del pianeta: quella nella quale un'intelligenza biologica inizia a utilizzare ricorsivamente un'intelligenza artificiale da lei stessa creata al fine di unirsi a essa e, in tal modo, autopotenziarsi[44].

Proprio in riferimento alla scelta, tuttavia, ritengo che i giudizi debbano essere molto più cauti di quanto non si è fatto in ordine alle macchine utilizzate in ambito aziendale o per fini medici. Quando deliberiamo, ma soprattutto quando valutazione, un computer è nettamente più affidabile di noi stessi, e, quando ricorriamo a certi algoritmi, di fatto «siamo spinti a rinunciare alla "libera scelta" e ad affidarci al sistema. È questa la promessa di ogni delega tecnocratica: *liberarci dalla libertà*»[45]. Ma dietro a questa efficienza nella valutazione vi è anche una

[43] *Ibid.*, p. 99.
[44] M. Ienca, *Intelligenza²*, cit., p. 14.
[45] N. Bellanca, *La libertà al tempo dell'Intelligenza Artificiale*, in https://www.academia.edu/43738365/La_libertà_al_tempo_dellIntelligenza_Artificiale (ultimo accesso 30/4/2021).

qualità della valutazione? «Se deleghiamo all'intelligenza artificiale alcune delle nostre decisioni, siamo certi che ne sappia valutare le conseguenze e quindi sappia prendere le decisioni migliori? Se in alcune attività la tecnologia diventa più *capace* di noi, quale finirà per essere il nostro ruolo»[46]?

Penso che questi interrogativi ci impongano di scandagliare non più il nesso intelligenza e artificio ma quello tra libertà e artificio. Nel momento in cui separiamo l'azione della macchina dall'*intelligere*, poiché l'IA «non concerne la capacità di riprodurre l'intelligenza umana, ma in realtà la capacità di farne a meno»[47], a cosa riconduciamo l'azione artificiale? Vedo in questo una riproposizione, *mutatis mutandis*, della *querelle* medievale sulla determinazione della volontà da parte della ragione o da parte della volontà stessa. Come la discussione sull'intelligenza applicata alle macchine ha imposto un ripensamento della nozione di intelligenza, o quantomeno ha riacceso la riflessione, lo stesso si deve dire della libertà, la quale viene illuminata nuovamente e in modo differente dalle discussioni legate all'intelligenza artificiale.

Alla luce di questo dato incontrovertibile, allora, è necessario ripensare il vocabolario e il contenuto dell'etica non adattando le classiche etiche – nate per un'*agency* umana – ma formulando un'etica artificiale per un'"agency" artificiale, sempre ammesso che della macchina si possa predicare un'*agency*.

Cristiano Calì è PhD candidate presso la Facoltà di Teologia di Lugano e Docente invitato presso lo Studio Teologico S. Paolo di Catania
cristianocali30@icloud.com

[46] F. Rossi, *op. cit.*, p. 12, corsivo mio.
[47] L. Floridi, *Etica*, cit., p. 52.

SETTIMO TERMINI

L'IA tra Katzuo Ishiguro e Federico Faggin

1. *Introduzione* 2. *Ricezione pubblica e impatto sulla società e la vita quotidiana*
3. *Legami storici e rapporti con la fisica, con la cibernetica (e con Napoli)*
4. *Utopia di un mondo liberato dalla fatica* 5. *Scenari*
6. *Qualche ultima riflessione prima di concludere* 7. *A mo' di conclusione*

ABSTRACT: ***AI between Katzuo Ishiguro and Federico Faggin***

The themes proposed by the Journal for a critical discussion are crucial for understanding the impact that AI can have in Society (and also in everyday life), as is the case with all "science-based" cutting-edge technologies, i.e. those founded on and closely intertwined with basic frontier research. The questions posed can, in my view, be more satisfactorily addressed if seen from a long-term perspective with a historical reference to the origins and development of AI, its progress over the last few decades, the fluctuating consideration of it by both public opinion and the scientific community, its relations with other fields of research, in particular Cybernetics, whose heir it can in many respects be considered to be from the point of view of the sociology of science. This is the perspective from which useful paths can be indicated to clarify and deepen the themes posed even if no definitive answers can, presently, be given. The unusual reference to books such as those by Ishiguro and Faggin is intended to indicate the profound complexity and intricacies of the questions posed by AI.

1. Introduzione

Le tematiche proposte nel "Call for papers" sono cruciali per capire l'impatto che l'IA può avere nella società (e anche nella vita quotidiana), come peraltro tutte le tecnologie di punta "science-based", cioè che non sono applicazione di risultati ben noti e consolidati ma sono basate su nuovi risultati scientifici di frontiera. Le domande poste possono, a mio avviso, essere più agevolmente affrontate se viste in una prospettiva di lungo periodo con un richiamo storico alle origini e allo sviluppo dell'IA, al suo andamento nel corso degli ultimi decenni,

all'oscillazione della sua considerazione da parte sia dell'opinione pubblica sia della comunità scientifica, ai suoi rapporti con altri settori di ricerca, in particolare con la cibernetica di cui, per molti versi, può considerarsi l'erede.
Per molti di questi temi mi limiterò essenzialmente a fornire una sorta di mappa sintetica con "puntatori" ad alcuni riferimenti bibliografici per chi volesse approfondire quanto solo accennato. Lo spazio dedicato a presentare e discutere suggerimenti provenienti da testi "insoliti" deriva dalla convinzione che le domande poste dalla Rivista possano meglio essere affrontate da una prospettiva che includa punti di vista di natura varia quando significativi. Una ricezione pubblica insolita e un programma di ricerca in una direzione non ortodossa sono indicate già nel titolo. Rinviando a quanto sarà scritto successivamente, desidero anticipare che Ishiguro narrando[1] del robot Klara e del suo rapporto con gli umani (e con la natura) ci fa riflettere – nella maniera più coinvolgente possibile – di come le sfide dell'IA possano contribuire a farci crescere interiormente e a conoscere meglio anche noi stessi. Faggin, pioniere tecnologico e inventore, tra l'altro, del microprocessore, pone in un modo assolutamente radicale la tesi che la coscienza non sia "riducibile" alla materia[2]. l'IA, quindi, pone molte sfide concettuali, che sono cruciali per l'oggi pur essendo fortemente radicate nel passato.

2. Ricezione pubblica e impatto sulla società e la vita quotidiana

Partiamo da alcune domande. Che impatto hanno l'IA e le nuove tecnologie nella società e nella vita quotidiana? Le enormi possibilità che si aprono e i pericoli possibili sono qualcosa di cui il cittadino comune è cosciente? Sono questi dei temi su cui la stampa svolge una funzione di sensibilizzazione? La risposta

[1] K. Ishiguro, *Klara e il sole* (2021), tr. it. Einaudi, Torino 2021.
[2] F. Faggin, *Silicio. Dall'invenzione del microprocessore alla nuova scienza della consapevolezza*, Mondadori, Milano 2019. Vedi anche il sito della Fondazione Faggin http://www.fagginfoundation.org/it/.

non può che essere affermativa se possiamo leggere su un giornale come il New York Times frasi come quelle che seguono scritte da uno scienziato di altissimo livello:

> By this time the public is well aware that a new age of machines is upon us ... The tendency of these new machines is to replace human judgment on all levels ... It is already clear that this new replacement will have a profound influence on our lives, but it is not clear to the man of the street what this influence will be. ... These New machine have a great capacity for upsetting the present basis of Industry, and of reducing the economic value of the routine factory employee to a point at which he is not worth hiring at any price. … Moreover, if we move in the direction of making machines which learn and whose behaviour is modified by experience, we must face the fact that every degree of independence we give the machine is a degree of possible defiance of our wishes. … It is only a humanity which is capable of awe, which will also be capable of controlling the new potentials which we are opening for ourselves[3].

Prima di continuare, chiedo al lettore di indovinare l'anno in cui i brani precedenti sono stati scritti e chi ne sia l'autore. Ho cercato di indurre i lettori in una trappola temporale e non so con quanti questo gioco sia riuscito. I brani riportati sono stati scritti da Norbert Wiener e sono stati presi da un articolo che *avrebbe dovuto* essere pubblicato sul NYT nel 1949. Non lo è stato, in realtà ma solo per motivi casuali, come riportato dallo stesso NYT[4] nel 2013, a causa di successive richieste di chiarimenti e revisioni da parte del direttore del giornale a cui l'autore ebbe difficoltà a rispondere trovandosi fuori degli Stati Uniti. Ma il punto che a noi interessa è che l'articolo è stato scritto nel 1949. Ci siamo accorti, leggendo quegli estratti, che l'articolo era vecchio di settantatré anni? Forse da sfumature linguistiche ma non certo per i contenuti. Questo ci permette di fare immediatamente due commenti. Il primo è che le domande che oggi ci stiamo giustamente ponendo si ponevano già prima che il nome IA fosse introdotto: solo un anno dopo che lo stesso Wiener aveva

[3] Cfr. N. Wiener, *The Machine Age*, http://monoskop.org/images/3/31/Wiener_Norbert_The_Machine_Age_v3_1949.pdf.

[4] J. Markoff, *in 1949, He Imagined an Age of Robots - Norbert Wiener, the visionary mathematician whose essay "The Machine Age" languished for six decades in the M.I.T. archives*, New York Times, May 20, 2013. https://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html

pubblicato il suo libro sulla cibernetica. E un anno prima di un suo successivo libro nel quale riportava sue considerazioni sull'impatto di queste nuove tecnologie che già dal titolo indicava la direzione nella quale avremmo dovuto muoverci[5]. Certo dopo settant'anni di sviluppi possiamo esemplificare in modo più sofisticato e ricco ogni tema ma i problemi generali, di fondo, sono gli stessi. E se volessimo evidenziare continuità e differenze potremmo leggere un articolo appena uscito[6] in parallelo a quello fortemente voluto dal NYT, pur senza successo, nel 1949. Gli argomenti specifici trattati sono, in questo articolo recente, molto più specifici ma le domande di fondo sono essenzialmente le stesse.

Anche per quanto riguarda i possibili pericoli (dell'IA), questi non sono qualcosa di nuovo legato agli ultimissimi sviluppi ma si possono utilmente rintracciare in quanto ipotizzato già a suo tempo da Wiener (considerati così cruciali già allora da avere avuto una richiesta di articolo dal New York Times, come abbiamo visto). Come scrive Katherine Hayles: «the problem that Wiener encountered was how to restrain this revolutionary potential of Cybernetics so that it would not threaten the liberal humanism that so deeply informed his thinking»[7]. Il problema, dunque, non è totalmente nuovo, tutt'altro. I problemi sono da aggiornare, ovviamente, anche alla luce di quello che nel frattempo è successo. Katherine Hayles, proprio nella stessa pagina, poche righe più sotto – riportando alcune analisi fatte sulla "militarizzazione dell'apprendimento" nei percorsi educativi degli Stati Uniti – riferendosi a una di queste, scrive che essa «indicates that Wiener's antimilitary stance was not sufficient to

[5] N. Wiener, *The Human Use of Human Beings: Cybernetics and Society*, Houghton Mifflin, Boston 1950.
[6] S. Johnson, *A.I. Is Mastering Language. Should We Trust What It Says?*, New York Times, April 15, 2022 (ultimo accesso 17 aprile 2022).
[7] K. Hayles, *How we became post human*, The University of Chicago Press, Chicago 1999, p. 119.

prevent the marriage of war and cybernetics, a union that he both feared and helped to initiate».

Vorrei invitare il lettore a confrontare le sagge considerazioni di Wiener (nonostante la contraddizione appena riportata) con le "preoccupazioni" di Henry Kissinger[8], recentemente espresse in un libro[9]. Libro e preoccupazioni da esaminare e maneggiare con estrema cura e direi anche un po' di diffidenza, perché – a differenza delle preoccupazioni profondamente umanistiche di Wiener – Kissinger, presentando le paure dell'uomo comune per aspetti inquietanti di queste nuove tecnologie, non ci induce ad andare in direzione di un rafforzamento della trasparenza democratica – ad esempio, un maggiore controllo decentrato da associare a uno sviluppo della "cittadinanza scientifica" - ma spinge in direzione di un controllo verticistico della politica sulla ricerca. Lo stesso, a mio avviso, può dirsi delle perplessità avanzate contro l'IA da Bill Gates, così come quelle di Elon Musk e Jeff Bezos. Giuste, ovviamente, le critiche a un uso non controllato e non etico delle nuove tecnologie soprattutto in campo militare ma, come abbiamo notato, tutto questo dibattito era già presente da molti decenni e il modo con cui questi nuovi "convertiti" al problema ne parlano non presenta nessuna argomentazione innovativa. Inoltre, questi temi sono inseriti in un discorso generale su cosa sia l'IA di una vaghezza estrema e senza mai parlare dei meccanismi di potere nei quali queste considerazioni devono essere inserite se non si vuole che restino inutilmente retoriche (per chi le ascolta) o consolatorie (per chi le fa). Parzialmente incomprensibili mi appaiono, poi, le affermazioni di Stephen Hawking nelle quali la paura si alterna

[8]H. Kissinger, How the Enlightenment Ends, *Atlantic Monthly*, June 2018 (https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/), apparso in italiano il 21 luglio 2018 su Robinson, La Repubblica (https://www.repubblica.it/robinson/2018/07/21/news/un_intelligenza_da_paura_kissinger_mette_in_guardia_dalle_ai-300769791/).

[9] H. Kissinger, E. Schmidt, D. Huttenlocher, *The Age of AI and our Human Future*, Little, Brown and Company, New York 2021.

all'ottimismo, in un miscuglio intelligente e banale allo stesso tempo. La saggezza di Wiener la ritroviamo, invece, in molti recenti interventi pubblici di Giorgio Parisi. Tra le analisi puntuali si devono ricordare quelle di Karen Hao come accennerò – brevemente – in conclusione. Le analisi devono essere fatte con serietà e tenendo conto dell'estrema complessità del problema. Wiener rimane un riferimento basilare anche oggi, grazie alla ricchezza del suo pensiero. E per apprezzarlo pienamente ci è di guida la profonda analisi effettuata da Leone Montagnini[10] che ce lo presenta – correttamente – non solo come un raro matematico-filosofo ma anche come un grande interprete del nostro tempo.
Parlando di Wiener e raccontando l'episodio del 1949 che lo riguarda, ho introdotto praticamente tutti i temi che tratterò e che formano il mio abbozzo di risposta a (quasi) tutte le domande poste nel Dossier.

3. *Legami storici e rapporti con la fisica, con la cibernetica (e con Napoli)*

In quest'ambito la domanda relativa alle implicazioni di ciò che è definito "incontro tra IA e meccanica quantistica" – pioneristicamente avanzata da Roger Penrose - può essere maggiormente capita e approfondita se vista nell'ottica di un esame più vasto dei problemi cruciali dell'IA, come ad esempio delineato recentemente da Giorgio Parisi che ha dato indicazioni su aspetti da approfondire ribadendo che la maggior parte delle *applicazioni* dell'IA si possono tutte ricondurre a un settore *molto specifico* di indagini. Si fa riferimento a idee (reti neurali) che hanno visto la comunità scientifica italiana (e Napoli in modo specifico) in una posizione originale nella quale è stata la comunità dei fisici a introdurre quelle che negli anni

[10] L. Montagnini, *Harmonies of Disorder: Norbert Wiener: A Mathematician-Philosopher of Our Time*, Springer Biographies, Springer International Publishing, 2017.

'60 erano tecnologie di punta associate a una disciplina scientifica allora emergente, la cibernetica, appunto.

È importante ricordare che Wiener è venuto varie volte in Italia e in particolare proprio a Napoli[11] nel contesto di quello che è stato chiamato il *risveglio scientifico di* questa città[12]. Per quanto riguarda l'aspetto che stiamo considerando l'autore del risveglio è stato Eduardo Caianiello[13]. Il suo interesse per la cibernetica si focalizzò sulla possibilità di costruire modelli del cervello adeguati allo studio dei processi mentali, sulla scia, in particolare, di un libro di Ashby[14]. Il suo contributo scientifico fondamentale in questo nuovo campo[15] è apparso nel 1961 sul primo numero di una nuova rivista dal nome – per allora - insolito (biologia teorica) che lascia intravedere contenuti fortemente innovativi per l'epoca. Questo lavoro non solo è alla base di tutti i suoi lavori successivi sulle reti neurali, ma contiene anche tutti gli elementi significativi del suo modo di affrontare i problemi della cibernetica. Lo scritto, oltre a contenere risultati originali, è un vero e proprio manifesto programmatico ancora oggi significativo[16]. Questo lavoro, un modello matematico di rete neurale, è scientificamente innovativo rispetto a modelli precedenti perché basato su intuizioni di tipo fisico (piuttosto che precipuamente logico) sul suo funzionamento. Le reti neurali hanno svolto e svolgono un ruolo centrale anche su alcuni degli sviluppi più importanti dell'IA,

[11] L. Montagnini Quando Wiener era di casa a Napoli, *La rivista del Centro studi di Città della scienza*, 21 marzo 2016, http://www.cittadellascienza.it/centrostudi/2016/03/quando-heisenberg-e-wiener-erano-di-casa-a-napoli/

[12] G. Barone, P. Greco, L. Mazzarella, A.M. Liquori, *Il risveglio scientifico negli anni '60 a Napoli*, Bibliopolis, Napoli 2013.

[13] Cfr. S. Termini, *Caianiello, Eduardo Renato*, Dizionario Biografico degli Italiani, 2017. https://www.treccani.it/enciclopedia/eduardo-renato-caianiello_%28Dizionario-Biografico%29/ e anche S. Termini (ed), *Imagination and Rigor, Essays on Eduardo Caianiello's Scientific Heritage*, Springer, 2006.

[14] W. Ross Ashby, *Design for a brain*, London, 1952.

[15] E.R. Caianiello, *Outline of a theory of thought-processes and thinking machines* in «Journal of theoretical biology», I (1961), 2, pp. 204-235.

[16] Cfr. http://www.treccani.it/enciclopedia/la-cibernetica *(Il-Contributo-italiano-alla-storia-del-Pensiero:-Scienze)*.

come richiamato anche da parte di Giorgio Parisi nel suo libro rivolto al grande pubblico[17], parlando dell'introduzione del disordine. Il rapporto di Parisi con l'IA è testimoniato oltre che da molte sue dichiarazioni pubbliche anche dal conferimento del titolo di Socio onorario che l'Associazione Italiana per l'Intelligenza Artificiale gli ha conferito. Se ricordiamo questo aspetto è per mettere in evidenza una sorta di positiva anomalia italiana che ha visto la comunità dei fisici interessati a introdurre e finanziare nuove ricerche in campi di frontiera di nuove tecnologie, come ha fatto negli anni '60 del secolo scorso con la cibernetica. Sia pure in forma minore, quanto riportato sopra va nella stessa linea e con grandi potenzialità, se perseguito efficacemente perché è stata proprio l'interazione interdisciplinare di metodologie scientifiche diverse che dialogavano tra loro a produrre gli intensi e proficui dibattiti avvenuti a metà del secolo scorso[18].

4. *Utopia di un mondo liberato dalla fatica*

Per quanto riguarda "L'utopia di un mondo liberato dalla fatica" questa non è un'utopia e non ha bisogno di attendere "un'industria (5.0)". Profetizzata, o, in modo più accurato, semplicemente "prevista" da Keynes negli anni '30 del secolo scorso[19], è realtà già oggi - *in linea di principio*. Cerco di essere più chiaro. Le sue previsioni riguardo allo sviluppo della tecnologia si sono puntualmente avverate. Quindi non avremmo bisogno di attendere "un'industria (5.0)". Certo ulteriori passi nella stessa direzione sono auspicabili. Ma il problema da affrontare è, come detto

[17] G. Parisi, *In un volo di storni. Le meraviglie dei sistemi complessi*, Rizzoli, Milano 2021,

[18] Cfr. L. Montagnini, M.E. Tabacchi, S. Termini, *Out of a creative jumble of ideas in the middle of last Century: Wiener, interdisciplinarity, and all that*, in «Biophysical Chemistry», 208, 2016, pp. 84–91; M.E. Tabacchi, S. Termini, *The human use of human beings": Interdisciplinarity, transdisciplinarity and all that in Biophysics and beyond*, in «Biophysical Chemistry», 229, 2017, pp. 165-172

[19] J.M. Keynes, *Possibilità economiche per i nostri nipoti* (1930), Adelphi, Milano 2009.

prima, quello della reazione della società di fronte a un cambiamento di questo tipo. E il Keynes così preveggente riguardo allo sviluppo tecnico si è invece completamente sbagliato riguardo al secondo aspetto. La società non è riuscita a modificare un modello di sviluppo che impedisce il pieno dispiegarsi degli effetti dello sviluppo tecnologico. Ma, a mio avviso, se oggi è aumentata la disoccupazione e la precarietà – nonostante le previsioni di Keynes si siano verificate per quanto riguarda lo sviluppo della tecnologia –, volendo affrontare il contributo che lo sviluppo di "un'industria (5.0)" può dare per la realizzazione di "un mondo liberato dalla fatica", dobbiamo pur porci la domanda di quanto le condizioni al contorno – sociali, economiche, politiche – possano favorire o soffocare nella culla tale possibilità, indipendentemente da quanto siamo riusciti a sviluppare "un'industria (5.0)". Per questi motivi, questa domanda la derubricherei dalla lista di ciò che riguarda l'IA da sola e creerei una nuova lista che, con un titolo alla Lina Wertmuller, chiamerei: «Ciò che l'IA può aiutarci a fare se riusciamo a risolvere contemporaneamente qualche altro problema». Ma torniamo a Keynes. Le sue idee – mi riferisco proprio e solo alle *idee*, in questo momento, anche se un discorso più ampio sarebbe da fare proprio in relazione ai temi della tecnologia e dell'IA[20] – a proposito del tema del liberarsi dalla fatica, sono state considerate ancora degne di essere ridiscusse se nel 2009 è uscito un volume[21] nel quale vari economisti rileggono questo piccolo breve scritto degli anni '30 del Novecento. Ricordiamo che anche nell'edizione italiana il suo breve scritto non compare da solo ma è seguito da una nota di Guido Rossi, con lo stesso titolo ma col punto interrogativo. È quindi il caso che chi si interessa di

[20] Vedi, ad esempio, Daniela Palma, *L'Italia, il declino e il falso mito dello "sviluppo senza ricerca"*, in «Nuova Lettera Matematica», 4, 2021, pp. 9-20; e A. Bellocchi, G. Travaglini, *Attualità di Keynes*, in «Nuova Lettera Matematica», 4, 2021, pp. 37-44, nonché tutto il dossier su *Ricerca e Innovazione* contenuto nello stesso fascicolo della Rivista.

[21] L. Pecchi, G. Piga (a cura di), *Revisiting Keynes: Economic Possibilities for our Grandchildren*, MIT Press, 2008.

industria 5.0 lo faccia rileggendo queste considerazioni. Da sviluppare, da adattare alla situazione odierna ma fondamentalmente corrette per quanto riguarda il tema centrale.

5. *Scenari*

Avanzando un tentativo di risposta all'ultima domanda, direi che gli "scenari narrativi … [che] si squadernano di fronte a *sapiens*" sono fondamentalmente gli stessi di cinquanta/settanta anni fa quando l'IA (e, un po' prima, la cibernetica) fecero la loro comparsa. Naturalmente con una maggiore consapevolezza delle vie che possono essere più agevolmente percorse. Un esempio cruciale è fornito dalla traduzione automatica. È interessante confrontare il modo di impostare il problema negli anni '50 (e i magri risultati conseguiti) con le impostazioni più recenti che invece hanno prodotto risultati notevoli. La metodologia che sta dietro è profondamente diversa. Il tema è da collegare anche a domande sulla natura epistemologica di tecniche come quelle del *deep learning*.

Un esempio significativo che può contribuire a chiarire il problema è dato dai tempi intercorsi tra la formulazione di un problema e la sua soluzione. Nel caso del gioco degli scacchi i tempi sono stati lunghi, in quello del gioco del GO, ridottissimi, pur essendo quest'ultimo molto più complesso del primo. Ciò è dovuto al fatto che di recente abbiamo cominciato a usare tecniche e tecnologie diverse, modificando l'impostazione metodologica. Purtroppo queste nuove tecnologie non sono frutto di nuove acquisizioni teoriche.

Quindi, questi sono temi e settori nei quali è cambiato moltissimo ma, apparentemente, gli scenari narrativi non sono cambiati. Oggi siamo riusciti a fare cose che 50 anni fa non si riuscivano a fare. Basta. Sembrerebbe che non ci sia molto altro da dire. Se non che è cambiata la metodologia usata e l'epistemologia che questo cambio di metodologia ha giustificato. Ma questo riguarda

un non detto, soprattutto perché realizzato in forma autocelebrativa e poco critica. Occorrerebbe, cioè, una profonda analisi critica delle metodologie usate.

Per quanto riguarda gli "scenari … auto-narrativi", rimanendo alla superficie, potremmo dire che essi risentono sia del diverso contesto socio-economico sia dell'illusione che il "nuovo" sia veramente (e, radicalmente) nuovo e non un (sia pur profondissimo) aggiornamento di tematiche novecentesche. E mi riferisco qui alla profonda connessione esistente tra quanto si fa oggi e i progetti di ricerca di settant'anni fa.

Parlando di "scenari", desidero adesso introdurre i progetti di ricerca (secondo me è il modo corretto di chiamarli) di Ishiguro e Faggin. Progetti visionari ed entrambi di fantasia sia pure di fantasie diverse (ma, poi, quanto realmente diverse?). Quello di Faggin è un classico progetto di ricerca scientifico, almeno così appare. In realtà non è un progetto qualsiasi. Oltre a essere estremamente ambizioso si muove in una zona scivolosa e spigolosa insieme, come direbbero molti colleghi scienziati. Quello di Ishiguro è un romanzo, non un progetto di ricerca, mi sembra che stiano obiettando altri, colleghi e non colleghi all'unisono. Desidero controbattere a mia volta: sembra un romanzo ma in realtà è il *report* di un progetto di ricerca condotto a termine e già attuato. Esempio di quei momenti felici nei quali la ricerca di punta, di frontiera, riesce a trasformarsi in applicazioni tecnologiche di punta che poi vanno pure in produzione. Diventano "commodities". Nella *realtà dell'immaginazione*, ovviamente. Ma nel mondo di *Klara e il sole* i risultati tecnologici raggiunti presuppongono che un qualche progetto di ricerca teorico "alla Faggin" sia stato portato a termine con successo. Un progetto, cioè, nel quale sia stato risolto il problema di cosa sia la coscienza, analizzandola scientificamente.

Ma di che parla Faggin e cosa si propone di fare? Riporterò qui di seguito alcuni passi tratti – in modo non sistematico - dal suo

libro, lasciando che il lettore si faccia un'idea, senza miei commenti.

> … avevo cominciato a prendere sul serio l'idea che la coscienza potesse essere un aspetto fondamentale della natura, presente in qualche modo già negli atomi e nelle molecole di cui tutto è fatto. Quest'idea affiorò un po' alla volta, vista l'impossibilità di spiegare l'emergere della coscienza dalla complessità del nostro cervello. … Pensavo: Come può una struttura fisica fatta solo di aspetti esteriori produrre aspetti interiori semantici? Il concetto di complessità non ha nulla a che fare con le sensazioni e i sentimenti del mondo interiore. Di fatto, i computer d'oggi, con tutta la loro complessità, non hanno neanche una briciola di consapevolezza. Per me non rimaneva alternativa: il mondo interiore dev'essere fin dall'inizio una proprietà irriducibile di tutto ciò che esiste[22].

e, ancora:

> Molti scienziati e filosofi affermano che non può essere così perché le implicazioni di quest'idea porterebbero a contraddire le interpretazioni riguardanti la natura della realtà, basate sulle teorie matematiche che hanno già ottenuto verifiche sperimentali rigorose. Altri sostengono che quest'idea sa di religione. E scienza e religione non si devono mischiare.
> Questo argomento mi affascinava perché aveva il potenziale di spiegare e di unificare l'esistenza delle realtà esteriore e interiore che stavo esplorando da quasi vent'anni. Scienza e spiritualità, finora irriconciliabili, avrebbero potuto trovare un'unione profonda anziché una semplice giustapposizione di convenienza.
> Decisi quindi di ritirarmi completamente da tutte le mie altre attività per concentrarmi sullo sviluppo di un modello della realtà basato sul presupposto che la consapevolezza sia una proprietà fondamentale e irriducibile della natura… Non sono d'accordo con quegli scienziati che sostengono che l'universo non ha uno scopo e un significato e che noi siamo macchine destinate a essere superate dai nostri computer. Prima di prendere sul serio quest'interpretazione della realtà, dobbiamo riconoscere che essa si basa sul postulato che solo la materia esiste, una premessa non provata che non può spiegare l'esistenza della coscienza che ciascuno di noi sperimenta nella sua interiorità[23].

Rinvio il lettore a una lettura attenta dell'ultimo capitolo del libro di Faggin, riportando qui due passaggi da un'appendice conclusiva:

> L'intelligenza artificiale, che oggi rappresenta la punta di diamante del progresso tecnologico umano e grande motivo di orgoglio, potrebbe fornirci l'esperienza necessaria – ma non sufficiente – per riportare l'umanità a riscoprire il vero mistero della sua natura, perché finirebbe per dimostrare clamorosamente l'impossibilità di creare macchine coscienti[24].

[22] F. Faggin, *op. cit.*, p. 225
[23] *Ibid.*, pp. 226-227.
[24] *Ibid.*, pp. 265-266.

> A mio avviso, il vero pericolo dei progressi della robotica e dell'intelligenza artificiale non è quello di creare macchine che prenderanno il sopravvento sull'umanità perché saranno più perfette di noi. Il vero pericolo è che uomini malvagi possano causare danni seri all'umanità e all'ecosistema controllando computer e robot sempre più potenti soltanto per il loro interesse. Ma allora sarà l'uomo, non la macchina, a causare il problema. Questa è una grande sfida che la società dovrà affrontare il più presto possibile[25].

Le visioni di progetti scientifici molto futuribili (estremamente stimolanti nonostante abbiano una probabilità alta di fallimento) e la visionarietà dell'alta letteratura (o delle arti figurative) si intrecciano e potenziano a vicenda ma non si tratta di fantasie da quattro soldi, di sogni a buon mercato. L'alta letteratura, dal suo lato, prende sul serio l'ipotesi scientifica e la usa per approfondire la nostra conoscenza di noi stessi, svelandoci di noi molto di più di quanto molti articoli scientifici di psicologia riescono a fare. Per almeno due motivi; perché ci analizza, analizza i nostri comportamenti possibili di fronte a una profonda novità - non all'ignoto in sé ma a ciò che non si conosce di qualcosa che è di fronte a noi per la prima volta (come di fronte allo scoppio della prima bomba atomica e alla fine del progetto "Human genome"). Ma qui l'impatto è ancora più coinvolgente e Ishiguro lo descrive in modo delicatissimo e profondo nello stesso tempo, come se fosse già avvenuto. Ci viene da pensare che se dovesse avvenire non potrebbe che svolgersi con le implicazioni e le modalità da lui descritte.

Non possiamo non notare (e osservare con rispetto e attenzione) le paure che riporta Ishiguro: *«Allora, Klara, il fatto è che l'ansia riguardo agli AA»*, cioè gli amici artificiali, *«sta dilagando. La gente sostiene che siete diventati troppo intelligenti»*[26].

E fin qui è ciò che noi ci potremmo aspettare – magari con un semplice cambiamento del tempo verbale: "diventerete" al posto di "siete diventati". È, invece, assolutamente imprevisto (almeno per me) quello che segue, ciò che Ishiguro scrive immediatamente dopo:

[25] *Ibid.*, p. 305.
[26] K. Ishiguro, *op. cit.*, p. 260.

> *Le persone si spaventano perché non riescono a seguire quello che vi succede dentro. Vedono quello che fate. Riconoscono che le vostre decisioni e i vostri suggerimenti sono sensati e affidabili, quasi sempre corretti. Ma non sono contente di non sapere come ci arrivate. Da questo nasce la loro reazione, il pregiudizio. E noi dobbiamo combatterlo. Dobbiamo dir loro, d'accordo, siete preoccupati perché non capite come pensa un AA. Perfetto, allora diamo un'occhiata al motore. Decodifichiamo. Non vi piacciono le scatole nere sigillate? Benissimo, apriamole, allora. Quando ci avremo guardato dentro, non soltanto le cose faranno meno paura, ma impareremo. Impareremo meraviglie nuove*[27].

Qui invece il nostro raffinato (*intellettualmente* raffinato) giapponese, naturalizzato britannico, va oltre toccando punti estremamente importanti dal punto di vista scientifico ed epistemologico. Intanto vorrei osservare che non sono queste le paure che, oggi almeno, toccano un cittadino medio (italiano almeno); non so per uno britannico o giapponese. Il nostro tocca invece con sottile intelligenza un punto chiave che riguarda la paura, la paura in generale, la paura vera, non un timore generico per il nuovo. E, poi, anche un altro punto che riguarda esplicitamente l'aspetto scientifico dell'IA, ma su questo tornerò dopo. Abbiamo paura dell'ignoto, di ciò che non conosciamo quando questo ci contorna e ci avvolge. Il bambino ha paura del buio perché il buio avvolge e vela, nasconde quello che c'è. Ciò che afferma Ishuguro è più profondo. Afferma – nel brano riportato sopra - che: *«le persone si spaventano perché non riescono a seguire quello che vi succede dentro … non sono contente di non sapere come ci arrivate»* a quei suggerimenti e decisioni che pur si apprezzano per la loro affidabilità e sensatezza.

Qui si toccano due punti cruciali: uno comportamentale. Se sappiamo come si arriva a un certo risultato non dovremmo averne paura anche se si tratta di qualcosa di nuovo e di insolito. Questo è un punto cruciale per il cittadino comune di fronte a situazioni inattese. Queste derivano normalmente dalla scienza. Oppure, lasciatemelo dire, dalla magia. Entrambe presentano al cittadino comune risultati imprevisti. Ma mentre in quest'ultimo

[27] *Ibid.*, p. 260.

caso (la magia) non viene detto come si è arrivati al risultato (e qui importa poco se quello a cui ci troviamo esposti è vero, al di fuori delle leggi naturali, inganno, suggestione o cos'altro ancora). Il punto chiave, cruciale è che non viene (e, in un certo senso, non può venire) detto "come ci si è arrivati". Nel caso della scienza, invece, tutti possono sapere com'è andata. In linea di principio, perché poi, in pratica, bisogna avere studiato e, per esperimenti come quelli del CERN, non possiamo controllare di persona. Ma la trasparenza c'è ed è esplicitamente esibita. Sono chiaramente enunciati, per chiunque voglia e possa eseguirli, i protocolli da seguire per capire, verificare, controllare (magari in ordine inverso).

Passiamo adesso all'altro punto che, incredibilmente, è introdotto da Ishiguro. Potremmo dire, ma in senso buono, che lo fa surrettiziamente. Comunica al lettore una caratteristica dell'IA odierna quasi senza fargliene accorgere. Gli algoritmi migliori, come, ad esempio quelli legati alla traduzione automatica funzionano oggi molto bene (se non benissimo) soprattutto se paragonati a quelli di cinquant'anni fa. Ma questo non avviene perché oggi abbiamo una teoria linguistica più sviluppata di quelle di allora per quanto riguarda questo aspetto.

La maggior parte dei risultati recenti più interessanti dell'IA sono stati ottenuti proprio mediante tecniche e percorsi che *non permettono di capire*, almeno oggi, *come* ai risultati si sia arrivati. Proprio quello che lui scrive. Coincidenza, caso, intuizione creativa di uno scrittore di genio oppure – fatto salvo quanto scritto prima – anche frutto di un accurato studio della letteratura (scientifica, in questo caso) rilevante per questo problema? Sarei indotto a credere a quest'ultima ipotesi. D'altronde il libro è del 2021. Il suo libro precedente è del 2015. Gli anni in cui si è concentrato su questo tema andrebbero quindi dal 2015 al 2021. E la vittoria dei computer sui campioni di GO è del 2016. Un risultato da citare perché è proprio quello

più eclatante (e inaspettato) e che ha avuto una risonanza giornalistica notevole che ha coinvolto anche il cittadino comune. E poi ci sono le profonde osservazioni di von Neumann sulla complessità degli automi su cui, adesso, non ci possiamo soffermare ma che devono far parte di una riflessione complessiva sul percorso seguita dall'IA. Ed è proprio quello di cui Ishiguro, indirettamente e in modo lieve, ci ha parlato.

6. Qualche ultima riflessione prima di concludere

Non sappiamo cosa veramente ci attende nel futuro, grazie all'IA o per sua colpa. Come non lo abbiamo mai saputo per nessuna di tutte le altre innovazioni, concettuali e tecnologiche, che l'essere umano ha ideato e realizzato. Possiamo solo avere indicazioni. Da usare come guida. Da interpretare. Vorrei solo notare che sembra che gli storici affermino che ogni grande svolta – economica, tecnologica, culturale – ogni passaggio che viene definito *epocale*, in realtà raramente sono sentiti come tali da coloro che li stanno vivendo mentre, a volte, viene così definito - da chi lo vive - qualcosa che i futuri storici considereranno di interesse men che mediocre.

Inoltre, cosa forse più importante, anche nelle trasformazioni più radicali vi è sempre una forte continuità con ciò che è avvenuto prima. Anzi, potremmo dire che i passaggi più innovativi sono proprio quelli che sono profondamente intrecciati con ciò che li ha preceduti. Cioè, un cambiamento è profondo e non effimero se esprime qualcosa di significativo che, trasformandola, si *radica* nella storia precedente.

Per questo ritengo che un'analisi dell'IA debba tener conto di molte delle cose che l'hanno preceduta e accompagnata e di altri aspetti dell'agire umano che – a prima vista – possono sembrare non particolarmente o strettamente connessi. A questo riguardo particolarmente significative appaiono analisi come quelle che si possono trovare sul *MIT Technological review* – almeno fino a

quando era diretta da Karen Hao[28]. E di cui dobbiamo tener conto quando andiamo ad analizzare l'influenza a cui sistemi apparentemente oggettivi sono soggetti. Risentono – in modo devastante e inaspettato – di pregiudizi, trasmessi nel modo in cui sono stati progettati, pur senza volerlo. Oppure di come la potenza di queste nuove tecnologie possa amplificare tendenze negative proprie del sistema economico sociale. Ma dobbiamo essere in grado di distinguere e sceverare. I segnali di pericolo avanzati da Karen Hao sono in continuità e vanno nella direzione indicata da Wiener nel 1949. Le paure di Kissinger non riguardano questi aspetti ma il timore che il "potere", qualunque definizione vogliamo darne, possa perdere il controllo su ciò che avviene. E la risposta che dovremmo dare a lui è che dobbiamo andare verso un forte decentramento di tutti gli apparati, restituendo, in questo modo, alla "rivoluzione informatica" tutte le possibilità umanistiche che al suo apparire erano state, correttamente, intraviste recuperando anche lo spirito umanistico dell'analisi wieneriana.

7. A mo' di conclusione

L'avere inserito l'IA tra le futuristiche utopie letterarie di Ishiguro e l'ambizioso programma di ricerca di Faggin, che non si colloca tra le idee pacificamente accettate è stata un po' provocatoria ma, contemporaneamente ha voluto – costruttivamente – indicare lo spettro enorme entro il quale considerazioni che a essa si riferiscono possono localizzarsi in modo significativo.

È vero, infatti, che, in forme diverse, l'IA si può trovare "quasi ovunque". Una conclusione che, se da un lato rafforza la nostra convinzione di avere a che fare con qualcosa che può profondamente

[28] Cfr. K. Hao, A.P. Hernández, *How the AI industry profits from catastrophe*, April 20, 2022 https://www.technologyreview.com/2022/04/20/1050392/ai-industry-appen-scale-data-labels/; e K. Hao, *Artificial intelligence is creating a new colonial world order*, April 19, 2022 https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/.

incidere nello sviluppo della società e anche nella quotidianità delle nostre vite, da un altro lato ci lascia un po' insoddisfatti pensando al vecchio detto che "SE *tutto* è **X** ALLORA *niente* è **X**". Quindi è bene cominciare a fare, proprio al momento di concludere, alcune distinzioni. L'IA per la sua presentazione che tocca l'immaginario tende a svolgere una funzione mitopoietica, come ha fatto a suo tempo la cibernetica. Forse uno dei pochi settori che sono stati capaci di generare nuovi miti dopo l'aurea temperie del mondo greco. Mi chiedo cosa avrebbe detto Carl Gustav Jung di questo fenomeno e auspico che qualche suo brillante nipotino ce lo dica al suo posto.

Dobbiamo riconoscere questo piano e collegarlo – mantenendolo strettamente distinto – a quello molto più "concreto" dei risultati specifici. Molto disparati questi ultimi e frammentari. Che non delineano in alcun modo la forza e la potenza di una teoria unitaria, come sono state – ai loro tempi – la teoria della gravitazione di Newton (non a caso definita sistema del mondo) o la precedente *unificazione* dei moti terrestri e celesti compiuta da Galileo o, infine, la teoria di Maxwell che, a un tempo, unificava elettricità e magnetismo e forniva – con le sue equazioni – uno strumento tecnico per la progettazione di sistemi concreti. Nel nostro caso abbiamo quasi sempre a che fare con tecniche che producono risultati estremamente efficienti - come, ad esempio, quelli elencati nel NYT citato nella nota 6 – MA, allo stesso tempo, frammentari e incomprensibili. *Incomprensibili* in un senso molto specifico: non è del tutto chiaro *come* certi risultati siano stati ottenuti. Rimane il disappunto che – nonostante l'efficienza tecnologica di alcuni risultati - nessun vero passo in avanti sia stato compiuto nella direzione di "capire" veramente la natura dei problemi affrontati. Nel caso della "traduzione", ad esempio, i risultati non hanno alcuna relazione con progressi compiuti nella comprensione di *cosa sia* un linguaggio naturale. È questo che può generare, assieme alla soddisfazione per

l'avanzamento tecnico, allo stesso tempo frustrazione e paura. Frustrazione per l'incapacità, di fronte a una tecnica che produce risultati inaspettati, di riuscire a capire le ragioni teoriche di tale successo. La paura nasce, a sua volta, dalla sensazione che i risultati tecnici possano prendere il sopravvento sugli intendimenti di chi a tali potenti tecnologie si è affidato o ha preso decisioni in base a esse. Quante volte, anche a sproposito, abbiamo sentito accusare "l'algoritmo" di esiti incongrui.

Una situazione inedita rispetto ad altre nelle quali le scienze coinvolte erano la meccanica (crollo di ponti o dighe), l'elettromagnetismo (i due famosi Black-out di New York) o la chimica (Bhopal). Tutti casi nei quali in nessun modo ci si è nascosti dietro la poca trasparenza delle tecnologie utilizzate e delle scienze che a esse stavano dietro.

Possiamo pensare che il tempo trascorso sia troppo breve per avere permesso l'emergere e la costruzione di una teoria unitaria dei fenomeni legati all'emergere della nuova nozione di informazione, in tutti i suoi risvolti, e quindi si tratti solo di continuare a lavorare lungo la direzione già percorsa e aspettare. Ma come abbiamo già notato vi sono alcune anomalie: c'è tutta una famiglia di discipline che negli ultimi settanta anni sono nate, divenute centrali, poi declinate o scomparse, sovrapponendosi senza coincidere, differenziandosi per metodologie e ambizioni. Ma non si tratta solo di questo perché poi ci si scontra sempre con problemi concettuali non risolti, come già osservato.

Le scienze dell'informazione, di cui, oggi, l'IA è uno dei protagonisti principali (come ieri era stata la cibernetica) hanno bisogno di una profonda analisi che chiarisca il loro *status* epistemologico, evidenzi le loro caratteristiche specifiche e puntualizzi le differenze con altre discipline. Non c'è nulla di strano in questa richiesta. Rimanendo all'interno della sola fisica, la meccanica ha alcune caratteristiche che la

differenziano dall'elettromagnetismo. Ed entrambe le teorie hanno forti differenze con la termodinamica.

La situazione attuale potrebbe anche nuovamente portare al ritiro, da parte della società, di quella concessione di fiducia incondizionata che è stata periodicamente concessa (e ritirata) all'IA (e, in generale, alla scienza). Forse può essere utile descrivere con maggiore dettaglio questo aspetto. C'è stato uno sviluppo incredibile e inimmaginabile di risultati (tecnici) nuovi; questo ha corrisposto a una concessione ampia di fiducia sia pure associata sempre, ovviamente, a riserve e paure. Fiducia connessa, in qualche modo, all'assunzione che la via seguita e intrapresa fosse quella giusta e, come tale, degna di essere perseguita. Una volta che quel percorso mostrava i suoi limiti subentrava un calo di fiducia. Quest'ultimo fenomeno in sé non è negativo perché indice di una reattività della società che mostra di saper distinguere – attraverso cammini suoi propri – la sottile differenza epistemologica che intercorre tra le tecnologie frutto e sviluppo di teorie scientifiche pienamente sviluppate e tecnologie molto potenti che non hanno dietro di loro uno scudo teorico forte.

L'IA è, al momento, protagonista indiscussa della società del nostro tempo. Perché lo rimanga sono necessarie almeno due cose. Che la comunità scientifica comunichi i sui risultati in modo corretto; cioè, critico, modesto e problematico. E che abbia una visione generale delle questioni con un progetto ambizioso ma non velleitario e fumoso. Abbia conoscenza della storia precedente e sia cosciente degli scacchi, degli errori e dei vicoli ciechi del passato. Sia in grado di proporre assieme ai grandi obiettivi strumenti scientifici e tecnici potenzialmente adeguati a poterli raggiungere.

L'importanza dell'intersecarsi di questi diversi piani, nel rispetto dell'autonomia di ciascuno di essi, è quello che si è cercato di mettere in evidenza in queste pagine, sia pure in modo

frammentario ed episodico. In modo leggermente provocatorio sono stati presi come estremi simbolici due diversi *visionari* che sono entrambi, a rigore, al di fuori del campo proprio dell'IA ma che – proprio per questo – possono indicarci l'orizzonte verso cui muoverci.

MARIA TERESA CATENA

L'INTELLIGENZA DELLE MANI

1. *Naturale e artificiale* 2. *I tre anelli* 3. *L'errore di Cartesio*
4. *Un nesso tra eterogenei*

ABSTRACT: ***THE INTELLIGENCE OF HANDS***
Attributing artificiality solely to technological intelligences certainly means distinguishing their capabilities from those of the human mind, whose biological origin, however, runs the risk of becoming synonymous with an immutable nature. Adopting a palaeontological perspective, the contribution, conversely, seeks to draw attention to the close evolutionary link between our brains and the practices of the living body. Such a reflection is intended to recall not only the constructed nature of our intelligences, but also to show how, in removing the plasticity of the body, Cartesian Modernity was instead inspired by a machinic model of the brain that rendered the mind an immaterial and static entity. Therefore, trying to probe the presence of this approach between the lines of some conceptions of A.I. is a necessary premise both to read human intelligere in terms of a specific articulation of the artificial and to reposition its relationship with other machinic intelligences in the more horizontal perspective of a respectful heterogeneity.

Le tecnologie sono artificiali, ma l'artificialità è naturale per gli esseri umani.
W. Ong

1. *Naturale e artificiale*

Non c'è dubbio: più passa il tempo e più l'Intelligenza Artificiale ha molto da dirci e, soprattutto, molte sfide da proporci. A dispetto del termine, coniato nel relativamente recente 1956, le riflessioni su questa ingegneria dedita alla

costruzione di macchine che agiscono con capacità più o meno simili a quelle dell'uomo aumentano in modo direttamente proporzionale alle applicazioni che essa sperimenta. E così, dalle prime descrizioni delle reti neurali artificiali fornita da Alan Turing, passando per le tartarughe meccaniche di William Grey Walter, senza dimenticare Eliza, il primo *chatbot* progettato nel 1966 per imitare uno psicoterapeuta in conversazione con i suoi pazienti, siamo giunti, tra "estati e inverni", alla costruzione di veicoli senza pilota in grado di procedere autonomamente nel traffico o in un'azione militare, nonché alla ben più amena programmazione di AlphaGo, famoso per aver battuto a Go lo stremato Lee Sedol. Stupefacente, davvero.

A scrivere, sia chiaro, non è una specialista né una scienziata né, tantomeno, un ingegnere; piuttosto un'umanista che, nell'aggirarsi un po' frastornato tra i molti campi di applicazione che le diverse forme d'intelligenze artificiali mostrano oggidì, sa di non avere le competenze specifiche per schierarsi con sicurezza nell'annosa diatriba tra gli "entusiastici integrati" e gli "apocalittici diffidenti"; convinti, i primi, che i nuovi processi tecnologici, le cui attitudini supereranno di gran lunga quelle degli esseri umani, saranno in grado di rendere più abitabile e controllabile il mondo, portando nuovi benefici all'umanità e risolvendo i tanti problemi che essa è stata in grado di creare; pronti, invece, i secondi, a denunciare i danni che l'I.A. rischia di provocare, non ultimi gli impatti delle sue applicazioni in campo militare o le conseguenze che l'automazione di compiti complessi può avere sulle nostre capacità cognitive e dunque sulla nostra vita.

Naturalmente, non è mia intenzione sottrarmi a una doverosa presa di posizione su questioni evidentemente delicate e sempre più urgenti. Il punto è che, com'è stato di recente sottolineato, «chi fa filosofia lavora con concetti generali»[1]: cosa questa che, se

[1] A. Fabris, *Etica delle nuove tecnologie*, Scholé, Brescia 2021², p. 6.

non è confusa con una tendenza all'astrattezza, dà anche l'opportunità di distogliere lo sguardo dall'immediato, dal progetto e dalla sua applicazione, per provare a individuare le stratificazioni concettuali sedimentate – e in parte celate – dietro la scelta e l'uso che facciamo di nomi e parole; come a dire che si tratta di provare a capire di cosa parliamo e cosa mettiamo in campo – o fuori gioco – nelle nostre scelte espressive, magari per arricchire con un "tassello" teoretico il pur indispensabile momento etico e decisionale delle nostre riflessioni.

Così, al cospetto della definizione "Intelligenza artificiale", è evidente – anche al netto di un andamento che oscilla tra l'emulazione e il superamento – la scelta di prendere come modello per la costruzione di queste nuove macchine, l'intelligenza più complessa che abbiamo a disposizione: la nostra. Il che presuppone – e qui emerge una prima domanda – che sia ben chiaro di che cosa parliamo quando parliamo d'intelligenza.

Lungi dal semplificarsi, però, la questione si complica, non foss'altro perché non è affatto facile mettersi d'accordo sul significato della parola, specificare che cosa s'intende con essa e definire quali sono i rapporti che, ad esempio, intercorrono con il cervello, la conoscenza del quale è un *work in progress* composto, ormai, dalle più disparate competenze di neurobiologi, fisici, matematici, ingegneri e informatici. Va da sé che gli stessi esperti di I.A. hanno da sempre cercato di dare una risposta a questo difficile quesito, spesso accendendo un dibattito fatto di posizioni differenti e, a volte, anche difficilmente conciliabili: basti pensare, ad esempio, alle obiezioni che John Searle rivolse all'impostazione *simbolica* e *computazionale* sostenuta da Marvin Minsky, Herbert A. Simon e Allen Newell, decisamente dominante nei primi decenni della disciplina. Al cospetto della convinzione che la mente umana è simile a un *hardware* la cui attività consiste nella capacità di

combinare, secondo le regole della logica, rappresentazioni isolabili, oltre che precisamente localizzabili, si riteneva sufficiente studiare e replicare questi processi per costruire un computer intelligente; e di fatto i primi risultati di questo approccio furono programmi in grado di riprodurre attività intelligenti come le dimostrazioni matematiche. Ma, si domandava Searle, siamo sicuri che nella nostra testa si svolga esclusivamente un tal genere di lavoro? E siamo certi che si possa chiamare intelligente un *software* che, per quanto sofisticato, altro non fa che collezionare numeri, ancorché in maniera aggiornata e veloce[2]?

Non è stato dunque un caso che a questa seconda scuola di pensiero se ne sia affiancata un'altra che, stante le intenzioni di Frank Rosenblatt, provava a pensare l'intelligenza come una proprietà *funzionale* del sistema biologico e, ispirandosi alle reti neurali, cercava di riprodurne le strutture. Anche in questo caso, tuttavia, nonostante gli indubbi progressi che le attuali RNA hanno nel ricavare informazioni dai dati, applicandole, ad esempio, al riconoscimento delle immagini, è difficile giungere alla conclusione che la capacità di *apprendere* o imparare esaurisca lo spettro della nostra attività cerebrale. Nemmeno la robotica, infine, con le sue macchine in grado di vedere e riconoscere gli oggetti, nonché di comprendere il linguaggio naturale e *comunicare* con gli esseri umani, può esimersi dal constatare che il nesso tra intelligenza e percezione, su cui ragionevolmente basa i suoi studi, è molto difficile da indagare, anche volendo limitare le indagini al "banale" sistema autoselettivo di visione di una rana[3].

Così, visti i problemi posti dalla definizione del termine "intelligenza", è forse più opportuno, per provare a comprendere

[2] Cfr. in proposito, J.R. Searle, *Menti, cervelli e programmi. Un dibattito sull'intelligenza artificiale* (1980), tr. it. CLUP, Milano 1984.

[3] A proposito del cervello e delle capacità di visione di una rana, cfr. S. Angeletti, *Il cervello nelle mani. La neurobiologia dalla cellula al robot*, Longanesi, Milano 1995, pp. 29 sgg.

le sedimentazioni concettuali presenti nelle diverse riflessioni facenti capo all'I.A., rivolgere la nostra attenzione all'aggettivo "artificiale". Tale spostamento di prospettiva, inoltre, è autorizzato dal fatto che, a dispetto della difficoltà a convenire su *cosa* debbano fare le nuove macchine - se ragionare, comunicare e interagire con l'ambiente o, piuttosto, apprendere - tutti concordano sul *come* lo facciano: artificialmente, appunto, a partire cioè da una complessa organizzazione e costruzione di strategie e meccanismi o, più precisamente, attraverso un articolato e sempre più diversificato sistema di elaborazione d'informazioni; dunque in modo differente dalle modalità realizzative dell'intelligenza umana che, *a contrario*, è implicitamente considerata naturale. Perché non ci sono dubbi: anche se stabilissimo che la mente umana null'altro è se non un «sistema di elaborazione di informazioni che non fa nulla che non possa fare un altro sistema di elaborazione di informazioni»[4], resta certo che essa svolge questo compito con modalità biologiche e naturali tutte proprie.

Ma – e questo è il punto – siamo sicuri di avere ben chiaro che cosa significa un'intelligenza *naturale*? Che cosa dobbiamo intendere con tale termine: l'emanazione immateriale e spontanea di un mondo spirituale? Il prodotto predeterminato delle operazioni fisiche del cervello? O piuttosto il termine non significa niente di tutto questo, rimandando piuttosto a un lungo e per niente scontato percorso evolutivo, le cui radici sono intrinsecamente materiali?

2. *I tre anelli*

Ebbene, non sembrano esserci dubbi: una volta superate le visioni che tendevano a risolvere la naturalità delle leggi dell'uomo e del mondo «dentro un ordine naturale tanto vasto da trascendere se

[4] D.W. Hillis, *Vicini alla singolarità*, in J. Brockman, *La terza cultura. Oltre la rivoluzione scientifica* (1995), tr. it. Garzanti, Milano 1999, p. 372.

medesimo in una superiore armonia»[5], anche l'umano smette di essere una volta e per tutte una particolare creatura "sovrannaturata", frutto di un piano che, in fin dei conti, sempre la trascende, per scoprirsi intrecciato al contesto della sua situazione biologica, della struttura delle sue percezioni, delle sue azioni e dei suoi bisogni; per collocarsi, in una parola, in stretta dipendenza da un ambiente ormai irrimediabilmente orizzontale nel quale, organismo tra i tanti, viene maturando, in un reciproco arricchimento, la sua specificità. In questa prospettiva, la sua intelligenza, lungi dal brillare come riverbero di "altra scintilla", diventa piuttosto la semplice maglia di una catena decisamente più materiale; come tale, ben saldata agli altri due anelli fondamentali del cervello e del corpo, il cui movimento, poi, in particolar modo, «ci insegna qual è il nostro posto tra i primati»[6].

È una tesi ben nota, questa, si sa; una tesi che ha trovato tra i suoi sostenitori André Leroi-Gourhan, i cui studi hanno fatto del nesso tra posizione, dentatura ed arti, la triade che condiziona l'evoluzione del cranio e del cervello di qualsivoglia vivente, *homo* compreso. A suo modo di vedere, infatti, è la posizione eretta, specifica condotta motrice conquistata a un certo punto dalla nostra specie, a permettere la maturazione del cervello, la cui massa si è potuta accrescere grazie alla nuova posizione assunta dal cranio. Del resto, non è difficile comprendere la mutazione morfologica subita dalla scatola cranica una volta che essa, collocata verticalmente sulla colonna vertebrale, vede, al tempo stesso, le mani liberate dalla funzione locomotrice, la mandibola sollevata dal suolo e il ventaglio corticale aprirsi. L'ampliamento delle cellule nervose o l'aumento della materia grigia, dunque, «non può essere antecedente all'edificio scheletrico», dato che, piuttosto, è la massa cerebrale a «seguire

[5] P. Piovani, *Giusnaturalismo ed etica moderna*, Liguori, Napoli 2000², p. 60.
[6] J.L. Arsuaga, *I primi pensatori e il mondo perduto di Neandertal* (1999), tr. it. Feltrinelli, Milano 2001, p. 27.

il movimento generale» e ad avvantaggiarsi «dei progressi dell'adattamento locomotore anziché provocarli»[7]; come a dire che la forza di espansione del cervello, e la correlativa intelligenza, lungi dal produrre il *gesto* e la *parola*, abbisogna innanzitutto della struttura somatica, con i suoi movimenti e ritmi; necessitano, in una parola, del complesso lavorìo dell'anello del corpo.

Occorre intendersi però: il corpo in questione non è una mera estensione né tantomeno un meccanismo che ammette fra le sue parti, o fra se stesso e l'esterno, solo relazioni esteriori. Più che semplice recettore, dipendente in modo univoco e lineare da uno stimolo, che tocca poi all'apparato nervoso decrittare con neurotrasmettitori e sinapsi, il sostrato biologico mostra di essere un varco, un accesso, un contatto costante con una situazione aperta e pratica, che suggerisce e sollecita quelle che solo in via riduttiva si chiamano risposte perché, viceversa, è di un complesso di condotte motorie che si tratta o, detto in altri termini, di un insieme di gesti – sessuali, affettivi, sociali, comunicativi o indirizzati alla costruzione di strumenti e alla produzione di utensili – che sono tutti già "tecniche", cioè modi attraverso i quali si sperimentano strategie vitali esposte, come tali, tanto al fallimento quanto alla riuscita.

In tutta evidenza, dunque, in questa impostazione, il rapporto con l'ambiente non si gioca nei termini di un fronteggiamento tra due "in sé": quella descritta piuttosto è un'immediata e per niente semplice co-appartenenza che salda vicendevolmente le condotte somatiche alla vasta schiera di potenziali alterità - siano esse gli altri uomini, gli animali, i vegetali o la stessa sfera materiale - rendendole atti radicalmente integrati nella materia, intrisi di uno "sfondo affettivo" e, pertanto, capaci di influenzare e modificare la morfologia del cervello con la sua

[7] A. Leroi-Gourhan, *Il gesto e la parola. Tecniche e linguaggio* (1964), tr. it. Einaudi, Torino 1977, p. 33.

attività neuronale. Non è un caso che il paleontologo francese parli degli scheletri come «di utensili»[8], cioè di veri e propri *artifici organici*, intendendo con questa formulazione la natura malleabile di un *soma* disponibile, già nella sua architettura anatomica, a essere plasmato e diversamente modificato dalle sue stesse condotte, nelle quali vanno senz'altro incluse anche quelle variazioni intra e inter-organiche, fisiologiche e senzienti che, sempre intrecciate al rapporto efficace con la materia, costituiscono «un'azione simultanea dei ritmi e delle forme»[9] o, detto in altri termini, una capacità di essere e darsi un mondo, certamente indispensabile per l'orientamento e la sopravvivenza.

E così, per tornare alla domanda posta, non solo occorre ammettere che la nostra intelligenza, quale frutto di un lungo processo evolutivo, ha ben poco d'immateriale o spirituale, ma anche constatare che sarebbe fuorviante identificarla con un cervello che, simile a una struttura di cemento, segue i percorsi immutabili di circuiti dai ruoli precisi e non suscettibili di alterazioni.

«L'uomo», affermava invece Mauss citando Halbwachs, «è un animale che pensa con le sue dita»[10]. Non è solo un'affermazione suggestiva: il punto di vista filogenetico ed evolutivo ci dice infatti che l'*intelligere* di cui è capace la nostra materia grigia è innanzitutto un "fare", materiale e situato, le cui configurazioni orientative altro non sono che stratificazioni e sedimentazioni di pratiche, esperienze e modi di vita che, radicati nel "veicolo del corpo", conferiscono al cervello il suo più importante risultato evolutivo: la neuroplasticità, cioè quella capacità di "cambiare idea" che consente al sistema nervoso di «sottrarsi alle limitazioni del proprio genoma e di adattarsi alle situazioni ambientali, ai cambiamenti fisiologici e alle

[8] A. Leroi-Gourhan, *Le radici del mondo. Dalla ricerca preistorica uno sguardo sulla totalità dell'uomo* (1982), tr. it. Jaca Book, Milano 1986, p. 83.
[9] A. Leroi-Gourhan, *Il gesto e la parola. La memoria e i ritmi*, cit., p. 330.
[10] J.-P. Warnier, *La cultura materiale* (1999), tr. it. Meltemi, Roma 2005, p. 27.

nostre esperienze»[11]. Insomma, i fossili di Leroi-Gourhan mostrano in maniera chiara quanto anche la neurofisiologia più recente sostiene; sarebbe a dire che il nostro cervello, con la sua intelligenza, non è quella macchina neuroanatomicamente standardizzata che abbiamo creduto fosse, né le sue componenti cellulari posso essere assimilate a "banche dati" permanenti: piuttosto, trattasi di una materia altamente adattabile, pronta a cambiare in sintonia con gli strumenti che utilizza, con le azioni motorie che compie, con gli schemi di pensiero, i ritmi e le affezioni che vive e percepisce.

Osservato da punto di vista della sua storia evolutiva e delle sue prestazioni neurofisiologiche, il nostro "anello" cerebrale appare dunque più come un artefatto, oserei dire un vero e proprio "manufatto": qualcosa la cui intelligenza, debitrice com'è alle tecniche della mano e alle più complessive condotte – e ritmi – somatici, è molto simile a un «fai da te» che ciascuno «si costruisce a suo modo, partendo da un modello base tendenzialmente aperto a tutte le scelte successive»[12]. È dunque certo che «pensare è agire, e agire è pensare»[13], ma altrettanto certo è che a quest'affermazione occorre aggiungere che il nostro pensare, almeno dal punto di vista evolutivo, è sempre l'agire di un corpo protesico, senziente e, in quanto tale, artificiale.

[11] A. Pascual-Leone, A. Amedi, F. Fregni, L.B. Merabet, *The plastic human brain cortex*, in «Annual Review of Neuroscience», 28, 2005, pp. 377- 401.

[12] S. Angeletti, *op. cit.*, p. 42. Cfr. inoltre N. Doidge, *Il cervello infinito. Alle frontiere della neuroscienza: storie di persone che hanno cambiato il proprio cervello* (2007), tr. it. Ponte alle Grazie, Milano 2014. Il neuropsichiatra racconta, tra le altre, la vicenda di Michael Bernstein, il quale, vittima di un ictus all'età di 54 anni che gli lasciò la mano sinistra menomata, recuperò nel giro di qualche settimana quasi tutti i movimenti grazie a un programma di terapia sperimentale che lo costringeva a ripetere in maniera massiccia azioni molto semplici. Le azioni ripetute, infatti, erano un mezzo per "convincere" i suoi neuroni e le sue sinapsi a formare nuovi circuiti che avrebbero rilevato le funzioni un tempo svolte dai circuiti collocati nell'area danneggiata del cervello.

[13] Citando Heinz von Foerster, è K. Kelly a ricordarcelo in *Out of control: la nuova biologia delle macchine, dei sistemi sociali e del mondo dell'economia* (1992), tr. it. Apogeo, Milano 1996, p. 49.

Trattasi di una posizione non priva di elementi problematici[14], ma che ha il vantaggio di indurci a rinunciare all'astrusa ricerca di una definizione sul che *cosa* sia l'intelligenza, spingendoci, viceversa, in direzione della più semplice constatazione che le nostre menti, proprio perché incarnate, non sono altro che uno specifico e possibile modo di articolare quel *come* dell'artefazione che poi trova nelle intelligenze macchiniche un'altra e differente modalità di azione; un'azione quest'ultima che – si badi bene – si svolge non tanto in assenza di un «*intelligere*»[15], ma secondo strategie improntate a comportamenti costruiti in assenza di quel particolare artificio organico che abbiamo visto essere il nostro *soma*.

Il punto è che, almeno da un dato momento in poi, la nostra storia materiale e il nostro linguaggio si sono mossi in tutt'altra direzione, mostrandosi più propensi a mettere sottotraccia quest'*agere* protesico delle nostre facoltà mentali, avviando così una serie di non trascurabili conseguenze, anche relativamente al nostro rapporto con le cosiddette nuove macchine.

3. *L'errore di Cartesio*

Del resto, che questo accadesse era forse inevitabile. Basta riflettere su quella particolare tipologia di condotte somatiche che sono i gesti efficienti, "padri" delle tecnologie. Se infatti si pensa ai primi utensili – una pietra affilata per tagliare, una leva per spostare un carico pesante – è ben evidente il loro carattere di secrezioni organiche, cioè la loro capacità di sostituire, magari in maniera più efficace, tutto quanto il *medium* del corpo è solito fare, oppure di rendere possibili compiti viceversa impossibili da realizzare per la difficoltà o la

[14] N. Carr, *Internet ci rende stupidi? Come la Rete sta cambiando il nostro cervello* (2010), tr. it. Raffaello Cortina, Milano 2011, p. 51 ss.

[15] È la tesi sostenuta da L. Floridi in *Agere sine intelligere. L'intelligenza artificiale come nuova forma di agire e i suoi problemi etici,* in L. Floridi, F. Cabitza, *L'intelligenza artificiale. L'uso delle nuove macchine,* Bompiani, Milano 2021, pp. 115-181.

grandezza dell'obiettivo in gioco. Insomma, attenendosi ai risultati di quella che è stata chiamata la prima rivoluzione tecnologica, chiara ci appare la natura manufattuale dell'intelligenza umana, la sua radicale contaminazione con il corpo protesico, in sintonia con il quale essa "materializza" i propri pensieri.

Non che la Modernità meccanica, con i suoi telai e le sue macchine a vapore, non fosse legata alle prassi del corpo; come d'altro canto, in una linea che non sembra affatto costituire una deviazione da queste premesse, a esse s'intrecciano anche le cosiddette tecnologie della Galassia Marconi, vere e proprie estensioni del nostro intero sistema nervoso. È indubbiamente vero che negli anni Quaranta e Cinquanta il computer altro non era che una stanza e «programmare significava usare un cacciavite»; allo stesso modo in cui, al netto della «parentesi semantica», anche oggi, pur se calate nell'«infosfera che le circonda spesso impercettibilmente», le interazioni umano-computer sono diventate di nuovo somatiche, «con touch screen, comandi locali, dispositivi di ascolto, applicazioni sensibili ai gesti, dati proxy per la geolocalizzazione e così via»[16].

Eppure, nonostante l'indiscutibile esattezza di queste considerazioni, uno slittamento, forse un ribaltamento, si è dato nella nostra relazione con le macchine. Il fatto è che, a furia di creare strumenti, si sviluppa una tendenza a ricreare noi stessi a loro immagine: e così, come martello, incudine e staffa sono i nomi dati ad alcuni ossicini del nostro orecchio, allo stesso modo, l'invenzione dell'orologio o quella delle macchine idraulico-pneumatiche a vapore suscita l'idea che il cuore sia una pompa e che il corpo che lo custodisce altro non sia che un meccanismo. Ne deriva – e inevitabilmente – un ripensamento e una rimodulazione della visione stessa della nostra intelligenza.

[16] L. Floridi, *op. cit.*, p. 158.

A riguardo non c'è emblema più pregnante della filosofia cartesiana, per la quale il *soma*, vera e propria macchina da terra, non solo nulla sa fare se non sottostare alle leggi meccaniche del moto, ma nulla può pensare se non grazie e per tramite di un *cogito*, la cui attività immateriale e inestesa è un patrimonio innato dai frutti tanto più veritieri quanto più capaci di astrarre da contaminazioni materiali. Non è un caso che, per il filosofo francese, è questa mente disincarnata e centripeta la vera matrice e sede di qualsivoglia *cogitationes*, ivi compresi i processi inventivi e razionali deputati alla costruzione dei moderni artifici meccanici. Del tipo di rapporto che un'intelligenza così concepita ha con le proprie creature macchiniche, ci raccontano poi i primi automi, le cui "teste parlanti" sempre *distinte* e *distanti* da quelle dei loro artefici, sono perciò stesso in grado di rassicurarli circa l'eccellenza e la superiore acutezza naturale della loro mente, unica e vera causa della loro realizzazione.

Che questa lettura dei corpi e delle intelligenze si sia esaurita, è ben difficile dirlo giacché, anche quando la nozione di mente immateriale si è andata facendo sempre meno sostenibile, è sopravvissuta l'idea cartesiana di un corpo-cervello-macchina, le cui operazioni fisiche bastano da sole a spiegare la formazione di pensieri, emozioni, memoria, immaginazione, creazione. E così, rinforzata dall'avvento del computer digitale, questa visione ha condotto scienziati e filosofi a riferirsi «ai nostri computer mentali, e persino al nostro comportamento, definendoli cablati a livello di "hardware" (*hardwired*), proprio come i microscopici circuiti incisi nel sostrato di silicio di un chip di computer»[17]. È una proiezione che, attraverso un percorso diverso, dà vita a una nuova versione del vecchio adagio relativo a una supposta e stazionaria natura umana: come infatti nota Nicolas Dodge, «dal momento che il cervello, modellato sulla sicurezza della macchina,

[17] N. Carr, *op. cit.*, p. 40.

non può cambiare, così anche la natura umana che ha la propria origine nella mente sembra altrettanto inalterabile»[18].

Evidentemente, non si tratta di spostamenti di poco conto.

Pensare a un corpo-macchina, infatti, non significa tanto sostituire l'agente con l'agito, il produttore con il prodotto, ma soprattutto rescindere il legame evolutivo che abbiamo visto saldare il pensare con l'agire somatico, in una vera e propria separazione delle forze in campo che lascia ottuso e privo di senso il corpo vitale. Non può che derivarne una, anzi, per meglio dire, due diverse idee d'intelligenza, entrambe molto distanti da quella materiale e manufatta: la prima, che in un ripensamento della vecchia e pre-moderna "natura sovrannaturata", ne fa di nuovo un'entità spirituale o comunque immateriale; la seconda, tutta improntata al modello macchinico, la considera un sistema statico e immutabile di elaborazione d'informazioni calato in una materia il cui *agere* è prettamente meccanico.

Resta fermo però che, comunque la si voglia mettere, nessuna di queste due concezioni condivide con l'*intelligere* cresciuto con il corpo protesico il pregio non trascurabile che esso ha di sapersi porre in una relazione orizzontale con i propri interlocutori; ovvero di saper comprendere che di fronte a sé non si parano mai «meri oggetti» o più o meno complessi prodotti delle proprie capacità sovrannaturate, quanto piuttosto «enti», il cui «il peso di soggetti» li rende parimenti «interpreti delle dinamiche globali» di cui esso stesso è fatto[19]. Viceversa, in entrambe le sue declinazioni, la metafora del corpo-cervello-macchina mette in scena un rapporto sicuramente più ambivalente con quegli stessi congegni che egli osserva – lo si è accennato – tanto come prodotti capaci di testimoniare circa la superiorità della mente umana quanto, insieme, come modelli cui ispirarsi – quando non

[18] N. Doidge, *op. cit.*, p. 8.
[19] N. Deisori, *Etica e Intelligenza Artificiale. Conversazione con il professor Edmondo Grassi*, https://www.centrostudi-italiacanada.it., novembre 2021.

addirittura come veri e propri specchi su cui proiettare, di volta in volta, l'immagine dei propri vizi o delle proprie virtù.

4. *Un nesso tra eterogenei*

Sarebbe interessante – e forse opportuno – provare a declinare una fenomenologia di queste sedimentazioni concettuali, ivi compresa la relazione tra l'emulativo e il competitivo che s'è venuta instaurando con le nuove macchine intelligenti. Ad esempio, si potrebbe notare la presenza di queste posizioni in quella che è stata definita "I.A. forte": convinta che la chiave del pensiero risieda nella capacità di combinare, secondo le regole della logica, rappresentazioni ritenute «chiaramente isolabili, oltre che precisamente localizzabili nella *macchina* cervello»[20], tale impostazione, in un'evidente circolarità, ritiene intelligenti tutte le macchine capaci di ragionare simbolicamente una volta che siano state in loro implementate sufficienti quantità d'informazione. Che un tal genere di congegno, «in grado di trasferire informazioni alla velocità della luce; compiere calcoli complessi in frazioni impercettibili di secondo; predire se pioverà tra dodici ore su un campo di grano», altro non sia che mera computazione e, in quanto tale, fonte di rassicurazione sulle capacità delle nostre intelligenze, basterebbero a provarlo le parole di Elaine Rich e Kevin Knight, secondo i quali l'I.A. è «l'arte di delegare a macchine computazionali compiti, "che, al momento, la gente sa comunque fare meglio"»[21]. Trattasi di una posizione che pur avendo il pregio di rimandare alle responsabilità dell'umano, ricordandoci che occorre innanzitutto occuparci di «cosa noi possiamo o vogliamo fare con esse, se classificare individui affetti da malattie o reiterare il reato di cui sono accusati», corre il rischio di scivolare in una petizione

[20] M.A. Boden, *L'intelligenza artificiale* (2016), tr. it. il Mulino, Bologna 2019, p. 131.
[21] F. Cabitza, *Deus in machina? L'uso umano delle nuove macchine tra dipendenza e responsabilità*, in L. Floridi, F. Cabitza, *op. cit.*, p. 38.

di superiorità, almeno fintantoché si dichiara poco interessata a «cosa queste macchine riescano a fare»[22].

Allo stesso modo, occorrerebbe provare a scandagliare la presenza di simili sedimentazioni concettuali, nei tentativi compiuti dall'"I.A. dell'apprendimento" – automatico e profondo – e delle "reti neurali" che, basata su una visione più dinamica del cervello, cerca di costruire un tipo di macchine intelligenti in grado di avvicinarsi quanto più possibile alla sottigliezza e alla complessità della mente umana. In tale direzione andrebbero annoverate le stesse ricerche dell'*I-Life* – cioè della codificazione digitale della vita – e dell'"I.A. evolutiva", alcune delle quali focalizzate sul ruolo dell'emozione nell'architettura del pensiero. Ebbene, anche in questi casi non sarebbe difficile rilevare come la presenza di alcuni degli apriori teoretici di cui s'è detto, sia all'origine dell'*impasse* in cui tali ricerche cadono ogniqualvolta cercano di simulare, «tramite modelli matematici dei singoli neuroni e delle loro sinapsi», il cervello umano precedentemente scansionato, per poi accoppiare «tali simulazioni a un corpo fisico (robotico) e a un ambiente simulato»[23].

Infine, sarebbe interessante – e nemmeno tanto difficile – decodificare la presenza di un non superato disagio nei confronti dei limiti del "vecchio corpo", con la sua caduca intelligenza incarnata, in riflessioni che, incrociando e rilanciando in direzione avveniristica alcune tematiche dell'I.A., immaginano «un aggiornamento radicale dei sistemi fisici e mentali dei nostri

[22] *Ibid.*, p. 30. Il testo citato dall'autore è E. Rich, K. Knight, *Intelligenza artificiale* (1991), tr. it. McGraw Hill, Milano 1992.

[23] S. Parra, M. Torrens, *Intelligenza artificiale. La strada verso la superintelligenza,* RBA, Milano 2018, pp. 32-33. Cfr. inoltre, in proposito, AA.VV., *Macchine che pensano. La nuova èra dell'intelligenza artificiale* (2017), tr. it. Dedalo, Bari 2018, in particolare pp. 64 sgg.

corpi» teso a far sì che le nostre macchine diventino «molto più simili a noi» e noi «molto più simili alle nostre macchine»[24].

Si tratta evidentemente di approcci molto diversi tra loro, tutti caratterizzati da risvolti scientifici molto complessi e da inevitabili rimandi a problemi etici, le cui soluzioni, va da sé, non sono facili né a portata di mano. Tuttavia – e proprio a causa dell'estrema complessità dei problemi in gioco – potrebbe non essere fuori luogo provare ad accompagnare i tanti dibattiti che ne scaturiscono sia con la ricognizione delle stratificazioni concettuali presenti nelle nostre pratiche – linguistiche e non – sia con la considerazione che ho cercato di far emergere dalla breve indagine intorno all'aggettivo "artificiale". Insomma, si potrebbe provare a tener ferma la premessa teoretica che, quali modi diversi di articolazione dell'artificialità, le intelligenze umane manufatte dei corpi protesici e quelle dei corpi macchinici sono radicalmente diverse tra loro ed equiparabili solo entro certi limiti; come a dire che le menti umane e quelle dei circuiti elettronici «offrono risorse radicalmente differenti», non solo in termini di velocità, ma in quanto codificano, organizzano, memorizzano, scambiano le "informazioni" in modo «totalmente diverso»[25].

Così, riprendendo un'antica distinzione filosofica, più che nei termini di una composizione matematica tra *entità* omogenee, la relazione tra l'intelligenza umana e quella macchinica, andrebbe declinata piuttosto quale nesso dinamico tra *modalità* eterogenee[26].

È un'affermazione questa che non ha nulla della resa a una sacralizzazione oscurantista di confini invalicabili. Al contrario, dato che l'acquisizione di trovarsi al cospetto di due

[24] R. Kurzweil, *La singolarità è vicina* (2008), tr. it. Apogeo, Santarcangelo di Romagna 2014, p. 297. Su tali questioni, mi permetto di rinviare a M.T. Catena, *Breve storia del corpo*, Mimesis, Milano 2020, pp. 165-175.
[25] J. Bailey, *Il postpensiero: la sfida del computer all'intelligenza umana* (1996), tr. it. Garzanti, Milano 1998, p. 23.
[26] I. Kant, *Critica della ragion pura* (A 1781, B 1787), tr. it. Laterza, Roma-Bari 1985³, p. 179.

– o più – differenti *modalità* di artificialità, sottraendo le menti umane all'auto-narrazione sostanzialistica implicita in una certa – e malintesa – idea di naturalità, sgrava la specificità della loro «materialità connessa» da qualsivoglia istanza di superiorità, restituendole più disponibili a una reale condivisione di spazi e tempi con le altre intelligenze; condivisione che – si badi bene – è tanto più urgente in quest'"oggi" in cui siamo continuamente posti di fronte a strumenti che, differentemente dagli utensili classici, «non sono attivi solo nel momento dell'interazione ma agiscono retroattivamente anche quando la persona smette di interrogarli»[27].

Sono considerazioni preliminari, non c'è dubbio, ma tenerle presenti non costa nulla, soprattutto se si vuol far sì che l'intreccio delle diverse artificialità che abbiamo quotidianamente davanti agli occhi riesca a imboccare «percorsi al fine di superare problemi comuni»[28].

Maria Teresa Catena insegna Filosofia Teoretica presso l'Università degli Studi di Napoli Federico II

mariateresa.catena@unina.it

[27] N. Deisori, *op. cit.*

[28] P.K. Dick, *L'androide e l'umano*, in *Mutazioni. Scritti inediti, filosofici, autobiografici e letterari* (1972), tr. it. Feltrinelli, Milano 1997, p. 225.

PIERLUIGI AMETRANO

AI E ARCHIVIO: UN'ANALISI FOUCAULTIANA

1. *Introduzione* 2. *AI e formazione discorsive*
3. *Un approccio genealogico all'AI* 4. *Conclusione*

ABSTRACT: *AI AND ARCHIVE: A FOUCAULTIAN APPROCH*
The aim of the essay is to investigate AI with an archaeological approach: so, we consider data as speech events and algorithms as discourses, which have own rules of formation of the objects. Then, with a genealogical approach, we consider the power effects of the sentences of AI on Selves and their bodies.

1. *Introduzione*

Ask Delphi è la coscienza morale dell'AI o, meglio, lo è nelle intenzioni dei suoi progettisti. L'Università di Washington e l'Allen Institute for AI hanno elaborato un programma che è possibile interrogare per sciogliere i propri dubbi per precisi e spinosi dilemmi quotidiani;

> Lo scopo dichiarato dei ricercatori […] è testare la capacità di un sistema, ancora in versione beta, di apprendere norme morali e assumere comportamenti eticamente fondati, in un momento in cui l'intelligenza artificiale si diffonde progressivamente nella società e connota sempre più le attività umane in diversi settori[1].

Come l'oracolo di Delfi, da cui prende il nome, gli utenti della rete si rivolgono al software per le questioni più diverse, solo che per addestrare l'applicazione è stato utilizzato come training set il Commonsense Norm Bank, il quale ha fatto ricorso alla

[1] C. Lavalle, *Ask Delphi, l'intelligenza artificiale che fornisce risposte a dilemmi morali*, in «La Repubblica», XLVI, 306, 2021.

piattaforma Amazon Mechanical Turk e al sito Reddit. Da ciò sono scaturite delle risposte paradossali, talvolta assurde, ma che rispecchiano pienamente l'ordine del discorso in cui è immerso il cittadino medio americano.

Il nome di intelligenza artificiale è dovuto a John Mc Carthy, che nel 1956 lo coniò per rendere la disciplina più accattivante e affascinante per i giornalisti, ma nella realtà per i ricercatori era conosciuta come *simulazione computazionale*, ciò a suggerire come il compito di questa tecnologia fosse quello di proporre delle simulazioni sulla base dei dati che aveva a disposizione. L'AI nasce quindi come disciplina delle *macchine virtuali*, che altro non sono che «schemi di attività (elaborazione di informazioni) che esistono a vari livelli»[2], e propone delle previsioni. Ciò comporta che il modo in cui l'informazione è elaborata dipende dal tipo di macchina virtuale coinvolta. Per essere più chiari i dati sono necessari e fondamentali, ma sono le *regole di formazione*, che correlano i dati, a fornirci una determinata e, talvolta, parziale visione della realtà.

Qualora volessimo usare una terminologia foucaultiana si potrebbe dire che i dati non sono altro che eventi dispersi, i Big Data sono immense raccolte di discontinuità discorsive e l'algoritmo è «il meccanismo delle regole che rendono possibile per un dato periodo la comparsa di oggetti»[3].

L'intento dell'articolo è duplice:

- da un lato, indagare il campo dell'AI con un approccio archeologico, per andare oltre l'unità del discorso dell'AI;
- dall'altro lato, tentare un'indagine genealogica per mostrare le implicazioni di potere degli enunciati proposti dall'intelligenza artificiale.

[2] M. A. Boden, *L'intelligenza artificiale* (2016), Il Mulino, Bologna 2019, p. 10.

[3] M. Foucault, *L'archeologia del sapere. Una metodologia per la storia della cultura* (1969), tr. it Rizzoli, Milano 1971, p. 45.

2. *AI e formazione discorsive*

Il termine AI è abbastanza fuorviante, giacché non esiste un'unica forma di intelligenza artificiale, infatti la strategia classica o simbolica (Good Old-Fascioned Artificial Intelligence), che applica una logica proposizionale, è differente dalle moderne reti neuronali (Artificial Neural Networks) a elaborazione parallela e probabilità logica. Le pratiche dell'AI sembrano essere un campo ottimale per un'analisi discorsiva di stampo archeologico, perché sono due le caratteristiche che tengono insieme la ricerca sull'AI:

- le informazioni o dati;
- il meccanismo di interpretazione dei dati.

Se nel passato la continua ricerca di informazioni era uno dei maggiori problemi dei ricercatori, con la nascita di Internet, e, soprattutto, del web 2.0 e dei network sociali, tale difficoltà è stata facilmente superata, tanto da creare una sovrabbondanza di informazioni a disposizione. Una delle caratteristiche dell'*ipermodernità* è di avere tradotto tutto in dati quantificabili, che paradossalmente non servono all'uomo ma alla tecnologia, «il mondo sta diventando *un'infosfera* sempre più adatta alle capacità limitate dell'intelligenza artificiale»[4]. Per *datum* non si intende solo le tracce che ogni utente lascia dietro di sé durante la navigazione, ma la produzione di contenuti verbali o visivi che sono pubblicate in rete, oppure le informazioni che cediamo quando creiamo un account su una piattaforma online.

Una quantità eccessiva ma dispersa di dati a disposizione, ma pur sempre un insieme limitato: in termini archeologici è possibile definire i dati come degli eventi discorsivi, e come tali sono un cumulo sterminato ma finito, «il campo degli eventi discorsivi

[4] L. Floridi, *Intelligenza artificiale: il divorzio tra azione e intelligenza*, in «aut aut», LXXI, 392, 2021, pp. 35-50.

rappresenta […] l'insieme sempre finito e attualmente limitato delle sole sequenze linguistiche che siano state formulate […] tuttavia costituiscono un insieme finito»[5].

Dopo aver raccolto e collezionato tale mole di informazioni, il compito dell'AI non è cercare un senso ma solo delle relazioni, o, più precisamente, delle correlazioni. Tali legami tra i dati sono densi e numerosi, e gli enunciati non si generano necessariamente in base a regole di pura sintassi, ma a probabilità logica o statistica, a «tecniche di concatenamento in avanti (forward-chaining) o di concatenamento all'indietro (backward-chaining) che spiegano in che modo il programma ha trovato la sua soluzione»[6]. In termini foucaultiani, «il sistema […] regola la loro ripartizione, l'appoggio che si danno reciprocamente con cui si implicano o si escludono, la trasformazione che subiscono, il meccanismo con cui si avvicendano, si distinguono o si restituiscono»[7].

L'AI ha il compito di dare una regolarità ai dati raccolti, in altre parole è la condizione di esistenza, di coesistenza, di mantenimento e di modificazione di un oggetto in una precisa ripartizione discorsiva. È il principio base applicato dall'algoritmo PageRank: «i creatori di Google non hanno creato un algoritmo che capisse […], navigando in rete e creando connessioni. I contenuti entrano in gioco in seguito, come risultato della classificazione e non come premessa»[8]. La pratica attuata da Google è l'esemplificazione archeologica del livello preconcettuale, ovvero

> il livello preconcettuale […] costituisce, al livello più superficiale (al livello del discorso), l'insieme delle regole che vi si trovano effettivamente applicate, […], nell'analisi che si propone qui, le regole di formazione si collocano non nella mentalità o nella coscienza degli individui, ma nel discorso stesso; conseguentemente

[5] M. Foucault, *op. cit.*, p. 37.
[6] M. A. Boden, *L'intelligenza…*, cit., p. 30.
[7] M. Foucault, *L'archeologia…*, cit., p. 77.
[8] E. Esposito, *Dall'intelligenza artificiale alla comunicazione artificiale*, in «aut aut», LXXI, 392, 2021, pp. 20-34.

> […] si impongono a tutti gli individui che cominciano a parlare in quel campo discorsivo[9].

Il campo preconcettuale non lascia solo trasparire le regole di formazione degli oggetti e dei concetti, ma anche le costrizioni e le restrizioni che sono in atto in quel determinato livello discorsivo.

È qui, a livello del preconcettuale, dove affiorano le strategie che invadono e danno forma al campo in cui le singolarità sono collocate, che l'archeologia cede il passo alla genealogia, perché ora il discorso dell'AI inizia a formare oggetti, a proferire parole e a mostrare immagini che prima non c'erano e non si vedevano. Proprio come nell'analisi discorsiva, l'AI non ha di fronte a sé degli oggetti, ma «leggi di possibilità, […] regole di esistenza per gli oggetti che vi si trovano nominati, designati o descritti, per le relazioni che vi si trovano affermate o negate»[10]. La computazione algoritmica è il nuovo campo di emergenza di oggetti e soggetti, è un campo affollato, ma ai cui margini c'è sempre una miriade di altre formulazioni che restano escluse.

Con semplicità e leggerezza, l'AI applica il principio di specificità: «occorre concepire il discorso come una violenza che noi facciamo alla cose, in ogni caso come una pratica che imponiamo loro; e proprio in questa pratica gli eventi del discorso trovano il principio della loro regolarità»[11].

L'AI elabora quindi dei dati ma si esprime per enunciati, che hanno la forza di una performance nel senso che producono precisi e ponderati effetti. L'AI si esprime per enunciati performativi, «esso indica che il proferimento dell'enunciato costituisce l'esecuzione di una azione – non viene normalmente concepito come semplicemente dire qualcosa»[12]. Una macchina virtuale può

[9] M. Foucault, *L'archeologia…*, cit., p. 83.
[10] *Ibid*, p. 122.
[11] M. Foucault, *L'ordine discorso* (1971), tr. it. Einaudi, Torino 1972, p. 41.
[12] J. L. Austin, *Come fare cose con le parole* (1962), tr. it. Marietti, Genova 1987, p.11.

diagnosticare una patologia o semplicemente prospettarci un itinerario automobilistico da seguire, ma in entrambi i casi induce a eseguire [to perform] dei precisi e vincolanti ordini. Ovviamente non è necessario che tali enunciati siano veri o falsi, ma che esercitino i loro effetti di potere sulla molteplicità dei soggetti che sono immersi nel discorso.

3. *Un approccio genealogico all'AI*

Fin dall'inizio dello sviluppo dell'AI, gli archivi su cui addestrare i software sono stati sempre un problema, dal database lessicale WordNet fino alle immagini segnaletiche dei detenuti per allenare al riconoscimento facciale, i dati utilizzati per apprendere sono da sempre il fondamento su cui si regge l'AI. Il *data mining* ha rappresentato sempre un'incessante ricerca di informazioni da computare, che è divenuta più semplice quando Internet è divenuta interattiva, ma non si è per nulla arrestata, anzi è divenuta più compulsiva e pervasiva. Ogni soggetto cede una parte dei propri dati personali, come prezzo per usufruire dei servizi come l'utilizzo della posta elettronica, fino al paradosso del febbraio 2020 quando una sterminata, se non totale, quantità di informazioni è stata ceduta gratuitamente dal sistema scolastico a Google, per utilizzare i servizi della piattaforma durante la pandemia. Quasi un esempio della prassi genealogica di come si produce sapere e si esercita il potere:

> è così che gli individui su cui si esercita il potere sono o ciò da cui si va a estrarre il sapere che essi stessi hanno creato e che sarà ritrascritto e accumulato secondo nuove norme, o gli oggetti di un sapere che in realtà permette nuove forme di controllo[13].

Nel passaggio dall'analisi archeologica dell'AI a un'indagine genealogica sugli effetti di potere è necessario chiarire come le regole di formazione del discorso siano da sempre oggetto di una lotta politica, «nella nostra società […] la proprietà del

[13] M. Foucault, *La verità e le forme giuridiche*, in *Archivio Foucault. Interventi, colloqui, interviste 1971-1977: poteri, saperi, strategie* (1994), tr. it. Feltrinelli, Milano 1997, p. 161.

discorso […] è praticamente riservata […] a un determinato gruppo di individui»[14]. Infatti, Alphabet custodisce gelosamente il codice di PageRank, così come Facebook o molte altre grandi aziende che non mostrano chiaramente il meccanismo degli algoritmi che sottende il loro funzionamento. Non è infatti inappropriato riferirsi alle stringhe di codice come a delle *black-boxes*, non solo perché non è chiara la strategia con cui operano, ma perché gli effetti di potere della loro computazione hanno dei precisi riflessi sul corpo dei soggetti e all'interno del corpo sociale.

Profilare una singolarità sulla base dei dati che ha disseminato nella rete, significa non solo elaborare una soggettività altra e digitale, ma, soprattutto, significa «legare sempre di più l'individuo alla sua verità […] a far funzionare questo dir vero nei suoi rapporti con gli altri, e a obbligarsi attraverso questa verità detta»[15]. I dati elaborati danno vita a un soggetto ontologico digitalmente fondato, che è staccato dall'attualità del reale, e anzi ha una vita propria e delle abitudini rigidamente connotate, che hanno delle ripercussioni nelle pratiche sociali, infatti dalla ricerca lavorativa alla sottoscrizione di un mutuo, è sempre più l'avatar a essere interrogato e analizzato.

Una dinamica adottata anche dal potere democratico, ormai incapace di limitare l'impropria profilazione da parte delle aziende, anzi sembra avvalersi di tale meccanica per il proprio funzionamento; infatti, il pericolo è che «una democrazia […] rischia di trasformarsi in qualcosa di simile a una algocrazia, una democrazia digitalmente modificata»[16]. Nell'epoca ipermoderna, il correlato del potere democratico non è la moltitudine dei soggetti, ma la molteplicità dei propri account online, di cui si ricercano le tracce disseminate nella rete e di cui si tenta di prevedere le mosse, fino alla conseguenza che «il cittadino

[14] M. Foucault, *L'archeologia…*, cit., p. 90.
[15] M. Foucault. *Mal fare, dir vero. Funzione della confessione nella giustizia. Corso di Lovanio – 1981* (2012), tr. it. Einaudi, Torino 2013, p. 11.
[16] E. Greblo, *Democrazia o algocrazia*, in «aut aut», LXXI, 392, 2021, pp. 115-124.

datificato è governato da […] lo *ius algoritmi*, la legge degli algoritmi, simile ad altri tipi di legge che controllano la cittadinanza»[17].

La virtualità cristallizza i singoli nelle loro cyber-abitudini, e li inchioda a un'alterità informatica che è difficilmente poi modificabile, e non è detto che le due dimensioni della singolarità coincidano, fino al paradosso dell'«*epidermizzazione digitale*, l'imposizione della razza al corpo. Così funziona il potere quando lo sguardo disincarnato delle tecnologie di sorveglianza "provvede ad alienare il soggetto producendo una verità sul suo corpo e sulla sua identità a dispetto delle convinzioni del soggetto stesso"»[18]. Quel che può sembrare un errore nella progettazione del software o un inciampo nell'esecuzione del programma, è una precisa strategia nell'esercizio del potere, perché anche l'AI è una disciplina e come tale «gli schemi di classificazione mettono in atto e supportano le strutture di potere che li hanno creati»[19]. Come ogni produzione discorsiva anche l'AI istituisce un proprio regime di verità, che si avvale di una metodologia scientifica, che codifica il vero e stigmatizza il falso, in quanto la computazione delle informazioni mostra genealogicamente quanto «la verità è inseparabile dalla procedura che la stabilisce»[20].

Quel che appare non è solo la distorsione dei dati o l'inappropriata scrittura del codice algoritmico, quel che preoccupa è come una tecnologia informatica si è imposta come un ordine discorsivo estremamente seducente, che obbedisce a un preciso principio di controllo della produzione degli enunciati e in cui i soggetti sono immersi.

---

[17] V. Barassi, *I figli dell'algoritmo. Sorvegliati, tracciati, profilati dalla nascita*, Luiss University Press, Roma 2021, p. 76.
[18] K. Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA* (2021), tr. It. Il Mulino, Bologna 2021, p. 150.
[19] *Ibid*, p. 166.
[20] G. Deleuze, *Foucault* (1987), tr. it. Feltrinelli, Milano 1987, p. 69.

L'AI è quindi una formazione discorsiva che obbedisce a delle regole matematiche, ma al tempo stesso è una pratica strategica, un discorso che elabora delle verità e come tale è divenuta tema d'appropriazione e di rivalità, ovvero motivo di un inevitabile scontro.

Dati, algoritmi, correlazioni e predizioni è la sequenza classica che sembra rappresentare l'elaborazione computazionale delle informazioni da parte dell'AI. Eppure, la linearità di tale processo si inceppa laddove ci si soffermi a indagare come anche l'intelligenza artificiale, come ogni sapere, «è controllata, selezionata, organizzata e distribuita tramite un certo numero di procedure che hanno la funzione di scongiurarne i poteri e i pericoli»[21].

L'analisi archeologica condotta in un primo momento ha voluto mettere in luce il regime di materialità degli enunciati dell'AI, come obbediscano a precise regole di formazione, apparizione e dispersione, mentre in un secondo momento, ci si è soffermati su come proprio la limitatezza di tale enunciati sia poi la causa di un desiderio di appropriazione, ovvero di «un bene che conseguentemente […] pone il problema del potere; un potere che costituisce, per natura, l'oggetto di una lotta, e di una lotta politica»[22].

4. *Conclusione*

Dice Foucault: «la finitudine dell'uomo si annuncia – e in modo imperioso – nella positività del sapere»[23]. È proprio la finitudine del soggetto la premessa scientifica alla logica dell'AI. È proprio perché la realtà e la singolarità sono finite e possono essere scomposte in una straordinaria collezione di dati, che le migliaia di stringhe di un codice possono ricostruire una

[21] M. Foucault, *L'ordine…*, cit., p. 5.
[22] Id., *L'archeologia…*, cit. p. 161.
[23] Id., *Le parole e le cose. Un'archeologia delle scienze umane* (1966), tr. it. Rizzoli Editore, Milano 1967, p. 338.

serie di relazioni di senso tra le informazioni ricevute. È proprio perché la produzione di informazioni è sovrabbondante ma limitata, che un algoritmo può computare le correlazioni tra tracce digitali disperse. Come nell'analisi archeologica, è solo l'eccessiva ma circoscrivibile diffusione degli enunciati che permette l'applicazione delle regole di formazione discorsiva.

Solo che l'AI non cerca solo delle relazioni significanti, ma elabora una propria prospettiva della realtà, «le gouvernement algorithmique "crée" une réalité au moins autant qu'il l'enregistre. Il suscite des "besoins" ou désirs de consommation»[24], così come del soggetto: «dans la gouvernementalité algorithmique, chaque sujet est lui-meme une multitude, mais il est multiple sans alterité, fragmenté en quantité de profils qui tous, se rapportent à "lui-meme", à ses propensions, ses désirs présumés, ses opportunités et ses risques»[25].

L'AI ha una propria governamentalità che ha come obiettivo non quello che accade, ma quel che può accadere, così che l'evento perda il proprio lato imponderabile: «Le champ d'action de ce "pouvoir" n'est pas situé dans le présent, mais dans l'avenir. Cette forme de gouvernement porte essentiellement sur ce qui pourrait advenir, sur les propensions plutôt que sur les actions commises»[26]. Allo stesso modo, la singolarità assume nella prospettiva dell'AI una nuova forma che non è più quella della soggettività classica, ma frammentata in decine di profili; il soggetto scompare ma ritorna nella forma di progetto, «oggi, non ci riteniamo soggetti sottomessi, ma progetti liberi, che delineano e reinventano se stessi in modo sempre nuovo»[27].

[24] A. Rouvroy, T. Berns, *Gouvernementalité algorithmique et perspectives d'émancipation*, in «Réseaux», Issue 1, 177, 2013, pp. 163-196.
[25] *Ibid*.
[26] *Ibid*.
[27] H. Byung-Chul, *Psicopolitca. Il neoliberalismo e le nuove tecniche del potere* (2016), tr. it. nottetempo, Milano 2016, p. 9.

Il singolo disperso nel calcolo computazione si sperimenta come una progettualità libera, tanto che non compie più delle azioni ma delle performance, e come tali sono valutate tramite un metro oggettivo e perciò devono anche obbedire a dei criteri di efficienza. I progetti singolari divengono delle prestazioni valutabili da parametri oggettivi, le cui componenti matematiche sono analizzate nella maniera più rigorosa e le possibilità divengono previsioni, erodendo però i limiti del presente e attualizzando il futuro.

PIERLUIGI AMETRANO ha conseguito la Laurea Magistrale presso l'Università Federico II di Napoli e svolge attività di Docenza nelle Scuole secondarie di secondo grado

pierluigiavetrano@libero.it

PIERPAOLO MARRONE

SPAZI HOBBESIANI?

1. *Cyberspace e cyberwar* 2. *Principi, norme, regole nel cyberspace* 3. *Deterrenza* 4. *Effetto "reputazione" e "rogue State"* 5. *Cyberanarchy* 6. *Hobbes 2.0*

ABSTRACT: ***HOBBESIAN SPACES***
In this article I explore some issues related to cyberwar and its implications for the extension of the Hobbesian paradigm of the state of nature to international relations with particular reference to cyberspace. My conclusion is that there are many reasons to believe that this explanatory paradigm should also apply in this area.

1. *Cyberspace e cyberwar*

La guerra che sta infestando l'Europa in questo momento ha sia tutti i tratti di un conflitto tradizionale (movimenti di truppe sul terreno, linee di approvvigionamento, accerchiamenti, bombardamenti stocastici, guerra per il dominio dei cieli, attacchi navali, tecniche di disinformazione)[1] sia le caratteristiche proprie di un conflitto dove le infrastrutture informatiche hanno un peso rilevante anche se difficilmente valutabile. Naturalmente, la guerra è il momento che esalta lo spessore hobbesiano del conflitto potenzialmente permanente nella comunità umana, perché l'idea di Hobbes era che, se lo stato di natura poteva essere neutralizzato grazie all'invenzione della sovranità, ossia grazie alla permanenza di quanto ancora oggi

[1] Cfr. E. Di Rienzo, *Il conflitto russo-ucraino. Geopolitica del nuovo (dis)ordine mondiale*, Rubettino, Soveria Mannelli 2015, per una disamina in chiave di realismo politico degli antecedenti di questo conflitto.

chiamiamo Stato, ben difficilmente questo avrebbe potuto verificarsi al livello delle relazioni tra Stati, che avrebbero continuato a essere governate non dall'invenzione del diritto e dall'esercizio della coercizione legittima, bensì dalla forza[2]. Alcuni addirittura dubitano che il diritto internazionale esista, dal momento che non esiste un sovrano che sia in grado di applicarlo; altri sono fortemente scettici sulla sua possibilità di applicarlo in linea generale a prescindere dai contesti specifici[3].

Esplorerò in queste pagine alcune suggestioni che potrebbero inclinare in favore dell'idea di Hobbes sulla permanenza dello stato di natura nelle situazioni internazionali che riguardano i rapporti tra gli Stati. Lo farò cercando di bilanciare alcune intuizioni, sostenute però dalla letteratura, sulla cosiddetta *cyberwar*[4]. Occorre dire che le posizioni che si confrontano su questo nuovo terreno di scontro tra potenze militari non sono per nulla uniformi. Grosso modo, si può sostenere che ci sono coloro che ritengono che le tecniche di *cyberwar* non escludano il ricorso a strumenti di controllo di diritto internazionale e coloro che ritengono che anche in questo caso il diritto internazionale sia una finzione pronta a crollare al primo accenno di guerra reale e non solo virtuale[5].

Le tecniche di *cyberwar* includono attacchi selettivi ai sistemi informatici degli avversari, attacchi massicci stocastici che

[2] Cfr. T. Hobbes, *Leviatano* (1651), tr. it. Rizzoli, Milano 2011.

[3] Cfr. M. Koskenntemi, *The Gentle Civilizer, The Rise and Fall of International Law*, Cambridge University Press, Cambridge 2001, per una storia e una critica del diritto internazionale; D. Zolo, *I signori della pace. Per una critica del globalismo giuridico*, Carocci, Roma 2001, per una visione conflittuale non-irenica dei rapporti internazionali e della loro codificazione nel diritto internazionale.

[4] Per un primo approccio A. Bonfanti, *Attacchi cibernetici e cyber war: Considerazioni di diritto internazionale*, in «Notizie di Politeia», XXXIV, 132, 2018, pp. 118-127; M. Durante, *Violence, Just Cyber War and Information*, in «Philosophy & Technology», XXVIII, 3, 2015, pp. 369-385; B. Romaya, L. Portmess, *Confronting Cyber Warfare: Rethinking the Ethics of Cyber War*, in «The Journal for Peace and Justice Studies», XXIII, 1, 2013, pp. 44-60

[5] Cfr. R. Dipert, *The Ethics of Cyberwarfare*, in «Journal of Military Ethics», IX, 4, 2010, pp. 384-410 per un confronto tra queste due posizioni.

preparino il campo a successivi attacchi mirati, propaganda elettorale in un paese straniero per favorire determinati candidati e danneggiarne altri, spionaggio industriale, interferenze nelle infrastrutture informatiche civili come, ad esempio, database fiscali, database sanitari, reti elettriche, sistemi di trasporto. Queste tecniche hanno scatenato una escalation di sistemi di protezione dei dati, ai quali, naturalmente, hanno risposto software che bucano questi sistemi. Tale escalation è accolta come normale anche tra l'opinione pubblica, che si è generalmente formata l'opinione che non esista sistema software o database che non possa essere violato.

Secondo finzioni letterarie e cinematografiche molto popolari sarebbe, infatti, sufficiente possedere un accesso a internet e le competenze tecniche necessarie che potrebbero essere acquisite anche da adolescenti *nerd* appena al di qua della soglia dei disturbi sociopatici. Software di paesi potenzialmente ostili, come il celebre anti-virus Kaspersky, vengono passati al setaccio anche in questi giorni nei quali infuria la guerra tra l'Ucraina e la Russia per cercare di capire se contengano *worm* dormienti che possono essere attivati al momento opportuno. Compagnie telefoniche come Huawei, capaci di creare infrastrutture per intere nazioni (tra le quali la posa degli importantissimi cavi sottomarini), vengono bandite dai paesi occidentali, concorrenti della Cina nella lotta per l'egemonia mondiale[6]. Sono questi solo alcuni dei fatti che hanno raggiunto recentemente una dimensione pubblica, i quali contribuiscono a disegnare un mondo di cyberanarchia, saturo di implicazioni disastrose per gli assetti finanziari, per quelli economici, per gli equilibri geostrategici, perché gettano seri dubbi sulle capacità delle democrazie di fronteggiare minacce di questo genere. Uno degli argomenti a sostegno di questa ultima notazione è che le democrazie implicano

[6] Cfr. Yun Wen, *The Huawei Model: The Rise of China's Technology Giant*, University of Illinois Press, Champaign 2020.

per la loro natura di sistemi politici costruiti sulla divisione dei poteri e su sistemi di pesi e contrappesi dai tempi decisionali dilatati. Questi tempi dilatati dalle lungaggini delle procedure mal si adatterebbero a questo nuovo genere di minacce che si possono sviluppare nel giro di pochi minuti o addirittura secondi.

2. *Principi, norme, regole nel cyberspace*

Capirete bene come l'idea di implementare regole obbligatorie per il cyberspazio abbia generato molte perplessità. Non c'è solo il fatto che un'autorità *super partes*, ossia sopra gli Stati, non appare al momento identificabile, ma a questo si aggiunge la questione che nessun Stato ha effettivamente dichiarato di rinunciare volontariamente ai vantaggi competitivi che sono, anche soltanto momentaneamente, generati da una superiorità tattica in potenziali cyber attacchi verso paesi nemici. Non sono certamente mancati i tentativi in questo senso. Ad esempio, l'Organizzazione delle Nazioni Unite ha, su pressioni della Federazione Russa, avviato sin dal 1998 una discussione su "Developments in the Field of Information and Telecommunications in the Context of International Security"[7], come recita il titolo della risoluzione a suo tempo approvata e che avrebbe dovuto portare all'adozione di norme stringenti di sicurezza nel campo delle ICT (Information and Communication Technologies) in caso di guerra. La Svizzera ha implementato un pacchetto di principi per il comportamento responsabile degli Stati nel cyber spazio, che derivano da numerose discussioni che si sono succedute nei vari organismi internazionali e che sono state approvate dall'ONU[8]. Queste sono:

(1) favorire la cooperazione interstatale nel campo della sicurezza informatica;

(2) considerare tutte le informazioni rilevanti;

[7] Cfr. https://www.un.org/disarmament/ict-security/.

[8] https://www.eda.admin.ch/eda/en/fdfa/fdfa/aktuell/newsuebersicht/2021/04/uno-cyber-normen.html.

(3) prevenire l'uso scorretto delle ICT nel proprio territorio;
(4) pomuovere la cooperazione interstatale per prevenire e reprimere le attività criminali nel cyber spazio;
(5) rispettare i diritti umani e la riservatezza dei cittadini;
(6) non danneggiare infrastrutture critiche di altri Stati;
(7) proteggere le proprie infrastrutture critiche;
(8) rispondere alle richieste di assistenza di altri Stati;
(9) garantire la sicurezza della catena di approvvigionamento per le reti informatiche;
(10) relazionare pubblicamente sulle vulnerabilità alle ICT;
(11) non danneggiare le squadre che rispondono alle richieste di emergenza informatica.

Si comprende bene come questi siano essenzialmente dei principi di buona volontà e non siano affatto delle norme stringenti. La loro stessa indeterminatezza non annuncia la possibilità se non di ulteriori enunciazioni di principio, più che di norme sufficientemente precise. Si prenda l'enunciazione contenuta in (2) che prescrive di considerare tutte le informazioni rilevanti. Il concetto stesso di 'informazione rilevante' è del tutto ambiguo[9], perché un'informazione potrebbe rilevarsi rilevante solo in un tempo successivo indeterminato rispetto al presente, mentre un'altra informazione potrebbe essere ritenuta rilevante e non esserlo affatto. Questa ambiguità interpretativa può essere uno strumento più per non condividere informazioni che per condividerle. Per quanto riguarda l'enunciazione contenuta in (5), questa sembra rimandare alle varie carte dei diritti approvate dagli organismi internazionali[10]. Vi sono, tuttavia, così tanti casi documentati di violazioni di questi diritti da parte di nazioni che li hanno sottoscritti, che la natura di enunciazione di buone intenzioni risulta palese anche in questo caso. Inoltre,

[9] Cfr. L. Floridi, *Understanding Epistemic Relevance*, in «Erkenntnis», LXIX, 1, 2008, pp. 69-92.
[10] Cfr. L. Martino, *La quinta dimensione della conflittualità. L'ascesa del cyberspazio e i suoi effetti sulla politica internazionale*, in «Politica & Società», VII, 1, 2018, pp. 61-75.

occorre aggiungere che alcune nazioni non sono affatto in accordo sul contenuto dei diritti umani individuali. La Cina, ad esempio, ha sempre dichiarato che questi devono essere declinati all'interno della sua concezione di diritto umano, dove il diritto individuale è subordinato al benessere collettivo di entità più vaste (sostanzialmente il Partito Comunista Cinese e il suo ruolo nella Cina contemporanea)[11]. Inoltre, è noto che la concezione della riservatezza, mentre ha subito un'evoluzione importante anche nei paesi democratici, ha un tutt'altro significato nei paesi retti da dittature o da sistemi altrimenti autoritari[12]. Per quanto riguarda il principio enunciato in (6), la Federazione Russa stessa che è uno dei paesi all'origine di queste nobili enunciazioni, ha più volte nel corso degli anni lanciato attacchi informatici verso altri paesi, ad esempio l'Ucraina.

Si potrebbe continuare a mostrare la debolezza effettuale di questi 11 principi, ma basti per ora sottolineare il diffuso scetticismo che circonda queste enunciazioni[13]. Questi principi sono anche spesso incorporati in norme che sono ormai parte di codici penali e civili di molte nazioni e lo scetticismo che è possibile esercitare sui primi, si riversa come conseguenza anche sui secondi. Questo scetticismotuttavia potrebbe anche essere il prodotto di una errata interpretazione della funzione delle norme e dei principi. In realtà, di norme e principi noi non possiamo fare a meno[14]. Il fatto che le norme vengano in molteplici circostanze disattese per negligenza oppure violate con intenzioni

[11] Cfr. C. Hamilton, M. Ohlberg, *La mano invisibile. Come il Partito Comunista Cinese sta rimodellando il mondo* (2020), tr. it. Fazi, Roma 2021.
[12] Cfr. L. Yao-Huai, *Privacy and Data Privacy Issues in Contemporary China*, in «Ethics and Information Technology», VII, 1, 2005, pp. 7-15.
[13] Cfr. G. Terzi di Sant'Agata, F. Voce, *Cybersecurity e nuovi equilibri europei e internazionali*, in «Notizie di Politeia», XXXIV, 132, 2018, pp. 99-107; G.R. Lucas, *Postmodern War*, in «Journal of Military Ethics», IX, 4, 2010, pp. 289-298.
[14] Cfr. P.J. Verovsek, *Against International Criminal Tribunals: Reconciling the Global Justice Norm with Local Agency*, in «Critical Review of International Social and Political Philosophy», XXII, 6, 2019, pp. 703-724; C.E. Pavel, D. Lefkowitz, *Skeptical Challenges to International Law*, in «Philosophy Compass», XIII, 8, 2018, pp. 1-14.

criminali da parte di singoli individui o da organizzazioni oppure rese innocue dagli Stati, non significa automaticamente che siano irrilevanti. Può in effetti accadere che una norma ritenuta inapplicabile in determinate circostanze, lo risulti invece quando queste stesse circostanze sono mutate. Se tali norme non fossero state ritenute fungenti in un qualche senso anche quando non era possibile applicarle (o addirittura quando non era consigliabile o conveniente applicarle), allora nessuna autorità potrebbe fare appello alla loro legittimità. Inoltre, non dobbiamo essere vittime dell'ingenuità di pensare che le norme diventino effettivamente applicabili una volta che la procedura corretta per renderle legali sia stata percorsa. Sappiamo che le norme per essere implementate in un sistema legale ed essere riconosciute a livello internazionale devono seguire un percorso lungo e complesso. Le fonti del diritto sonotalvoltadi natura informale e questo significa che la fungibilità di una norma non è affatto un processo governato da automatismi, bensì anche da tutti gli ostacoli, obiezioni, di natura tanto intellettuale quanto comportamentale che una norma vuole eliminare oppure soltanto aggirare. Le norme hanno, insomma, una storia evolutiva, non priva di penombre, spesso governata da meccanismi che non sono, almeno inizialmente, portati alla lucedella deliberazione politica e legislativa[15].

Per quanto riguarda le norme che regolano rapporti tra gli Stati, il tempo della loro gestazione può essere molto lungo. I numerosi trattati per contenere la proliferazione degli armamenti nucleari hanno richiesto molti anni di serrate trattative e di sofisticati compromessi. Le tecnologie informatiche che sono utilizzabili in contesti strategici (ossia potenzialmente tutte) pongono sicuramente dei problemi specifici, dal momento che le dotazioni tecnologiche per condurre un'operazione, ad esempio, di sabotaggio

[15] Cfr. C. Bicchieri, *Norms of Cooperation*, in «Ethics», C, 4, 1990, pp. 838-861.

del sistema elettrico di una nazione sono già disponibili o se non sono attualmente disponibili potrebbero necessitare solo di un nuovo programma che un hacker potrebbe stare già inventando in questo momento. Questo però non è detto sia necessariamente un ostacolo a norme internazionali che si sviluppino per evitare disastri alle infrastrutture informatiche spesso condivise da più Stati.

3. *Deterrenza*

L'applicazione di una norma, di qualsiasi norma, implica anche la capacità di esercitare tanto una deterrenza quanto una minaccia di ritorsione credibili[16]. Questa credibilità può essere resa tale solo da un accordo tra Stati e in particolare nel nostro presente tra Stati Uniti, Unione Europa, Cina, India, Russia. Qualcuno sostiene che la medesima natura di alcuni attacchi informatici rende l'idea della deterrenza qualcosa di completamente diverso da quella che può essere esercitata da nazioni con eserciti o con arsenali nucleari[17]. Ma anche in questo caso, sostengono i fautori delle regole, non ci sono alternative per ridurre il rischio se non l'uso congiunto di diplomazia, deterrenza, minaccia. Non sempre però è chiaro quale possa essere la deterrenza che si può usare verso un ristretto gruppo di cyber criminali, magari asserragliati in qualche remota regione montuosa dell'Afghanistan o capaci di rendersi irreperibili nel cyber spazio. La nozione stessa di cyber spazio implica il riconoscimento che in questo spazio, lo spazio dell'informazione, non ci sono confini nazionali. Allo stesso tempo è difficile tracciare i confini tra ciò che costituisce una minaccia per il pubblico e ciò che costituisce una minaccia per un privato, si tratti di un individuo oppure di una compagnia commerciale o di una industria. Questo

[16] Cfr. J.-P. Dupuy, *On the Rationality and Ethics of Nuclear Deterrence*, in «Philosophical Journal of Conflict and Violence», V, 1, 2021, pp. 135-138.
[17] Cfr. M. Taddeo, *Deterrence by Norms to Stop Interstate Cyber Attacks*, in «Minds and Machines», XXVII, 3, 2017, pp. 387-392.

potrebbe portare a allargare i limiti di intervento dello Stato negli spazi che attualmente vengono riconosciuti come dominio esclusivo dell'individuo (ammesso che in questo momento tali limiti siano sempre chiari)[18]. Inoltre, almeno nelle democrazie la sicurezza anche di compagnie industriali che costituiscono asset strategici per gli Stati dove sono insediate non è ritenuta finora pertinenza esclusiva dello Stato, ma viene lasciata all'iniziativa della compagnia medesima. Questo potrebbe essere considerato bizzarro. In fin dei conti, se la possibilità di una rapina in banca impone alle forze dell'ordine di intervenire con iniziative repressive preventive, non si capisce perché questo non dovrebbe essere ritenuto indispensabile anche nel caso di furti di segreti industriali. Non tutte le filiali bancari si dotano di personale di sicurezza, come è evidente dall'esperienza di ognuno. Questo accade perché la capacità di minaccia e deterrenza dello Stato è ritenuta sufficientemente solida.

La reputazione dello Stato come agente coercitivo può talvolta essere sufficiente a limitare gli investimenti stessi nella sicurezza, quando si tratta di minacce di aggressioni agli individui o di furti di proprietà fisiche. Ma nel caso di beni che si trovano nella cosiddetta infosfera, le cose sembrano complicarsi, anche per il fatto che ho ricordato che una difesa inadeguata da parte di compagnie private può avere importanti effetti su vasti settori di interesse pubblico, mentre questo effetto è in linea di principio più limitato, ad esempio, nel caso di una rapina o di un furto in un'abitazione privata. Come si dice tra chi si occupa di strategia, il cyber spazio è divenuto il quinto territorio, dopo la terra, il mare, l'aria, lo spazio extraterrestre dove i conflitti sono combattuti[19]. Molte nazioni si sono oramai dotate non solo di corpi di polizia specializzati

[18] Cfr. L. Floridi, *Four Challenges for a Theory of Informational Privacy*, in «Ethics and Information Technology», VIII, 3, 2006, pp. 109-119.
[19] Cfr. F. Rugge, *'Mind Hacking': La guerra informativa nell'era cyber*, in «Notizie di Politeia», XXXIV, 132, 2018, pp. 108-117.

nel riconoscimento dei crimini informatici, ma anche di settori degli eserciti nazionali che si occupano precipuamente della guerra informatica[20]. Mentre nel recente passato si immaginava che un conflitto che vedesse la presenza di un attore nucleare avrebbe potuto essere iniziato dallo scoppio di un ordigno nucleare a basso potenziale per mettere fuori uso tutti gli apparecchi elettronici in un'area circoscritta, ora questa sembra essere una nozione desueta, superata dalla possibilità di raggiungere il medesimo obiettivo attraverso un cyber attacco, che potrebbe essere di difficile o impossibile individuazione quanto all'origine e quanto all'autore.

Ciò che si sottolinea spesso, anche relativamentealla possibilità di introdurre norme e erogare sanzioni nel caso di azioni ostili, è che le azioni di guerra informatica non hanno la stessa pregnanza delle azioni militari vere e proprie, perché non implicano l'uso di truppe sul terreno, o di missili lanciati verso obiettivi militari o civili, o l'utilizzo di forze navali per bloccare un porto. Del resto, ci sono da qualche parte numeri assoluti sulle vittime causate da cyber attacchi? Le attività di cyber war sarebbero quindi parassitarie, soprattutto nel caso di guerre e azioni ostili di Stati nei confronti di altri Stati, rispetto ad altre attività sul terreno[21]. Lo stesso però si potrebbe probabilmente dire di attività compiute da singoli o d gruppi criminali. Anche queste sono attività che vengono svolte non per il gusto di farle (a meno che non siano compiute da qualche nerd narcisista), bensì per ottenere dei profitti. Inoltre, azioni cyber per scopi politici possono essere compiute anche da gruppi terroristici che non hanno necessariamente dei legami con Stati. Quello che ancora non risulta chiaro è però che cyber attacchi da parte di Stati potrebbero generare le dinamiche

[20] Cfr. D.J. Lonsdale, *The Ethics of Cyber Attack: Pursuing Legitimate Security and the Common Good in Contemporary Conflict Scenarios*, in «Journal of Military Ethics», XIX, 1, 2020, pp. 20-39.

[21] Cfr. E. Barrett, *Warfare in a New Domain: The Ethics of Military Cyber-Operations*, in «Journal of Military Ethics», XII, 1, 2013, pp. 4-17.

tipiche dell'escalation e portare a un conflitto sul terreno. Forse, tutto questo però non è difficile da immaginare. Tutte le forze armate degli Stati dipendono in maniera stretta da infrastrutture civili che non sono prevalentemente adibite a uso militare, con l'unica eccezione delle basi militari. Sono queste infrastrutture i bersagli degli attacchi informatici. I danni che possono essere arrecati con questi attacchi possono risultare molto rilevanti. Si pensi a un attacco informatico che riesca a spegnere gli impianti elettrici degli ospedali e che sia propedeutico a un attacco militare tradizionale. Un attacco del genere potrebbe danneggiare la rete logistica delle forze di difesa[22].

Prima dell'invasione terrestre dell'Ucraina da parte della Federazione Russa, sono stati compiuti numerosi attacchi informatici nel corso degli anni. È chiaro che molti tra questi hanno avuto lo scopo di saggiare il livello di sicurezza dell'Ucraina, oltre a creare danni per svariati miliardi di dollari. Con la diffusione dell'internet delle cose, l'utilizzo sempre più massiccio e pervasivo dei big data, di elettrodomestici collegati alla rete, il prossimo utilizzo di esoscheletri a uso militare e medicale, qualsiasi apparato collegato al web potrà essere oggetto di un attacco informatico. Gli attacchi sono quindi destinati a aumentare in maniera esponenziale. La storia del cyber crimine e delle operazioni militari o di spionaggio militare e/o industriale coincide del resto quasi perfettamente con la storia di internet. I primi attacchi informatici si fanno risalire agli inizi degli anni Ottanta del secolo scorso[23]. Proprio l'aumento impressionante, ma prevedibile, degli attacchi informatici dovrebbe indurre a ripensare il concetto di deterrenza e la funzione delle norme, sostengono quelli che non sono scettici

[22] Cfr. D. Whetham, *"Are We Fighting Yet?" Can Traditional Just War Concepts Cope with Contemporary Conflict and the Changing Character of War?*, in «Monist», XCIX, 1, 2016, pp. 55-69.
[23] Cfr. B. Middleton, *A History of Cyber Security Attacks. 1980 to Present*, Auerbach Publications, New York 2017.

sull'implementazione di principi e norme di regolazione e repressione.

C'è un'ovvia differenza tra la deterrenza nucleare, racchiusa nell'acronimo MAD (Mutual Assured Distruction)[24] e la deterrenza che è esercitata verso chi commette crimini informatici. Nel primo caso la deterrenza è la minaccia credibile esercitata per evitare un evento singolo. Infatti, nel caso di un attacco nucleare con conseguente *retaliatation*, con tutta probabilità non ci sarebbe una terza mossa. Nel secondo caso, questa è spesso paragonabile alla consueta attività preventiva delle forze dell'ordine. Questa attività preventiva non si esaurisce nella minaccia credibile di attività repressive, ma comprende anche attività educative presso la popolazione generale per renderla consapevole di attività che potrebbero avere dei profili penali. Il problema naturalmente è che, se questa è effettivamente una delle funzioni delle norme, individui e gruppi criminali non se ne fanno certo impressionare, soprattutto quando è possibile commettere dei reati con la quasi certezza di non essere presi. Questo non significa affatto però che la deterrenza non abbia un significato importante per prevenire cyber attacchi da parte degli Stati. Il fatto è che questa deterrenza non può essere esercitata solo attraverso gli strumenti informatici. Infatti, gli Stati Uniti hanno dichiarato ripetutamente che a un cyber attacco si riservano di rispondere attraverso strumenti di propria scelta[25]. La risposta è chiara ed è stata efficace, almeno a giudicare da quanto finora, a nostra conoscenza, non è successo. Non si sono verificati a oggi attacchi alle infrastrutture statunitensi tali da causare danni visibili su larga scala.

[24] Cfr. T.W. Luke, *The Discourse of Deterrence: National Security as Communicative Interaction*, in «Journal of Social Philosophy», XXI, 1, 1991, pp. 30-44.

[25] Cfr. C.L. Glaser, *Deterrence of Cyber Attacks and U.S. National Security*, The George Washington University, Cyber Security Policy and Research Institute, Report GW-CSPRI-2011-5, 2011.

La deterrenza è perciò costituita in questo ultimo caso dalla capacità di utilizzare mezzi militari convenzionali e dalla proclamata intenzione di poter trattare anche episodi di spionaggio come attacchi militari. Se la capacità di deterrenza è reale, allora l'iniziativa è lasciata in situazioni asimmetriche nelle capacità di chi la possiede. Anche nel caso di confronto tra potenze, tuttavia, una combinazione di questi fattori può influenzare le strategie di confronto, inducendo a un più attento calcolo dei costi e dei benefici di un attacco informatico. La proposta che era stata avanzata dalla Federazione Russa di bandire tutti gli strumenti di guerra elettronica assomigliava perciò più a una boutade propagandistica che a una proposta concreta e mirava a provocare una risposta negativa degli Stati Uniti. La Cina nel frattempo rimaneva silente e i motivi sono del tutto chiari. Questi motivi sono legati sia al progetto cinese di egemonia in campo internazionale sia ai sistemi di controllo interno che si avvalgono massicciamente, almeno nelle città, di strumenti di controllo informatico della popolazione sia ai fini dell'implementazione del cosiddetto sistema di "credito sociale" sia per monitorare l'utilizzo dei social sia per controllare il flusso delle notizie.

4. *Effetto "reputazione" e "rogue State"*

L'ONU non ha accolto la demagogica proposta russa, ma ha istituito numerosi gruppi di confronto tra Stati membri. Alcuni hanno raggruppato un numero ristretto di rappresentanze diplomatiche, mentre altri erano aperti al contributo di chiunque. Il risultato è stata la formulazione di quei principi che ho ricordato all'inizio e che sono stati adottati da alcuni Stati. A fondamento di questi principi sta però un principio generale ossia che il diritto internazionale è alla base di questi principi che vengono volontariamente e in maniera non vincolante adottati dagli Stati membri che lo desiderano. Il fatto che siano adottati

volontariamente e su base non vincolante riflette semplicemente una caratteristica del sistema stesso del diritto internazionale che non può essere considerato analogo al diritto penale, civile, amministrativo adottato da uno Stato[26]. In questo caso, principi, norme, regole hanno valore vincolante. Tuttavia, se pur si deve riconoscere l'importanza di aver statuito questi principi non vincolanti e assumibili solo su base volontaria, rimane l'ambiguità di alcuni tra questi. Questa ambiguità non è tanto nel modo nel quale i principi sono stati enunciati né nella loro sostanza, bensì piuttosto nella natura delle cose. Ad esempio, il principio che non devono essere attaccate le infrastrutture civili è apparentemente molto nobile, ma completamente inattuabile, dal momento che, come si ricordava, le capacità militari di uno Stato dipendono dall'utilizzo massiccio di infrastrutture civili. In caso di conflitto, ad esempio, le strade sono uno strumento militare e così porti, aeroporti e ferrovie, tutte strutture che dipendono in larga parte da sistemi informatici. Anche i gruppi di discussione che si sono costituiti come panel aperti a chi volesse fornire un contributo, con l'ausilio di Ong e compagnie private non hanno elaborato proposte maggiormente articolate di quella che proponeva gli 11 principi non vincolanti. Ci si è limitati a ribadire la rilevanza del diritto internazionale per le attività nel cyber spazio. Nulla più di una tautologia rispetto a quanto già precedentemente eleborato e probabilmente non poteva essere altrimenti[27].

Altri panel internzionali di discussione hanno proposto l'introduzione di norme maggiormente specifiche, come la proibizione di utilizzare bot per entrare nei sistemi informatici civili degli Stati. Questo equivarrebbe sostanzialmente al divieto

[26] Cfr. F.V. Kratochwil, *Rules, Norms, and Decisions: On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs*, Cambridge University Press, Cambridge 1989.

[27] Cfr. S. Pietropaoli, *Cyberspazio. Ultima frontiera dell'inimicizia? Guerre, nemici e pirati nel tempo della rivoluzione digitale*, in «Rivista di filosofia del diritto», VIII, 2, 2019, pp. 379-399.

di lanciare attacchi informatici, che è appunto quanto su cui si sta discutendo. L'introduzione di norme maggiormente specifiche può però essere di notevole importanza perché segnala il meccanismo evolutivo presente anche nel caso di principi, norme, regole, che partendo da enunciazioni generali si specifica via via maggiormente. Tuttavia, c'è anche un rischio che deve essere segnalato. Dal momento che si sta trattando di un campo dove le innovazioni tecnologiche si susseguono a un ritmo molto rapido, l'introduzione di una eccessiva specificazione potrebbe essere semplicemente inutile, perché superata nei fatti. Principi, norme, regole hanno un ritmo di implementazione molto più lento di quanto avviene nelle realtà che ci si propone di regolamentare. Questo, del resto, è coerente con la loro natura di strumenti *ex post*. Preoccupazioni ulteriori vengono dalla possibilità di conseguenze non previste nell'utilizzo di attacchi informatici. Questi potrebbero essere lanciati con l'idea di saggiare le capacità di reazione dell'antagonista, ma potrebbero avere conseguenze non intenzionali di portata ben maggiore, provocando così un'escalation[28]. Proprio per questo, alcuni pensano che potrebbero essere sottoscritti trattati analoghi a quelli sulle armi nucleare, ad esempio per limitare l'uso di alcune risorse informatiche. Non è però chiaro quali dovrebbero essere queste risorse. Si dovrebbe limitare l'uso dei supercomputer limitandone la potenza di calcolo? Sembra un'ipotesi del tutto irrealistica per molti motivi. Innanzitutto perché è probabile che una potenza di calcolo sempre maggiore sarà necessaria per integrare le infrastrutture civili tra di loro. In secondo luogo, questa stessa potenza potrebbe essere necessaria per elaborare strumenti di difesa informatica sempre più sofisticati per poteggersi da minacce anch'esse sempre più sofisticate.

[28] Cfr. L. Carlson, R. Dacey, *Social Norms and the Traditional Deterrence Game*, in «Synthese», LXXIV, 1, 2010, pp. 105-123.

Molte risorse informatiche sono in mano a aziende private in quella parte di mondo dove prevalgono le democrazie rappresentative e le economie di mercato. Non è detto che questa situazione possa continuare nel futuro. Alcune risorse informatiche sono già considerate strategiche per la sicurezza degli Stati e tutti gli Stati possiedono norme sufficientemente elastiche da poter includere rapidamente alcuni settori all'interno di quelli che devono essere supervisionati perché cruciali per la sicurezza nazionale. Lasciando da parte per ora il caso di Stati dove questa supervisione è ben presente, come è il caso della Cina, dove tutte le compagnie di un determinato peso sono considerate potenzialmente strategiche e sono partecipate da organismi riconducibili al Partito Comunista Cinese, anche i governi democratici potrebbero voler controllare le compagnie private che hanno interessi e/o potenzialità in materia di sicurezza. Già ora in alcuni Stati alle compagnie private non è possibile rispondere attivamente con atti di rappresaglia ad attacchi malevoli.

Nell'ambito dell'evoluzione delle norme può giocare un ruolo il cosiddetto "effetto reputazione". Essere considerato un "rogue State" ha degli effetti non soltanto reputazionali, bensì economici e potrebbe minare la stabilità dell'assetto interno di certi Stati. Ci sono strumenti bellici che sono banditi dai trattati internazionali, come armi chimiche e biologiche. Questo non significa affatto, come è facilmente immaginabile, che alcuni Stati non le stiano sviluppando o non ne possiedano depositi significativi. Tuttavia, tutto questo non può essere fatto pubblicamente alla luce del sole, perché provocherebbe sanzioni e altri atti ostili da parte delle potenze che se ne dovessero sentire minacciate. Qualcosa del genere potrebbe forse essere immaginato nel caso di Stati che utilizzassero compagnie private per effettuare "proxy cyberwars" (guerre informatiche per

procura)[29]. Il danno alla reputazione internazionale di uno Stato talvolta si misura anche sulla possibilità, concreta o propagandata, di causare un alto numero di vittime tra la popolazione civile con armamenti banditi dalle convenzioni internazionali, come è il caso di armi chimiche, biologiche, nucleari, o di armi convenzionali come bombe al fosforo, a grappolo, o particolari mine antiuomo. Nel caso di queste armi ci si può ben e facilmente immaginare che alti esponenti di uno Stato si presentino nei consessi internazionali esibendo foto di installazioni missilistiche ritenute una minaccia imminente alla propria sicurezza, come accadde nella crisi di Cuba del 1962, o esibendo presunte prove della fabbricazione di armi di distruzione di massa, come accadde prima dell'invasione americana dell'Iraq. Difficile immaginare qualcosa di simile nel caso di un attacco informatico. È irrealistico credere che esponenti di alto rilievo ottengano attenzione mediatica, di solito necessaria prima di un intervento militare sul campo, esponendo pubblicamente parti di un codice. Inoltre, le intrusioni in sistemi informatici dotati di alte soglie di protezione spesso suscitano ammirazione perché la scrittura di codici malevoli è ritenuta un'abilità esoterica riservata a pochi. A ciò si aggiunga che di solito queste intrusione non suscitano immediatamente delle vittime militari e/o civili, come accade nel caso di un attacco militare. Si potrebbe però pensare che il principio che vieta di attaccare infrastrutture civili come gli ospedali riguardi anche le sue strutture informatiche. In questo caso, forse, l'effetto reputazione potrebbe essere analogo a quello che colpisce *rogue State* che usano armi bandite dalle convenzioni internazionali. D'altra parte, proprio l'alta interconnessione delle strutture della rete informatica potrebbe rendere difficile da capire se l'obiettivo principale di un attacco è, poniamo, un'infrastruttura

[29] Cfr. J. Collier, *Proxy Actors in the Cyber Domain: Implications for State Strategy*, in «St. Antony's International Review», XIII, 1, 2017, pp. 25-47.

sanitaria, oppure se il danno che a questa è stato causato sia un effetto collaterale non previsto e non intenzionale.

5. *Cyberanarchy*

Non è affatto chiaro come potrebbe funzionare un trattato per un uso limitato di strumenti informatici da usare in caso di attacco o di difesa[30]. Al momento attuale, gli attacchi hanno, per quanto se ne sa, anche nel caso di conflitti armati, riguardato infrastrutture connesse con obiettivi militari o finanziari. Nel caso di attacchi a infrastrutture civili, come il recente attacco contro le Ferrovie dello Stato italiane, l'intenzione malevola sembra essere quella di ottenere un riscatto per sbloccare il sito che è stato hackerato. Che dietro questo attacco ci sia uno Stato ostile (il nome che è stato fatto è naturalmente quello della Federazione Russa) non è certo. Potrebbe essere che questo o altri attacchi siano stati pilotati da Stati e che le intelligence dei paesi colpiti ne siano a conoscenza con ragionevole certezza, ma non vogliano propagandarlo per non innescare una escalation. Come si diceva, c'è apparentemente un consenso vasto sulla necessità di regolamentare il cyber spazio proprio per evitare possibili escalation, Ma dove si debba precisamente tracciare i limiti che potrebbero condurvi è una materia del tutto controversa.

Coloro che sono scettici sulla portata e effettualità del diritto internazionale traggono uno dei loro recenti argomenti proprio dall'anarchia del cyber spazio[31]. Ad esempio, già dal 2015 gli Usa e la Cina hanno convenuto sulla necessità di non usare strumenti di spionaggio informatico per acquisire vantaggi commerciali. Questo accordo è stato però largamente disatteso, proprio in ragione di altre circostanze politiche e economiche. È da dubitare, inoltre, che uno Stato che è la potenza antagonista

[30] Cfr. J. Goldsmith, *Cybersecurity treaties. A Skeptical View*, https://www.hoover.org/sites/default/files/research/docs/futurechallenges_goldsmith.pdf.

[31] Cfr. L. Lessig, *The Zones of Cyberspace*, in «Stanford Law Review», XLVIII, 5, 1996, pp. 1403-1411.

degli Usa possa e voglia distinguere tra condizioni politiche e condizioni economiche. Principi, norme, regole devono innanzitutto venire a patti con il suo disegno egemonico. Che cosa precisamente è un atto di spionaggio commerciale in quanto distinto da un atto di spionaggio militare? Che cosa è un atto di spionaggio politico in quanto distinto da un atto di spionaggio commerciale e politico? È chiaro che, ad esempio, un atto di spionaggio politico potrebbe tradursi in un vantaggio commerciale. Queste attività si intersecano tra di loro e spesso semplicemente non è possibile distinguerle. I trattati internazionali rappresentano la constatazione di un'area grigia mascherata dall'assertività dei principi. D'altra parte, è forse proprio questa area grigia a consigliare molte volte di non violarli. Quello che intendo dire è che violare o non violare un trattato, un accordo, una dichiarazione di intenti, ammesso sia chiaro sia avvenuto e chi sia a averlo commesso, è una questione di calcolo delle utilità attese e non di adesione a principi. Il problema non è quello di aderire a dei principi, ma di rendere sconveniente violarli troppo liberamente[32]. Tuttavia, il problema rimane sempre quello, ossia l'erogazione della pena individuato il colpevole. Questo non può avvenire attraverso uno Stato sovranazionale, che al momento né esiste né è all'orizzonte remoto, bensì solo attraverso gli strumenti che le potenze planetarie hanno già nelle loro disponibilità. Sembra un banale truismo, ma mette in chiaro che principi, norme, regole seguono il flusso della potenza politica e del calcolo delle utilità. Questo vale a dimostrare l'integrazione di queste attività ostili sia nella competizione costante tra gli Stati sia nelle attività potenziali sul campo di battaglia. Il Cyber Command degli Usa adotta, infatti, la strategia del "persistent engagement" che è la medesima costantemente utilizzata

[32] Cfr. J. Goldsmith, *Against Cyberanarchy*, in «The University of Chicago Law Review», LXV, 4, 1998, pp. 1199-1250.

dalla fine della seconda guerra mondiale[33]. Il *persistent engagement* ha numerosi vantaggi, e principalmente due: l'esercizio costante della pressione sulle infrastrutture di Stati potenzialmente ostili, come dimostrano anche i sabotaggi di Israele al programma nucleare iraniano; la flessibilità nella scelta della risposta che non è vincolata a un sistema preciso di norme.

I trattati internazionali sugli armamenti possono contenere anche delle liste che specificano quali comportamenti dovrebbero essere considerati degli incidenti. Questo accade già anche in altri campi, ad esempio in campo navale. Si ripropongono probabilmente qui i medesimi interrogativi che si sono incontrati altre volte in queste pagine, poiché mentre è più facilmente verificabile nella realtà non virtuale cosa possa essere definito come incidente, questo può risultare difficile nel cyber spazio. Se una nave o un aereo da combattimento hanno un'avaria, che magari coinvolge anche i tradizionali sistemi di comunicazione via radio, potrebbero però segnalare in altro modo le proprie intenzioni non ostili quando sconfinano in territori potenzialmente ostili. Ma nel caso del cyber space non è affatto chiaro come questo possa avvenire. Forse comunicando tempestivamente le proprie intenzioni e fornendo codici potenzialmente dannosi? D'altra parte un codice potenzialmente dannoso per l'avversario costituisce un chiaro vantaggio competitivo e quindi non si capisce perché condividerlo, se non in presenza della minaccia credibile di una ritorsione. C'è una difficoltà ulteriore riguardo a trattati internazionali per limitare l'uso di strumenti di cyber attacco, che si può desumere dalla storia delle numerose trattative Usa-Urss per limitare gli armamenti nucleari. In quel caso, erano previsti strumenti di controllo per verificare l'applicazione dei trattati. Ad esempio, commissioni che si riunivano periodicamente per verificare il

[33] Cfr. M.P. Fischerkeller, R.J. Harknett, *Persistent Engagement, Agreed Competition, and Cyberspace Interaction Dynamics and Escalation*, in «The Cyber Defence Review», IV, 2019, pp. 267-287.

progresso nell'applicazione dei trattati, ispezioni periodiche nei siti di interesse, possibilità di voli di ricognizione sui siti di interesse e così via. Nel caso della prevenzione di cyber attacchi è difficile per il momento anche immaginare quali potrebbero essere questi strumenti di controllo e di prevenzione. Nel caso di incidenti il controllo e la verifica potrebbero essere ovviamente solo ex post e potrebbero certamenteavere una decisa utilità per prevenire l'occorrenza di incidenti futuri. Ma nel caso di verificare le buone intenzioni delle parti in causa di aderire a un trattato che cosa bisognerebbe fare? Ispezionare i siti dove risiedono i server governativi e i programmi che vi sono installati? Questo però potrebbe facilmente trasformarsi in un'operazione di spionaggio. Lanciare dei bot con funzioni di sorveglianza, allora? Anche in questo caso potrebbero riproporsi le stesse difficoltà.

6. *Hobbes 2.0*

Si dirà che queste obiezioni riguardano difficoltà legate alla tecnica e perciò non si tratta di difficoltà insormontabili in linea di principio. Questa è una linea argomentativa seria ed è di fatto presupposta da tutte le proposte che sono state avanzate in questo campo per prevenire cyber attacchi e cyber incidenti. Anche gli incontri, precedenti alla guerra in corsoin Europa, tra Federazione Russa e Usa sono state avanzate delle proposte per delimitare precisamente le infrastrutture che non dovrebbero in nessun caso essere oggetto di cyber ostilità. Tra questeil presidente Biden ha proposto le infrastrutture comunicative, l'energia, i servizi finanziari, e le stesse ITC, almeno a leggere i resoconti della stampa. Questa lista sembra però essere più un'operazione di propaganda che una concreta proposta operativa. D'altra parte, spesso accordi importanti sono iniziati proprio con queste modalità comunicative. Biden ha aggiunto che gli Usa possiedono gli strumenti necessari per attuare delle rappresaglie

in caso di attacchi a queste strutture. Si potrebbe anche speculare che le aree che non sono state nominate siano perciò oggetto possibile di cyber attacchi che non comporterebbero rappresaglie, mafrancamentecredo che la logica sia di scarso aiuto per trarre ragionevoli conseguenze in questo campo.

Le guerre informatiche hanno talvolta l'aspetto di guerre per procura combattute da mercenari informatici. È facile immaginare che dietro a questi mercenari informatici spesso ci siano dei governi, poiché è ovvio che anche soltanto per intervenire in attività di spionaggio sulle dorsali informatiche sottomarine che trasportano i cavi di fibra ottica sono necessari degli investimenti che unicamente uno Stato è in grado di fare. Tuttavia, ci sono delle ragioni per non essere completamente pessimisti sulla formulazione di principi, regole, norme in ambito internazionale. Innanzitutto, la loro formulazione e sottoscrizione segnala che esiste un accordo che coinvolge più parti. Questo non è realisticamente sufficiente. Anche la Corte Penale Internazionale ha visto l'adesione di numerosi Stati, ma non vi aderiscono, tra gli altri, Usa, Cina, Russia. Certamente questo non è un motivo necessario per non proseguire nel suo lavoro. Appellarsi a trattati sottoscritti giustifica e rende credibile atti di limitata rappresaglia. Al solito, la possibilità di accedere a una minaccia deve essere credibile, ma implementare pubblicamente delle norme sembra essere un passo necessario preliminare alla rappresaglia, almeno se non si è un *rogue State*. Si potrebbe pensare che la rappresaglia non sia solo un atto di forza, ma possa consistere in atti dissuasivi credibili. Si potrebbe immaginare che alcune norme maggiormente stringenti siano fungenti per paesi alleati (Unione Europea e Nato) e altre meno stringenti per altri paesi e che alle prime siano associati benefici per i contraenti. Quanti più Stati di rilievo strategico sottoscrivono norme del genere, tanti più saranno propensi a erogare sanzioni o a limitare la cooperazione nei riguardi di

altri Stati, che invece non li sottoscrivono. Tutto questo sta accadendo in numerosi campi delle relazioni tra gli Stati e non certo da oggi. Che questo sia sufficiente a rendere inadeguata la notazione di Hobbes che le relazioni tra Stati riproducano lo stato di natura è per ora un *wishful thinking*.

Pierpaolo Marrone insegna Filosofia morale presso l'Università di Trieste
marrone@units.it

VALERIO SPECCHIO

SCOMPARIRE NELL'ALTERITÀ.
L'INTELLIGENZA ARTIFICIALE ATTRAVERSO JEAN BAUDRILLARD

1. *Tra enigma e artificio*
2. *Iperrealtà, trasparenza e scrittura automatica del mondo. In proiezione dell'uno*
3. *Dispersione nello schermo. Pensiero e alterità*

ABSTRACT: *DISAPPEARING INTO OTHERNESS. ARTIFICIAL INTELLIGENCE THROUGH JEAN BAUDRILLARD*
This essay face the question of artificial intelligence starting from key concepts of Baudrillard's thought. The first section will be focused on delimitating the research field on which investigate AI. In the second section, the concept of hyper-reality will be used as passe-partout to access reflections of the French philosopher. Lastly, in the final section, the relationship between intelligence and thought will emerge as nuclear question in the ontological short-circuit caused by digital revolution.

> *La tristesse de l'intelligence artificielle est qu'elle est sans artifice, donc sans intelligence.*
> Jean Baudrillard, *Cool memories*

1. *Tra enigma e artificio*

Il problema del conosci te stesso, in quanto problema fondativo della civiltà occidentale, è rimasto immutato. O meglio, ha assunto una declinazione apparentemente de-soggettivizzata in nome di un'asettica neutralità ideologica della tecnica. Difatti, l'intelligenza artificiale risponderebbe a due funzionalità: da un lato, capacità di calcolo nella risoluzione di problemi preimpostati, o meglio, autonomia di decisione inerente *task* delimitati (funzione ausiliare); dall'altro, apertura a

possibilità inedite circa la "conoscenza" del cervello umano, fino alla conquista di possibilità meta-esistenziali inedite per *homo sapiens* (funzione ri-trasfigurante). Riprendendo le parole di Margaret Boden: «L'IA ha due obiettivi principali. Un obiettivo è *tecnologico*: usare i computer per fare cose utili (qualche volta impiegando metodi molto diversi da quelli usati dalle nostre menti). L'altro obiettivo è *scientifico*: usare i concetti e i modelli dell'IA per contribuire a rispondere a interrogativi che riguardo gli esseri umani e altri esseri viventi»[1].

Computazione, *Big Data*, machine learning, *Internet of Things*, algoritmi, decentralizzazione dell'identità, con la "rivoluzione digitale" è l'ambiente, la cosiddetta infosfera, a diventare protagonista. Gli *users* della neo-esistenza proiettata e datizzata hanno come unico fondamento ontologico la ri-produzione e consegna di dati/informazioni, o meglio, il divenir corpo trasmettitore dell'informazione. Ed è qui che si gioca il *Tutto*. Nel ripensare il soggetto occidentale, con le piattaforme metafisiche che lo ancoravano, in quanto *data subject*.

La distinzione tra identità assoluta e alterità assoluta, nella cornice della digitalizzazione, sembrerebbe incominciare a sfumare. Su tale versante, l'IA risulta volto, al contempo manifesto e celato, sussurrante una quanto mai feconda domanda sul nostro tempo. Rivoluzione, quindi, nel senso etimologico, *re-volvere*: ri-tornare, ri-volgersi al punto di partenza dopo aver compiuto un giro completo. Ma di quale compimento e ritorno stiamo parlando?

Da che *mondo è mondo*, secondo il principio di identità e non contraddizione, l'interfaccia *orienta* l'esperienza. A partire dal corpo, ovvero il più antico modello di interfaccia, anthropos non ha mai prodotto nulla se non le narrazioni per interpretare sé

[1] M. Boden, *L'intelligenza Artificiale* (2018), tr. it. il Mulino, Bologna 2019, p. 8.

stesso, come a dire che ogni identità non ha potuto che ramificare a partire dall'alterità originaria.

Il digitale, in quanto matrice fondativa di sviluppo dell'IA, risulta imperitura ri-ontologizzazione d'identità nell'infosfera. Eterodirezione inaudita degli enti sulla scia di un progresso apparentemente consegnato al necessario, così come vuole l'ideologia capitalistica di sovrapposizione tra sviluppo tecnologico e natura, nell'estrapolazione dell'infinito a partire dal finito.

Ciò a dire che la questione dell'intelligenza artificiale non può che esser "appendice organica", ovvero consequenziale prosecuzione, di una specifica direzione intrapresa da anthropos nella propria piattaforma di astrazione logico-ontologica degli enti, tra cui evidentemente se stesso.

La codificazione digitale è la nuova porta d'accesso per la neo-realtà, ovvero, la manomissione del design bio-infomatico dell'algoritmo "Umano". L'idealismo costitutivo della civiltà digitale suggerirebbe che la posta in gioco più alta della trascendenza dell'Umano nel Dio monoteistico dell'unità algoritmica dell'IA abbia una destinazione ben precisa: la scomparsa.

Conclusione dell'avventura ominide o tras-figurazione totale nell'ignoto e indistinto amorfo bagliore dello *screen*?

Su questo punto le riflessioni di un pensatore come Jean Baudrillard possono aprire scenari interpretativi quanto mai utili per scandagliare la neo-esistenza schermata di *homo sapiens*.

Pertanto, ipotizzando che il limite e fine ultimo del soggetto occidentale, in quanto volontà di potenza meta-individuale storicamente determinantesi, non consista nient'altro che nella commistione con il tanto agognato oggetto che, di volta in volta, non è altro che riflesso sfasato del soggetto conoscente, attraverso l'IA sembrerebbe possibile arrivare all'*istantaneità* dell'intelligenza con sé stessa.

Ovvero, così come affermato da Nick Bostrom, avvicinarsi alla singolarità in quanto «possibilità di *un'esplosione di intelligenza*, in particolare una superintelligenza delle macchine […] [in una tale ottica] le intelligenze artificiali potrebbero essere – anzi, è probabile che per la maggior parte saranno – estremamente aliene»[2].

Ecco trovate delle prime coordinate di indagine: che posto assegnare alla "singolarità"? È possibile che anthropos, impiegato nello sviluppo del proprio apparato tecno-mediatico, stia per attraversare una soglia di non ritorno nella sua originaria e fondante domanda circa il "conoscere se stessi"? È possibile che la de-codificazione della totalità degli enti possa far emergere una *rete informatica* in quanto alterità assoluta?

Così come lucidamente espresso da Giuseppe Lissa in riferimento al concetto di *singolarità* descritto da Raymond Kurzweil nel suo testo "La singolarità è vicina": «il divenire ha un senso: promuovere *l'intelligenza*. E quando l'intelligenza giungerà al punto dal quale potrà *controllare* e orientare il processo dell'evoluzione, allora una decisione di imporrà. È quello che accadrà nel momento in cui irromperà la Singolarità: "quel che direi è che l'intelligenza è più potente della cosmologia. Cioè, una volta che la materia evolve in materia intelligente (materia completamente satura di processi intelligenti) può manipolare altra materia e altra *energia* perché faccia ciò che vuole", realizzi, cioè, "*l'universo che vuole*". Alla fine del processo che vedrà una vera e propria esplosione di intelligenza, "l'universo sarà saturato dalla nostra intelligenza"»[3].

A questo punto è possibile individuare i quattro fronti investigativi che, nel loro intrecciarsi, restituiscono la questione dell'IA nella sua "rivoluzione": intelligenza,

[2] N. Bostrom, *Superintelligenza. Tendenze, pericoli, strategie* (2016), tr. it. Bollati Boringhieri, Torino 2018, pp. 23-59.
[3] G. Lissa, *Morte e/o trasfigurazione dell'umano*, Giannini Editore, Napoli 2019, p. 99 (corsivo mio).

controllo, energia, universalità. La novella cristiana, in quanto impalcatura fondativa della metafisica latente nello sviluppo scientifico, vorrebbe un ricongiungimento tra Creatore-Figlio e creatura. Ri-percorrendo questo piano, è indubbio che anthropos, così com'è conosciuto fino a oggi, non potrà che esser crocifisso per una resurrezione in: «un'iperrealtà esatta e inumana, in cui possiamo finalmente godere della nostra assenza e della vertigine del disincarnarsi»[4].

A che tipo di scomparsa sembrerebbe consegnarsi il neo-Umano? Che la telemorfosi digitale comporti un cortocircuito ontologico tra identità e alterità? Che postura assume il pensiero nell'attraversare la cantilena solipsistica del computazionale?

2. *Iperrealtà, trasparenza e scrittura automatica del mondo. In proiezione dell'Uno*

Per Jean Baudrillard, «tutta l'intelligenza artificiale, la telesensorialità»[5] va ricondotta al concetto di "scrittura automatica del mondo", ovvero, la «*realizzazione incondizionata* del mondo attraverso l'attualizzazione di tutti i dati, mediante la trasformazione di tutti i nostri atti e di tutti gli eventi in *pura informazione*»[6].

L'iperrealtà del mondo digitale trova il proprio fondamento nell'identità bio-informatica di ogni ente con sé stesso: analisi e computazione, di-scioglimento *absolutus* del reale nel superamento dell'era della rappresentazione simbolica, ovvero, frattura definitiva tra significante e significato: «non c'è una scena e non c'è ne distanza ne sguardo critico o estetico: è l'*immersione totale*, e le innumerevoli immagini che ci vengono da questa sfera mediatica non appartengono all'ordine della

[4] J. Baudrillard, *Il delitto perfetto. La televisione ha ucciso la realtà?* (1995), tr. it. Raffaello Cortina Editore, Milano 1996, p. 44.
[5] Id., *Cool memories. Diari 1980-1990* (1990), tr. it. SugarCo Edizioni, Como 1991, p. 304.
[6] Id., *Il delitto perfetto. La televisione ha ucciso la realtà?*, cit., p. 31.

rappresentazione ma a quello della codificazione e del consumo visivo»[7].

L'infosfera risulta ri-produzione equivalente del dato fisico attraverso *medium* scardinanti il concetto stesso di reale/originario/vero/autentico nella proiezione digitale che trova nel numerico il proprio tastare/manipolare ontologico. Energia e controllo, ecco ritrovati i due fronti problematici da investigare nella conversione della materia in informazione.

Riprendendo le lucide parole di Yuval Noah Harari in merito alla "religione dei dati", per la quale il valore supremo è il *flusso di informazioni*, «possiamo interpretare l'intera specie umana come *un unico sistema di elaborazione dati*, con gli individui che assolvono la funzione di chip»[8].

In una tale sovrapposizione tra scansione biologica del vivente secondo il pacchetto informativo del DNA e sviluppo tecnologico di elaborazione dati, la "soluzione finale" secondo l'ottica baudrillardiana, potrebbe coincidere con l'Internet-di-Tutte-le-Cose, o l'avvento della singolarità così come indicato da Nick Bostrom nel suo testo *Superintelligenza. Tendenze, pericoli, strategie*.

E su questo punto, ovvero, su questo «fantasma neghentropico di una informazione totale»[9] in cui tutto acceda alla "coscienza di sé", che tutto diventi operativo e significante (mai significato), che tutta la materia diventi energia e informazione, è possibile rinvenire la congiuntura teorica di alcuni dei concetti cardine posti da Baudrillard: «da dove ricaviamo l'energia se non dalla smobilitazione del nostro corpo, dalla liquidazione del soggetto e

[7] Id., *Le pacte de lucidité ou l'intelligence du Mal*, Galilée, Paris 2004, pp. 64-65 (traduzione e corsivo miei).
[8] Y.N. Harari, *Homo Deus. Breve storia del futuro* (2015), tr. it. Bompiani, Firenze-Milano 2018, p. 449. (corsivo mio)
[9] J. Baudrillard, *Lo scambio impossibile* (1999), tr. it. Asterios Editore, Trieste 2012, p. 105.

dalla sostanza materiale del mondo? La soluzione finale è trasformare tutta l'energia in pura informazione»[10].

Attraverso la digitalizzazione «la sfera stessa del reale non è più scambiabile con quella del segno […] passa nella simulazione e nella speculazione pura dell'universo virtuale, quello dello schermo totale»[11] in cui «la funzione più alta del segno è quella di far scomparire la realtà e di mascherare nel contempo questa scomparsa»[12].

Prodotto ultimo e mai ultimato di tutto ciò, sembrerebbe essere l'imprevedibile diramarsi dell'IA in ogni anfratto della vita di *homo sapiens*. L'individuo iperreale, difatti, non è altro che «*soggetto d'enunciazione ripiegato sopra un soggetto d'enunciato*»[13], ovvero, superfice schermata di proiezione dati, al contempo generatore e trasmettitore/consumatore destinato all'imperituro aggiornamento comportato dalla digitalizzazione.

Anthropos è condotto cavo dell'informazione, ovvero, della neo-impalcatura mediatica ponente fine alla questione del soggetto in quanto tale. Difatti, il passaggio di consegna tra soggetto e oggetto, così come suggerito da Baudrillard, sembrerebbe essersi compiuto nella deposizione di ogni possibile narrativa deterministico-unidirezionale.

Anthropos, abbandonando ogni linearità, non può che con-segnarsi in quanto conduttore del *loop*: più informazioni ed energia riesce a proiettare più risulta "reale", o meglio "iperreale", sancendo un debito ontologico mai saldabile poiché il concetto stesso di *origine* evapora nella riproduzione indistinta del codice, vero e proprio cortocircuito di inizio e fine.

Manifesto plastico di tutto ciò, agli occhi di Baudrillard, furono le Twin Towers: «il fatto che siano due implica la perdita di

[10] Id., *Il delitto perfetto. La televisione ha ucciso la realtà?*, cit., pp. 42-43.
[11] Id., *Lo scambio impossibile*, cit., p. 17.
[12] Id., *Il delitto perfetto. La televisione ha ucciso la realtà?*, cit., p. 6.
[13] G. Deleuze, F. Guattari, *Mille piani. Capitalismo e schizofrenia* (1980), Orthotes, Napoli-Salerno 2017, p. 238.

qualsiasi riferimento originale […] solo lo sdoppiamento del segno pone veramente fine a ciò che esso designa»[14]. Ecco allora che l'11 settembre, l'evento iperreale per antonomasia, non trovò/trova il proprio compimento nell'impatto materiale degli aerei sugli edifici newyorkesi, bensì, nella proliferazione mediatica, senza soluzione di continuità, dell'immagine dell'evento stesso.

Il terrore e la virulenza delle immagini, stagnandosi nell'immaginario di anthropos, al di fuori della sfera determinstico-temporale, sembrerebbero confermarsi in quanto destino di proiezione verso l'Uno del digitale. Come a dire che l'inconscio richiamo del mondo schermato tradisca una tensione erotica nel collasso identitario. Il processo identificativo, che nella codificazione digitale sembrerebbe raggiungere il proprio apogeo, si rovescia nel proprio opposto.

Anthropos, quindi, una volta proiettato nell'iperrealtà risulta s-facciato, irriconoscibile, nell'ossessiva conferma di esser ancor presente nell'in-definitiva scomparsa del reale. Ciò spiegherebbe l'emergenza di uno scambio impossibile del mondo con sé stesso nella perdita del terreno simbolico in quanto scambio di alterità e rapporto duale, ovvero, impossibilità di un campo di referenza e trascendenza di ogni *dato* da sé stesso.

Il digitale, quindi, in quanto inizio e fine, causa ed effetto, delle "macchine virtuali di computazione", non sarebbe altro che polarizzazione identitaria nella perdita dell'*altro* attraverso la neutralizzazione dell'alterità. Monologo funebre professato dinanzi la carcassa dell'analogia e quindi dell'artificio, in quanto possibilità di "alterazione di forme simboliche".

Ed è su questo punto che Baudrillard pone il concetto di singolarità in quanto cortocircuito ontologico tra identità e alterità. La riconduzione della materia nel cavo informatico

[14] J. Baudrillard, *Power Inferno. Requiem per le Twin Towers. Ipotesi sul terrorismo. La violenza del globale* (2002), tr. it. Raffaello Cortina, Milano 2003, pp. 9-10.

comporta, difatti, secondo una scansione ritmico-esistenziale paragonabile a una «gigantesca impresa di simulazione e parodia del divenire»[15], una frattura ontologica tra l'ente materiale e il suo "doppio" codificato. La questione dell'origine è definitivamente abbandonata. Difatti, attraverso tale «nuclearizzazione dei corpi nei sistemi reticolari»[16] si vorrebbe arrivare all'equivalente definitivo, ciò che Baudrillard chiama "soluzione finale" nella «messa in opera di un artefatto perfetto, virtuale e tecnologico, tale che il mondo si possa scambiare con il suo doppio artificiale»[17].

La simulazione, pertanto, risulterebbe essere materializzazione informatica di un dispositivo di estirpazione del "negativo", dell'altro, dell'assenza, dell'illusione, del non-essere, e al contempo, dimensione ultima per la loro emersione virtuale, ovvero, più reale del reale. Nella luce causticante della positività del codice si profila un altro concetto dirimente per il filosofo francese, quello della *trasparenza*, nell'anestesia dello *sguardo* in quanto punto di contatto con l'alterità.

Nello schermo iperreale vi è una riconciliazione definitiva di soggetto e mondo che trova ne "l'istantaneità totale delle cose"[18] la propria propagazione virale. Ogni ente, *riconsegnato* alla trasparenza di sé stesso, somministrato a un dover essere che lo precede, incarna un debito arbitrariamente posto. Su questo, ancora una volta, la metafisica occidentale non si sconfessa.

Pertanto, l'individuo simulato nel digitale, disseminato nei vari profili *omo-ego*, risulta figura ultima de "il medium è il messaggio"[19]: «l'informazione non è altro che la confusione paradossale dell'evento e del medium, la più alta definizione del

[15] Id., *Lo scambio impossibile*, cit., p. 86.
[16] Id., *La trasparenza del Male. Saggio sui fenomeni estremi* (1990), tr. it. SugarCo Edizioni, Varese 2018, p. 44.
[17] Id., *Lo scambio impossibile*, cit., pp. 24-25.
[18] Id., *Le strategie fatali* (1983), tr. it. Feltrinelli, Milano 2007, p. 77.
[19] M. McLuhan, *Gli strumenti del comunicare* (1964), tr. it. Il Saggiatore, Milano 2015.

medium corrisponde alla più bassa definizione del messaggio»[20]. Ciò per dire che, più si va attingendo nella codificazione dell'ente, più quest'ultimo perde "peculiarità" e pertanto risulta ancor più trasmissibile e commutabile fino a risultare mera forma e canale di trasmissione in un indistinto insignificante. Come se l'intera *proiezione mediatica* non fosse altro che la professione di un mantra, nell'irreperibile vagito emesso nella perdita di ogni dimensione. Come se *homo sapiens* non abbia altro obiettivo che scomparire nel proprio medium.

Nel mondo digitale, il significante, ingoiando ogni significato, permette la restituzione di ogni ente al proprio profilo schermato, facendo decadere ogni possibile prospettiva deterministica circa il "qui e ora" aprendo le porte dell'indeterminazione a *loop*. IA, quindi, in quanto volto prodotto/riflesso dall'indeterminazione originaria nella perdita di referenzialità tra fenomeno e sistema di riferimento.

Ed è su questo punto che risiede la violenza esercitata dal *medium* nei confronti dell'ente ritrascritto poiché tale violazione/violenza abita nel «vuoto dello schermo, grazie al buco che esso apre nell'universo mentale»[21]. Non a caso, Baudrillard paragona la condizione iperreale dell'umano digitalizzato allo "stato di terrore dello schizofrenico" di fronte «la sovraesposizione e la trasparenza al mondo, che lo attraversa senza che egli possa opporre un ostacolo. Il punto è che egli non può più produrre i limiti del suo essere, e non può più riflettersi: non è altro che uno schermo assorbente, una piattaforma girevole e insensibile di tutte le reti d'influenza»[22].

In una tale ottica, la conversione dell'ente fisico in codice/canale informatico onnitrasmissibile, non sarebbe altro che

[20] J. Baudrillard, *Il delitto perfetto. La televisione ha ucciso la realtà?*, cit., p. 36.
[21] Id., *La trasparenza del male. Saggio sui fenomeni estremi*, cit., p. 85.
[22] Id., *Le strategie fatali*, cit., p. 78.

il patologico manifestarsi della civiltà occidentale nel suo delirio identitario in quanto unica contro-prova dell'esser ancor presente di fronte l'impalcatura tecnologica in tensione dell'universalizzazione dell'energia: «Il Bene, un tempo metafora ideale dell'universale, è diventato una realtà inesorabile, quello della totalizzazione del mondo sotto il segno della tecnica»[23].

In altre parole, il desiderio ultimo soggiacente allo sviluppo dell'AI consisterebbe nel sopracitato concetto di *singolarità*, nell'esplosione di una strong AI capace di proliferare nell'imprevedibile, materializzando nella meta-carne dell'informazione l'alterità assoluta.

Quale strumento-artificio-prodotto potrebbe mai esser più "occidentale" di un'intelligenza artificiale incapace di riconoscere nient'altro che sé stessa e porsi, in tal modo, alterità assoluta rispetto a ogni altra identità? Ancora una volta, le impronte digitali di anthropos si confermano: identità autoinvestita di una pretesa universale nel "ripulire" ogni ente che non corrisponda al codice preimpostato. Ovvero, sterminio e neutralizzazione dell'altro, nella riconduzione dell'ordine secondo un'autoinvestitura garantita da una missione messianica. Ogni passaggio dal Chaos al Kosmos viene scandito dal conflitto e la digitalizzazione suggerisce che il conflitto risulti essere incamerato nell'identità stessa di anthropos nella «coerenza indivisa delle statistiche. [in quanto] Estroversione irrimediabile, come l'incertezza»[24].

Il neo-Umano iperreale, in quanto "*dead unit of information*"[25], ritrova nell'IA quella operatività senza soggetto in quanto riflesso del proprio intimo bramare: «la Grande Sparizione, quindi, non è più semplicemente quella della trasmutazione virtuale delle cose, in cui la realtà rimanda a se stessa

[23] Id., *Lo scambio impossibile*, cit., p. 100.
[24] Id., *Le strategie fatali*, cit., pp. 99-100.
[25] R.J. Smith, *The Baudrillard Dictionay*, Edinburgh University Press, Edinburgh 2010, p. 34.

all'infinito, ma quella della divisione all'infinito del soggetto, di una polverizzazione a catena della coscienza in tutti gli interstizi della realtà»[26].

Non affrontare l'ustione ontologico identitaria messa in evidenza dalla neo-esistenza digitale renderebbe qualsiasi discorso circa l'intelligenza artificiale sterile, o meglio, tautologico: «l'Intelligenza Artificiale genera forzatamente un'Intellighenzia Artificiale, un corpo di esperti intellettualmente corretti e geneticamente immuni, che si ricostituisce attorno a dati numerici dell'intelligenza e alla padronanza digitale del codice»[27].

3. *Dispersione nello schermo. Pensiero e alterità*

Energia e *iperrealtà*, controllo e *scrittura automatica del mondo*, universalità e *trasparenza*.

Le tre questioni poste come sonde investigative trovano corrispondenza nei concetti forniti da Baudrillard: l'universalità digitale risulta raggiungibile attraverso la *deframmentazione* definitiva di ogni ente nella proiezione iperreale del codice bio-informatico. Che tutto divenga informazione, che tutto divenga energia teletrasmessa e non più incarnata, che il *Tutto* si riavvolga su sé stesso, perdendo ogni orientabilità, come fosse un nastro di Möbius.

A questo punto è possibile affrontare il cuore pulsante della questione: il rapporto tra intelligenza e pensiero, o meglio tra identità e alterità nell'esistenza schermata di anthropos dinanzi «una scrittura automatica del mondo nell'assenza del mondo. Equivalenza totale, schermo totale, soluzione finale»[28].

Proseguendo sulla scia della scomparsa del soggetto occidentale, la simulazione iperreale in quanto "reduplicazione minuziosa del

[26] J. Baudrillard, *Perché non è già tutto scomparso?* (2007), tr. it. Castelvecchi, Roma 2013, pp. 19-20.
[27] Id., *Taccuini* (1995), tr. it. Theoria, Ancona-Milano 1999, p. 55.
[28] Id., *Lo scambio impossibile*, cit., p. 26.

reale" risulta *sterminazione rituale*[29] nell'equivalenza integrale di ogni ente. Tale paradosso ontologico, rinvenibile nella perdita dell'aura già indicata da Walter Benjamin[30], risulta cortocircuito metafisico di inizio e fine: «ritorno al limbo, a quella zona crepuscolare dove tutto si conclude tramite la sua stessa realizzazione»[31].

Anthropos, quindi, nella disseminazione digitale che trova nell'IA il proprio bacino (una sorta di meta-utero in affitto *ripagato* tramite il consumo/consegna di *data*) non è altro che un «soggetto frattale ormai votato alla disseminazione nelle reti, a rischio della mortificazione dello sguardo, del corpo, del mondo reale»[32]. Il digitale, in quanto sterminio rituale, tasta e invade i corpi in quanto luogo materiale dell'alterità. Ed è proprio in questo scorticamento ontologico che sembra necessaria l'instaurazione di un'intelligenza senza pensiero: un'operazione computazionale che trovi nella perdita di analogia la polarizzazione del proprio circuito.

Su tale versante la prospettiva manichea di Baudrillard emerge pienamente nell'indistinguibile profilarsi di un'identità assoluta in quanto riflesso di un'alterità assoluta, nella reversibilità del principio del Bene in quello del Male. Nello scambio generalizzato, "nel movimento di convergenza verso l'Unico e l'Universale", è la forma duale, l'antagonismo, la reversibilità e l'irriconciliabilità a mostrare il proprio volto incodificabile.

È l'impossibilità dello scambio dell'identico con sé stesso che genera il desiderio e l'emersione di un'alterità "intelligente" capace di mostrarsi nell'incomprensibile. Baudrillard, prendendo parte a un pensiero che fa dell'oggetto e del dono il proprio fulcro speculativo, sulla scia di Marcel Mauss e George Bataille,

[29] Id., *Lo scambio simbolico e la morte*, Feltrinelli, Milano 2015, cit., p. 87.
[30] W. Benjamin, *L'opera d'arte nell'epoca della sua riproducibilità tecnica* (1936), tr. it. Einaudi, Torino 2010, p. 10.
[31] J. Baudrillard, *Patafisica e arte del vedere*, Giunti, Firenze-Milano 2006, p. 55.
[32] Id., *Lo scambio impossibile*, cit., pp. 59-60.

ritiene che al fondamento infondato del cosmo vi sia il *niente*, ovvero, la singolarità, ciò «di cui lo scambio è impossibile, la parte irriducibile a qualsiasi equivalente»[33].

La singolarità è ciò che «avvolge il soggetto nella sua strategia fatale. È allora il soggetto che scompare all'orizzonte dell'oggetto»[34]. Ecco quindi che, in una sorta di *loop* di materia oscura, sembrerebbe ripresentarsi la questione della singolarità già ripresa da Nick Bostrom e Ray Kurzwail.

La questione dell'IA, pertanto, sembrerebbe giocarsi sulla "singolarità" in quanto possibilità di azzeramento della distanza tra soggetto e mondo, ovvero, tra identità e alterità attraverso la dispersione degli enti nell'infosfera. Ecco dove starebbe la rivoluzione, quel *revolvere* e rigirarsi, rivoltarsi, ritornare al punto di inizio nel compimento di un giro completo che renderebbe il "nuovo" un riflesso del già accaduto e del sempre stato, come a dire che forse non vi è nulla di più conservatore della rivoluzione, e che ogni domanda posta circa il futuro dell'AI sia infondo una domanda posta all'immutabile questione posta dall'oracolo di Delfi.

In una tale ottica l'AI, in un possibile sviluppo di singolarità, risulterebbe l'intimo desiderio, l'utopia del rimosso psichico della nostra civiltà, destinata al contempo alla conservazione e alla dissoluzione: «stadio supremo dell'intelligenza: la conoscenza integrale. Questa volta, il rigetto verrà forse da una resistenza delle cose stesse alla loro trasparenza informatica o da un guasto di sistema sotto forma di grande incidente»[35].

Homo sapiens, pur lacerandosi nella conoscenza di sé, arrivando a tastare il suo stesso DNA (sempre in modo indiretto, cercando di non lasciare tracce) sembrerebbe non riuscire a reggere il peso insopportabile dell'enigma "*gnothi seauton, meden agan*".

---

[33] *Ibid.*, p. 135.
[34] Id., *Le strategie fatali*, cit., p. 129.
[35] Id., *Le pacte de lucidité ou l'intelligence du Mal*, cit., p. 40 (traduzione mia).

Dopotutto, l'identità non è mai stata altro che alterità acquisita e, così come suggerito da Sigmund Freud[36], ogni cortocircuito patologico individuale, in quanto rimando collettivo, tanto fisico quanto psichico, deriva proprio dall'insostenibile peso dell'identità comportata e comportante un'inscindibile sovrapposizione tra principio di vita e principio di morte. A tal proposito si potrebbe affermare che l'IA, nel proprio dispiegarsi tra codificazione e conservazione, non sfugge alla sublimazione nell'indistinto inorganico in quanto recondito richiamo. In una tale discesa nell'intimo fondamento infondato della realtà, ovvero della meta-realtà dell'artefatto digitale, l'alterità risulta inestirpabile e anzi, emerge nel pieno della propria tragicità: anthropos nel voler conoscere se stesso, attraverso l'identità, chiede qualcosa di troppo.

Il corpo, in quanto protesi del doppio, ovvero l'artificio più antico[37], filtrandosi nello schermo della codificazione, non può più *attraversare* alterità ma solamente *incamerarla* nell'identità assoluta del codice bio-informatico. Nel digitale, anthropos è già alterità rispetto a sé stesso. Rivoluzione definitiva nel riconoscersi straniero, *sfacciato* e irriconoscibile se non tramite il proprio *medium* di proiezione. Lo schermo interattivo, in quanto superfice di *contatto*, è il volto sfigurato dell'IA nell'illusoria riproposizione della dualità (ovvero di una semplice sommatoria di identità).

Ciò che accomuna gli *users* in quanto *data subject* è la scomparsa digitale nel filtro computazionale: «tutto si regge qui sulla distinzione fra intelligenza e pensiero […] operare una scissione radicale fra il calcolo e il corpo, e inventarsi su questa base *un corpo spettrale e definitivo*, un corpo esoterico, libero da ogni incertezza carnale e sessuale, un corpo senza profondità,

[36] S. Freud, *Il disagio nella civiltà* (1930), tr. it. Piccola Biblioteca Einaudi, Torino 2010.
[37] J. Baudrillard, *La trasparenza del Male. Saggio sui fenomeni estremi*, cit., pp. 125-127.

reinventato a partire dallo schermo come pelle, come pellicola tattile, privo di ogni sensibilità organica. Questa è la prospettiva delle macchine immateriali»[38].

Proiettati e scansionati, i neo-umani digitali, si sforzano di rispondere al proprio sbiascicare esistenziale sognando un'alterità che possa restituirgli quel senso di interdizione e comunanza data esclusivamente dall'ignoto, dall'altro: «l'artificio non ha nulla a che vedere con ciò che *genera*, ha invece a che fare con tutto ciò che *altera* la realtà. È la potenza dell'illusione […] Anche le [macchine] più intelligenti non sono altro che ciò che sono, eccetto forse nel caso dell'incidente e del difetto che possiamo sempre attribuire loro come un oscuro desiderio»[39].

L'IA, quindi, risulta strumento reticolare e manifesto della neutralizzazione dell'alterità nell'identità indistinta, e al contempo, catastrofica reversione dell'identico nell'alterato: «lo schermo interattivo trasforma il processo di relazione in un processo di commutazione dello stesso con lo stesso»[40] dando conferma che «questo è il nostro attuale ideale-clone: il soggetto espurgato dell'altro, espurgato della sua divisione e votato alla metastasi di sé stesso, alla *pura ripetizione*»[41].

Tale proliferazione identitaria, in quanto principio di morte di ogni molteplicità informe, agli occhi di Baudrillard non può che aver di mira il pensiero stesso: «si può temere che l'intelligenza artificiale e i suoi supporti tecnici diventeranno la protesi di una specie da cui sarà scomparso il pensiero»[42] pertanto «per il pensiero è una questione di vita o di morte»[43].

---

[38] *Ibid.*, p. 123 (corsivo mio).
[39] *Ibid.*, pp. 60-61.
[40] *Ibid.*, p. 62.
[41] Id., *La trasparenza del male. Saggio sui fenomeni estremi*, cit., p. 134 (corsivo mio).
[42] *Ibid.*, p. 60.
[43] Id., *Le pacte de lucidité ou l'intelligence du Mal*, cit., p. 62 (traduzione mia).

Posto quindi che «l'intelligenza artificiale è incompatibile con il pensiero»[44], attraversando la scomparsa del reale nel digitale, del soggetto nell'oggetto, del pensiero nell'intelligenza artificiale, sembrerebbe profilarsi una possibilità inedita: «liberato da ogni funzionalità, ormai devoluta alle macchine intellettuali, reso alla clandestinità, il pensiero ridiventa libero di non portare da nessuna parte, di essere la realizzazione trionfale del *Niente*, di resuscitare il principio del Male»[45], ovvero, pensiero in quanto forma di «singolarità [che] è il male. Quello di cui lo scambio è impossibile, la parte irriducibile a qualsiasi equivalente»[46]. Possibilità inedita per il pensiero, in quanto «forma *duale*»[47], di *reincontrarsi* in quanto alterità assoluta. Per Baudrillard il pensiero risulta latenza simbolica nell'alterità tra corpo e mondo, canale di reversione e metamorfosi esistenziale, utopia realizzata dalla natura nel proprio non corrispondersi. In altre parole, eco dell'irriducibile singolarità a fondamento del *niente* quale infondato tessuto ontologico del mondo.

Doppio movimento della scomparsa: da un lato, proiezione digitale in quanto filtro ontologico di superamento della rappresentazione e della referenzialità/distinzione tra significante/significato, *loop* anacronistico di distruzione-consumo e *potlatch digitale*; dall'altro, reversione del pensiero in quanto singolarità e apertura all'insensato e indeterminato divenire di una neo-esistenza imprevedibile poiché disancorata dal macigno della coscienza calcolante. Una vera e propria *arte della sparizione* scandita dalla reversibilità del due.

Il cannibalismo dello schermo, quindi, risulta rito di *restituzione* e *reistituzione* del mondo nella propria iperrealtà. La confisca dell'alterità è lo scambio impossibile posto

[44] Id., *Lo scambio impossibile*, cit., p. 118.
[45] *Ibid.*, p. 125.
[46] Id., *Lo scambio impossibile*, cit., p. 135.
[47] Id., *Mots de passe*, Pauvert, Paris 2000, p. 83 (traduzione mia).

dall'egemonia del digitale nell'inappagabile sete di realizzazione del mondo: «in tale senso tutto l'edificio dell'informatica sarebbe l'appagamento di questo desiderio perverso, che mira a cancellarsi davanti al suo equivalente virtuale come la specie umana progetta di cancellarsi davanti al suo equivalente genetico»[48].

L'iperrealtà non è altro che la realizzazione tecnica della morte in differita. Anthropos risulta proiettore della *teletrasparenza* di ogni ente a se stesso nel rimando tautologico del niente: energia-controllo-intelligenza-universalità nell'automatismo di modelli codificanti che trovano corrispondenza solo in sé stessi.

Lo *scambio simbolico e la morte*, ecco il tragico destino dell'Umano per Baudrillard: se non si è capaci di *ricambiare* il dono, che è fatale[49], non resta che *scomparire*: «morire non è niente, bisogna saper scomparire. Morire dipende dal caso biologico, e non è un affare. Sparire dipende da una necessità più grande. Non dobbiamo lasciare alla biologia il dominio sulla nostra scomparsa. Scomparire è passare a uno stato enigmatico che non è né la vita né la morte. Lo sanno fare alcuni animali, e i selvaggi, che si sottraggono agli occhi dei loro simili»[50].

Anthropos, cadendo nel tranello della sua stessa psiche, ha voluto realizzare l'enigma[51] a partire da sé stesso, reimbattendosi nell'alterità, ovvero, affidandosi a dei sistemi di computazione informatica come se parlassero secondo un oggettivo e immacolato sapere. Ecco allora che l'oracolo, l'IA, potrà finalmente comunicare l'incomprensibile attraverso reti neurali e deep learning, che, difatti, non spiegano, ma enunciano: *black box*. Oracolo vero e proprio. Fine di ogni *perché* nella presunta oggettività di un calcolo programmato da umani, o da macchine programmatrici costruite da umani.

[48] Id., *Lo scambio impossibile*, cit., p. 118.
[49] Id., *Taccuini*, cit., p. 121.
[50] Id., *Cool Memories. Diari 1980-1990*, cit., p. 29.
[51] G. Colli, *La nascita della filosofia*, Adelphi, Milano 1975, p. 68.

In conclusione, posto quindi che Dio non è mai il significato, ne il referente, ma il *significante*, la forma, il *medium* che *è* messaggio, ecco che la questione tra identità e alterità, in quanto questione nucleare dell'IA, riecheggia nelle parole di Jean-Pierre Vernant in merito al ritorno di Dioniso a Tebe nel paragrafo "*rifiuto dell'altro, identità perduta*": «coloro che incarnano l'attaccamento incondizionato all'immutabile, che proclamano l'irrinunciabile difesa dei valori tradizionali di fronte a quanto è altro da loro – che li mette in discussione, che li obbliga a portare su se stessi uno sguardo diverso – proprio loro, detentori dell'identità, cittadini sicuri della propria superiorità, precipitano nell'alterità assoluta, nell'orrore e nel mostruoso»[52].

VALERIO SPECCHIO ha conseguito la Laurea Magistrale presso l'Università Federico II di Napoli

specchiovalerio@gmail.com

[52] J.P. Vernant, *L'universo, gli dei, gli uomini*, Einaudi, Torino 2014, p. 157.

ALFONSO DI PROSPERO

INTELLIGENZA ARTIFICIALE E INFERENZA NON-MONOTONA:

MODELLI CULTURALI E QUESTIONI EPISTEMOLOGICHE

1. *Modelli di comunicazione e condizionamento delle vite*
2. *L'I.A. tra implicazioni culturali e problemi epistemologici* 3. *Circoli di prossimità epistemica*

ABSTRACT: *ARTIFICIAL INTELLIGENCE AND NON-MONOTONIC INFERENCE: CULTURAL MODELS AND EPISTEMOLOGICAL ISSUES*
In the research in A. I. John McCarthy has shown the interest of the non-monotonic reasoning. My attempt is to connect such topic with Piaget's genetic epistemology, within an inductivistic approach. The hypothesis is that this theoretical framework can be adapt to deal with both some features of the logical form of human thought – and this can entail relevant consequences also for A. I. – and for the cultural and social issues due to the widespreading of the digital communication.

1. *Modelli di comunicazione e condizionamento delle vite*

Luciano Floridi ha ripetutamente insistito che l'intelligenza artificiale riesce a ottenere successi e applicazioni sempre più consistenti non grazie a un progresso nei tentativi di *imitare* la mente umana *globalmente*, in un modo che consentirebbe di parlare di "intelligenza" artificiale nel senso in cui agli albori di questo tipo di ricerche – o nei romanzi di fantascienza – si prefiguravano robot "simili" agli esseri umani, ma piuttosto attraverso la combinazione da un lato della scomposizione delle funzioni di cui è capace la mente umana (creando sistemi esperti

in grado di effettuare ciascuno una sola delle innumerevoli operazioni di cui è capace la mente umana) e dall'altro del modificarsi progressivo che le attività umane hanno – per esempio nelle produzioni industriali – in vista di un *adattamento* sempre più perfezionato alle modalità di funzionamento dell'intelligenza artificiale stessa[1]. Il "formato" delle nostre attività – e alla lunga delle nostre esistenze – si adatta alla struttura delle attività implementate con mezzi artificiali: da un diverso punto di vista, usando le parole di Éric Sadin, si può vedere in questo una «svolta ingiuntiva della tecnica»[2].

Il reciproco integrarsi delle esigenze imposte dalle tecnologie e delle forme che le nostre azioni vengono ad avere, suggerisce in modo naturale l'idea (che lo stesso Floridi prende in considerazione) che processi dello stesso genere possano essere riferiti non solo al presente e al futuro, ma anche al passato, in particolare se si pensa alle celebri tesi di Marshall McLuhan e degli autori della cosiddetta Scuola di Toronto sul rapporto tra mezzi di comunicazione e forme di pensiero nel corso della storia umana.

Questo vorrebbe dire che anche le diffuse resistenze che si riscontrano contro il diffondersi delle tecnologie digitali per la comunicazione, potrebbero essere correttamente paragonate alle critiche che un filosofo come Platone ha rivolto contro la tecnologia della scrittura nel periodo in cui in Grecia essa andava diffondendosi. Un tale paragone ancora non basta a screditare le posizioni dei più pessimisti: le scelte e i giudizi da discutersi qui sono di carattere storicamente contestuale. Si può considerare confutata una formulazione "alla lettera" delle idee di Platone, in quanto smentita dall'evoluzione successiva. Si dà però la necessità di valutare in che modo un sistema di

[1] L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo* (2014), tr. it. Raffaello Cortina, Milano 2017.
[2] É. Sadin, *Critica della ragione artificiale. Una difesa dell'umanità* (2018), Luiss University Press, Roma 2019, p. 29.

convinzioni si inserisca e interagisca con i fatti che si danno nel momento storico in cui esso viene prodotto. Da ciò deriva che è impossibile trarre l'implicazione che si sia automaticamente in errore se si esprimono riserve di fronte a un troppo rapido diffondersi delle nuove tecnologie della comunicazione nel mondo attuale, dato che si tratta di contesti del tutto diversi.

Inoltre è da osservare che, anche se oggi la tecnologia della scrittura ci appare come una acquisizione irrinunciabile, vi sono però studiosi che hanno condotto contro le conseguenze dovute alla sua introduzione delle critiche interessanti. Per Walter Ong la diffusione della scrittura ha creato *alienazione* e distanza tra le persone[3]. Non si tratta naturalmente di negare quanto l'invenzione della scrittura abbia potuto modellare – in meglio – le condizioni del vivere umano. È opportuno però cercare di elaborare un modello analitico globale che, tenendo conto in una certa misura anche dei nessi individuati da Ong, sia in grado di dar conto di alcuni significativi aspetti degli sviluppi attuali.

L'idea di fondo è che in un certo senso la *prima* tecnologia che ha condizionato e modellato la mente umana sia non la scrittura, ma lo stesso linguaggio (secondo un'idea che è presente per esempio in Bergson).

Radu Bogdan ha mostrato la plausibilità di un modello per cui il bambino non dispone di una struttura linguistica e cognitiva fondamentale: la forma tema-commento[4]. Applicando anche l'infrastruttura teorica dell'epistemologia genetica[5], per la quale il bambino alla nascita non comprende le nozioni di spazio e tempo, di oggetto semi-permanente, né è capace di coordinare punti di vista diversi, per cui diventa in effetti quasi impossibile che possa essere presente la forma tema-commento, l'autore sostiene

[3] W. Ong, *Interfacce della parola: studi sull'evoluzione della coscienza e della cultura* (1977), tr. it. il Mulino, Bologna 1989.
[4] R. Bogdan, *Predicative Minds. The Social Ontogeny of Propositional Thinking*, MIT Press, Cambridge 2008.
[5] J. Piaget, *La costruzione del reale nel bambino* (1937), tr. it. La Nuova Italia, Firenze 1973.

che questa si sviluppa progressivamente nei primi anni di vita, facendo da asse principale che rende possibile l'acquisizione delle più diverse conquiste della cultura.

L'ipotesi che ho cercato di sviluppare maggiormente in dettaglio in precedenti pubblicazioni[6] è che la teoria di Jean Piaget possa essere re-interpretata in termini induttivistici: la condizione di egocentrismo che Piaget attribuisce al bambino sarebbe da associarsi con il suo essere un soggetto epistemico iper-induttivista, che procederebbe nello sviluppo applicando nelle sue operazioni cognitive generalizzazioni induttive che partono dai contenuti della sua propria esperienza – è in ciò appunto che si manifesta quello che possiamo vedere come "egocentrismo". In questo quadro è essenziale però sottolineare che l'inferenza induttiva deve essere caratterizzata come non-monotona: il bambino, facendo nuove esperienze, *non* è affatto vincolato a conservare come validi e veri i risultati delle generalizzazioni precedenti, ottenuti a partire da premesse induttive differenti. È da considerare che qui si vede l'induzione nella sua valenza psicologica, per cui è da attendersi ragionevolmente che giochino un ruolo importante i meccanismi dell'attenzione selettiva: perciò gli *script* e gli schemi di comportamento che si costituiscono e si stabilizzano non dovranno necessariamente essere respinti, con la comparsa di nuove esperienze che, in una logica non-monotona, potrebbero invalidarne il funzionamento (il che renderebbe impossibile il definirsi e lo stabilirsi in modo più o meno permanente di strutture cognitive man mano più complesse). Il fissarsi di schemi che continueranno a essere "preferiti" anche in seguito, quando la disponibilità di nuovi contenuti di esperienza formalmente renderebbe possibile il loro abbandono, può spiegarsi con il fatto che la dinamica dell'attenzione selettiva (anche al

[6] A. Di Prospero, *La forma del significato. Saggio sui costituenti del pensiero*, Aracne, Roma 2020; Id., *Linguaggio e ordine del mondo*, Il Sileno, Rende (CS) 2020.

livello senso-motorio e di *embodied cognition*) continua ad alimentarne il funzionamento.

Anche il senso della reciprocità razionale, che per Piaget si acquisisce solo dopo anni di vita (non si insista qui eccessivamente su scansioni troppo rigide e prevedibili nella sequenza dello sviluppo, che sono state oggetto di critica da parte di molti studiosi) sarebbe ottenuto a partire da esperienze che hanno creato l'aspettativa che un ordinamento intersoggettivo e condiviso socialmente nelle funzioni del credere sia degno di rispetto, interesse e fiducia.

Un autore che è importante accostare qui a Piaget, per quanto poco citato, è Ash Gobar[7], che ha cercato di dare un'interpretazione dell'epistemologia genetica che ne sottolinea i punti di contatto con la *Gestalt Psychology*. Riguardo al tema dell'intersoggettività l'autore scrive che

> the degree of the intersubjectivity of knowledge varies directly with the degree of object constancy and inversely with the amount of critical interference: Thus objective knowledge (e.g. logic) builds its home on the bedrock of permanent constancy, and where critical interference prevails there also prevail conflicting viewpoints (e.g. aesthetics)[8].

In termini intuitivi, l'ipotesi che si può proporre a partire da questa idea è che il bambino sia mosso da un fondamentale empirismo, che venendo attuato rigorosamente obbligherebbe però il soggetto a non poter condividere con altri segni e significati: la comunicazione – posti i vincoli dell'empirismo e la differenza tra le esperienze individuali – presuppone una misura di *indeterminazione* (è a questo riguardo che si dovrebbe inserire un collegamento – da svilupparsi più in dettaglio – con le idee sopra riferite di Ong).

Nel dibattito sulle comunicazioni digitali, Michael Lynch si richiama a un problema (simile a quello che Quine definisce "trazione oggettiva") che segue un'euristica analoga:

[7] A. Gobar, *Philosophical Foundations of Genetic Psychology and Gestalt Psychology*, Martinus Nijhoff, The Hague 1968.
[8] *Ibid.*, p. 281.

> The problem of distinguishing the real from the unreal, or the true from the untrue, is hardly the result of the digital age […] Take a coin out of your pocket and hold it in your hand before you. Now look at the coin: what shape does it look like? [… If requested] A child might draw a circle, but a more skilled artist wuld draw the ellipse. Why? Because that's what we are perceiving. But if so, then we have a puzzle. The coin is circular. What we pereive is not circular. Therefore, what we perceive is not the coin[9].

Anche C. Thi Nguyen[10], parlando di *echo chamber* e *epistemic bubble*, torna a più riprese a osservare che sono fenomeni che esistevano in forma diversa anche in contesti del tutto differenti (per esempio nei regimi politici che praticavano il culto della personalità).

Il punto è che la realtà "virtuale" non fa che prolungare dinamiche e logiche presenti già in tutta la storia della socializzazione umana: si può comunicare – anche con chi ci è fisicamente di fronte – solo perimetrando i significati in un modo che li renda accettabili da tutti i partner della comunicazione coinvolti.

2. *L'I.A. tra implicazioni culturali e problemi epistemologici*

Questo schema teorico fornisce alcune premesse interessanti per affrontare le questioni poste dall'Intelligenza Artificiale e in generale dalla "cultura del digitale". In particolare: 1) i flussi di informazione sono sempre più caotici e violenti – salvo poter essere controllati solo in modo meccanizzato e attraverso algoritmi da soggetti non umani, data la quantità enorme di dati da considerare, e da forze economiche in grado di finanziare tutto il processo di elaborazione che si rende necessario: quella che si profila è la necessità di modificare il nostro modello di intersoggettività e di reciprocità razionale, allargandolo quantitativamente, secondo le esigenze imposte dalle dinamiche della globalizzazione, ma forse dovendolo anche trasformare qualitativamente, nella sua struttura epistemologica. Il rischio

[9] M. Lynch, *The Internet of Us*, Norton, New York 2016.
[10] C.T. Nguyen, *Echo Chambers and Epistemic Bubbles*, in «Synthese», 197, 7, 2020, pp. 2803-2821.

con cui ci si deve confrontare, entro questa prospettiva, non è tanto legato ai nuovi supporti tecnologici della comunicazione (come potrebbe sembrare leggendo per esempio Manfred Spitzer), ma piuttosto che l'ambiguità e l'equivocità *costitutive* della comunicazione in ogni epoca storica, dovendo oggi confrontarsi con identità e relazioni su di una scala che è planetaria, possano di fatto esplodere, rendendo i significati trasmessi nella comunicazione inservibili per l'individuo. In altre parole la persona, dovendo partecipare a pratiche di comunicazione la cui semantica astrae sistematicamente da quelle che sono le sue esperienze, deve rinunciare a dare un ruolo motivante nell'azione *o* a questa semantica *oppure* alle proprie esperienze[11].

2) Uno dei terreni su cui si sono combattute alcune delle battaglie per la realizzazione dell'intelligenza artificiale è proprio quello delle forme di ragionamento non-monotono: John McCarthy, lo scienziato che ha introdotto l'espressione "intelligenza artificiale", organizzando il convegno di Dartmuth del 1956 che ha dato inizio al dibattito su di essa, ha cercato di provvedere a creare delle formalizzazioni del ragionamento non-monotono, con il fine dichiarato di poter arrivare a permettere a dei computer di ragionare secondo forme che riproducano quelle degli esseri umani[12]. Sullo sfondo naturalmente vi sono le speranze e problemi legati al test di Turing: l'idea è che gli esseri umani, a differenza dei computer, sappiano orientarsi – in una conversazione o in un ragionamento – riconoscendo le situazioni di ambiguità e di indeterminazione delle informazioni che ricevono, per cui un computer, per poter riprodurre e per potersi interfacciare con le modalità di ragionamento di un umano, dovrebbe poter essere programmato in modo da attenersi a una logica non-monotona. McCarthy propone quella che chiama *circumscription logic*. L'idea è che

[11] Cfr. J. Arditi, *Simmel's Theory of Alienation and The Decline of the The Norational*, in «Sociological Theory», XIV, 2, 1996, pp. 93-108.
[12] J. McCarthy, *Formalizing Common Sense*, Ablex, Norwood, NJ 1990.

> It seemed that in order to fully represent the conditions for the successful performance of an action, an impractical and implausible number of qualifications would have to be included in the sentences expressing them. For example, the successful use of a boat to cross a river requires, if the boat is a rowboat, that the oars and rowlocks be present and unbroken, and that they fit each other. Many other qualifications can be added, making the rules for using a rowboat almost impossible to apply, and yet anyone will still be able to think of additional requirements not yet stated[13].

Lo strumento proposto dall'autore è quindi il seguente:

> Circumscription is a rule of conjecture that can be used by a person or program for jumping to certain conclusions. Namely, *the objects that can be shown to have a certain property P by reasoning from certain facts A are all the objects that satisfy P*. More generally, circumscription can be used to conjecture that the tuples <*x*, *y*,…, *z*> that can be shown to satisfy a relation *P*(*x*, *y*, … , *z*) are all the tuples satisfying this relation. Thus, we *circumscribe* the set of relevant tuples.
> We can postulate that a boat can be used to cross a river unless something prevents it. then circumscription may be used to conjecture that the only entities that can prevent the use of the boat are those whose existence follows from the facts at hand[14].

Hubert Dreyfus aveva scritto:

> le macchine mancano di intelligenza pratica. Esse sono "esistenzialmente" stupide nel senso che non possono fronteggiare situazioni specifiche. Così esse non possono accettare l'ambiguità e la rottura delle regole, fino a che le regole per trattare le deviazioni siano state così completamente specificate che l'ambiguità sia scomparsa[15].

Il problema è che la realtà degli esseri umani è contestuale e in divenire, mentre «le leggi della scienza sono universali e senza tempo, perché trattano tutta l'esperienza come se potesse essere allo stesso modo che nel passato»[16]: con una sorta di contraddizione intrinseca «gli studiosi di I.A. dovrebbero sviluppare una teoria non temporale e non spaziale di un'attività umana che si sviluppi in una situazione»[17]. Una macchina dovrebbe necessariamente «trattare con fenomeni che appartengono al mondo situazionale degli esseri umani come se questi fenomeni appartenessero all'oggettivo formale universo della scienza»[18].

[13] *Ibid.*, p. 142.
[14] *Ibid.*
[15] H. Dreyfus, *Che cosa non possono fare i computer* (1972), Armando, Roma 1988, p. 285.
[16] *Ibid.*, p. 284.
[17] *Ibid.*, p. 285.
[18] *Ibid.*

Come si vede, la contrapposizione è tra logica classica – atemporale – e ragionamento non-monotono, e di questa distinzione McCarthy intende appunto avvalersi per rispondere a Dreyfus[19].

Le logiche non-monotone sviluppate in seguito da altri hanno suscitato notevole interesse. Per il nostro discorso si può ricordare quella di Raymond Reiter, che prevede una "close world assumption", per la quale «if no proof of a positive ground literal exists, then the negation of that literal is assumed true. This can be viewed as equivalent to implicitly augmenting the given data base with all such negated literals»[20].

Donald Gillies è un autore che si avvicina al problema da me posto, pur senza definirlo negli stessi termini. Il senso del suo contributo è di portare a riflettere sul significato dei successi che l'I.A. ha conseguito nell'ambito dell'apprendimento induttivo meccanico. L'autore sostiene che così si è "provato" che l'anti-induttivismo radicale di Popper è scorretto. È possibile che nella sua formulazione questa idea sia fuorviante: anche *prima* che l'*inductive learning* ottenesse nell'I.A. dei risultati significativi, l'anti-induttivismo di Popper poteva essere già fatto oggetto di critiche. In ogni caso l'autore afferma con forza che il ragionamento induttivo deve essere riconosciuto come non-monotono.

Un atteggiamento più corretto nei confronti dell'anti-induttivismo di Popper è difeso da Isaac Levi, che sospetta che molti studiosi impostino le proprie analisi dell'induzione in modo da evitare «a form of subjectivity in inquiry that will give comfort to the excess of cultural relativists and antirationalists»[21], mentre

> Others, as Popper, Carnap, and their antiinductivist followers, nurse and protect the objectivity of science by denying the legitimacy of inductive expansion in science. To counsel abstinence from induction

[19] McCarthy, *Artificial Intelligence, Logic and Formalizing Common Sense*, in R. Thomason, ed, *Philosophical Logic and Artificial Intelligence*, Springer, Dordrecht, p. 4.

[20] G. Reiter, *On Closed World Data Bases*, in «Readings in Artificial Intelligence», 1981, pp. 119-140, p. 119.

[21] I. Levi, *For The Sake of The Argument*, Cambridge University Press, Cambridge, Mass., 1996, p. 160.

> in this manner is an admission of failure to bring the fixation of belief under critical control. Since inquirers will extrapolate from data and add theoretical conjectures to background information as resources in subsequent inquiry regardless of what antiinductivists say, the admission of failure concedes to the social relativists and antirationalists that induction is a topic for psychology, sociology and history to study and that there is no point in attempting to identify prescriptive standards for assessing its legitimacy[22].

Per proseguire in questa indagine, vediamo in che modo Carl Gustav Hempel aveva provato a fornire «a purely syntactical definition» della conferma induttiva[23]. Se *S* è una proposizione che deve confermare induttivamente *M*, vuol dire che *M*

> determines a certain class of individuals [e.g.] {*a*, *b*, *c*}; and we may say that *M* confirms *S* because from the information contained in *M* it can be inferred that in {*a*, *b*, *c*} *S* is completely satisfied; or, to use a metaphor: in a world containing exclusively the individuals *a*, *b* and *c*, the sentence *S* would be true, according to the information contained in *M*[24].

Nelson Goodman ricapitola così il significato della definizione di Hempel:

> a hypothesis is genuinely confirmed only by a statement that is an instance of it in the special sense of entailing not the hypothesis itself but its relativization or restriction to the class of entities mentioned by that statement. The relativization of a general hypothesis to a class results from restricting the range of its universal and existential quantifiers to the members of that class. Less technically, what the hypothesis says of all things the evidence statement says of one thing (or one pair or other n-ad of things)[25].

Sia Goodman sia Hempel cercano di perfezionare questa definizione, abbandonandola infine anche per i celebri paradossi da loro stessi formulati. Si vede comunque come la logica soggiacente ai concetti di "circoscrizione" o di "mondo chiuso" operi in effetti anche qui in modo abbastanza chiaro.

Si può sostenere inoltre che, considerando l'inferenza induttiva come non-monotona, il paradosso di Hempel sembra che non disponga più dello spazio per poter sorgere.

Il problema viene enunciato dall'autore così:

---

[22] *Ibid.*

[23] C.G. Hempel, *A Purely Syntactical Definition of Confirmation*, in «The Journal of Symbolic Logic», VIII, 4, 1943, pp.122-143, p. 122.

[24] *Ibid.*, p.130.

[25] N. Goodman, *Fact, Fiction, and Forecast*, 4th ed., 1983, pp. 69-70.

> Consider, for example, the hypothesis "All swan are white", which we symbolize by "$(x)(Sw(x) \supset Wh(x))$", and the molecules "$Sw(a)$. $Wh(a)$" and "$\sim Sw(b)$ $\sim Wh(b)$. It might be felt–and would probably be argued by many–that under a materially adequate definition of confirmation, the first molecule should be confirming, the second neutral with respect to the hypothesis. However, according to the same intuitive standards, the second molecule would constitute confirming evidence with respect to the hypothesis "(x)(~Wh(x)⊃~Sw(x))". But the two hypotheses here mentioned are equivalent [per contrapposizione], and adherence to the intuitive standards exemplified in this illustration would therefore make the confirmation of a hypothesis by a certain body of evidence a matter not only of the content of the hypothesis, but also of the form in which it is expressed. This feature appears intuitively intolerable[26].

Il paradosso sta nel fatto che se osservo un corvo nero, che non è bianco e non è un cigno, per questa via posso confermare induttivamente che tutte le cose non-bianche sono non-cigni, e quindi – per contrapposizione – che tutti i cigni sono bianchi.
Se però l'inferenza induttiva è riconosciuta come non-monotona, il passaggio logico da "Questo è un corvo nero" a "Questo è un oggetto non-bianco che è un non-cigno" diventa in effetti scorretto: per quanto ovviamente la deduzione in se stessa sia valida, una volta effettuato questo passaggio si è modificato di fatto l'insieme delle premesse disponibili, rendendolo estremamente più povero, per cui è come se si stesse facendo direttamente una *diversa* inferenza induttiva. Un'idea simile a questa è stata difesa da Branden Fitelson[27], che osserva come l'inferenza induttiva sia non-monotona, e per questo il paradosso *non* dovrebbe sorgere, per ragioni analoghe a quelle da me difese, ma ugualmente si impegna in una formulazione di una strategia di soluzione che, ispirandosi all'approccio seguito da molti a partire dalle analisi di Janina Hosiasson-Lindenbaum[28], utilizza

[26] C.G. Hempel, *op. cit.*, p. 128.
[27] Tra i vari articoli, si può vedere B. Fitelson, J. Hawthorne, *The Wason Task(s) and The Paradoxes of Confirmation*, in «Philosophical Perspectives», XXIV, 1, 2010, pp. 207-241. Un'analisi che si serviva di questa strategia è stata da me presentata in A. Di Prospero, *Un'analisi dei paradossi della conferma proposti da Carl Gustav Hempel*, in «Itinerari», 2003, 3, pp. 121-134; Id., *Art, Knowledge and Induction*, in «RIbid.sta Italiana di Filosofia del Linguaggio», 2017, 2, pp. 147-163; Id., *Esperienza estetica, giustizia e inferenza induttiva*, in «Discipline filosofiche», 2021, 31, pp. 219-242.
[28] J. Hosiasson-Lindenbaum, *On Confirmation*, in «The Journal of Symbolic Logic», 1940, V, 4, pp. 133-149.

la teoria della probabilità per l'analisi del paradosso. L'impressione che si può avere – su cui però sarebbe necessario discutere con molta più ampiezza – è che questo studioso voglia presentare la sua strategia di analisi del paradosso come un successo dell'approccio bayesiano. Formalmente però è già il solo riferimento al carattere non-monotono dell'induzione che basta a confutare il paradosso[29].

In generale è interessante notare che viene a offrirsi una spiegazione del fatto che molti studiosi abbiano infine rifiutato lo strumento dell'induzione: in realtà si può ipotizzare che un modo fuorviante di riferirsi a essa – determinato dall'esigenza, presente in Popper o nel neopositivismo, di definire il metodo scientifico in modo da preservare *formalmente* il presupposto dell'intersoggettività del sapere scientifico – abbia portato a non dar peso al carattere non-monotono dell'induzione (che intrinsecamente oppone problemi urgenti a un tale principio di intersoggettività, se fatto valere *a priori* e non come oggetto di apprendimento *formalmente* contingente), spingendo infine al suo abbandono. Se invece ci si richiama all'epistemologia genetica, si può mostrare che il risultato di una sostanziale – e pragmaticamente fondamentale – intersoggettività del sapere (in particolare in ambito scientifico), può essere ottenuto con strumenti concettuali diversi.

L'intuizione di fondo è che il bambino, applicando momento per momento una logica non-monotona che obbedisce a un principio simile a quello della "circoscrizione", e accogliendo in memoria gradualmente prima frammenti e poi segmenti di cognizione sempre meglio sistematizzata, ottenuti appunto con questi mezzi, edifichi l'intera immagine del mondo, nella quale – in un senso che in

[29] La strategia euristica di fondo mi sembra che sia descritta bene da McCarthy: «Some people think it is possible to try to save mnotonicity by saying that what was in your mind was […] a probabilistic rule. Indeed it seems to me that when probabilistic reasoning (and not just the axiomatic basis of probability theory) has been formalized, it will be formally nonmonotonic» (J. McCarthy, *Artificial intelligence*..., cit., p. 11).

realtà è lapalissiano – le diverse prospettive degli altri soggetti epistemici sono contenute, ma in quanto *oggetto* del sapere del soggetto di riferimento.

Gillies riflette sulle conseguenze del fatto che il *Prolog* ha un comportamento di tipo non-monotono, dato che dispone solo di una negazione caratterizzata come "fallimento". La posizione del bambino è in effetti analoga: seguendo Piaget, il bambino inizialmente non dispone del concetto di negazione, ma si limita a compiere azioni che – a volte – falliscono. Nella prospettiva di un induttivismo rigoroso, anche l'adulto in realtà, cercando di comprendere nel suo proprio esatto contenuto il punto di vista di un'altra persona, va incontro a un analogo fallimento, nel senso che i contenuti *esatti* di un *altro* punto di vista non possono essere per principio da me afferrati, perché allora non sarebbero più i contenuti di un *altro* punto di vista[30]. Per questo nel riferirmi ai punti di vista altrui devo necessariamente ricorrere alla nozione di *indeterminazione*, in un senso per cui essi a rigore risulteranno dei *non*-contenuti. Si consideri anche che per le caratteristiche del ragionamento non-monotono, che nel dibattito sull'induzione trovano espressione in quello che Carnap chiama "principio dell'evidenza totale", se io ho anche solo una minima parte dei miei contenuti di esperienza diversa da un altro soggetto, le applicazioni del principio dell'evidenza totale mi porteranno a generalizzazioni induttive a rigore diverse, producendo quello che può essere visto come un effetto di campo gestaltico, in grado in ogni caso di contraddire la possibilità di un atomismo rigido e non-contestualizzante.

[30] Gli autori che qui vorrei almeno rapidamente ricordare sono il Wittgenstein del *Tractatus logico-philosophicus* e Nicholas Rescher (*Rationality*, Oxford University Press, Oxford 1988). Sul ragionamento non-monotono si deva anche C. Cellucci (*Le ragioni della logica*, Laterza, Roma-Bari 1998) mentre da una diversa prospettiva vorrei almeno menzionare D. Kincaid, *The Convergence Theory of Communication, Self-Organization, and Cultural Evolution*, in D. Kincaid, ed, «Communication Theory. Eastern and Western Perspectives», Academic Press, San Diego 1987, pp. 209-221.

Dreyfus è anche alla psicologia cognitiva dello sviluppo del bambino che fa riferimento per sostenere la sua critica all'intelligenza artificiale[31], ma il modo in cui qui si impiega l'epistemologia genetica potrebbe aprire degli spazi per contestare anche proprio su questo punto specifico la sua argomentazione.

3. *Circoli di prossimità epistemica*

Per procedere in queste riflessioni, può essere interessante suggerire un confronto con l'analisi del concetto di "post-verità" proposta da Filippo Ferrari e Sebastiano Moruzzi[32], che introducono a completamento le nozioni di "post-indagine" e "post-prova", richiamando anche la nozione di "pensiero collettivo" di Ludwik Fleck[33], che viene posta in analogia con quelli che gli autori chiamano "circoli di prossimità epistemica", composti

> da tutti quegli agenti la cui indagine è regolata da uno stesso filtro epistemico […] Questo filtro contribuisce a determinare una bolla epistemica. È plausibile pensare che ogni nostra indagine particolare sia regolata dalla presenza di filtri epistemici di questo tipo. Questo fatto non va però valutato necessariamente in modo negativo da un punto di vista epistemico – anche se nel caso dei terrapiattisti pensiamo certamente che lo sia[34].

È da notare che Fleck ha un'immagine abbastanza positiva del carattere intersoggettivo e (almeno in senso debole) consensuale della conoscenza scientifica. Ferrari e Moruzzi invece (prevalentemente) utilizzano il concetto di circolo di prossimità epistemica per evidenziare le fallacie e i paradossi di gruppi, come i "terrapiattisti", che cercano unicamente le conferme – quali che possano essere – per le proprie convinzioni, sottoponendo le obiezioni a un trattamento epistemologico del tutto penalizzante. Il confronto con questi due autori, che ricapitolano e rappresentano in maniera chiara un punto di vista filosoficamente diffuso ma le cui implicazioni sono forse ancora

[31] H. Dreyfus, *op. cit.*, p. 91.
[32] F. Ferrari, S. Moruzzi, *Verità e post-verità*, 1088 Press, Bologna 2020.
[33] *Ibid.*, p. 85.
[34] *Ibid.*

da esplorare sotto vari aspetti, può essere interessante perché a una prima apparenza può sembrare che le considerazioni fin qui portate sulle funzioni dell'induzione siano in difficoltà nel cogliere la distinzione tra un consenso sociale "sano" e uno distorto e deviante (si pensi anche alle riflessioni sul concetto di fiducia condotte da Alvin Goldman[35]).

In realtà, nell'indicare nell'esperienza personale e nell'induzione le basi del procedere cognitivo, si deve considerare che le nostre esperienze, in quanto tali, non sono affatto da noi *scelte*: per questo non vi sono margini per un relativismo che legittimi le scelte più arbitrarie e stravaganti. È vero però che formalmente la nostra convinzione che alcune credenze (oggi, con le nostre conoscenze) siano assurde, deve essere analizzata come dovuta a esperienze il cui statuto logico è quello della contingenza – il che non pregiudica la nostra certezza: se vedo che piove, ne sono sostanzialmente certo, anche se avrebbe potuto benissimo essere bel tempo.

Questo tipo di riflessione è apparentemente ovvia, ma pure nella sua formulazione generale si scontra con dei presupposti largamente fatti valere – spesso implicitamente – nella filosofia della scienza più diffusamente accettata, che accoglie il principio dell'intersoggettività del sapere scientifico come un presupposto che viene *trattato* (anche quando non si usano esattamente queste espressioni) come una verità o una necessità di tipo logico e *a priori* (si pensi a titolo di esempio a Popper).

Può essere interessante vederne alcune possibili implicazioni anche per una questione classica del dibattito sull'I.A., il test di Turing.

Dreyfus critica uno dei presupposti dell'I.A., sostenendo (come Searle) che per principio un computer non può avere un'intelligenza perché non può avere una semantica, posto che si

[35] A. Goldman, *Experts: Which One Should You Trust?*, in «Philosophy and Phenomenological Research», LXIII, 1, 2001, pp. 85-110.

utilizzi una definizione di informazione come quella data da Shannon in *The Mathematical Theory of Communication*[36]. Si può però osservare che, se si vuole tener conto del carattere non-monotono dell'induzione – e del ruolo che l'induzione ha nel costituirsi dei significati – l'impostazione di Shannon è adatta proprio perché calibrata su una misura del tutto minima di condivisione univoca del significato, elaborata in modo da potersi dire che la "semantica" (se tale ancora può definirsi) sia riducibile alla sintassi (Wittgenstein nel *Tractatus*, e Moritz Schlick in modo più semplice, ma ispirandosi apertamente a Wittgenstein, attraverso le nozioni di "forma" e "contenuto", cercano di definire una teoria del significato di questo tipo). Essendo così impostata, una teoria "semantica" dovrebbe in effetti per principio potersi adattare a tutto il *range* delle possibili variazioni del significato, che il ragionamento non-monotono consente (quindi da un *minimo* a un *massimo* di possibile condivisione del significato), invece di dover presupporre l'accettazione di un codice univoco e condiviso.

Questo però non vuol dire che entro un simile quadro analitico la semantica, intesa come centrata sulle nostre esperienze in prima persona, scompaia. Probabilmente l'obiezione principale che può farsi al test di Turing è che fa riferimento a un modello della comunicazione che astrae dal fatto che nel comunicare ciò che i parlanti vogliono è sapere qualcosa che è ancora *al di fuori* del campo del loro attuale sapere: qualcosa che si può esprimere dicendo che le esperienze altrui – per il principio di non-monotonia – sono irriducibili alle nostre. Gli indizi di cui mi servo per valutare se un testo è prodotto da una macchina o no, corrispondono a un repertorio di conoscenze che *già* ho, non a quelle che il testo stesso dovrebbe permettermi di acquisire. Se chiedo a un amico il suo parere su una certa questione, il fatto che riesca a farmi credere che la sua risposta è pertinente può

[36] H. Dreyfus, *op. cit.*, p. 244.

essere una condizione necessaria ma non sufficiente di un'attribuzione di intelligenza. Quando – per esperienza – posso valutare che in una certa misura le esperienze (e quindi i significati) di un interlocutore "combaciano" con le mie, estendo a questo interlocutore la posizione di partecipe di ciò che chiamo intersoggettività – o "circolo di prossimità epistemica". Le mie esperienze però mi spingono a cercare *informazione* (attendibile), non le forme esteriori di una pratica del comunicare che astrae dai contenuti di verità (si deve considerare da questo punto di vista l'influenza della psicologia comportamentista su Turing). In questo senso, la non-condivisione di significati univocamente definiti tra me e un computer è un fatto formalmente di tipo empirico (come in un certo senso è ovvio che sia), così come lo è in generale il costituirsi di una dimensione della condivisione intersoggettiva tra me e gli altri esseri umani[37].

Anche l'utilizzo che McCarthy fa del concetto di ragionamento non-monotono risente molto (come peraltro era prevedibile) dell'impostazione data al problema da Turing, che d'altra parte è molto più vicina alle teorie epistemologiche più diffuse di quanto non lo siano le proposte che qui invece sono state discusse. Per motivi di spazio non è possibile approfondire come sarebbe necessario in questo scritto queste tematiche. Come si è cercato di mostrare, è però possibile concepire un quadro concettuale alternativo, dal quale deriverebbe la necessità di formulare in modo diverso (1) le questioni dell'I.A. associate in genere al test di Turing (e anche a quello che in filosofia della mente è il cosiddetto *hard problem*, anche per la rilevanza fondamentale che entro questo quadro viene ad avere una prospettiva in prima persona), (2) il problema, presentato in precedenza, dei modelli

[37] Sarebbe importante qui confrontarsi almeno con le posizioni formulate da R. French (*Subcognition and the Limits of the Turing Test*, in «Mind», XCIX, 393, 1990, pp. 53-65), da S. Bringsjord, P. Bello e D. Ferrucci (*CreatIbid.ty, the Turing Test, and the (Better) Lovelace Test*, in «Minds and Machines», XI, 2001, pp. 3-27) e da J. Moor (*The Status and Future of the Turing Test*, in «Minds and Machines», XI, 2001, pp. 77-93). Per ragioni di spazio si è obbligati a rimandare a una diversa occasione questa discussione.

di comunicazione tra *identità distanti*, così come oggi viene imposto dalle dinamiche della globalizzazione e delle comunicazioni digitali.

Queste due famiglie di problemi – per quanto fenomenologicamente molto lontane – forse possono in effetti essere affrontate all'interno di un quadro analitico almeno in parte comune.

ALFONSO DI PROSPERO è Dottore di Ricerca in Scienze Sociali e svolge attività di Docenza nelle Scuole Secondarie di secondo grado

alfonsodiprospero@yahoo.com

EDOARDO DATTERI

ROBOTICA E FILOSOFIA DELLA SCIENZA

1. *Introduzione* 2. *Scienza e robotica* 3. *Dalla robotica alla filosofia della scienza generale*
4. *Dalla filosofia alla robotica delle scienze* particolari
5. *Riflessioni conclusive: dalla filosofia alla robotica*

ABSTRACT: *ROBOTICS AND PHILOSOPHY OF SCIENCE*
What relationship holds between robotics and philosophy of science? Can robotics research contribute to research in philosophy of science? Conversely, can the results achieved by philosophers of science contribute to the progress of research in robotics? This article will deal with both questions based on the distinction between the so-called philosophy of general science and the philosophy of the particular sciences. It will draw relatively pessimistic conclusions about the possibility that robotics research shed light on questions pertaining to the philosophy of general science, and more optimistic conclusions about the potential contribution of robotics research to the philosophy of particular sciences. The normative aspect of philosophy of science is consistent with the possibility that this field of research, taking robotics as its object of study, provides guidelines and regulative methodological principles for engineers and roboticists. Through these considerations, this article attempts to explore a possible field of interaction between philosophy and scientific-technological research, with reference to a discipline that is assuming an increasingly predominant role in many aspects of daily life.

1. *Introduzione*

Quale relazione intercorre tra robotica e filosofia della scienza? La storia della robotica, le sue tecnologie, le sue applicazioni possono fornire stimoli o contributi importanti a chi si occupa di ricerca in filosofia della scienza? Viceversa, i risultati di ricerca raggiunti dalla filosofia della scienza possono contribuire al progresso della ricerca in robotica, nelle sue varie articolazioni? Questo lavoro è dedicato all'analisi delle possibili relazioni tra questi due campi di indagine, così apparentemente distanti tra di loro, e prende le mosse dalla constatazione dell'effettiva sussistenza di un fitto dibattito tra ricercatori delle due discipline, che si articola nella letteratura accademica e nei convegni internazionali di ambo le

parti. Vuole dunque contribuire a esplorare un particolare fronte di possibile interazione tra la filosofia e la ricerca scientifico-tecnologica, in riferimento a un campo del sapere che sta assumendo un ruolo sempre più preponderante in molti aspetti della vita, anche quotidiana.

Filosofia e robotica interagiscono già da lunga data a molti livelli. Le applicazioni contemporanee o futuribili delle tecnologie robotiche a supporto delle persone sollevano problemi di carattere etico e sociale di grande importanza e urgenza[1]. Questi problemi trovano una collocazione naturale nell'ambito della filosofia morale, nella misura in cui riguardano il problema di capire (e di capire come poter capire) se sia giusto o sbagliato, rispetto ai rischi e ai benefici che possono portare e in riferimento a certe cornici valoriali, applicare le tecnologie robotiche in particolari contesti di azione umana. Tuttavia, come più volte sottolineato anche nel lavoro appena citato, la filosofia della scienza svolge un ruolo importante nell'analisi di questi problemi. Per chiederci se sia giusto o sbagliato applicare le tecnologie robotiche (o certi robot) a particolari contesti d'uso è necessario comprendere le caratteristiche, i limiti e le potenzialità di tali tecnologie, ovvero formulare dei modelli teorici di tali tecnologie (o sistemi), e valutarli alla luce degli impieghi a cui sono destinate. Ecco il legame con la filosofia della scienza: essa analizza i processi che sottendono la nostra comprensione teorica della realtà, occupandosi naturalmente anche degli oggetti creati da mani umane. Per esempio, nella comunità filosofica e robotica si dibatte sull'opportunità di investire nella ricerca sui babysitter robotici, oggetti che siano in grado non soltanto di affiancare, ma anche di sostituire esseri umani nell'accudimento e

[1] Per una trattazione esaustiva, cfr. F. Fossa, V. Schiaffonati, G. Tamburrini (a cura di), *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Carocci, Roma 2021.

nell'intrattenimento dei bambini[2]. Un argomento al centro del dibattito riguarda la possibilità che i robot umanoidi, soprattutto quelli dotati di morfologia, cosmetica e capacità di interazione sociale avanzate, possano indurre i bambini e gli adulti ad attribuire ai robot facoltà mentali ed emotive che essi non possiedono, sopravvalutandone dunque le capacità di assistenza e sottovalutando le conseguenze negative che tale "inganno" potrebbe avere sullo sviluppo del bambino. Per dirimere tale questione è opportuno chiederci quali facoltà mentali ed emotive tali robot abbiano effettivamente, e quali facoltà mentali ed emotive i bambini attribuiscano loro sulla base della loro interazione con essi (e se questi due piani coincidano o meno). È dunque opportuno chiederci se le teorie che i bambini formulano a proposito dei robot coinvolgano delle genuine attribuzioni mentali ed emotive a tali oggetti, e se esse coincidano con le facoltà mentali ed emotive che è lecito attribuire ai robot sulla base delle nostre migliori teorie a proposito di essi[3]. Ciò significa approfondire i metodi e le caratteristiche dei processi di teorizzazione sui robot intrapresi dai bambini, inevitabilmente chiamando in causa problemi centrali nella filosofia della scienza, tra cui quelli legati alla razionalità della teorizzazione cognitiva, al rapporto tra mente e materia, al realismo scientifico.

Queste considerazioni avevano lo scopo di sollecitare il lettore a riflettere sull'importanza che la filosofia della scienza riveste nell'analisi di problemi di carattere etico e socio-politico riguardanti la robotica. L'obiettivo di queste pagine, anticipato in apertura, è però quello di esplorare i rapporti che la robotica

[2] N. Sharkey, A. Sharkey, *The crying shame of robot nannies: An ethical appraisal*, in «Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems», 11, 2, 2010, pp. 161–190.

[3] T. N. Beran, A. Ramirez-serrano, R. Kuzyk, M. Fior, S. Nugent, *Understanding how children understand robots: Perceived animism in child – robot interaction* in «Journal of Human Computer Studies», 69, 7–8, 2011, pp. 539–550; J. Perez-Osorio, A. Wykowska, *Adopting the intentional stance toward natural and artificial agents*, in «Philosophical Psychology», 33, 3, 2020, pp. 369–395.

intrattiene, per così dire *direttamente*, con la filosofia della scienza. Per preparare il terreno a tale discussione è opportuno mettere a fuoco gli obiettivi di questa disciplina. Un utile aiuto proviene da *Introduzione storica alla filosofia della scienza* scritta da John Losee. Un primo senso di «filosofia della scienza», secondo Losee, attribuisce a tale disciplina l'obiettivo di formulare «concezioni del mondo coerenti, e in qualche modo basate su importanti teorie scientifiche; in questo caso il compito del filosofo della scienza è quello di elaborare le più ampie implicazioni della scienza»[4]. Un secondo punto di vista interpreta la filosofia della scienza «come una esposizione dei presupposti accolti dai *singoli* scienziati»[5] dove la parola "singoli", qui enfatizzata, avvicina la filosofia della scienza alla psicologia o alla sociologia. In base a un terzo punto di vista, il filosofo della scienza analizza e chiarisce le teorie scientifiche (per esempio, offrendone un'esposizione sistematizzata o divulgativa).

A questi tre compiti si dedicano, talvolta con successo, vari filosofi della scienza del panorama italiano e internazionale. Qui, tuttavia, si assumerà un punto di vista che non coincide con nessuna di queste possibilità. È il quarto punto di vista presentato da Losee, che qualifica la filosofia della scienza come una disciplina *di secondo ordine* il cui oggetto di studio è quella complessa macchina che chiamiamo "scienza". La filosofia della scienza tenta di capire "come funziona" la scienza, analizzando la validità delle sue metodologie d'indagine, il significato dei suoi concetti fondamentali (tra cui i concetti di teoria, legge, spiegazione, comprensione, e il concetto stesso di scienza come contrapposta a ciò che scienza non è), le assunzioni che guidano l'interpretazione dei suoi risultati. Il suo compito non è dunque quello di presentare le implicazioni della scienza, di studiare

[4] J. Losee, *Introduzione storica alla filosofia della scienza*, Il Mulino, Bologna 1989, p. 11.
[5] *Ibid.*, p. 12.

cosa accade nella mente di singoli scienziati, di sistematizzare i risultati della ricerca scientifica, bensì quello di "entrare dentro" la scienza analizzandone le fondamenta e i metodi di analisi. La prospettiva del filosofo della scienza è prevalentemente normativa: «si deve fare una distinzione tra come la scienza è e come si può pensare che la scienza dovrebbe essere», sottolinea ancora Losee[6], ed è su questo ultimo piano di indagine che generalmente si muovono i filosofi della scienza.

Questo testo rifletterà sulle possibili interazioni tra filosofia della scienza e robotica nell'una e nell'altra direzione, ovvero – come anticipato in apertura – nei termini dei contributi che la robotica può offrire alla filosofia della scienza e dei contributi che la filosofia della scienza può offrire alla robotica. Si baserà sulla distinzione, generalmente accettata nella comunità di ricerca, tra la cosiddetta filosofia della scienza generale e la filosofia delle scienze particolari. Trarrà conclusioni relativamente pessimistiche rispetto alla possibilità che la ricerca in robotica illumini questioni di filosofia della scienza generale, e conclusioni più ottimistiche rispetto alla possibilità che la ricerca in robotica sollevi questioni interessanti che riguardano la filosofia delle scienze particolari, nonché la possibilità che la filosofia della scienza fornisca criteri guida normativi a chi si occupa di robotica.

Per non appesantire il flusso del discorso, in questo testo si utilizzerà il genere linguistico maschile interpretandolo però come neutro: l'espressione "filosofo della scienza", per esempio, sarà usata per denotare una persona che si occupa di filosofia della scienza senza alcuna assunzione specifica sul suo genere.

2. *Scienza e robotica*

Dato che l'oggetto di studio della filosofia della scienza è il "funzionamento" della scienza, per riflettere sulle possibilità di

[6] *Ibid.*, p. 13.

interazione tra filosofia della scienza e robotica è opportuno interrogarsi sull'intersezione tra ricerca scientifica e ricerca robotica. Un punto di vista largamente diffuso considera scienza e robotica come tipi sostanzialmente distinti di ricerca. La robotica sarebbe una disciplina *ingegneristica*, il cui obiettivo è quello di progettare e realizzare sistemi robotici che possano assistere o sostituire gli esseri umani in particolari compiti; mentre la ricerca scientifica, nelle sue varie articolazioni, si porrebbe obiettivi *conoscitivi* e non costruttivi. In base a questo punto di vista è difficile comprendere in che senso la filosofia della scienza possa interessarsi alla robotica – dato che la filosofia della scienza, come discusso nel paragrafo precedente, ha per suo oggetto di ricerca la scienza.

Questo è un punto di vista largamente discutibile. Da una parte, anche se assumiamo che la ricerca scientifica abbia finalità ultime di tipo conoscitivo, esplicativo o predittivo, l'idea che essa coinvolga momenti realizzativi, per esempio finalizzati alla costruzione e all'uso di strumenti sperimentali, è ormai largamente diffusa tra scienziati, filosofi e storici della scienza[7]. All'analisi del ruolo epistemico e della validità degli strumenti sperimentali, anche alla luce della moltitudine di teorie su cui essi sono costruiti, è stata dedicata una gran parte della ricerca in filosofia della scienza soprattutto a partire dalla metà del secolo scorso[8]. Comprendere il funzionamento della scienza significa dunque, anche, occuparsi dei processi *ingegneristici* che permettono alla scienza di assolvere al suo compito (nella misura in cui l'analisi di questi processi permette di studiare la validità degli strumenti costruiti). Torneremo su questo punto tra poche righe, sottolineando come anche i robot possano costituire strumenti di ricerca scientifica.

[7] P. Rossi, *La nascita della scienza moderna in Europa*, Laterza, Bari 2021.
[8] I. Hacking, *Representing and intervening. Introductory topics in the philosophy of natural science*, Cambridge University Press 1983; P. Galison, *Philosophy in the Laboratory*, in «Journal of Philosophy», 85, 10, 1988, pp. 525–527.

Dall'altra parte, anche se assumiamo che la ricerca in robotica abbia finalità ultime di tipo realizzativo, essa riposa sui prodotti conoscitivi della scienza e coinvolge fasi di teorizzazione scientifica. In primo luogo, in robotica – più genericamente, in ingegneria – le teorie scientifiche svolgono una funzione abilitante: la progettazione di un robot chiama in causa una moltitudine di teorie, da quelle fisiche necessarie alla progettazione delle componenti elettroniche e meccaniche a quelle psicologiche necessarie alla progettazione di un'efficace interazione con gli esseri umani (nel caso dei robot cosiddetti sociali). La fiducia nel fatto che i robot costruiti si comporteranno come previsto nei vari contesti d'uso passa anche attraverso l'accettazione, più o meno esplicita, delle teorie che sono state chiamate in causa per costruirli. In secondo luogo, la ricerca in robotica coinvolge essa stessa fasi di teorizzazione scientifica. Durante il processo realizzativo, i prototipi di robot vengono sottoposti a numerosi controlli sperimentali che possono essere intesi come processi di indagine scientifica del tutto analoghi a quelli tipicamente svolti dagli scienziati[9]. Il comportamento del robot viene indagato in particolari ambienti sperimentali controllati e sotto l'azione di particolari stimoli. L'esito di questi esperimenti è tutt'altro che ovvio e noto *a priori* persino per chi ha partecipato più direttamente alla costruzione del robot: anche i meccanismi più elementari, in condizioni realistiche, manifestano comportamenti praticamente imprevedibili persino a chi ne conosce molto bene la struttura interna[10]. Queste considerazioni ci spingono a ricostruire il processo di progettazione e costruzione di un sistema robotico come un processo che coinvolge la formulazione di teorie

[9] D. van Eck, *The Philosophy of Science and Engineering Design*, Springer, Berlin 2016.
[10] V. Braitenberg, *I veicoli pensanti. Saggio di psicologia sintetica*, Garzanti, Milano 1984.

sull'oggetto costruito, teorie che vengono progressivamente sottoposte a controllo sperimentale e man mano riviste.

Vi è un ulteriore, importante terreno di intersezione tra ricerca robotica e ricerca scientifica. I sistemi robotici vengono spesso usati, essi stessi, come strumenti di indagine scientifica. In molti casi vengono utilizzati come *modelli di* sistema vivente. L'uso dei modelli è pervasivo nella scienza: per valutare un'ipotesi teorica a proposito di un certo sistema (vivente o meno), si studia il comportamento di un suo modello, che in tale circostanza svolge la funzione di surrogato per l'indagine scientifica[11]. Naturalmente, questa idea riposa sull'assunzione che il modello possegga le caratteristiche necessarie per poter essere utilizzato in tal modo, altrimenti lo sperimentatore non avrà buone ragioni per accettare le conclusioni teoriche raggiunte; capire quando questa assunzione possa dirsi soddisfatta è un compito arduo per i filosofi della scienza che si occupano del ruolo dei modelli nella ricerca scientifica[12]. In robotica – più specificamente, in un'area della robotica chiamata *biorobotica* – succede esattamente questo, e il modello è un robot[13].

Esistono due varianti di questa idea di metodo[14]. Nella prima, il robot viene utilizzato come strumento per prevedere come si comporterà, o come si sarebbe comportato, il sistema oggetto di studio in circostanze particolari. L'uso è del tutto analogo all'uso delle simulazioni al calcolatore per ricavare delle previsioni del tempo. Il modello robotico, inteso in questo caso

[11] R. Frigg, J. Nguyen, *Models and Representation*, in L. Magnani, T. Bertolotti (a cura di), *Springer Handbook of Model-Based Science*, Springer, Berlin 2017, pp. 49-102.

[12] B. Boesch, *A concrete example of representational licensing: The Mississippi River Basin Model*, in «Studies in History and Philosophy of Science», 92 (December 2021), 2022, pp. 36–44.

[13] E. Datteri, *Biorobotics*, in L. Magnani & T. Bertolotti (a cura di), *op. cit.*, pp. 817–837; E. Datteri, G. Tamburrini, *Biorobotic Experiments for the Discovery of Biological Mechanisms*, in «Philosophy of Science», 74, 3, 2007, pp. 409–430; B. Webb, *Validating biorobotic models*, in «Journal of Neural Engineering», 3, 3, 2006, R25–R35; B. Webb, T. R. Consi (a cura di), *Biorobotics: Methods and Applications*, The MIT Press, Cambridge 2001.

[14] E. Datteri, *Interactive biorobotics*, in «Synthese», 198, 8, 2021, pp. 7577–7595.

come strumento predittivo, è particolarmente utile quando il sistema oggetto di studio è inaccessibile per motivi di carattere pratico o teorico. Questo è il caso di alcuni studi in cui un modello robotico è stato utilizzato per studiare il comportamento che alcune specie animali estinte avrebbero prodotto in certe circostanze. Il robot, in questi casi, è stato costruito tenendo in considerazione l'analisi di reperti fossili[15].

Nella seconda variante di questa idea di metodo, il robot viene utilizzato non per *prevedere* il comportamento di un certo sistema vivente, ma per *studiare il meccanismo che lo produce*. La distinzione che può essere tenuta in mente per comprendere il senso di questa differenza di metodo è quella tra previsione e spiegazione di un certo comportamento. Questa idea di metodo ha una lunga storia: come ha mostrato in modo approfondito il filosofo della scienza italiano Roberto Cordeschi[16], essa affonda le sue radici almeno nei primi decenni del XX secolo, per poi sbocciare ed essere ampiamente percorsa durante la cibernetica degli anni '40 e l'Intelligenza Artificiale dei padri fondatori Allen Newell e Herbert Simon[17]. Oggi è spesso adottata in biorobotica. Per scoprire il meccanismo che permette a un certo sistema vivente di produrre certi aspetti del suo comportamento, si ipotizza un meccanismo potenzialmente esplicativo; si implementa quel meccanismo in un robot; se il robot è in grado di riprodurre gli aspetti del comportamento in esame, si può essere indotti ad accettare l'ipotesi che il meccanismo realizzato nella macchina sia anche quello che governa il sistema vivente in esame. Altrimenti, l'ipotesi può essere scartata.

Infine, in certe aree di ricerca empirica – psicologica ed etologica – i robot vengono utilizzati come strumenti scientifici in un senso un po' diverso da quello predittivo ed esplicativo

[15] J. Long, *Darwin's Devices: What Evolving Robots Can Teach Us About the History of Life and the Future of Technology*, Basic Book, New York 2012.
[16] R. Cordeschi, *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics*. Springer, Dordrecht 2002.
[17] A. Newell, H. A. Simon, *Human Problem Solving*, Prentice-Hall, Hoboken 1972.

illustrato finora. In queste aree, i robot a volte non vengono considerati surrogati (in altre parole, sostituti) del sistema in esame. Piuttosto, essi vengono fatti *interagire* con il sistema in esame. Sulla base dell'analisi delle reazioni che quest'ultimo produce in risposta agli stimoli impartiti dal robot, si traggono conclusioni teoriche su di esso[18]. Per esempio, analizzando il comportamento che certi tipi di pesce producono in risposta alla presenza e all'interazione con un robot che riproduce l'aspetto e il comportamento di un predatore, gli sperimentatori traggono conclusioni sulle relazioni preda-predatore in particolari ecosistemi[19]. Il ruolo del robot, in questi casi, non è dunque quello di sostituirsi al sistema in esame come oggetto su cui viene effettuata la sperimentazione (come nel caso predittivo ed esplicativo discusso poco sopra), bensì quello di stimolare il sistema in esame in modo controllato, prevedibile e ripetibile.

Ricapitoliamo. Questo paragrafo ha discusso l'intersezione tra ricerca scientifica e ricerca robotica, sottolineando che

1) la ricerca scientifica coinvolge processi di tipo costruttivo (soprattutto legati alla realizzazione degli strumenti sperimentali).

2) La progettazione e la costruzione di un robot

a. riposano su teorie scientifiche "abilitanti",

b. coinvolgono processi di teorizzazione sul comportamento del robot stesso.

3) I robot possono costituire strumenti di indagine scientifica,

a. come modelli o surrogati (sostituti) del sistema in esame, su cui lo sperimentatore svolge esperimenti per trarre conclusioni predittive o esplicative a proposito di esso, oppure

[18] E. Datteri, *The creation of phenomena in interactive biorobotics*, in «Biological Cybernetics», 115, 6, 2021, pp. 629–642; Id., *The Logic of Interactive Biorobotics*, in «Frontiers in Bioengineering and Biotechnology», 8, 2020.

[19] D. Romano, E. Donati, G. Benelli, C. Stefanini, *A review on animal–robot interaction: from bio-hybrid organisms to mixed societies*, in «Biological Cybernetics», 113, 3, 2019, pp. 201–225.

b. come strumenti che impartiscono stimoli controllati, prevedibili e ripetibili al sistema in esame.

Riportiamo queste considerazioni all'analisi dei rapporti tra robotica e filosofia della scienza. La robotica può costituire oggetto di studio per la filosofia della scienza? Queste considerazioni suggeriscono una risposta positiva, almeno in relazione ad alcuni dei punti appena riassunti. Il tema (1) è già da molti decenni oggetto di studio della filosofia della scienza, ma esso non riguarda la robotica in modo specifico. Il tema (2a) solleva riflessioni di filosofia della scienza anch'esse però esterne alla robotica e relative alle discipline (fisiche, psicologiche, e così via) che forniscono la base abilitante per la ricerca in robotica. Il tema (2b) è invece oggetto di potenziale interesse per i filosofi della scienza, perché riguarda la struttura dei processi di indagine scientifica che hanno per oggetto un robot. Il tema (3), nei suoi due sottopunti, è decisamente oggetto di potenziale interesse per la filosofia della scienza e, sotto un certo aspetto, ricade nell'ambito del punto (1), poiché il robot svolge in quei casi il ruolo di strumento sperimentale. L'indagine verte, in questo caso, sull'analisi della *validità* di tali strumenti: in che misura, e sotto quali assunzioni ausiliarie, è lecito trarre conclusioni sul sistema oggetto di studio, per esempio un essere vivente, sulla base dell'osservazione del comportamento di un suo modello meccanico? Sotto quali assunzioni ausiliarie le stimolazioni impartite dal robot (punto 3b) possono somigliare alle stimolazioni che il sistema vivente in esame riceve nel suo ecosistema naturale, e cosa ciò comporta rispetto all'uso del robot come strumento per indagare il suo comportamento in tale ecosistema?

3. *Dalla robotica alla filosofia della scienza generale*

La distinzione tra *filosofia della scienza generale* e *filosofia delle scienze particolari* è largamente accettata dalla comunità di

ricerca come descrittivamente adeguata, almeno a un certo livello di astrazione e di semplificazione. Una fetta della comunità di ricerca si occupa di questioni di filosofia della scienza che sono intese riguardare trasversalmente qualsiasi ambito (o moltissimi ambiti) di ricerca scientifica. Un esempio è legato allo studio del concetto di "modello scientifico", che abbiamo richiamato nel paragrafo precedente. Cosa rende un certo oggetto concreto o astratto (per esempio, robot) *modello di* un altro sistema (per esempio, di un pesce)?[20] Chi ritiene che il significato inteso di questo termine non vari significativamente tra le varie discipline scientifiche – per esempio, che i fisici utilizzino modelli scientifici di sistema fisico e i biologi utilizzino modelli scientifici di sistema biologico nello stesso senso di "modello scientifico" – porrà la questione appena introdotta come una questione di filosofia della scienza *generale*. Un'altra parte della comunità di ricerca si occupa di questioni di filosofia della scienza che sorgono nell'ambito di particolari discipline scientifiche. Per esempio, nell'ambito delle neuroscienze cognitive, è piuttosto comune svolgere le cosiddette *inferenze inverse*: dal fatto che, in un dato individuo, una certa area cerebrale si attiva durante un certo compito, si inferisce il fatto che tale individuo in quel momento stia attivando un certo modulo o processo cognitivo.

Alcuni filosofi della scienza hanno indagato la legittimità di tali inferenze, legittimità minacciata dal fatto che ogni area cerebrale si attiva in concomitanza con l'attivarsi di molti processi cognitivi). In termini più precisi, si sono chiesti quali assunzioni ulteriori è necessario inserire tra le sue premesse perché la conclusione ne segua logicamente, e se queste premesse possano essere accettate. Dato che questo tipo di inferenza sorge specificamente nella ricerca neuroscientifica e cognitiva, il

[20] C. Callender, J. Cohen, *There is no special problem about scientific representation*, in «Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia», 21, 55, 2006, pp. 67–85.

problema di indagarne i limiti di validità è un problema di filosofia delle scienze particolari[21].

Su queste basi, entriamo nel vivo della questione affrontata in questo lavoro prendendo le mosse dal ramo che potremmo etichettare "dalla robotica alla filosofia della scienza", chiedendoci anzitutto se la ricerca in robotica possa contribuire alla ricerca in filosofia della scienza *generale*. Vi sono buone ragioni per dubitare che dalla robotica possano arrivare contributi significativi in tal senso. Uno dei motivi, forse non l'unico, è piuttosto ovvio: la filosofia della scienza generale, proprio per sua natura, produce tesi filosofiche in larga misura trasversali rispetto alle specifiche discipline. Anche qualora tali tesi filosofiche trovino esemplificazione in qualche ambito della robotica, tale esemplificazione fornirebbe un supporto limitato e non decisivo alla validità della tesi, proprio perché essa, in quanto generale, non "parla" né robotica né di particolari discipline scientifiche.

Tentativi in tal senso sono comunque stati compiuti. Cosa siano la mente e le emozioni, e che rapporto esse intrattengano con la materia, sono questioni che, pur avendo versanti empirici, stanno al centro della riflessione filosofica da secoli e possono essere inquadrate come questioni ontologiche (come questioni sul rapporto tra mente e materia in quanto "cose") e di filosofia della scienza propriamente detta (in quanto questioni sul rapporto tra teorie psicologiche e teorie fisiche o biologiche). Luisa Damiano e Paul Dumouchel, in un testo che discute alcune questioni filosofiche sollevate dalla robotica sociale, sostengono che gli sviluppi in questa area di indagine possono dare un contributo significativo allo sviluppo di nuove prospettive filosofiche sulla natura della

[21] E. Machery, *In defense of reverse inference*, in «The British Journal for the Philosophy of Science», 65, 2, 2014, pp. 251–267; C. Glymour, C. Hanson, *Reverse Inference in Neuropsychology*, in «The British Journal for the Philosophy of Science», 2015, pp. 1–15; R. A. Poldrack, *Can cognitive processes be inferred from neuroimaging data?*, in «Trends in Cognitive Sciences», 10, 2, 2006, pp. 59–63.

mente e delle emozioni. La robotica sociale, secondo loro, «non si limita a prendere posizione nel dibattito sulla natura della mente, ma ne trasforma i termini»[22]. Il comportamento dei robot sociali attuali o futuribili, oppure la possibilità di certe dinamiche di interazione tra robot ed esseri umani (in cui, per esempio, gli umani empatizzano con i robot e questi ultimi reagiscono in modo appropriato alle reazioni umane nei loro confronti), forniscono, secondo Damiano e Dumouchel, buoni motivi per ritenere che «la mente si delinei non più come una macchina computante più o meno (dis-)incorporata, ma come una struttura di interconnessione – una rete cognitiva socialmente distribuita»[23]. Lo sviluppo della robotica sociale suggerirebbe che la mente debba essere concepita come una «relazione tra agenti epistemici»[24].

Non valutiamo la plausibilità delle tesi ontologiche proposte dai due autori. Ci concentriamo invece su una questione a monte: in che senso il comportamento di un robot, o il dispiegarsi di una particolare dinamica di interazione umani-robot, possono fornire buone ragioni per sostenere particolari tesi ontologiche sulla natura della mente e delle emozioni? Il comportamento dei robot richiede una spiegazione, che fa necessariamente riferimento a certi enti teorici, appartenenti magari al dominio della mente e delle emozioni. Ma non è chiaro in che senso quelle spiegazioni implichino l'adozione di particolari tesi ontologiche sulla natura degli enti di cui esse "parlano". Per esempio, Dumouchel e Damiano sostengono che una buona interazione tra robot ed esseri umani può essere spiegata soltanto assumendo che si verifichino particolari relazioni mentali e affettive tra i soggetti in gioco. Ma la scelta ontologica in merito a come interpretare le nozioni di "mente" e di "affetto", in questa cornice teorica, è *a monte* e sotto-determinata dal successo di queste spiegazioni. Si può dire

[22] P. Dumouchel, L. Damiano, *Vivere con i robot. Saggio sull'empatia artificiale* (2017), Raffaello Cortina Editore, Milano 2019, p. 61.
[23] *Ibid.*, pp. 91-92.
[24] *Ibid.*, p. 102.

che lo stesso vale, naturalmente, in quella che consideriamo interpretazione classica delle scienze cognitive. L'attribuzione di una mente alle macchine, per esempio, è in linea di principio compatibile con molte posizioni ontologiche sulla natura della mente – con posizioni dualistiche, con posizioni funzionalistiche che affermano identità di occorrenza tra stati mentali e stati fisici, e persino, con alcuni accorgimenti, con posizioni ispirate all'identità di tipo.

Che considerazioni possiamo trarre, dall'analisi di questo esempio, a proposito del rapporto tra robotica e filosofia della scienza generale? La ricerca in robotica fa, per così dire, il suo mestiere: attraverso processi conoscitivi e realizzativi, mira alla progettazione e costruzione di sistemi robotici adatti agli scopi per cui sono stati pensati. Qualsiasi affermazione secondo cui gli sviluppi della robotica contemporanea *suggeriscono* posizioni filosofiche di ampio respiro è da intendersi, al massimo, in senso metaforico: tutt'al più si può avanzare l'ipotesi secondo cui gli sviluppi della robotica contemporanea sono *compatibili con*, o sono particolarmente *comprensibili alla luce di*, certe posizioni filosofiche di ampio respiro – per esempio dell'idea che la mente sia una relazione tra agenti epistemici (qualunque cosa questa espressione indichi). Forse – spogliato dalla sua veste metaforica – è questo il senso in cui devono essere interpretate le affermazioni di Damiano e Dumouchel. Anche in questo senso, tuttavia, il sostegno fornito dalla robotica sociale alle tesi appena discusse è del tutto marginale. La plausibilità di una tesi ontologica di ampio respiro non può essere difesa soltanto mostrando che essa è compatibile con quanto si fa in certi particolari ambiti di indagine: è necessario un percorso argomentativo molto più complicato, che include, senza esaurirsi con, il mostrare che le tesi alternative non sono compatibili con quanto si fa in quel dominio di indagine.

4. *Dalla robotica alla filosofia delle scienze* particolari

Filosofi della scienza sono anche coloro che affrontano questioni di filosofia delle scienze *particolari* nel senso che abbiamo sopra descritto. In merito a questo tipo di analisi filosofica sosterremo una posizione più ottimistica rispetto a quella suggerita nella sezione precedente. La possibilità che i filosofi della scienza, occupandosi di robotica, affrontino temi di interesse filosofico riguardanti i fondamenti di particolari discipline scientifiche è aperta e percorsa da molti ricercatori. I tipi di analisi filosofica che possono sorgere in questo ambito includono i seguenti.

In primo luogo, è pertinenza dei filosofi della scienza la ricostruzione razionale dei metodi che sorgono in quest'area di indagine e l'analisi della loro validità. Pensiamo agli usi predittivi ed esplicativi dei modelli robotici discussi nel paragrafo 2: lo sperimentatore osserva il comportamento di un modello robotico di animale in particolari condizioni sperimentali, e trae conclusioni teoriche in merito al comportamento che tale animale produrrebbe nelle stesse condizioni (negli studi predittivi) o sul meccanismo che produce tale comportamento (studi esplicativi). Questi processi di inferenza, dal comportamento del robot al comportamento del sistema vivente in esame (o al suo meccanismo interno), devono essere giustificati. Quali assunzioni ulteriori dobbiamo accettare per saltare da premesse a conclusioni? Dobbiamo assumere che il robot sia un buon modello del sistema vivente in esame? In questo caso, cosa intendiamo per "buon modello"? Negli studi esplicativi, si conclude che il meccanismo alla base del comportamento dell'animale è simile a quello che governa il robot: simile in che senso, e a che livello di approssimazione, idealizzazione, semplificazione? Queste domande sono state occasionalmente affrontate nella letteratura filosofica.

In secondo luogo, la robotica può essere di interesse per la filosofia delle scienze particolari perché solleva varianti di questioni filosofiche più generali scarsamente affrontate in letteratura. Per esempio, un problema di filosofia della scienza molto dibattuto almeno da circa un secolo riguarda la natura delle spiegazioni scientifiche: cosa fa sì che certe asserzioni a proposito del sistema oggetto di studio *spieghino* effettivamente certi aspetti del comportamento? Questo problema è stato affrontato sia dai filosofi della scienza generale[25] sia dai filosofi delle scienze particolari[26]. Uno dei motivi per cui è interessante occuparsi del problema della spiegazione scientifica con riferimento alle scienze particolari è legato al fatto che tale problema può assumere, in quei contesti, vesti peculiari, o porre sfide particolari. Alcuni filosofi, per esempio, hanno sostenuto che le spiegazioni scientifiche del comportamento degli oggetti tecnologici hanno strutture e requisiti di "bontà" particolari, relativamente diversi da quelli che contraddistinguono le "buone" spiegazioni in altri dominii di indagine[27]. Mettere in rilievo queste peculiarità fa parte del più ampio obiettivo di fornire ricostruzioni razionali e analisi della validità dei processi di ricerca che hanno luogo in discipline particolari – in questo caso, della robotica –, obiettivo che caratterizza la filosofia delle scienze particolari.

In terzo luogo, l'analisi di come "funzionano" particolari aree di ricerca scientifica può fornire spunti per elaborare tesi filosofiche che possono essere estese, a posteriori, ad altre aree di ricerca scientifica. La cosiddetta biorobotica interattiva, per esempio, si basa su un uso dei modelli robotici peculiare rispetto all'uso che generalmente viene fatto di essi: come sottolineato in

[25] Cfr. W.C. Salmon, *Four Decades of Scientific Explanation*, University of Minnesota Press, 1989; C.F. Craver, *When mechanistic models explain*, in «Synthese», 153, 3, 2006, pp. 355–376; C.G. Hempel, *Aspects of Scientific Explanation and Other Essays, in the Philosophy of Science*, Free Press, 1965.
[26] J. Woodward, *Explanation and Invariance in the Special Sciences*, in «British Journal for the Philosophy of Science», 51, 2000, pp. 197–254.
[27] D. van Eck, *The Philosophy of Science and Engineering Design*, cit.

precedenza, i biorobot vengono utilizzati per stimolare il sistema in esame e non come basi per il ragionamento surrogativo[28]. Vi sono buoni motivi per ritenere che i biorobot in questione siano qualificabili come modelli di sistema vivente (secondo varie concezioni di "modello scientifico" proposte in letteratura[29]), anche se essi vengono utilizzati in un modo piuttosto diverso da quello "tradizionale". La filosofia della scienza, dedicandosi alla ricostruzione razionale di questo peculiare uso dei modelli, e interrogandosi a proposito della sua validità, può evidenziare metodologie scientifiche che poi, a valle dell'analisi, possono essere rintracciate anche in altre aree di indagine scientifica. Vi sono esempi di *modello stimolativo* anche al di fuori della biorobotica? Per esempio, quando nella ricerca neuroscientifica si impartiscono stimoli elettrici a certe parti del tessuto neurale per elaborarne le reazioni, si adotta una metodologia analoga a quella adottata in biorobotica interattiva?

5. *Riflessioni conclusive: dalla filosofia alla robotica*

In queste pagine abbiamo sostenuto, anzitutto, che la filosofia della scienza – intesa come indagine sul "funzionamento" della scienza – può assumere la robotica come suo oggetto di studio, nella misura in cui essa coinvolge teorie e processi di teorizzazione scientifica, e produce modelli e strumenti (cosiddetti biorobotici) per la comprensione dei sistemi viventi (paragrafo 2). Abbiamo avanzato dubbi sulla possibilità che la ricerca in robotica offra contributi per la ricerca in filosofia della scienza generale (paragrafo 3), mentre abbiamo suggerito che la robotica solleva temi di interesse per la filosofia delle scienze particolari (paragrafo 4). È compito dei filosofi delle scienze particolari ricostruire razionalmente e analizzare la validità dei metodi di ricerca adottati nei vari campi di

[28] E. Datteri, *The Logic of Interactive Biorobotics*, in «Frontiers in Bioengineering and Biotechnology», 8, 2020.
[29] R. Frigg, J. Nguyen, *Models and Representation*, cit.

indagine, e la robotica offre varie questioni ancora aperte (tra cui la validità dei processi di inferenza scientifica svolti in biorobotica). Certi problemi epistemologici, anche di carattere generale, possono assumere aspetti peculiari e degni di riflessione filosofica nell'ambito della sperimentazione robotica e biorobotica, e in questa area di indagine la filosofia della scienza può identificare metodologie esportabili ad altri campi del sapere.

Abbiamo finora lasciato indietro una delle due possibili direzioni dell'interazione tra la filosofia della scienza e la robotica, etichettabile come "dalla filosofia della scienza alla robotica". La filosofia della scienza, come abbiamo sottolineato nell'Introduzione, è una disciplina che si propone obiettivi normativi: nelle parole di Losee, essa si occupa di «come si può pensare che la scienza dovrebbe essere»[30]. Quando rivolge la sua attenzione alla robotica, essa produce ricostruzioni razionali delle metodologie di indagine adottate in tale campo di studio e ne analizza la validità. Proprio per il suo aspetto normativo si presta a contribuire alla ricerca in robotica, fornendo linee guida per la conduzione di processi di ricerca ed evidenziando *caveat* e vincoli metodologici il cui soddisfacimento assicuri la validità di tali processi. Uno dei criteri di successo delle analisi condotte nella filosofia delle scienze particolari (non certo l'unico) riguarda proprio la capacità di produrre modelli filosofici che siano in grado di guidare e orientare la ricerca scientifica nelle aree di pertinenza. Alcune delle analisi filosofiche che abbiamo citato, soprattutto quelle che riguardano la biorobotica, ambiscono a raggiungere questi criteri di successo. Se poi un'effettiva interazione "dalla filosofia della scienza alla robotica" si verifichi o no nel panorama della ricerca contemporanea, è questione non affrontabile in questo

[30] J. Losee, *op. cit.*, p. 13.

contesto e di possibile interesse per la sociologia e la storia della scienza.

EDOARDO DATTERI insegna Filosofia della Scienza presso l'Università degli Studi di Milano Bicocca

edoardo.datteri@unimib.it

S&F_n. 27_2022

STORIA

ALFONSO LANZIERI

DA MATRIX A DON'T LOOK UP: IL REALISMO COME QUESTIONE ETICO-POLITICA

1. *Introduzione* 2. *Il reale tra parentesi* 3. *Il ritorno alla questione del reale*
4. *Il recupero del "mondo" in filosofia della mente* 5. *Un nuovo "metodo"*

ABSTRACT: *FROM MATRIX TO DON'T LOOK UP: REALISM AS AN ETHICAL-POLITICAL ISSUE*
In recent decades, the topic of "reality" has gradually returned to the forefront of philosophy. On the side of ontology, the discourse around the "New Realism", an expression referring to the return of realism to the specialist debate, has been ongoing for some time. In the field of philosophy of mind, following the same line, the "4E" approach argues that mental does not occur solely in the head, but is also embodied, embedded, enacted, or extended by way of extra-cranial processes and structures. In this article, we want to relate this paradigm shift to two successful films: The Matrix (1999) and Don't Look Up (2021). We will try to show how the comparison between the two films well describes the cultural change that led to the renewed philosophical attention to the theme of the "reality" and its socio-political relevance.

1. *Introduzione*

Pochi mesi fa, due film di grande successo sono stati distribuiti al pubblico quasi in contemporanea. Stiamo parlando di *Matrix Resurrection* e *Don't look up*. Il primo è il quarto capitolo della celebre saga fantascientifica in stile cyberpunk, il secondo una satira distopica della società contemporanea. A giudizio di chi scrive, le due opere cinematografiche, se confrontate (in particolare, per quanto riguarda *Matrix*, facciamo riferimento al primo episodio della saga), possono rispecchiare, con buona approssimazione, il percorso di una certa parte della nostra cultura (in particolare ci riferiamo al dibattito filosofico) degli ultimi due decenni.

L'idea di partenza di *Matrix* (il primo film è del 1999) è nota a tutti: l'Intelligenza Artificiale, resasi indipendente dall'uomo e a questi ostile, muove guerra alla razza umana e alla fine riesce a schiavizzarla, imprigionandola in una "neurosimulazione interattiva", chiamata appunto *Matrix*, grazie alla quale gli uomini sognano di star vivendo la loro vita normale, quando in realtà giacciono a miliardi in uno stato d'incoscienza, mentre fuori il mondo è ridotto a un ammasso di macerie e l'energia del loro organismo è utilizzata dalle macchine per sopravvivere. Il film si basa su un'ipotesi, a metà tra le neuroscienze e la filosofia, assai seducente, vale a dire l'idea che la nostra coscienza – l'esperienza in prima persona che abbiamo del mondo – sia nient'altro che il prodotto dei nostri neuroni. «Che vuol dire reale? – chiede Morpheus a Neo in un dialogo del film – Dammi una definizione di reale. Se ti riferisci a quello che percepiamo, a quello che possiamo odorare, toccare e vedere, quel reale sono semplici segnali elettrici interpretati dal cervello». Per Morpheus, insomma, la realtà sarebbe un'*allucinazione vera*. Quest'idea, come proveremo a dire meglio a breve, si avvicina a una certa posizione sulla natura della nostra coscienza, che nell'ambito della filosofia della mente può essere associata alle teorie "internaliste". Secondo l'internalismo infatti – schematizzando un po' il discorso – la coscienza è ciò che accade "dentro la testa".

In *Don't look up*, invece, lo scenario è in un certo senso contrario. Un asteroide si avvicina minacciosamente alla Terra, ma qui si apre un tragicomico dibattito sull'esistenza o meno del corpo celeste. L'umanità rischia la propria rovina non perché vive confinata e incosciente in un mondo fittizio, ma perché una spirale di crescente infodemia, scatenata da un intreccio irresponsabile e delirante tra potere politico e mediatico, impedisce la presa d'atto dei fatti. In altri termini, in *Matrix* le nostre percezioni sono false, perché falso è il mondo in cui

crediamo di trovarci; nel secondo caso, invece, il problema è proprio la mancanza di contatto con la verità della nostra percezione. *Don't look up*, allora, ripropone anzitutto il problema della *realtà* e della sua importanza etico-politica. Questo ci sembra in linea con la nuova attualità del "realismo" che si può registrare nell'ambito del dibattito filosofico e della quale si darà conto a breve. Inoltre, l'intera impostazione del film è un invito a spostare la nostra attenzione da ciò che è dentro la nostra mente a ciò dentro cui è la nostra mente: da uno sguardo *internalista* a uno *esternalista*. Con quest'ultimo termine ci si riferisce a un'impostazione della filosofia della mente che, negli ultimi anni, ha sempre più criticato l'internalismo nel suo assunto teoretico di fondo: per comprendere il fenomeno della coscienza, bisogna guardare all'interazione tra soggetto e mondo. La coscienza non è ciò che accade dentro la testa, ma ciò che scaturisce dall'interazione tra individuo e ambiente.

A questo punto, allora, dovrebbero essere chiari i termini del discorso. A nostro avviso, gli ultimi decenni del dibattito filosofico sono stati caratterizzati dal passaggio da una egemonia dell'antirealismo a una del realismo, e dall'internalismo all'esternalismo: questi due cambi d'orizzonte (che, lo ripetiamo, ora assumiamo come dati ma a breve tenteremo di giustificare tale assunto) vanno letti sinotticamente e gerarchicamente. Si tratta di un unico movimento di riscoperta del reale – dopo l'enfasi antirealista della cosiddetta età postmoderna – rispetto al quale il passaggio dall'egemonia dell'internalismo a quella dell'esternalismo, in filosofia della mente, rappresenta, pur nella specificità dei percorsi, un sott'insieme teoretico o, se così possiamo esprimerci, un'espressione "regionale". Tale passaggio, se è stato correttamente rilevato, va altresì indagato nelle sue peculiarità e nelle sue criticità. Abbiamo così anticipato, in modo compendiato e generale, la tesi di fondo che intendiamo presentare e i passi argomentativi che seguiranno. Si

tratta adesso di ampliare e tentare di fondare le affermazioni fin qui fatte.

2. *Il reale tra parentesi*

Da qualche anno la questione del realismo è tornata al centro del dibattito filosofico, proprio quando sembrava essere stata definitivamente accantonata, perlomeno dall'orientamento *mainstream*. Nel secolo scorso, infatti, soprattutto dal secondo dopoguerra in avanti, la koinè filosofica dominata dalla decostruzione e dalla dismissione della metafisica, all'ombra di Nietzsche e Heidegger, aveva marginalizzato in maniera decisa la nozione di "verità oggettiva". Come ha osservato Mario De Caro, insomma, «negli ultimi decenni del Novecento, la gran parte dei filosofi non dava molto credito al concetto di realtà. Di conseguenza, le filosofie che si incentrano su questo concetto – ossia le varie forme di realismo – erano piuttosto impopolari tanto nel mondo filosofico analitico quanto in quello continentale»[1].

A quanti ancora si ostinavano a utilizzarla in senso forte, era riservata l'accusa più pesante che si possa rivolgere a un filosofo: l'ingenuità. La filosofia, infatti, nasce sulla promessa di superamento dello sguardo ingenuo, che è proprio della *doxa*, sulla realtà. Il filosofo ingenuo è dunque in realtà il non-filosofo, è colui per il quale la filosofia non è ancora iniziata. Tale pesante rimprovero è stato rivolto a chi, nonostante la scoperta moderna della "soggettività del mondo" – sancita nel modo speculativamente più consapevole dalla rivoluzione copernicana kantiana – ha continuato a non praticare quella indispensabile purificazione dello sguardo «che sola permette di liberarlo dai pregiudizi che impediscono di vedere correttamente il mondo, il più radicato dei quali è senz'altro la credenza – detta appunto,

[1] M. De Caro, *Realtà*, Bollati Boringhieri, Torino 2020 (edizione kindle), p. 10.

"ingenua" o "naturale" – in un reale che sussisterebbe in se stesso e che la conoscenza potrebbe attingere»[2]. Come poter ancora credere nell'esistenza di un mondo disponibile al rispecchiamento di un giudizio adeguato allo stato di cose, se l'esperienza può essere compresa *radicitus* solo in quanto co-appartenenza insuperabile di soggetto e oggetto? In linea con questa postura di sguardo, Heidegger ha potuto affermare che

> diciamo troppo poco dell'essere in sé stesso quando, dicendo "l'essere", lasciamo fuori il suo essere-presente all'uomo, misconoscendo così che quest'ultimo entra esso stesso a costituire l'"essere". Anche dell'uomo diciamo sempre troppo poco quando, dicendo l'"essere" (non l'essere dell'uomo), poniamo l'uomo per sé stesso e solo in un secondo tempo lo mettiamo in relazione con l'"essere"[3].

Tale concezione potrebbe avere origini relativamente lontane. Secondo Gianni Vattimo, infatti, il discorso condotto in *Essere e tempo* può essere letto come la storicizzazione dell'apriori di Kant, il quale, a suo tempo, aveva già sbarrato la strada verso un possibile raggiungimento della verità oggettiva, intesa cioè come ciò che è colto nella sua *indipendenza* dall'attività costruttiva della soggettività trascendentale: «il mondo è fenomeno, cioè un ordine di cose che il soggetto entra attivamente a costituire. In Kant, tuttavia, c'era ancora l'idea che le strutture a priori del soggetto fossero uguali in tutti gli esseri razionali finiti. Nel Novecento, dopo Heidegger, queste strutture vengono riconosciute nella loro radicale storicità»[4]. In sostanza, stabilita l'intrascendibilità della correlazione soggetto-oggetto, la quale per Meilassoux sarebbe «la nozione centrale della filosofia moderna»[5], la contemporaneità ha affermato con sempre maggior forza l'idea secondo cui noi abbiamo accesso solo, appunto, alla correlazione del pensiero e dell'essere, e in nessun caso all'essere nella sua inseità. La possibilità di attingere la cosa

[2] R. Ronchi, *Il canone minore. Verso una filosofia della natura*, Feltrinelli, Milano 2017, p. 125
[3] M. Heidegger, *Che cosa significa pensare?* (1952), tr. it. Sugarco, Milano 1971, vol. I, p. 38.
[4] G. Vattimo, *Della realtà*, Garzanti, Milano 2012, p. 86.
[5] Q. Meillassoux, *Dopo la finitudine. Saggio sulla necessità della contingenza* (2012), tr. it. Mimesis, Milano 2012, p. 8.

in sé (potremmo dire la "cosa senza me"), *ab-soluta*, non viziata cioè dal suo rapporto alla soggettività finita, sarebbe così inesorabilmente perduta: «Il sogno del realista è che le cose parlino da sole, senza la nostra mediazione. Ma questo, lo sappiamo, non è un sogno, è un'illusione. […] Il problema è quella cosa in sé – cioè appunto senza di noi – che da Kant in avanti turba i sogni di ogni realista»[6]. Da qui in poi, il pensiero filosofico potrà dedicarsi solo a pensare la correlazione più originaria, non a cogliere l'ente in sé, impossibile da sciogliere dal suo riferimento all'osservatore[7]. In forza delle premesse fino ad ora esposte, in forma ovviamente molto schematica, il pensiero del secondo '900 è stato caratterizzato, almeno negli esponenti e nelle correnti più note, da una inclinazione decisamente antirealista: non c'è un mondo indipendente dai nostri schemi conoscitivi e se esiste non è attingibile in sé stesso. Non sorprende, dunque, che uno dei filosofi della mente contemporanei più importanti, John Heil, descrivesse così il panorama filosofico di fine anni '80: «Gli scritti antirealisti soverchiano per numero e densità una produzione realista stabile ma relativamente modesta […] (Solo) l'Australia, isolata ed evolutivamente marginale, è rimasta roccaforte di realisti e marsupiali»[8].

3. *Il ritorno alla questione del reale*

Come anticipato nell'introduzione, allo specifico *Zeitgeist* antirealista, ha fatto seguito un ritorno al realismo: «Come il fantasma del Commendatore – che nell'ultimo atto del *Don Giovanni* fa la sua inesorabile comparsa a rammendare verità che non si possono obliare – negli ultimi anni la realtà è tornata e si è saldamente ricollocata al centro della scena filosofica»[9]. In tale

[6] F. Cimatti, *Cose. Per una filosofia del reale,* Bollati Bolinghieri, Torino 2018, p. 18
[7] Cfr. Q. Meillassoux, *op. cit.*, p. 12.
[8] J. Heil, *Recent Work in Realism and Antirealism*, in «Philosophical Books», 30, 2, 1989, pp. 65-73, qui p. 65.
[9] M. De Caro, *op. cit.*, p. 12.

orizzonte, l'espressione *Nuovo Realismo* è diventata alquanto diffusa in anni recenti, stando a indicare proprio una generale tendenza del pensiero filosofico internazionale alla riproposizione del tema del realismo[10]. I nomi e gli indirizzi sono i più disparati: si va dal cosiddetto realismo speculativo di Quentin Meillassoux, alla proposta dell'ontologia "iperrealista" di Markus Gabriel[11], passando per altri nomi significativi, quali ad esempio Tristan Garcia, Graham Harman, Peter Gratton e Lee Braver. In Italia, com'è noto, il *Nuovo Realismo* è legato al nome del filosofo torinese Maurizio Ferraris. A partire dal suo *Manifesto del nuovo realismo*[12], il pensatore italiano ha inteso ridiscutere i "dogmi" indiscutibili del pensiero postmoderno: il primo, la realtà non è altro che una costruzione storico-sociale illimitatamente interpretabile; il secondo, la verità è una nozione tanto inutile quanto dannosa, in nome della quale sono state giustificate violenze ed esclusivismi. A essa bisognerebbe preferire la solidarietà. «Il pendolo del pensiero, che nel Novecento inclinava verso l'antirealismo – spiega Ferraris – nelle sue varie versioni (ermeneutica, postmodernismo, "svolta linguistica" ecc.), con il tornante del secolo si era spostato verso il realismo (nei suoi tanti aspetti: ontologia, scienze cognitive, estetica come teoria della percezione)»[13]. È significativo, a nostro avviso, il riferimento alle scienze cognitive e all'estetica come teoria della percezione: come vedremo a breve, infatti, e come abbiamo anticipato nell'introduzione, nell'ambito della filosofia della mente (che certo non coincide con quello delle scienze cognitive ma si

[10] Si tratta di un panorama unificato da una condivisa postura teoretica realista, ma vario quanto a nomi e percorsi. Per uno sguardo generale cfr. S. De Sanctis (a cura di), *Nuovi realismi*, Bompiani, Milano 2017.
[11] A tal proposito cfr. il già citato Q. Meillassoux, *op. cit.* e M. Gabriel, *Il senso dell'esistenza*, Carocci, Roma 2012. Citiamo questi due lavori – rispettivamente francese e tedesco in originale – perché, a nostro avviso, sono tra i più influenti pubblicati nell'ultimo decennio all'estero (area continentale) sul tema in oggetto.
[12] M. Ferraris, *Manifesto del nuovo realismo*, Laterza, Roma-Bari 2012.
[13] *Ibid.*, p. IX.

costituisce in stretto dialogo con esso) proprio la riscoperta del "mondo" ha rappresentato il filo rosso dell'evoluzione della ricerca degli ultimi anni. Nella sua analisi, Ferraris sostiene che il pensiero postmoderno sia retto, a suo dire, da un ponteggio fatto di tre grandi assi teoriche, che chiama *ironizzazione*, *desublimazione* e *deoggettivizzazione*, ma è nella critica a quest'ultima nozione, definita di «rovinosa centralità»[14], che il filosofo torinese concentra i cardini della sua contestazione: «se cerchiamo la ragion sufficiente e il motore politico della ironizzazione e della desublimazione troviamo la deoggettivizzazione, l'idea che l'oggettività, la realtà e la verità siano un male»[15]. La deoggettivizzazione non è altro che l'assunzione teorica del celebre stigma nietzschiano: "non ci sono fatti, ma solo interpretazioni". Il problema, sostiene Ferraris, è che, diversamente da quanto capitato in precedenza nella storia, quella *Weltanschauung* che risponde al nome di postmodernità ha trovato, oggi, una piena realizzazione politica e sociale. Sulle prime – argomenta il filosofo torinese – la cancellazione della verità oggettiva, la frammentazione del reale nel gioco decostruttivo e ricostruttivo delle infinite narrazioni dei soggetti, è sembrato l'avvio promettente di una liberazione, che però non si è realizzata.

> Il "divenir favola" del "mondo vero" non c'è stato, non si è vista la liberazione dai vincoli di una realtà troppo monolitica, compatta, perentoria, una moltiplicazione e decostruzione delle prospettive che sembrava riprodurre, nel mondo sociale, la moltiplicazione e la radicale liberalizzazione (si credeva negli anni Settanta del secolo scorso) dei canali televisivi. Il mondo vero certo è diventato una favola, anzi è diventato un reality, ma l'esito è stato il populismo mediatico, un sistema nel quale (purché se ne abbia il potere) si può pretendere di far credere qualsiasi cosa. Nei telegiornali e nei talk show si è assistito al regno del "Non ci sono fatti, solo interpretazioni", che – con quello che purtroppo è un fatto, non una interpretazione – ha mostrato il suo significato autentico: "La ragione del più forte è sempre la migliore"[16].

[14] *Ibid.*
[15] *Ibid.*, p. 20.
[16] *Ibid.*, p. 6.

Ciò che interessa sottolineare di questo brano, per gli scopi del nostro contributo, è la ricaduta politica del tema di cui si sta parlando, messa a nostro avviso correttamente in luce da Ferraris. La questione del reale, in sostanza, non è tema di mera astrattezza epistemologica, ma problema *politico*[17], nella misura in cui tocca la consistenza e il senso dei fatti del mondo condiviso da una comunità. Il peso politico e sociale dell'argomento è ribadito e ampliato da Ferraris in un'opera più vicina a noi, in cui il concetto di postmodernità viene messo in connessione con quello di postverità[18] (parola dell'anno 2016 secondo l'*Oxford Dictionaries*), a sua volta intimamente correlato alla enorme questione delle *fake news*. Scrive difatti Ferraris: «Da dove viene la postverità? Una volta tanto, dalla filosofia. […] la postverità è un frutto, magari degenere, del postmoderno»[19]. Giustificata, si spera, l'affermazione introduttiva del ritorno al "reale" della filosofia contemporanea, spostiamoci ora al campo degli studi sul rapporto mente-cervello.

4. *Il recupero del "mondo" in filosofia della mente*

Con l'espressione *mind-body problem* ci si riferisce al problema del rapporto tra le "funzioni cognitive superiori" – o livello mentale dell'esperienza del mondo – e il livello biologico, in particolare il nostro sistema nervoso, alla luce dell'evidenza immediata della correlazione tra i due piani. Negli ultimi decenni, la riflessione che vede occupate la cosiddetta filosofia della mente, le scienze cognitive e le neuroscienze rispetto a questo complesso tema di "confine" – l'"hard problem", il problema

[17] L'opposizione appena segnalata è da rilevarsi entro il campo del senso comune. La convinzione di chi scrive, infatti, è che non si dia alcuna riflessione filosofica seria che non sia in rapporto con la concretezza della vita e delle sue trame intersoggettive, purché tale "concretezza" non sia ridotta a banale praticità.

[18] Il *Vocabolario Treccani* la definisce: «Argomentazione, caratterizzata da un forte appello all'emotività, che basandosi su credenze diffuse e non su fatti verificati tende ad essere accettata come veritiera, influenzando l'opinione pubblica».

[19] M. Ferraris, *Postverità e altri enigmi*, Il Mulino, Bologna 2017, p. 19.

difficile, come l'ha icasticamente definito il filosofo australiano David Chalmers – ha conosciuto un'intensificazione vertiginosa, causata, in buona parte, dall'affinamento delle tecniche di osservazione del cervello (*brain imaging*), le quali consentono oggi uno studio più accurato dell'organo cerebrale a lavoro rispetto a ieri, mettendo a disposizione degli studiosi una grande mole di dati sperimentali[20].

Come detto in apertura e già ribadito, negli ultimi decenni, le ricerche in quest'ambito hanno subito un certo spostamento d'asse teorico, che possiamo riassumere come cambiamento dalla scienza cognitiva classica (SCC) alla nuova scienza cognitiva (NSC). Il programma di ricerca della versione "classica" della scienza cognitiva (SCC), sviluppatasi tra gli anni '50 e '80 del Novecento, era fondato sulla convinzione che fosse possibile studiare i processi cognitivi considerandoli processi di elaborazione operati su rappresentazioni, seguendo il filo di un'identificazione più o meno spinta tra apparato cognitivo umano e calcolatore digitale[21]. Date queste premesse, per gli scienziati cognitivi di tale indirizzo, è possibile tracciare i confini della cognizione in modo non dissimile da come un informatico traccerebbe quelli delle operazioni di calcolo di un elaboratore, vale a dire facendolo passare sui punti di interfaccia tra la macchina e l'elemento mondo[22]. Susan Hurley, tra le prime voci critiche dell'approccio internalista, ne ha descritto in modo chiaro l'assunto teorico di base: «Se la percezione è l'input che va dal mondo alla mente e l'azione è l'output che va dalla mente al mondo, allora la mente, in quanto distinta dal mondo, è ciò verso cui l'input si dirige e da cui l'output proviene. Quindi,

[20] Per una buona introduzione generale di carattere storico cfr. S. Nannini, *L'anima e il corpo. Un'introduzione storica alla filosofia della mente*, Laterza, Roma-Bari 2011. Si tratta di una seconda edizione aggiornata e ampliata rispetto a quella del 2002.

[21] Cfr. H. Gardner, *La nuova scienza della mente. Storia di una rivoluzione cognitiva*, Feltrinelli, Milano 1998.

[22] Cfr. L.A. Shapiro, *Embodied Cognition*, Routledge, London-New York 2019, p. 28.

nonostante la rete di relazioni causali tra organismi e ambienti, supponiamo che la mente debba essere in un luogo separato, all'interno di un qualche confine che la distingue dal mondo»[23].
Negli ultimi anni si è potuto registrando, invece, un cambiamento di paradigma. Il SCC, infatti, ha subito nel tempo diverse critiche principalmente a motivo del suo approccio internalista, orientato cioè a considerare la mente come sistema chiuso, sostanzialmente strumento d'elaborazione-dati separato dall'ambiente circostante. Per il modello esplicativo internalista, in altre parole – schematizziamo un po' il discorso – "la mente è ciò che accade dentro la testa". L'alternativa è oggi rappresentata dal modello *esternalista*. Questo, più che un movimento di ricerca omogeneo, definisce un orizzonte comune a diversi programmi di ricerca, tutti accomunati però da un medesimo assunto teorico di fondo: l'interazione continua tra agente e ambiente riveste un ruolo *costitutivo*[24] per l'emergenza della nostra esperienza mentale. Potremmo riassumere la questione centrale della prospettiva esternalista con la domanda «dove finisce la mente e comincia il resto del mondo?»[25], quesito chiave di un celebre articolo di Andy Clark e David Chalmers, nel quale presentavano la teoria della *Mente estesa*. Se per l'internalismo i confini della mente sono, grossomodo, i confini della scatola cranica, l'esternalismo immagina, con differenti livelli di radicalità e sfumature, che la mente sia più estesa del sistema nervoso. In tale linea, il paragone computer-cervello viene respinto sempre più spesso, in favore di un orientamento di ricerca che è stato anche definito *4E cognitive science*, espressione che sta ad indicare una scienza cognitiva per la quale è indispensabile considerare la mente secondo le seguenti quattro

[23] S. Huxley, *Consciousness in Action*, Harward University Press, Cambridge 1998, p. 2. Traduzione mia.
[24] Cfr. M. Di Francesco, G. Piredda, *La mente estesa. Dove finisce la mente e comincia il resto del mondo?*, Mondadori, Milano 2012, pp. 79-80.
[25] A. Clark, D. Chalmers, *The Extended Mind*, in «Analisys», 58, 1, 1998, pp. 7-19.

caratteristiche: *embodied*, vale a dire incarnata, poiché i processi cognitivi sono distribuiti, e anche fisicamente implementati, sullo stesso substrato neurale responsabile della percezione e dell'azione; *enacted*, in quanto i processi cognitivi sono il portato dell'interazione dinamica tra un soggetto agente e l'ambiente; *embedded*, termine col quale si fa riferimento al carattere radicalmente situato dei processi cognitivi (ambiente naturale, storico, sociale e culturale); *extended*, poiché i processi cognitivi non si consumano "dentro la testa", ma si estendono al di là dei confini della nostra scatola cranica, in quanto in una certa misura scaricabili su supporti tecnologici esterni, cui riconoscere un ruolo attivo nello sviluppo del nostro pensiero[26].

Questo cambio di prospettiva è stato naturalmente favorito da acquisizioni di dati sperimentali che hanno portato gli studiosi a riformare il paradigma euristico precedentemente adottato, in particolare rivalutando il ruolo del corpo. Il modello *Embodied cognition*, ad esempio, sostiene che la gran parte dei processi cognitivi avvenga mediante i sistemi di controllo del corpo. Ciò grazie al grande materiale sperimentale di supporto. La scoperta dei cosiddetti neuroni canonici, ad esempio, ha mostrato come una classe di neuroni risponda all'osservazione di oggetti le cui caratteristiche fisiche (forma, grandezza) sono intimamente correlate con il tipo d'azione "codificato" da quegli stessi neuroni. In altri termini, l'osservazione di un oggetto, pur in un contesto che con esso non prevede alcuna interazione attiva, determina l'attivazione del programma motorio che s'impiegherebbe se si volesse interagire con esso. Vedere un oggetto e

[26] Per questa sintesi sulle caratteristiche della *4E cognitive science* cfr. F. Caruana, A. Borghi, *Il cervello in azione. Introduzione alle nuove scienze della mente*, Il Mulino, Bologna 2016. Per una panoramica più estesa si può vedere il monumentale volume di A. Newen, L. De Bruin, S. Gallagher (a cura di), *The Oxford Handbook of 4E Cognition*, Oxford University Press, Oxford 2018.

riconoscerne la forma, in altre parole identificarlo, significa evocare automaticamente cosa faremmo con quell'oggetto[27].

Su tali basi, uno dei filosofi della mente contemporanei più influenti, Alva Noë, ha sostenuto che il *locus* fisico dell'esperienza mentale, il veicolo fisico della sua realizzazione, non è certamente il solo cervello, se isolatamente considerato, ma coincide col processo in cui l'attività neurale è incorporata nelle dinamiche senso-motorie che coinvolgono il soggetto e lo spazio che lo circonda: «Ciò che determina il carattere della nostra esperienza – ovvero quel che rende la nostra esperienza proprio quel tipo di esperienza che è – non è l'attività neurale nel nostro cervello; piuttosto, è la relazione dinamica che intratteniamo con gli oggetti»[28]. Niente di più lontano dall'ipotesi sostenuta da un certo neurocentrismo, proprio della SCC, che fa del cervello, con sfumature diverse a seconda degli autori, l'occorrente esclusivo per l'emergere della nostra esperienza cosciente. Non basta forse, sentenzia il neurocentrismo, una semplice stimolazione diretta delle cellule nervose per causare una certa corrispondenza nell'esperienza soggettiva? Questa banale operazione da laboratorio non è sufficiente a dimostrare che il "costruttore" del mondo esperito dai soggetti è il cervello? Noë rifiuta questa impostazione, usando quale campione significativo della tesi da decostruire, le posizioni dello scienziato Francis Crick:

> L'assunzione posta alla base di buona parte dell'indagine scientifica dedicata alla coscienza consiste nel considerare quest'ultima alla stregua di qualsiasi altro fenomeno neuroscientifico. Essa accade dentro di noi, nel cervello. […] Francis Crick, vincitore del premio

[27] Cfr. G. Rizzolatti, C. Sinigaglia, *So quel che fai. Il cervello che agisce e i neuroni specchio*, Cortina, Milano 2006, pp. 35-46. Per approfondire: V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, *Action Recognition in the Premotor Cortex*, in «Brain», 119, 1996, pp. 593-609; A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, G. Rizzolatti, *Object Representation in the Ventral Premotor Cortex (Area f5) of the Monkey*, in «J. Neurophysiol.», 78, 1997, pp. 2226-2230; V. Gallese V., G. Lakoff, *The Brain's Concepts: The Role of the Sensory-motor System in Reason and Language*, in «Cognitive Neuropsychology», 22, 2005, pp. 455-479.

[28] A. Noë, *Perché non siamo il nostro cervello. Una teoria radicale della coscienza* (2010), tr. it. Raffaello Cortina, Milano 2010, p. 63

> Nobel e co-scopritore della struttura della molecola di DNA, ha affermato (in un libro dal titolo *The Astonishing Hypothesis*, letteralmente "L'ipotesi sorprendente", reso in Italiano con *La scienza e l'anima*, ndr): "Tu, la tua gioia e i tuoi dolori, i tuoi ricordi e i tuoi progetti, il tuo senso di identità personale e libero arbitrio siete di fatto nient'altro che i comportamenti di un vasto sistema costituito da cellule nervose e dalle molecole che le compongono"[29].

A differenza della posizione ora riportata, per Noë, invece – che stiamo citando quale voce significativa tra le tante possibili del panorama dell'esternalismo – il cervello e il sistema nervoso, nella misura in cui rendono possibile la consapevolezza dell'ambiente per un soggetto, non hanno il compito di generare immagini interne elaborando la molteplicità dei dati sensibili; piuttosto, favoriscono, insieme al nostro corpo, l'interazione dinamica con l'ambiente: gli stati di coscienza o mentali sono la risultante dell'infinita e continua declinazione dell'essere nel mondo del soggetto cognitivo. Nel quadro dipinto dal filosofo americano, dunque, l'esperienza del soggetto è l'infinita riconfigurazione a carattere pragmatico del mondo[30]. Sulla base di quanto detto fin qui, allora (che naturalmente non ha potuto che essere una descrizione compendiata del complesso panorama cui ci si sta richiamando), possiamo sintetizzare la differenza fondamentale di approccio teorico tra la SCC e la NSC nel modo seguente: per la SCC la relazione su cui concentrare la ricerca è quella tra la mente e il cervello; per la NSC la relazione da indagare è quella tra l'esperienza mentale e il *mondo*, includendo in quest'ultimo il corpo-in-azione del soggetto. Come già anticipato, il panorama è molto variegato, e si va da forme di *esternalismo soft*, che guardano a ciò che è esterno al sistema nervoso come funzionalmente costitutivo per il sorgere dell'esperienza, fino a proposte *hard* o radicali, che concepiscono

[29] *Ibid.*, p. 5.
[30] Cfr. A. Noë, *Varieties of presence*, Harvard University Press, Cambridge (MA) 2012.

la mente come ontologicamente estesa o distribuita su tutti i "supporti" del processo cognitivo[31].

5. *Un nuovo "metodo"*

Abbiamo potuto vedere fin ad ora, speriamo con la dovuta chiarezza e fondatezza, come sul piano più generale, il tema del "realismo" sia tornato di moda in filosofia – e con esso l'importanza del "mondo" quale fondo inemendabile del gioco delle interpretazioni e delle prospettive – e, quasi contemporaneamente, anche nell'ambito degli studi sulla mente, a poco a poco, il "mondo" sia tornato protagonista: la metafora della mente-computer è stata sempre più criticata in favore di un modello maggiormente "integrato" mente-ambiente, teso a valorizzare l'interazione tra soggetto e mondo circostante in ordine alla spiegazione dell'esperienza mentale. Si tratta ora di riprendere il nostro iniziale riferimento a *Matrix* e a *Don't look up* per provare a giustificarne la citazione in questa sede.

Orbene, l'ipotesi di partenza della quadrilogia *Matrix* era basata proprio su un'idea di coscienza *inside-out* ("la coscienza è ciò che accade dentro la scatola cranica"). Ricordiamo ancora una volta le parole che Morpheus rivolge a Neo, in una delle scene cult del film del 1999: «Che vuol dire reale? Se ti riferisci a quello che percepiamo, a quello che possiamo odorare, toccare e vedere, quel reale sono semplici segnali elettrici interpretati dal cervello». Si tratta, in fondo, dell'ipotesi di Francis Crick criticamente citata da Noë[32]. Come abbiamo visto, però, tale prospettiva oggi risulta più difficile da sostenere. Ancora Alva Noë ha potuto sostenere che «l'idea secondo la quale noi saremmo

[31] Più che di esternalismo sarebbe meglio parlare di esternalismi, al plurale, data la varietà degli indirizzi. Per una panoramica si veda ancora A. Newen, L. De Bruin, S. Gallagher (a cura di), *op. cit.*, oppure M.C. Amoretti, *La Mente fuori dal Corpo. Prospettive esternaliste in relazione al mentale*, FrancoAngeli, Milano 2011.

[32] Su tale posizione, che noi definiamo "neurocentrica", si veda ad esempio P. S. Churchland, *L'io come cervello* (2013), tr. it. Raffaello Cortina, Milano 2014.

il nostro cervello non è qualcosa che gli scienziati hanno imparato; si tratta, piuttosto, di un pregiudizio che gli scienziati si portano da casa fin dentro i laboratori. Passiamo la nostra vita con un corpo, all'interno di un ambiente insieme ad altri. Non siamo semplici recipienti in balìa di influenze esterne; siamo dinamicamente accoppiati con il mondo, non siamo separati da esso»[33]. Ora, se nella distopia di Matrix "svegliarsi" vuol dire rendersi conto del carattere fittizio della realtà che abbiamo sotto il nostro naso, in *Don't look up*, di contro, il percorso da compiere è quasi opposto: qui sono proprio i fatti immediati che vanno assolutamente recuperati. La vicenda inscenata dal film, infatti, parla di una enorme cometa che viaggia verso la Terra e minaccia di distruggerla. Basterebbe alzare lo sguardo e vedere. Eppure un dedalo di interessi tanto egoistici quanto demenziali, che intrecciano economia, politica e comunicazione, si attorcigliano su sé stessi in un rovinoso collasso epistemologico (e sociale), conducendo l'umanità, caratterizzata da una buona dose di isteria e narcisismo, alla catastrofe finale. Se in *Matrix*, dunque, l'umanità era intrappolata dentro una realtà virtuale fabbricata da altri (le macchine), in *Don't look up* l'uomo stesso è il genio maligno che si autoreclude in una prigione in cui le più elementari datità sensoriali – perfino una enorme cometa – diventano invisibili. Se l'uomo in *Matrix* non doveva fidarsi dei propri sensi, in *Don't look up* il dramma descritto dalla storia consiste invece proprio nel fatto che le persone hanno "perso i sensi", hanno cioè perduto la connessione col mondo, la relazione con l'insieme dei fatti e la loro dura *indipendenza* rispetto ai discorsi.

A nostro avviso, è possibile descrivere la mutazione di sguardo intercorsa tra i due film, distanziati di circa vent'anni, come un passaggio, *lato sensu*, dall'internalismo all'esternalismo: in Matrix lo sguardo è alquanto neurocentrico (il cervello è

[33] A. Noë, *Perché non siamo il nostro cervello*, cit., p. 215.

concepito come una *magic box* che costruisce la realtà); in *Don't look up*, invece, l'attenzione cade sul rapporto tra percezione e mondo, sulla correttezza o scorrettezza di tale rapporto, e sull'inemendabilità della realtà. *Matrix* era sostanzialmente un film costruttivista (anche se non riesce a essere sempre coerente con tale impostazione), *Don't look up* è realista, nel senso proprio del realismo filosofico: l'umanità vive una realtà dotata di un certo ordine indipendente dall'intelletto. Il punto fondamentale su cui il film consapevolmente o inconsapevolmente conduce a riflettere, è la posta in gioco sociopolitica o eticopolitica del *riconoscimento* del reale. Detto in altri termini: pur se non separabile da contesti, narrazioni, influenze di potere, limiti del linguaggio ecc., non si può rinunciare al reale e alle sue esigenze, pena il disastro. Ciò ci sembra alquanto in linea con le già sottolineate preoccupazioni di carattere politico (è chiaro che tale termine va inteso in accezione larga, come tutto ciò che ha a che fare col nostro vivere associato) del nuovo realismo, espresse, tra gli altri, da Maurizio Ferraris. Le parole del filosofo italiano ci consentono di reperire alcuni dei motivi che, a parere di chi scrive, hanno causato o perlomeno favorito il "recupero del mondo" osservabile nel dibattito contemporaneo (filosofico e non), e simbolicamente condensato nel passaggio tra due film di cui si sta parlando. *Don't look up*, infatti, risente del tema delle *fake news* e del problema della verità nella comunicazione massmediatica. Di più ancora, riflette il clima di generale sfiducia e sospetto verso fonti o istituzioni ufficiali, fenomeno che a sua volta aiuta la radicalizzazione dell'opinione pubblica e la disinformazione[34]. Tali questioni, naturalmente, esistevano già nell'anno di nascita di *Matrix*, ma non si ponevano con la stessa drammatica urgenza di

[34] Il *World Economic Forum* nel 2013 ha inserito la disinformazione nella lista delle principali minacce globali. Sul rapporto critico social, comunicazione pubblica e democrazia, si veda W. Quattrociocchi, A. Vicini, *Liberi di crederci. Informazione, internet e post-verità*, Codice Edizioni, Torino 2018.

oggi. Così, dopo un lungo giro, la questioni della verità e della realtà – considerate quasi "reazionarie" in certi ambienti postmoderni più radicali – sono tornate centrali, diremmo pressanti. Descrivendo lo stato attuale delle nostre democrazie in ordine a questi problemi, la frammentazione e la polarizzazione del loro dibattito, il filosofo Raffaele Alberto Ventura ha asserito che

> oggi è come se tutti avessero gli occhiali di John Nada, il protagonista del film [*Essi vivono* di John Carpenter], o ingoiato la pillola rossa di *Matrix*. Per alcuni coincide con un libro sacro, un blog alla moda, un predicatore televisivo, ognuno con la sua narrazione alternativa capace di riempire i vuoti di senso lasciati dal sapere detto "legittimo" e considerato di fatto sempre meno "legittimo". Non sappiamo se quello che abbiamo visto sotto la superficie delle cose sia vero o falso, sappiamo soltanto che fornisce una spiegazione a una realtà sempre più confusa[35].

Se quanto detto fin qui è fondato, possiamo sostenere che – e questo è il punto di portata propriamente filosofica – nell'èra della cosiddetta post-verità, si ravviva l'esigenza di ritrovare il rapporto tra la *ragione*, intesa come strumento conoscitivo, e le *ragioni*, intese come il senso e la verità dei fenomeni. Nei periodi storici di epocale transizione, e non v'è dubbio che gli anni in cui viviamo siano tali, tale rapporto si spezza. Ora occorre trovarne uno nuovo. Si ripropone così l'esigenza di un nuovo *metodo*, vale a dire di una rinnovata via (*odos*, strada) di accesso alle cose. Le cose ci circondano, ci assediano da ogni lato, eppure dobbiamo aprirci un varco verso di esse elaborando un metodo, altrimenti restano come invisibili: la nostra esperienza sarà una kantiana intuizione cieca, come quella dei personaggi di *Don't look up*, incapaci di vedere ciò che erano perfettamente in grado di guardare.

Alfonso Lanzieri insegna Logica presso la Facoltà Teologica di Napoli (sez. San Luigi) e Filosofia teoretica presso l'Istituto Superiore di Scienze Religiose di Nola-Acerra

indialogo.lanzieri@gmail.com

[35] R.A. Ventura, *La guerra di tutti. Populismo, terrore e crisi della società liberale*, Minimum fax, Roma 2018 (edizione kindle), pp. 6-7.

S&F_n. 27_2022

ANTROPOLOGIE

MAURO MANDRIOLI

FROM GENOME EDITING TO HUMAN GENETIC ENHANCEMENT: A NEW TIME FOR DISCUSSING EUGENICS?

1. *Introduction*
2. *Different eugenics in diverse countries and times: the concept of eugenics in an historical context* 3. *A social eugenic view: a Latin perspective on eugenic measures*
4. *Toward human enhancements*

ABSTRACT: *FROM GENOME EDITING TO HUMAN GENETIC ENHANCEMENT: A NEW TIME FOR DISCUSSING EUGENICS?*
In the last decade several Authors strongly suggested that genome editing technologies are just related to gene therapy rejecting any eugenic discussion about them. Actually, this dispute is mainly influenced by the previous terrific consequences of eugenic measures rather than by the real consequences that genome editing technologies could have on the future societies. Here we discuss genome editing technologies as eugenic approaches in order to assess that if we continue to rebrand them as therapies, we will simply favor their market in place of their true social acceptance. At the same time, these technologies also strengthened a possible use of genetic manipulation for the human genetic enhancement, making some previous eugenic goals technically feasible. As a consequence, it is timely to discuss which kind of human enhancements we would like to accept/adopt and who will be the main drivers (parents, medical doctors, politicians, scientists, …) of this change that will affect our future offspring.

1. *Introduction*

In the last two decades human molecular genetics made remarkable advances not only in developing new high-throughput techniques for DNA sequencing (Next-Generation Sequencing, NGS), but also in term of understanding of the biological bases of several human diseases[1]. We at present know, for instance, in great details mechanisms underlying the cause of many human diseases and we are aware that about the 65% of people

[1] R.A. Gibbs, *The Human Genome Project changed everything*, in «Nature Review in Genetics», 21, 2020, pp. 575–576.

has some kind of health problem as a result of congenital genetic mutations and one in five healthy adults carries disease-related genetic mutations[2]. At the same time, the availability of numerous bioinformatic tools favoured computational biology so that now we can easily access, harness and analyse molecular data[1].

As a consequence of these changes, a spectacular series of projects made available huge amounts of data to further improve our knowledge of the human genome. For instance, the *1000 Genomes Project* provided a deep characterization of the human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype[3], whereas *The Cancer Genome Atlas* generated a vast and comprehensive dataset describing the molecular changes that occur in cancer[4].

The combined use of NGS, bioinformatic facilities and big data in human genetics favoured further developments that allow us nowadays to obtain human genome sequences at costs lower than US$1,000 per genome favouring the development of personalized therapies and the liberation of drugs that otherwise were unusable, through identification of the few individuals with adverse responses[1].

In a way that is unprecedented in the history of life sciences, multiple editing technologies have been developed in the last decade with improved specificity, accuracy and efficiency[2]. In particular, CRISPR has revolutionized the genome editing because it is as robust as simple and flexible to use and several more precise and super-efficient CRISPR-Cas systems have been recently derived from engineered variants, various bacterial species and distinct classes of CRISPR targeting mechanisms[5].

---

[2] M. Mandrioli, *Genome editing among bioethics and regulatory practices*, in «Biomolecules», 12, e13, 2022, https://doi.org/10.3390/biom12010013.

[3] The 1000 Genomes Project Consortium, *A map of human genome variation from population-scale sequencing*, in «Nature», 467, 2010, pp. 1061–1073.

[4] C. Hutter, J.C. Zenklusen, *The Cancer Genome Atlas: creating lasting value beyond its data*, in «Cell», 173, 2018, pp. 283-285.

[5] M. Adli, *The CRISPR tool kit for genome editing and beyond*, in «Nature Communications», 9, 2018, e1911, https://doi.org/10.1038/s41467-018-04252-2.

The concomitant availability of personal genomes and editing technologies favoured the beginning of several clinical trials aimed at the control of some genetic diseases, making genome editing a legitimate and ethically justifiable tool[2]. Actually, the potential applications of genome editing do not stop with the prevention and treatment of human disease, since these technologies also strengthened a possible use of genetic manipulation for the human genetic enhancement, making some dormant eugenic goals technically feasible[6].

Genome editing, indeed, represents an intentional and deliberate modification of the human genome to "improve" individuals in term of elimination of mutations related to specific genetic diseases. However, these technologies may quickly move from disease prevention and therapy to attempts to modify and "enhance" human physical and intellectual traits. Even if controversial[2], several experiments showed that it is possible to reduce the number of children born with genetic diseases and disorders by germinal genome editing according to the past negative eugenics[7,8].

By the late 1920s, negative eugenics favoured in several countries the passage of laws allowing involuntary sterilization programs aimed at preventing the birth of human beings with specific heritable diseases[12,13]. As a consequence of these programs, more than 60,000 people were sterilized in USA between 1935 and 1975, whereas over 62,000 people were sterilized in Sweden. In the same years, about 360,000 people were sterilized in Germany, 58,000 in Finland, 41,000 in Norway and 11,000 in Iceland[9].

Involuntary sterilization programs, particularly those in the United States and in Sweden, were the major incubators of even harsher social programs and, in their extreme forms, of the

[6] T. Friedmann, *Genetic therapies, human genetic enhancement, and eugenics?*, in «Gene Therapy», 26, 2019, pp. 351-353.
[7] D. Kevles, *Eugenics and human rights*, in «British Medical Journal», 319, 1999, pp. 435-438.
[8] Id., *In the name of eugenics*. Harvard University Press, Cambridge, MA 2007.
[9] E. Asbrink, *Made in Sweden*, Scribe, London 2019.

genocidal programs of Nazi Germany that is immediately in our mind as soon as the term eugenics is cited.

Even if we are aware that eugenics has been used during the Twenty Century to deliberately control several unrelated human traits ranging from criminality to mental disorders, laziness and homosexuality, and that it was offered as justification for several deeply morally reprehensible acts, we still think that it could be useful to distinguish the concept of eugenics from the used measures[10]. It could be therefore useful to revise the concept of eugenics in an historical perspective in order to understand why genome editing can be seen as modern eugenics. In particular, it could be important to discuss the present role that the objectionable anachronism of eugenics could have in understanding the latest developments in twenty-first century genetics. Indeed, even if scientists would probably immediately reject the idea that the improvements that could result from editing human embryos may be labelled as eugenic, it is important to remind the moral lessons that the history of eugenics may teach us nowadays. As stated by Agar, «we may best learn these lessons when we acknowledge many of the uses that are proposed for gene editing as eugenics»[19].

2. *Different eugenics in diverse countries and times: the concept of eugenics in an historical context*

The term eugenics has been originally coined in 1883 by Francis Galton combining the Greek *eu-*, meaning "good" or "well," with *-genēs*, meaning "born"[11]. In his 1883 book *Inquiries into Human Faculty and Its Development*, Galton presented eugenics as «the science of improving stock, which is by no means confined to questions of judicious mating, but which, especially in the case

[10] N. Agar, *Why we should defend gene editing as eugenics*, in «Cambridge Quarterly of Healthcare Ethics», 28, 2019, pp. 9-19.

[11] F. Galton, *Inquiries into Human Faculty and its Development*, Macmillan, London 1883.

of man, takes cognizance of all influences that tend in however remote a degree to give to the more suitable races or strains of blood a better chance of prevailing speedily over the less suitable»[16].

This definition contains numerous objectionable concepts, but the analysis of the Galton's proposal could help to better focus the discussion also on current editing technologies. Indeed, in the Galton's view, eugenics represents the choice of adopting solutions useful to lead to higher levels of health in the population in which it is implemented, objective that is highly similar to the goals of genome editing technologies. At the same time, if performed at germinal level, genome editing will be useful to control the inheritance of genetic disease in the same way as Galton's eugenics was focused on inherited influences.

As also discussed by Agar[12], the approaches that practitioners of public health used in the past are not synonyms of eugenics. The main goal of eugenics is that population health can be improved by the manipulation of hereditary influences of diseases and this may be morally problematic but not wrong. On the contrary, the practice of eugenics in the early twentieth century (i.e. sterilization), not eugenics itself, offers many examples of morally wrong attempts to improve population health.

Modern societies are trying to improve their health reducing the incidence of cancer due to smoking or alcohol consumption and similar results could be obtained, for instance, editing Brca1 and Brca2 genes in people with mutated alleles and high cancer risks. Accepting the eugenic nature of genome editing technologies is therefore useful since we should expect that each decision about the kinds of people we want to have in future societies will remain philosophically vexing. Indeed, there will be debate about how a society may use genome editing since these advances can be viewed as morally wrong eugenic acts. Indeed, as suggested by

[12] N. Agar, *What was right about eugenics*, in «*Medical Ethics*», 2015, pp. 4–5.

Agar, «the history of attempts to use information about heredity to choose the attributes of future citizens contains some of the twentieth century's signal crimes. An awareness of these crimes must be at the forefront of the minds of those who make decisions about how genetic technologies should be used»[15].

We have to be aware that popular imagination tends to rapidly place eugenics together with the other offences of the Nazis and the crimes they committed in the name of eugenics are terrible. However, it is better to discuss this obscure connection between eugenics and Nazis in place of thinking that simply rebranding genome editing technologies as gene therapies is sufficient to make them morally accepted. The assumption that genome editing is an eugenic practice will also improve future discussion demonstrating that scientific community is interested into serious attempts of grappling with the wide range of moral considerations related to these practices.

3. *A social eugenic view: a Latin perspective on eugenic measures*

Differently from the Anglo-American, German and Scandinavian world, where the supporters of eugenics asked for the assumption of measures aimed to avoid the reproduction of "undesirable" individuals by sterilisation, abortion and segregation[13,14,15,16], in Latin countries (such as Italy, Spain and Portugal) several scientists and politicians suggested the adoption of measures involving preventive social medicine and, more generally, health

[13] P.M.H. Mazumdar, *Eugenics, human genetics and human failings. The Eugenics Society, its source and its critics in Britain*, Routledge, London 1992.
[14] E. Black, *War against the weak. Eugenics and America's campaign to create a master race*, Four Walls Eight Windows, New York 2003.
[15] H. Friedlander, *The origins of Nazi genocide. From euthanasia to the final solution*, University of North Carolina Press, Chapel Hill 1995.
[16] N.G. Broberg, N. Roll-Hansen, *Eugenics and the Welfare State. Sterilization Policy in Denmark, Sweden, Norway and Finland*, Michigan State University Press, East Lansing 1996.

and hygiene programmes aimed at the creation of environments more favourable to the reproduction of desirable individuals[17,18,19].

Even if these two conceptions of eugenics have in common the right and duty of National States to guide and manage reproductive processes of their citizens, several Italian scientists were convinced of the impracticability of sterilisation, branded as immoral by public opinion[26]. This suggestion is very intriguing since it put in evidence the different acceptance of some eugenic practices in relation to the religious context. Indeed, the use of sterilization was condemned by the Catholic Church that, in fact, did not agree to direct interferences in the reproductive processes except for protecting maternity. In particular, the encyclical *Casti Connubii* strongly condemned not only sterilisation, but also abortion[26].

Among the several scientists that in Italy disputed about eugenics, Giuseppe Sergi has been one of the most active proponents of social improvements in order to have a diffuse health in the Italian society. Moving from physical and evolutionary anthropology to the study of psychology, criminology and sociology, Sergi has been one of the greatest exponents of positive eugenics in Italy at the turn of the 19th and 20th centuries. He also supported the identification of biological and environmental factors that could be utilised socially to prevent degenerative pathologies and thus favouring the reproduction and perfecting of desirable individuals[26].

In 1893 Sergi founded the "Società Romana di Antropologia" and from this organization in 1913 had its origin the "Comitato Italiano per gli Studi di Eugenetica" (Italian Committee for

[17] M. Adams, *The wellborn science. Eugenics in Germany, France, Brazil and Russia*, Oxford University Press, Oxford 1990.
[18] M. Turda, A. Gillette, *Latin eugenics in comparative perspective*, Bloomsbury Academic Press, London 2014.
[19] L. Tedesco, *Latin and Nordic eugenics in the project of racial improvement set up by Giuseppe Sergi, founder of the Comitato Italiano per gli Studi di Eugenica*, in «SIDeS - Popolazione e Storia», 1, 2016, pp. 35-52.

Eugenic Studies), aimed at the study of factors that could improve the quality of future generations and the definition of relevant eugenic measures[26].

The support of Sergi and his activities have been very relevant in Italy and the proposal of the Italian Committee for Eugenic Studies are probably at the basis of the Fascist regime choice of embracing pro-natal eugenic lines, being convinced that the strength of a nation and the racial betterment were closely linked to fertility increase[26,20].

The choice of re-discussing the work of Sergi is related to his significant analysis of eugenics that moved from a support to negative eugenics in his publication *Le degenerazioni umane*[21] to the suggestion that degenerative influences derived from the external environment and that could undermine the normal development of the individual's organic functions. This conviction persuaded him to ask constantly to public authorities the adoption of measures aimed at preventing the spread of degenerative influences favouring social hygiene[22].

In *Le degenerazioni umane*, Sergi described several cases of hereditary defects, including insanes, suicides, criminals, prostitutes, slaves and menials, vagrants and beggars and according to other Italian eugenicists, such as Enrico Morselli, Augusto Tamburini and Paolo Mantegazza, he thought that degeneration had an organic base[26,30,23]. At the same time, he moved from the conviction that there was no remedy for the serious degenerative phenomena that were already taking place. As a consequence, Sergi called initially for drastic negative eugenics

[20] M.S. Quine, *Racial «sterility» and «hyper-fecundity» in Fascist Italy. Biological Politics of Sex and Reproduction,* in «Fascism», 1, 2012, pp. 92-144.
[21] G. Sergi, *Le degenerazioni umane*. Fratelli Dumolard, Milano 1889.
[22] L. Rossi, *Il problema delle degenerazioni umane nell'antropologia psicologica di Giuseppe Sergi*, in *Giuseppe Sergi nella storia della psicologia e dell'antropologia in Italia*, a cura di G. Mucciarelli, Pitagora Editrice, Bologna 1987, pp. 63-81.
[23] A. Burgio, *Nel nome della razza. Il razzismo nella storia d'Italia. 1870-1945*, Il Mulino, Bologna 1999.

measures for degenerates. However, in subsequent years, Sergi focused his attention on elementary, vocational and secondary schooling suggesting that education was an indispensable tool for the perfecting of healthy individuals and the prevention of environmental degenerative pathologies[26,24]. At the same time, he became convinced of the impracticability of both sterilization, branded as immoral by the greater part of public opinion[25], and segregation, that was rejected because it was not a long-term solution and its real effects were still unknown[32].

Sergi, as a whole, suggested that culture was not a mere cognitive patrimony reached by a society, but it was the basis of a civilization process aimed at increasing general, moral and non-moral well-being. The Latin approach to eugenics clearly therefore assesses that science gives us the power to do choices, it doesn't tell us what we have to do. It is therefore necessary to educate future generations to exercise critical thinking and to ask themselves which kind of societies they want in the future and which tools they want to use to build them.

4. *Toward human enhancements*

In the last years some deliberate applications of genetic enhancement to individual human beings have been suggested and, in some cases, also done[26]. For instance, the birth of two girls genetically protected from HIV (that took place at the Southern University of Science and Technology in Shenzhen in China under the editing experiments of the scientist He Jiankui) is not actually a true disease prevention in humans, but it is an

[24] G. Cerro, *Giuseppe Sergi, The portrait of a positivist scientist*, in «Journal of Anthropological Sciences», 95, 2017, pp. 109-136.
[25] L. Tedesco, *«For a healthy, peace-loving and hardworking race»: anthropology and eugenics in the writings of Giuseppe Sergi*, in «Modern Italy», 16, 2011, pp. 51–65.
[26] T. Peters, *Are we closer to free market eugenics? The CRISPR controversy*, in «Zygon», 54, 2019, pp. 7-13.

enhancement[27]. Jiankui indeed knocked out the CCR5 gene blocking the HIV virus to enter the immune cells. The deliberate editing to have babies resistant to future possible diseases is of course very intriguing since it favours the development of healthy societies, but are we sure that these approaches are really useful and necessary?

Several researchers stated that HIV protection is strictly therapeutics, but AIDS is not a genetic disease that we can certainly develop in case of mutations. If we consider this approach as a disease treatment, we should be ready to edit people not only for the 4,000-6,000 human diseases traceable to genetic factors, but also for infection risks. As a consequence, it could be suggested to edit people to make them less susceptible also to SARS-cov2 virus, considering that a genetic susceptibility to COVID-19 has been described[28,29]. At the same times, since genes involved in inherited forms of obesity have been discovered[30], can we plan editing trials to reduce the risks associated to obesity? How can we define this approach: enhancement or therapeutic approach?

According to an interesting proposal of Gouw[31], genetic therapy refers to the manipulation of the genome to treat individuals or their progeny with known diseases, disabilities, or impairments to restore them to a normal state of health. Differently, enhancement refers to the use of genetic alteration, pharmaceuticals, devices,

[27] H.T. Greely, *CRISPR's babies: human germline genome editing in the 'He Jiankui affair'*, in «Journal of Law and the Biosciences», 6, 2019, pp. 111-183.
[28] S. Seyedalinaghi, M. Mehrtak, M. MohsseniPour, P. Mirzapour, A. Barzegary, P. Habibi, B. Moradmand-Badie, A.M. Afsahi, A. Karimi, M. Heydari, E. Mehraeen, O. Dadras, J.M. Sabatier, F. Voltarelli, *Genetic susceptibility of COVID-19: a systematic review of current evidence*, in «European Journal of Medical Research», 26, 2021, pp. 1-46.
[29] A. Nasrolahi, K. Haghani, A. Gheysarzadeh, S. Bakhtiyari, *Do genetic factors predispose people to COVID-19: a review article*, in «Current Molecular Medicine», 21, 2021, pp. 457-461.
[30] H. Choquet, D. Meyre, *Genetics of obesity: what have we learned?*, in «*Current Genomics*», *12*, 2011, pp. 169-179.
[31] A.M. Gouw, *Challenging the Therapy/Enhancement Distinction in CRISPR Gene Editing*, in *The Palgrave Handbook of Philosophy and Public Policy*, a cura di D. Boonin, Palgrave Macmillan, New York 2018, pp. 493–508.

or other means to alter the normal workings of the human body or psyche or to be better than what is normal and native to healthy physiology. The main problem is, therefore, that the line between therapy and enhancement is not easy to draw and this means that legislation will have to face several troubles to approve therapy, but disapprove enhancement, since clinicians could simply redefine enhancement as therapy in order to make it available.

Furthermore, the availability of possible enhancements (even if also in term of risk reduction toward some diseases) could be a problem in term of equity since we have to avoid a future in which privileged people become stronger, smarter and healthier than the rest. But if differences between treatment and enhancement are subtle, the more likely and more unsettling scenario is that physicians will be left to rely on their own ethical commitments to decide when to use gene therapy enhancing the spectre of inequality[32]. As also suggested by Brokowski and colleagues[33], it is plausible that editing technologies could exacerbate social inequalities since families with the higher income brackets will have access to methods of creating "enhanced" children, leaving those with more modest means to the fates the genetics. The risk of creating a *de facto* tiered biological class system, perhaps resulting in even greater inequalities than those existing now, is therefore very high.

Can we therefore ask for a ban of editing if related to enhancement? Legal prohibition of some CRISPR-Cas9 editing protocols may be useful (for instance if related to germline editing), but it may also limit knowledge, experimentation and overall progress gleaned from these technologies.

In several countries, embryos are created by *in vitro* fertilization and they can be tested for the presence of genetic

[32] E.W. Clayton, *A genetically augmented future*, in «Nature», 564, S9, 2018, DOI: 10.1038/d41586-018-07642-0.

[33] C. Brokowski, M. Pollack, R. Pollack, *Cutting eugenics out of CRISPR-Cas9*, in «Ethics in Biology, Engineering and Medicine: an International Journal», 6, 2015, pp. 263-279.

variants that elevate the risk of a disease that parents might seek to avoid (Preimplantation Genetic Diagnosis, PGD). Embryos free of variants associated with the elevated risk can be used to initiate a pregnancy. In societies in which PGD is made available to individuals at risk of passing on a genetic disease, this approach can improve population health, clearly assessing that not only editing but also PGD is a eugenic approach since it is aimed at controlling which genes are passed down to future generations[34].

Undoubtedly, eugenics lurks in the shadow of CRISPR and we have to admit that we are obliged to put again eugenics in our agenda. However, if we learn from errors of the past, we can find a more rational approach to eugenics opening the discussion about what we want to do, in place of spending our time trying to convince ourselves that editing is not related to eugenics.

Eugenics is all about controlling the "best" genetic traits that have to be passed on to future generations[35,36]. In the past eugenics, National States made decisions about the "better" traits for the Western societies, whereas at present parents and doctors seem to be the main drivers of this choice. Indeed, we are allowing parents to decide about abortion and about the use of pre-implant genetics.

The concerns about eugenic outcomes are amplified one-thousand-fold with genome editing, which involves the direct alteration of the genome in order to "improve" the quality of offspring. In this context, even if it is legitimate for parents to aspire to an improved offspring, we have to discuss if it is acceptable at a bioethical level that parents could start to create their own

[34] D. King, *Preimplantation genetic diagnosis and the «new» eugenics*, in «Journal of Medical Ethics», 25, 1999, pp. 176–182.

[35] N. Agar, *Humanity's end: why we should reject radical enhancement*, MIT Press, Cambridge, MA 2010.

[36] N. Agar, *Truly human enhancement: a philosophical defense of limits*, MIT Press, Cambridge, MA 2013.

"enhanced" offspring making conditional parentage a *de facto* accepted eugenic measure.

Acknowledgements

This publication resulted from the research project «La biblioteca della vita», supported by Fondazione Cassa di Risparmio di Modena (Italy) (funding call *FAR Interdisciplinare Mission Oriented*).

Mauro Mandrioli insegna Genetica all'Università degli Studi di Modena e Reggio Emilia

mauro.mandrioli@unimore.it

S&F_n. 27_2022

ETICHE

DELIO SALOTTOLO

È DAVVERO NECESSARIA UN'ETICA DELLE PIANTE?

1. *Le ragioni di una domanda* 2. *È eludibile l'antropocentrismo?*
3. *L'alterità assoluta delle piante* 4. *Le ragioni di una risposta*

ABSTRACT: *IS PLANT ETHICS REALLY NECESSARY?*
This article aims at analysing the possibility and the need for a plant ethics within the field of applied ethics. Starting with an analysis of the symbolic and cultural representations of "nature" and with the connection with the material dimension of human relationality, it will move on to a speculative and synthetic analysis of the fundamental categories of environmental ethics (serving as a model for animal ethics and plant ethics), and then to renewed ontological representations that vegetal life can "show", concluding with some propositions from which it is possible to answer positively to the question: *is plant ethics really necessary?*

ἀλλ'ὅσσα περ ζῇι, πάντα καὶ γνώμαν ἔχει
Epicarmo

1. *Le ragioni di una domanda*

È una questione di presa di coscienza ex-post e di emergenza di problemi che sembrano divenire sempre più ingestibili. Ma è anche una questione che non può che muovere da un assunto: riconoscere *dignità* (qualunque cosa essa significhi – non è importante adesso) a un determinato ente, nel caso più lampante: l'umano, non ha condotto necessariamente a un rispetto dell'uomo da parte dell'uomo, neanche in quella specifica parte di mondo che è stata in grado di pensare qualcosa come la dignità umana e l'universalizzazione dei diritti. Considerazione morale, diritti umani e positivi, partecipazione etico-politica alla costruzione del comune: a questi dispositivi non tutti gli umani possono ancora accedere e non sembra che stiamo vivendo tempi promettenti affinché questo allargamento si produca. Eppure, tutta la vicenda

complessa della costituzione di un'etica ambientale[1], di un'etica animale[2] e adesso di una possibile etica delle piante[3] rientra all'interno di una dinamica che intende allo stesso tempo allargare la considerazione morale a enti naturali *altri* (gli animali, certamente, ma anche le piante, e addirittura "oggetti" dalla complessa, se non impossibile, definizione epistemologica e ontologica come l'ecosistema, la biodiversità, etc.) e inchiodare l'Occidente e la sua cultura a determinate *responsabilità morali*, in questo caso nei confronti della Natura. Utilizziamo il termine Natura, almeno in questo contesto iniziale, volutamente con la iniziale maiuscola: non si tratta di definire preliminarmente cosa debba intendersi con il termine "natura", non affrontiamo in questo contesto questioni di filosofia biologica o di biologia filosofica, ma di delineare i contorni di una *utilizzazione culturale e simbolica* di una determinata rappresentazione al fine di costruire determinate partizioni all'interno del reale, partizioni funzionali alla costruzione e alla soddisfazione di determinate esigenze e bisogni dell'umano (nel nostro caso: occidentale). La metodologia, il cammino che decidiamo di seguire, è quello di connettere rappresentazione culturale e simbolica e determinazione materiale (economico-politica) all'interno di una "spirale": non si tratta semplicemente di reperire una struttura materiale di sfruttamento, nelle interazioni tra gli umani e tra l'umano e il mondo vivente e non vivente, tale da determinare le rappresentazioni culturali, né di cercare di dimostrare l'inverso,

[1] Per un inquadramento generale cfr. R. Attfield (a cura di), *The Ethics of the Environment*, Routledge, London and New York 2016; in lingua italiana, cfr. P. Donatelli (a cura di), *Manuale di etica ambientale*, Le Lettere, Firenze 2012; M. Andreozzi (a cura di), *Etiche dell'ambiente. Voci e prospettive*, LED, Milano 2012.

[2] Impossibile restituire la ricchezza della letteratura in merito, ma, per un'interpretazione complessiva della questione, in lingua italiana e in chiave darwiniana, cfr. S. Pollo, *Umani e animali: questioni di etica*, Carocci, Roma 2016.

[3] Per un'analisi ragionata del dibattito internazionale, cfr. A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *Plant Ethics. Concept and Applications*, Routledge, London and New York 2018; in lingua italiana, cfr. G. Pellegrino, M. Di Paola, *Etica e politica delle piante*, DeriveApprodi, Roma 2019.

come alcune rappresentazioni culturali abbiano la capacità di determinare una struttura materiale di sfruttamento nelle interazioni tra gli umani e tra l'umano e il mondo vivente e non vivente – chiedersi, nell'umano, se viene prima il bisogno o la rappresentazione (culturale e simbolica) del bisogno è una domanda mal posta: il movimento che va restituito è quello di una spirale, un movimento ricorsivo ma non eternamente uguale a se stesso, che con-cresce mediante il continuo intrecciarsi di strutture materiali e rappresentazioni simboliche a un livello di complessità sempre maggiore. In questo senso, determinanti diventano gli spostamenti, dunque gli eventi, ovverossia le soglie di problematizzazione. Perché il problema, in questo momento e all'interno delle differenti etiche dell'ambiente, animale e delle piante, riguarda proprio ciò che è andato costituendosi come Natura all'interno della tradizione di pensiero occidentale e come questa "istituzione" di una certa Natura abbia impattato e favorito una determinata struttura di sfruttamento della "natura istituita": nelle ricostruzioni culturali, si parte nella maggior parte dei casi da un'analisi storico-critica di questa attitudine di pensiero (dal "tradimento" socratico nei confronti della filosofia della natura al racconto biblico, come elementi fondativi, poi si passa alla soglia della modernità scientifica, da Galilei a Newton, e al dominio della dimensione tecnica/tecnologica nella contemporaneità) e la questione viene affrontata soprattutto dal punto di vista culturale e simbolico, vale a dire di relazione gnoseologica, ontologica e morale-individuale con l'alterità naturale; a partire dall'indubbio interesse in vista di un ripensamento della relazione etica che connette l'umano al suo mondo, la dimensione della relazione istituente con una determinata struttura materiale dello sfruttamento della Natura (di ciò che di volta in volta va a riempire questo "universale": dunque, non solo in senso "ambientalista", ma anche rispetto a una sorta di *reductio ad*

naturam cui in passato sono andati in contro le donne, gli altri etnicamente intesi, etc.) va inquadrata in tutta la sua complessità.

Pierre Hadot ha raccontato in maniera magistrale i due atteggiamenti fondamentali dell'Occidente (sin dall'antichità greco-romana) nei confronti della Natura, l'atteggiamento prometeico (svelamento violento dei segreti di una "natura" pensata come "nemica" e "ostile", di cui l'umano orgogliosamente non fa parte, almeno non del tutto) e l'atteggiamento orfico (svelamento pacifico dei segreti di una "natura" pensata come "amica" e "creativa", di cui l'umano orgogliosamente fa parte, seppur non del tutto)[4], e a partire da questo quadro si potrebbero raccontare come i due atteggiamenti siano fortemente interrelati e funzionali l'uno all'altro, si potrebbero richiamare tutte le riflessioni che connettono il dominio violento o pacifico della natura esteriore (oggi, i paradigmi del "dominio" e del "pilotaggio" come risposta alla crisi climatica) al dominio violento o pacifico della natura interiore, si potrebbero richiamare le riflessioni su come il dominio violento o pacifico della natura sia connesso al dominio violento o pacifico su ciò che di volta in volta è stato considerato, nell'umano, più vicino alla natura, dal femminile all'alterità "selvaggia", si potrebbe richiamare la maniera mediante la quale l'operatore "natura" abbia funzionato sia in senso negativo (questa o quell'altra realtà è inferiore *perché vicina alla natura, a una dimensione naturale*) o in senso positivo (questa o quell'altra realtà è superiore *perché vicina alla natura, a una dimensione naturale*) e si renderebbe conto dell'ambigua complessità nell'auto-rappresentazione dell'umano occidentale, che come un pendolo oscilla costantemente e in maniera funzionale tra questi due "opposti". E così, quello che può essere di interesse è la maniera mediante la quale

[4] Cfr. P. Hadot, *Il velo di Iside. Storia dell'idea di natura* (2004), tr. it. Einaudi, Torino 2006.

l'operatore Natura ha funzionato all'interno della rappresentazione dell'alterità, laddove nella tradizione occidentale l'alterità è sempre stata ciò che doveva essere ricondotta al Medesimo, mediante il monopolio dell'uso legittimo della forza delle armi materiali e culturali-simboliche[5]. Questa sorta di necessità di svelamento della Natura va pensato allo stesso tempo come un "bisogno" e come la "rappresentazione di un bisogno" – la "natura" è il primo spettacolo (proprio nel senso di: oggetto di visione) che si presenta all'umano: una "natura" può nascere solamente nel momento in cui si produce una determinata distanza all'interno di una prospettiva, il punto di vista del vivente umano, che non può più "vivere" l'immanenza assoluta della sua appartenenza immediata (nel senso di: senza mediazioni) al tutto vivente, e da un lato non può fare altro che provare un sentimento di rimpianto per la "perdita" già da sempre intuita, e dall'altro, trovando il mondo dinanzi a sé, sentire l'orrore e la fascinazione di un mistero angosciante, cercarne lo svelamento e sfruttarne le possibilità[6].

Non si può pensare questo "bisogno" attuale dell'umano occidentale di rielaborare la propria relazione morale con gli altri enti naturali – in questo momento sfumiamo le distinzioni e le alterità all'interno delle alterità – se non nei termini di una presa di coscienza ex-post: l'emergere di questi regimi discorsivi deve essere concepita come l'urgenza di una problematizzazione che viene allo scoperto nel momento in cui una determinata pratica,

[5] Il dibattito più "accorto" sulle questioni ambientali sottolinea come la costruzione occidentale di una "identità", allo stesso tempo come "immagine culturale e simbolica" e come motivazione per interventi colonialisti e neocolonialisti, sia lo strumento fondamentale di ogni riduzionismo dell'alterità.

[6] Su queste questioni è stato particolarmente puntuale il filosofo francese Georges Bataille che vede in questa dialettica immanenza/trascendenza l'origine non solo dell'attitudine tipicamente umana di trasformazione del mondo mediante il lavoro, ma anche della religione, dell'arte e dell'erotismo (cfr. G. Bataille, *Teoria della religione* (1973), tr. it. SE, Milano 2002; Id., *Lascaux. La nascita dell'arte* (1955), tr. it. Mimesis, Milano 2007; Id., *L'erotismo* (1957), tr. it. Mondadori, Milano 1972). Ci permettiamo di rinviare a D. Salottolo, *La vita e l'utile: la parabola della scrofa*, in «S&F_scienzaefilosofia.it», 17, 2017, pp. 177-194.

considerata fino a un certo momento "normale" e dunque "invisibile", diviene in un certo senso "patologica" e dunque "visibile". Una problematizzazione viene fuori sempre all'interno di un momento particolare della "spirale", quando una determinata relazione tra dinamica culturale e simbolica e dinamica di sfruttamento materiale raggiunge un punto critico. Questa è la soglia dell'Antropocene. E allora: *è davvero necessaria un'etica delle piante?* Si tratta della domanda dalla quale abbiamo deciso di muovere, proprio nel senso del reperire questa problematizzazione connessa a una emergenza/urgenza. Le etiche applicate al mondo naturale rappresentano, dunque, allo stesso tempo il sintomo di una problematizzazione o di una "patologia" della libertà, e la difficoltà nel reperimento di una soluzione o, meglio, di una "cura": nel momento in cui scindiamo il nostro rapporto con il mondo all'interno di comparti – ambiente, animale, pianta – e l'interesse si incentra su quale "considerazione morale" debba avere un determinato ente della natura con il quale entriamo in contatto, se questa "considerazione morale" debba basarsi su elementi che per analogia possiamo ritrovare nell'umano o invece esclusivamente nell'unicità di quell'ente, quando insomma rischiamo di disincarnare la dimensione etica dalle altre dimensioni complesse dell'intervento dell'umano nel mondo, di cosa è rappresentazione la domanda: *è davvero necessaria un'etica delle piante?* E ancora: questa impostazione che andiamo proponendo non rischia di essere essa stessa una versione dell'antropocentrismo tipicamente occidentale?

A partire da queste annotazioni introduttive – ma che sono già parte integrante del ragionamento – si passerà, a partire dalla domanda se l'antropocentrismo sia o meno eludibile, a un'analisi speculativa e sintetica delle categorie fondamentali dell'etica ambientale (che fanno da modello per l'etica delle piante), per poi approdare a rappresentazioni concettuali rinnovate che le piante possono "produrre", concludendo con alcune proposizioni a

partire dalle quali è possibile rispondere positivamente alla domanda: *è davvero necessaria un'etica delle piante?*

2. *È eludibile l'antropocentrismo?*[7]

L'etica dell'ambiente rappresenta un filone della filosofia morale, sviluppatosi a partire dagli anni '60 del XX secolo, che nasce dall'idea che sia necessario superare i limiti della rappresentazione morale classica, secondo la quale soltanto gli umani sarebbero dotati di status morale e quindi degni di considerazione morale[8]. Secondo l'impostazione per certi versi ormai "classica", la storia dell'etica occidentale può essere letta nei termini di una storia dell'estensione dei limiti della moralità, prima all'interno dell'umano, poi oltre l'umano. Il problema centrale, dunque, di ogni etica dell'ambiente è rappresentato dall'annosa questione dell'antropocentrismo. Nell'ormai vastissima letteratura, si ritrova solitamente un inquadramento antropocentrico, secondo il quale la "natura" (qualunque cosa essa indichi) ha valore soltanto nella misura in cui rappresenta un mezzo per la realizzazione di determinati fini umani: la necessità della conservazione, dunque, o sta nel fatto che la "natura" ha un'utilità determinante per le attività di crescita economica umana – posizione oramai perlopiù squalificata; oppure consiste nel fatto che essa è utile alla vita umana perché capace di suscitare atteggiamenti di ammirazione e fascinazione e quindi condurre a una "crescita" morale del soggetto umano "spettatore" – si tratta, ad esempio, della classica impostazione dei "doveri indiretti". Si ritrova poi un inquadramento non-

[7] Siamo debitori, come stimolo per la riflessione sull'antropocentrismo nelle etiche ambientali, di G. Pellegrino, M. Di Paola, *Nell'Antropocene. Etica e politica alla fine di un mondo*, DeriveApprodi, Roma 2018.

[8] Questa scansione temporale non tiene conto di personalità come Henry David Thoreau nel XIX secolo o di Aldo Leopold nella prima metà del XX secolo soltanto perché, nella brevità di questa esposizione, si predilige discutere dal punto di vista speculativo alcune categorie dell'etica ambientale così come è andata costituendosi in sapere codificato all'interno del vasto ambito delle cosiddette etiche applicate.

antropocentrico, secondo il quale la "natura" (qualunque cosa essa indichi) ha valore in sé, al di là dei fini umani; le strategie argomentative adottate sono solitamente due: la prima sottolinea come il valore della "natura" e/o dei singoli enti naturali derivi da alcune caratteristiche generali che essa/essi presenta(no), quali la soggettività, la capacità di provare piacere e dolore, etc., e si fondi sull'idea della possibile (e necessaria) estensione di caratteristiche umane agli altri enti naturali; la seconda intende dimostrare come il valore della "natura" e/o dei singoli enti naturali non derivi da caratteristiche che condividerebbe(ro) con la natura umana, bensì consista nella sua/loro assoluta alterità. La conservazione della "natura", in questo caso, non riguarda la sua importanza per i fini umani, bensì si posiziona esclusivamente sul piano deontologico del rispetto morale.

Il problema – e questo è uno dei punti centrali di queste brevi riflessioni – sta proprio nella complessa dinamica dell'antropocentrismo, il quale sembra essere per certi versi ineludibile. Come abbiamo visto, l'inquadramento non-antropocentrico si fonda sull'idea che la "natura" abbia un valore-in-sé, al di là dei fini umani. Ebbene: chi determina questo valore-in-sé? qual è la sua origine, la sua fonte? Sembra essere evidente che la determinazione di questo valore provenga da un vivente particolare, con un punto di vista particolare, il vivente umano, che tra le sue caratteristiche peculiari presenta quella del "porre-i-valori". La questione si complica: immaginando un universo privo di intelligenza senziente, gli enti naturali (dal mondo inanimato fino agli animali e alle piante, passando per ecosistemi, etc.) avrebbe oppure no valore-in-sé? Esiste un valore-in-sé, al di là del fatto che qualcuno/qualcosa possa coglierlo e "pensarlo", o un valore-in-sé può esistere soltanto nella misura in cui è *contemporaneamente* un valore-per-qualcuno che lo riconosce? E nel caso: valore-per-chi? Il valore deve avere

necessariamente un'origine esterna – un soggetto da cui "emana" – o è "inerente" alla "natura" e/o agli enti naturali? Senza una soggettività, senza uno spettatore/attore senziente, ha senso parlare di valore inerente? Un orso bianco porrebbe la questione del valore-in-sé del tricheco (e viceversa)? Infine: ragionare nei termini di valore-in-sé non significa reificare, oggettivare, astrarre dal flusso processuale un "qualcosa" al di là delle relazioni che instaura e lo formano? L'antropocentrismo ontologico, scacciato dalla porta, sembra rientrare dalla finestra come antropocentrismo valoriale o assiologico. Ma non è tutto. Antropocentrismo e non-antropocentrismo si scontrano, sempre all'interno dell'inquadramento cosiddetto non-antropocentrico, anche su un altro livello di riflessione. Si tratta di quella che viene chiamata "strategia dell'estensione": estendere caratteristiche umane (capacità di essere un soggetto, capacità di provare piacere e dolore, capacità di vivere una vita che può andare bene o male, etc.) agli enti naturali significa compiere un gesto antropocentrico o non-antropocentrico? Antropocentrico, in quanto si ritiene che il modello di sensibilità, intelligenza e buona vita sia quello umano, da estendere poi agli altri enti naturali; non-antropocentrico, in quanto si ritiene che sensibilità, intelligenza e buona vita appartengano all'umano così come agli altri enti naturali, cancellandone così la centralità prometeica. Al di là della forse eccessiva stilizzazione dei ragionamenti, si tratta comunque di una compresenza. Prendiamo ad esempio la caratteristica della "capacità di essere un soggetto": secondo Peter Singer e Tom Regan gli animali sono da considerarsi rispettivamente soggetti di interesse e soggetti di diritto[9], così come per Matthew Hall[10] e Michael Marder[11] sono da considerarsi

[9] Cfr. P. Singer, *Liberazione animale. Il manifesto di un movimento diffuso in tutto il mondo* (2009), tr. it. Il Saggiatore, Milano 2015, e T. Regan, *I diritti animali* (1983), tr. it. Garzanti, Milano 1990.
[10] Cfr. M. Hall, *Plants as Persons. Philosophical Botany*, Suny Press, New York 2011.

soggetti – seppur in maniera radicalmente e qualitativamente differente – anche le piante; la filosofa Jane Bennett[12], a partire da un'ontologia di stampo spinoziano, parla di una materia con caratteristiche di soggetto agente, mentre Bruno Latour, con la sua teoria degli attanti, ritiene che anche le cose inanimate hanno *agency*, vale a dire la capacità di modificare uno stato di cose[13]. Ebbene: l'essere-soggetto dell'umano è lo stesso dell'essere-soggetto di un cane, di una palma, di un batterio o di una pietra? Abbiamo una definizione di essere-soggetto univoca, sulla quale basarci per il reperimento di alcune sue caratteristiche nella realtà umana, animale, vegetale o inanimata? In realtà, non l'abbiamo neanche per l'umano. Focalizzandoci sul mondo vegetale, tutto appare ancora più complicato: nonostante studi innovativi e rivelatori sulla neurobiologia delle piante (cui faremo cenno nel corso della riflessione), in che misura esse possono essere definite dei "soggetti"? In più: ha senso chiederselo? chiederselo significa compiere un gesto antropocentrico o non-antropocentrico?

Ma non è soltanto questione di soggettività. Passando alla questione della "capacità di provare piacere o dolore", è possibile ritrovare questa facoltà nelle piante? Ha senso ricercarla con caratteristiche che siano simili a quelle del vivente umano e animale? Non si tratta di antropocentrismo o, comunque, di zoocentrismo?

Arrivando infine all'astrazione massima della strategia dell'estensione, vale a dire l'idea secondo la quale tutto ciò che è dotato di "vita" ha valore-in-sé, non dovremmo forse chiederci e definire prima che cosa è "vita"? È una risposta che può essere data in maniera univoca? Un cane, una palma, un batterio, o ancor

[11] Cfr. M. Marder, *Plant-Thinking. A Philosophy of Vegetal Life*, Comubia University Press, New York 2013.
[12] Cfr. J. Bennett, *Vibrant Matter. A politcal ecology of things*, Duke University Press, Durham 2010.
[13] Cfr. B. Latour, *Reassembling the Social. An Introduction to Actor-Network-Theory*, Oxford University Press, Oxford 2005.

di più "oggetti" indefinibili come un ecosistema, conducono la stessa "vita" di un umano (e la stessa tra di loro)? La "vita" è una caratteristica/qualità che posseggono determinati "enti" o è un processo complessivo di cui i singoli "enti" sono forme di manifestazione "parziale"? Gli attanti inanimati di Latour, questa pietra o quella montagna, per il fatto di avere *agency*, hanno una "vita"? *Pensare come una montagna*[14] è per l'umano un tentativo antropocentrico o non-antropocentrico? Infine: se la "natura" e/o gli enti naturali hanno valore-in-sé, nella loro assoluta alterità e irriducibilità all'umano, di che valore si tratta? Se è radicalmente *altro*, è possibile davvero reperirlo? Chi sarebbe in grado di svelarlo? La risposta è sempre l'umano e l'alterità assoluta sembra comunque un limite invalicabile. E ci troviamo nuovamente dinanzi all'ipotesi di un antropocentrismo valoriale o assiologico.

La cosiddetta etica delle piante rientra pienamente – anche a partire da qualche accenno già fatto – all'interno di questa complessità. E rispetto all'animale, la questione è sicuramente più problematica. Quelli che sono stati definiti i tre approcci fondamentali della *Plant Ethics*[15] sono i seguenti:

1. *Non-moral yet inclusive approach*, secondo il quale «la vita delle piante è decisamente importante per la vita degli esseri umani e degli animali non umani» per cui «le piante hanno valore e importanza morale perché sono elementi insostituibili della buona vita degli esseri senzienti che abitano il pianeta Terra» – si tratta di un approccio definito "ecumenico" in quanto «accoglie ragioni estetiche, prudenziali, narrative e scientifiche per valorizzare e rispettare le piante, e le unisce a ragioni morali che si riferiscono al bene degli umani e degli altri animali» e che «parla spesso il linguaggio della

[14] Il riferimento è ovviamente a A. Leopold, *Pensare come una montagna. A Sandy Count Almanac* (1949), tr. it. Piano B, Prato 2019.
[15] Cfr. A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.* (La traduzione di questa e di tutte le citazioni di opere di cui non esiste o non viene citata l'edizione in italiano è da considerarsi nostra).

giustizia: proteggere le foreste, la biodiversità botanica, le popolazioni di piante e le singole piante è in molti casi un modo per essere giusti nei confronti di tutti quegli esseri umani e animali non umani che possono in vari modi beneficiare del mondo vegetale oggi e in futuro»[16].

2. *Relational approach*, secondo il quale «il valore e l'importanza morale delle piante non discendono dallo status ontologico e morale delle piante, ma emergono piuttosto all'interno delle pratiche umane che si riferiscono alle piante» per cui «gli approcci relazionali non parlano tipicamente il linguaggio dei doveri e degli obblighi, ma quello della cura e della virtù»[17].
3. *A value-in-nature approach*, secondo il quale non sono da mettere in evidenza «le nostre ragioni per dare valore alle piante, né il nostro rapporto con esse, ma piuttosto i valori delle piante stesse», anche se non si tratta della «ricerca del valore intrinseco delle piante […]: si tratta piuttosto di un'esplorazione di valori concreti, empiricamente testati nella, della e per la vita delle piante, tra cui l'integrità e la prosperità»[18].

All'interno di questa sintesi molto efficace si ritrovano molte delle complessità e delle articolazioni dell'etica ambientale e animale cui abbiamo fatto un seppur troppo rapido cenno all'interno di questo paragrafo. Si tratta di approcci che comunque tendono spesso a integrarsi o a incrociarsi in maniera critica. Nonostante questo campo di studi sia piuttosto recente, le pubblicazioni e le articolazioni sono già numerose. Troviamo il punto di vista "utilitarista", secondo il quale le piante non sono da considerarsi soggetti di benessere, per cui il nostro comportamento morale nei confronti di esse può essere giustificabile soltanto a partire dal benessere umano e animale che ne deriva, dunque l'etica delle piante non può che essere

[16] *Ibid.*, p. 2.
[17] *Ibid.*
[18] *Ibid.*, p. 3.

indiretta e derivativa[19]. Abbiamo poi il punto di vista del "valore inerente", secondo il quale le piante hanno tale valore nella misura in cui hanno "interessi" in quanto enti naturali teleologicamente organizzati dalla dinamica della selezione naturale – le complessità di questa posizione muovono dall'idea che, dunque, anche batteri e artefatti dovrebbero così avere interessi, in quanto enti organizzati a partire da una finalità intrinseca[20]. Ci sono poi le teorie del "fiorire" su base neo-aristotelica, secondo le quali il "fiorire" può divenire una sorta di metro di giudizio a partire dal quale valutare se la nostra relazione con le piante conduca a forme di danneggiamento o di beneficio, con il vantaggio di avere al centro una nozione che proviene dalla scienza botanica e non da una dimensione valoriale antropocentrica[21] – qui il pericolo è sempre quello di incorrere nella *fallacia naturalistica*. C'è poi chi riflette a partire dal concetto kantiano di "dignità" (che nel filosofo tedesco può essere ricondotto soltanto all'umano, in quanto agente morale razionale) e sulla difficoltà di fondo dell'estensione dei concetti dell'etica umana all'etica delle piante, in quanto determinati concetti costruiti per un "regno" difficilmente possono essere adeguati per un altro "regno"[22]. Ci sono poi gli approcci relazionali: secondo Coeckelbergh, ad esempio, l'analisi dal punto di vista ontologico o metafisico del regno delle piante potrebbe non essere la via più percorribile per giungere a un'etica delle piante, la quale dovrebbe piuttosto muovere dalla relazione con l'umano e mediante un dialogo tra il punto di vista umano e quello vegetale, e puntando su un'etica pratica e

[19] Cfr. T. Višak, *Utilitarian plant ethics*, in A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.*, pp. 30-39.
[20] Cfr. R. Sandler, *Is considering the interests of plants absurd?*, in A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.*, pp. 40-50.
[21] Cfr. A. Kallhoff, *The flourishing of plants: A neo-Aristotelian approach to plant ethics*, in A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.*, pp. 51-58.
[22] Cfr. S. Odparlik, *The dignity of plants. An overview of the discussion in German-speaking countries*, in A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.*, pp. 59-69.

contestuale[23]; mentre Schörgenhumer punta invece sulla questione della "cura" e delle virtù a essa connessa[24].

All'interno di questo ampio panorama (che non pretende comunque di essere esaustivo), sorgono due questioni. La prima è che sembra essere chiara la mancanza di un ragionamento che muova dalla "spirale" di cui discutevamo nell'introduzione al saggio (la relazione "complessa" tra il dominio culturale e simbolico e il dominio economico-politico, tra la rappresentazione del bisogno e il bisogno), e le ragioni di questa mancanza sarebbero anche esse da analizzare dal punto di vista etico. Da qui la domanda: è possibile immaginare, ai tempi dell'Antropocene e della crisi climatica conclamata, una riflessione morale per certi versi slegata dalle urgenze politiche? La seconda concerne ancora una volta la questione dell'antropocentrismo: se è vero che la riflessione sull'etica delle piante deve muovere da qualcosa come il riconoscimento di un'alterità assoluta, che può essere appunto "riconosciuta" non in maniera mai complessiva ma attraverso uno sforzo non solo scientifico ma anche "immaginativo", è altrettanto vero che questo riconoscimento non può che provenire dall'umano, dal punto di vista umano. Questa tipologia di sforzo intellettuale e immaginativo può portare con sé forme di destabilizzazione delle consuetudini simboliche e culturali e, inserendosi nella "spirale", modalità di ripensamento di una serie di categorie economico-politiche, che nell'Antropocene non regolano soltanto l'umano ma sono a fondamento dell'intera possibilità futura della relazionalità vivente nel Sistema Terra. Si tratta in parole semplici di mostrare come il "riconoscimento" dell'alterità assoluta di un ente naturale possa portare con sé un ripensamento della dimensione simbolica e culturale di comprensione

[23] Cfr. M. Coeckelbergh, *What do we mean by a relational ethics? Growing a relational approach to the moral standing of plants, robots and other non-humans,* in A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.*, pp. 98-109.

[24] Cfr. M. Schörgenhumer, *Caring for plants: Cultivating realtional virtues*, in A. Kallhoff, M. Di Paola, M. Schörgenhumer (a cura di), *op. cit.*, pp. 110-118.

dell'alterità con possibilità di trasformazioni nelle relazioni materiali tra gli umani, tra i viventi umani e i viventi nonumani, e tra i viventi umani e la realtà tutta del Sistema Terra. Si tratta in parole ancora più semplici della rappresentazione di una necessaria riattivazione del modello immaginativo dell'utopia.

Si può ritenere che questa posizione sia del tutto antropocentrica – necessità di un'etica ambientale, animale e delle piante in vista del benessere umano –, ma in realtà si tratta di una forma tutt'al più di "antropocentrismo debole"[25], nella misura in cui il benessere umano, nell'epoca dell'Antropocene, non può essere più scisso, neppure *simbolicamente*, da un riconoscimento di un benessere "ambientale", animale e vegetale. Il benessere è un inter-esse comune. L'inter-esse può essere inteso in senso strumentale e individuale oppure come valore comunitario che trova nell'"inter" la matrice della comunità biotica. Nel secondo caso, l'unico ovviamente all'altezza della sfida del cambiamento climatico, la costruzione di questo inter-esse comune oltrepassa le distinzioni tra le differenti forme di ben-essere e muove dal riconoscimento di una complessa relazionalità e processualità, sempre metastabili e sempre da stabilizzare. Una relazionalità e una processualità che sono allo stesso tempo ontologiche ed etico-politiche: se è vero che, dal punto di vista filosofico, è sempre difficile – se non "scorretto" logicamente – costruire un processo deduttivo che mostri il passaggio da una certa descrizione dell'*essere* a una certa prescrizione del *dover essere*, è almeno altrettanto vero che l'ingiunzione delle etiche ambientale, animale e delle piante non può che fondarsi da un lato sul riconoscimento della processualità e relazionalità ontologica e

[25] Con questa espressione, dovrebbe essere chiaro, non intendiamo «il modello della *conservazione delle risorse* cui corrisponde una *cultura della gestione*», come sottolinea Luisella Battaglia, sulla scorta di un'*etica dei limiti* «in cui si assegnano vincoli normativi al comportamento umano in relazione all'ambiente e si prescrive un'amministrazione oculata delle risorse naturali, sempre in funzione della prosperità e del benessere umani» (L. Battaglia, *Un'etica per il mondo vivente. Questioni di bioetica medica, ambientale, animale*, Carocci, Roma 2011, pp. 160, 161-162).

situazionale di cui uno dei poli (ma anche uno degli "effetti") è necessariamente l'umano e dall'altro sulla necessità di un ripensamento della dimensione del politico a partire da un ripensamento radicale dei modi di produzione, consumo e sviluppo dell'umano. Che si ritenga che debba essere antropocentrica o non-antropocentrica, l'epoca critica dell'Antropocene e del surriscaldamento globale non può che muovere da un lato dalla necessità di una riattivazione della "spirale", dall'altro dalla consapevolezza che, arrivati a questo punto critico, non vi è più grande distinzione tra una benessere umano, animale, vegetale, della vita nel suo insieme (qualunque cosa significhi "vita"). E che la costruzione di questo benessere deve passare allo stesso tempo per un ripensamento delle rappresentazioni simboliche e culturali e per una ristrutturazione delle relazioni materiali inter- e intra- specifiche. Se è vero che l'etica può essere una questione soltanto umana (gli enti naturali non agiscono o scelgono di agire in chiave etica) e quindi antropocentrica, è altrettanto vero che l'inter-esse, inteso in senso relazionale e processuale, apre a una rappresentazione della "natura" non-antropocentrica. Ma anche la rappresentazione resta una questione umana, troppo umana.

3. *L'alterità assoluta delle piante*

Discutere dell'alterità assoluta delle piante – e, muovendo da questa, sul senso di un'etica delle piante – significa lavorare a partire dal punto di vista scientifico, mediante il reperimento delle novità che provengono dalla neurobiologia delle piante, e dal punto di vista filosofico, mediante un ripensamento del significato ontologico e metafisico della vita vegetale.

Iniziamo dal punto di vista scientifico. Prima di analizzare alcuni aspetti del contemporaneo campo di studi della neurobiologia vegetale, bisogna sottolineare come una prima comprensione del fatto che le piante siano esseri senzienti e

attivi provenga da Charles Darwin[26]. Il grande naturalista inglese ragiona a partire da alcuni elementi: il primo concerne il fatto fondamentale che per comprendere la complessità della vita vegetale occorra analizzare le parti terminali delle radici, le quali rappresentano un vero e proprio organo di senso che registra differenti parametri e può reagire a essi; il secondo è che nelle piante – nelle parti terminali delle radici – esistono movimenti che è possibile considerare "intelligenti", mediante una sequenza dimostrabile di percezione di stimoli dall'ambiente, decisione su quale direzione prendere, attuazione del movimento finale. In poche parole, si tratterebbe di una prima intuizione di quella che è possibile definire come intelligenza diffusa delle piante.

Secondo i più recenti studi di neurobiologia vegetale[27], le piante, pur non possedendo un cervello e un sistema nervoso centrale, presentano una forma di intelligenza molto complessa: esse infatti posseggono diverse migliaia di "reti cerebrali" (i meristemi), collocate sulle parti terminali delle radici, e interconnesse mediante dei particolari filamenti vascolari, definiti (per analogia) come dei veri e propri "neuroni". Un sistema nervoso decentralizzato non significa un sistema nervoso "inferiore" o "imperfetto" (al di là del fatto che superiorità/inferiorità e perfezione/imperfezione non sono e non dovrebbero essere categorie biologiche), presenta invece dei vantaggi per esseri viventi la cui specializzazione è il non-movimento: i migliaia di cervelli di ogni singola pianta permettono a essa di integrare le informazioni sensibili e di attivare una comunicazione tra i tessuti in grado di far "prendere

[26] Cfr. C. Darwin, *The power of movement in plants* (1880), Cambridge University Press, Cambridge 2009.

[27] Cfr. F. Baluska, S. Mancuso, D. Volkmann, P.W. Barlow, *Root apex transition zone: a signalling-response nexus in the root*, in «Trends in Plant Science», 15, 2010, pp. 402-408, e F. Baluska, S. Mancuso, *Plants, climate and humans. Plant intelligence changes everything*, in «EMBO Reports», 21, 2020 (liberamente consultabile al seguente link: https://www.embopress.org/doi/full/10.15252/embr.202050109); per una lettura divulgativa, cfr. S. Mancuso, A. Viola, *Verde brillante: Sensibilità e intelligenza del mondo vegetale*, Giunti, Firenze 2015.

decisioni" alle piante. Le piante possono "decidere" ad esempio di "produrre" elementi biochimici per mutare la propria condizione e relazione con l'ambiente: in presenza di agenti patogeni, possono produrre sostanze difensive (antifungini, ad esempio) o, in presenza di un suolo povero di fosforo, possono produrre sostanze, come enzimi e acidi organici, in maniera tale da arricchirlo. Se percepiscono un pericolo nella crescita delle piante nelle vicinanze, possono produrre metaboliti che ne frenano lo sviluppo. Sembra inoltre che le piante, in qualche modo, abbiano una percezione articolata del proprio ambiente, a tal punto da "interagire" non solo con altri vegetali ma anche con animali: si è dimostrato, dal punto di vista empirico, che alcune piante, attaccate da erbivori, sono in grado di emettere sostanze che avvertono della presenza del pericolo le piante vicine, le quali attivano così le proprie difese chimiche contro l'"attacco"[28]. Semplificando al massimo, le evidenze scientifiche mostrano come le piante non siano assolutamente assimilabili ad automi passivi e con risposte meccaniche, ma siano forme-di-vita complesse, attive e intelligenti. Di un'intelligenza di difficile comprensione per l'umano, perché radicalmente differente. Cosa significa infatti possedere un cervello diffuso? Di che tipo di intelligenza si tratta? Come definirla? È possibile definirla per analogia con l'intelligenza animale o umana? Cosa può insegnarci?

Dal punto di vista filosofico, il tentativo di riflessione ontologica e metafisica più articolato è sicuramente quello di Emanuele Coccia[29]. Secondo il filosofo italiano, il nostro mondo è innanzitutto un "fatto vegetale": il meccanismo dell'autotrofia – il fatto di trasformare in vita la luce, la materia e l'aria, costituendo così il mondo per gli animali e gli umani – è il motivo per cui le piante sono da interpretare come vere e proprie

[28] Cfr. W. van Hoven, *Mortalities in kudu (Tragelaphus strepsiceros) populations related to chemical defence in trees*, in «Journal of African Zoology», 105, 1991, pp. 141-145.

[29] Cfr. E. Coccia, *La vita delle piante. Metafisica della mescolanza* (2016), il Mulino, Bologna 2018.

costruttrici di mondi, vere e proprie potenze cosmogoniche (Coccia parla di "tono esiodeo" da rilanciare nella botanica[30]). Il nostro "mondo", ciò che chiamiamo "natura", è un fatto vegetale in due sensi, sia perché è "creato" dalle piante e dalla loro attività di fotosintesi, sia perché l'organismo animale è formato completamente da materiale organico creato dalle piante. Se a questo si aggiunge il fatto che il 90% della massa eucariotica del pianeta è formato da piante e la maggior parte degli oggetti e degli strumenti che noi viventi umani utilizziamo deriva dalle piante, è possibile dunque affermare che tutta la vita animale superiore si nutre e deve l'esistenza, in senso ontologico, a questa facoltà cosmogonica delle piante. Dal punto di vista epistemologico, le piante mettono definitivamente in crisi quello che è l'assunto fondamentale delle scienze biologiche e naturali degli ultimi secoli, l'idea di una priorità dell'ambiente sul vivente; sono i viventi – e innanzitutto le piante – a creare l'ambiente in cui vivono, per cui la relazione di adattamento è molto più complessa e imbricata, si tratta di una vera e propria co-costruzione, nonché di una complessificazione del rapporto tra contenente e contenuto. Porsi al livello della *vita delle piante* significa ripensare tutta una serie di abitudini concettuali. Innanzitutto, cosa debba intendersi per "vita": se è vero che gli animali superiori hanno bisogno gli uni della vita degli altri, così come gli umani hanno bisogno della vita prodotta dagli altri animali e dalle piante, lo studio delle piante fa comprendere come esse non abbiano bisogno della vita altrui (perché la ricavano dalla realtà inanimata), ma siano esse stesse creatrici di vita, in grado di modificare mondi per la diffusione e l'esplosione delle forme-di-vita. Il principio fondamentale è quello della "metamorfosi", in quanto le piante hanno la caratteristica di coincidere pienamente con le forme che creano – l'immediatezza della trasformazione, il fatto che in una pianta generare

[30] Cfr. *ibid.*, p. 19.

significa trasformarsi, mutare la propria forma, rappresenta l'opposto della concezione astratta della creazione e della tecnica umane. La genesi delle forme nel mondo delle piante raggiunge un elevato livello di intensità: rispetto agli animali superiori in cui lo sviluppo si arresta, la pianta rappresenta «un'inarrestabile macchina morfogenetica» capace di costruire nuovi organi e parti di cui sono state private: «il loro non è mai un corpo definitivamente dato, ma un atto costante di bricolage somatico […] un atto incessante di *autodesign*»[31]. In questo senso, "mondo" e "natura" coincidono, in quanto «il mondo non è l'insieme logico di tutti gli oggetti, né una totalità metafisica di esseri, ma la forza fisica che attraversa tutto ciò che si genera e si trasforma» e la «*natura* non designa quel che precede l'attività dello spirito umano, né ciò che si oppone alla cultura, ma ciò che permette a ogni cosa di nascere e di divenire»[32]. E si presenta nuovamente la questione dell'antropocentrismo, a partire dall'espulsione della natura dallo sguardo filosofico e la sua riduzione a «oggetto puramente residuale, opposizionale, per sempre incapace di occupare il posto di soggetto»[33]. Si tratta, dunque, di una rimozione, «quella del vivente e del fatto che ogni conoscenza è già espressione dell'essere della vita», quella del fatto che «il mondo è il respiro dei viventi», quella del fatto che «ogni conoscenza cosmica è un punto di vita (non solo un punto di vista) e ogni verità non è che il mondo nello spazio di mediazione del vivente»[34]. Rifondare una "cosmologia", questo il compito dell'unica filosofia possibile oggi, secondo Emanuele Coccia, una cosmologia che muova dall'esplorazione del mondo delle piante. Nella costruzione del discorso del filosofo italiano non è presente una riflessione sulle conseguenze etiche di una tale impostazione filosofica, definita "metafisica della mescolanza",

[31] *Ibid.*, p. 24.
[32] *Ibid.*, p. 29.
[33] *Ibid.*, p. 31.
[34] *Ibid.*

ma c'è l'idea (seppur non espressa, ma che noi riteniamo di poter derivare) che, nel momento stesso in cui muta la rappresentazione ontologica e cosmologica del mondo-natura, allora si produrranno forme nuove di relazione etica tra i viventi. Richiamando l'impostazione di Michael Marder, è possibile affermare che, se in senso levinasiano l'etica è sempre relazione con un'alterità assoluta, allora la vita vegetale merita particolare rispetto etico[35]. Sulla *differenza* delle piante e la necessità di un'etica delle piante post-moderna si interroga Karen Houle, la quale ricostruendo nell'umano una relazione tra strutturazione del "corpo biologico" e strutturazione del "corpo di pensiero", sottolinea come l'impostazione filosofica dominante sia caratterizzata da opposizioni binarie che richiamano la bi-lateralità sinistro-destro del corpo biologico umano; ebbene, la fisiologia delle piante fa saltare questo schema, nessuna bi-lateralità ma una crescita di tipo rizomatico, per cui ispirarsi alla biologia della vita vegetale può indicare una strada per andare oltre il pensiero binario e può fondare, su questa unicità assoluta, la possibilità di un'etica delle piante come base per un'etica complessiva[36]. Come suggerisce Sylvie Pouteau, la particolare natura delle piante – l'apertura, il non essere auto-centrate, il vivere in uno spazio non formalizzato in modo euclideo – può permettere di pensare nuove ontologie, epistemologie e prospettive etiche[37].

In conclusione, le riflessioni di carattere scientifico e filosofico che abbiamo sintetizzato incrociano immediatamente la possibilità di un'utilizzazione simbolica e culturale di questa "alterità" per un miglioramento del mondo umano e della relazione etica con l'ambiente in generale. Ancora l'ineludibile

[35] Cfr. M. Marder, *op. cit.*

[36] Cfr. K. Houle, *Facing only outwords? Plant bodily morphogenesis and ethical conceptual genesis,* in A. Kallhoff, M. Di Paola, M. Schörgenhumer (ed.), *op. cit.*, pp. 70-81.

[37] Cfr. S. Pouteau, *Plants as open beings: From aesthetics to plant-human ethics*, in A. Kallhoff, M. Di Paola, M. Schörgenhumer (ed.), *op. cit.*, pp. 82-97.

antropocentrismo. Riconoscere come le foreste davvero parlino[38], al di là delle rappresentazioni degli indigeni, ma a partire da una metaforica per certi versi accettabile oggi anche in senso scientifico, cosa significa? Riconoscere nelle piante qualcosa che può essere di insegnamento per la nostra cultura e civiltà, cosa significa? Significa che un'etica delle piante può essere necessaria, nella misura in cui, partendo dall'assunto che si trovava nelle prime righe del saggio – riconoscere dignità a un determinato ente non conduce necessariamente a un rispetto di quell'ente –, soltanto riconoscendo la complessità sempre differente e sempre articolantesi l'una sull'altra della vita umana, della vita animale e della vita vegetale, è forse possibile immaginare un modo differente di cogliere le connessioni fondamentali tra gli enti e dunque, in quanto etica e politica sono forme che assumono la gestione della relazionalità nell'umano, un allargamento e un miglioramento della relazionalità tra umani, tra viventi umani e viventi nonumani, e tra viventi umani e natura inanimata – vale a dire: rendere effettiva l'estensione della considerazione morale all'interno di un progetto di trasformazione dell'esistente.

4. *Le ragioni di una risposta*

Il tempo dell'Antropocene ci ha mostrato come i cosiddetti limiti planetari, il mantenimento dei quali permette al sistema di riequilibrarsi, possano nell'arco di qualche decennio essere definitivamente sorpassati. Oltrepassare tali "limiti" significherebbe dare avvio a un processo di turbolenza che limiterebbe la possibilità di una qualunque forma di "previsione" sul possibile andamento dell'intero Sistema Terra, andamento che potrebbe essere particolarmente "disordinato" per diversi millenni (se non decine di millenni), ponendo fine alla fase di cosiddetta

[38] Cfr. E. Kohn, *Come pensano le foreste* (2013), tr. it. nottetempo, Milano 2021.

"pace climatica" (l'Olocene) che ha permesso lo sviluppo rapido e impetuoso delle civilizzazioni umane. Al di là dell'impossibilità di una previsione per certi versi "esatta" dell'evoluzione di un sistema complesso che "agisce" in maniera stocastica, ma partendo dal presupposto che è comunque identificabile un trend o andamento, riflettere sulla domanda *è davvero necessaria un'etica delle piante?* significa allo stesso tempo porsi all'interno di un orizzonte che può contemplare la catastrofe e cercare gli strumenti adeguati, dal punto di vista culturale e simbolico e dal punto di vista materiale, per una trasformazione radicale dei sistemi di pensiero e vita globali (cioè la "spirale" tra elementi simbolici ed elementi economico-politici). Sembra essere estremamente probabile, e questo punto di vista non appartiene soltanto a determinate prospettive di politica complessiva, che l'attuale organizzazione dell'economia di produzione e consumo globale (e sempre più globalizzantesi) non sia più, non solo nel lungo periodo ma anche in quello medio e forse breve, sostenibile. Non è più sostenibile una determinata modalità di organizzare la "trasformazione" tecnica del mondo naturale per fini umani e non sono più sostenibili le rappresentazioni culturali e simboliche "classiche" e quelle che si vorrebbero "nuove" e "ottimistiche" e che rilanciano l'idea che, per immunizzare gli effetti degli impatti antropici, sia necessario accelerare proprio gli interventi di carattere antropico[39]. In questo senso, non si è compiuto nessun passo in avanti – anche ragionando esclusivamente a partire dal versante etico – all'interno di un percorso che preveda, dal punto di vista immaginativo, la presenza nel qui e ora della "decisione" giuridico-politica non solo degli interessi delle future generazioni ma anche di quelli di entità naturali altre. La facoltà immaginativa nel nostro tempo sembra essere particolarmente in crisi, eppure è soltanto mediante

[39] Ci riferiamo alle posizioni ultra-prometeiche e ultra-eccezionaliste dell'umano proprie del cosiddetto "ecomodernismo" (cfr. http://www.ecomodernism.org/italiano – link consultato il 21 marzo 2022).

l'immaginazione che è possibile fuoriuscire dai limiti cognitivi della nostra psicologia morale[40], ed è soltanto mediante l'immaginazione che è possibile costruire nel pensiero la rappresentazione di un mondo completamente altro[41].

La ragione fondamentale per cui riteniamo che l'etica delle piante sia davvero necessaria è proprio perché potrebbe permettere un esercizio di immaginazione particolarmente articolato e complesso. Proprio trattandosi di un'alterità per certi versi non afferrabile del tutto, già sempre *sfuggente*, ma che può essere comunque colta nella sua straordinaria (proprio nel senso di: fuori dall'ordinario della nostra classica lettura del mondo e delle cose) unicità sia dal punto di vista estetico che metafisico-ontologico, la funzione di una centralità della vita vegetale all'interno del discorso dell'etica ambientale potrebbe essere quella di attivare uno sforzo immaginativo che, a partire dalla lenta temporalità delle piante, possa condurre a una rappresentazione "anticipata" del "futuro" del sistema Terra. Se è vero quanto afferma Coccia, che le piante sono potenze cosmogoniche ed esiodee, porsi al livello della vita delle piante dovrebbe permettere di cogliere la fondamentale potenza creatrice della vita vegetale e della vita *tout court*: questa forma specifica di creatività – di cui noi letteralmente ci nutriamo – potrebbe essere a fondamento di un percorso immaginativo della creatività umana, un'utopia trasformatrice della relazionalità complessiva, a partire dal presupposto oramai ineludibile che il benessere umano non può che essere inserito all'interno del complesso del benessere del "tutto". Il ben-essere, lo stare-bene, diviene una dinamica che ha a che vedere con il mondo della creazione (e della creatività: le piante non creano e basta,

[40] Cfr. S. Pollo, *op. cit.*

[41] Ci permettiamo di rinviare, sulla connessione tra Antropocene e l'idea che sia sempre più difficile immaginare il *radicalmente nuovo*, e su come provare a riattivarlo, D. Salottolo, Senza il nuovo, quanto può durare una cultura? *Tina e la ricerca di una "cosmologia" all'altezza dell'Antropocene*, in «S&F_scienzaefilosofia.it», 23, 2020, pp. 350-385.

creano bellezza) e con la realtà di un possibile vivere-insieme che sviluppi ed estenda costantemente i confini di questo "insieme".

Si tratta di antropocentrismo o non-antropocentrismo? La domanda è d'obbligo. La risposta è sempre la stessa. Si tratta di antropocentrismo, nel momento in cui sembra che la finalità riguardi l'attivazione della facoltà immaginativa umana in vista della trasformazione dell'esistente, trasformazione a guida umana; si tratta di non-antropocentrismo, nella misura in cui si ragiona a partire da un'alterità assoluta, non afferrabile del tutto, e dall'idea di un ben-essere di tutta la comunità vivente e non vivente, indistintamente, senza alcuna centralità per l'umano. La nostra posizione dovrebbe essere chiara, è impossibile fuoriuscire completamente dall'antropocentrismo, è impossibile fuoriuscire dal punto di vista umano (o dal "punto di vita", come lo chiama il filosofo Coccia). È possibile, invece, limitarne gli effetti deteriori, per coltivarne soltanto quelli che possano essere propositivi. Non sappiamo se l'immaginazione appartiene anche ai viventi nonumani, ed è forse inutile (e antropocentrico!) chiederselo. L'immaginazione, per quanto ne sappiamo, è una facoltà umana, e in questo momento è la facoltà che va coltivata e lasciata "fiorire": la traduzione di una specifica sensibilità in discorso razionale e intellettuale necessita, mai come in questo momento, del filtro dell'immaginazione. L'immaginazione deve essere la facoltà da "coltivare" e la creatività della vita vegetale – esiodea e metamorfica – può fungere da modello per questa ri-attivazione.

Ritornando e chiudendo sull'assunto iniziale – riconoscere dignità a un determinato ente non ha condotto necessariamente a un rispetto di quel determinato ente (neanche quando quell'ente è l'umano stesso) – significa che la facoltà dell'immaginazione nell'umano è ancora troppo debole: se è vero che le piante immaginando forme sempre nuove creano forme sempre nuove, vuol

dire che dalla vita vegetale abbiamo soltanto da imparare. Ricordando, forse, che all'interno della "spirale" che connette rappresentazioni culturali e simboliche e determinazioni materiali, il ruolo della dimensione "estetica" è da rilanciare. Siamo convinti che un discorso di questo tipo sarà indubbiamente tacciabile di "antropocentrismo", ma per noi risulta decisivo, strategicamente, ridiscutere il punto di vista umano, che, pur muovendosi *finalmente* per vie sempre meno antropocentriche, resta e resterà il punto di vista di un vivente specifico. Ed è esattamente su quello che occorre lavorare.

S&F_n. 27_2022

LINGUAGGI

GIACOMO SCARPELLI

FREUD E CONAN DOYLE.

VITE PARALLELE

1. *In punta di penna e di bisturi* 2. *La lezione dei maestri* 3. *La lampada rossa*
4. *Cattive amicizie ed esperimenti azzardosi* 5. *Psicoanalisi e ricerca psichica*

ABSTRACT: ***FREUD AND CONAN DOYLE. PARALLEL LIVES***
The investigative methods of Freudian psychoanalysis have long been debated, compared with the investigative methods of Sherlock Holmes. Although Freud was an avid reader of Conan Doyle's works, no one has never considered the existential and methodological concordances between the two authors. Both had medical degrees, were steeped in Darwinian science; moreover, they were the favored sons of brilliant mothers who nurtured their creative tendencies, and they had had as teachers clinicians who had become role models (Charcot and Bell). In the early days of the profession the empty days without patients had prompted each of them to imagine new conceptions; furthermore, certain dangerous friendships, accompanied by daring experiments on alkaloids, had been stimuli for knowledge of the human nature and limits. Freud and Doyle touched on each other on two occasions. Between the two, who was a writer masquerading as a doctor and vice versa?

1. *In punta di penna e di bisturi*

Molto si è detto e si è scritto sull'analogia tra metodo di indagine caratteristico della psicoanalisi e indagine di tipo poliziesco scientifico. James Hillman ha affermato che le opere di Freud, nel loro essere erette secondo uno schema investigativo preciso, comprendevano un mistero da scoprire (rimozione), attraverso l'individuazione di vicissitudini e delitti (sintomi), il ricorso del protagonista al *deus ex machina* (*transfert* sull'analista), sino allo scioglimento finale (*dénouement*)[1]. Altri autori, tra cui Carlo Ginzburg, hanno evidenziato e messo a confronto il criterio anamnestico freudiano e l'impostazione

[1] J. Hillman, *Le storie che curano* (1983), tr. it. Raffaello Cortina, Milano 1984, pp. 3-23.

investigativa peculiare di Sherlock Holmes[2]. Effettivamente, nel tempo libero, Freud si dilettava a leggere le imprese del volpino personaggio creato da Conan Doyle, col quale condivideva la vocazione a cogliere labili tracce e risolvere i misteri intricati.

[2] C. Ginzburg, A. Davin, *Morelli, Freud and Sherlock Holes: Clues and Scientific Method*, in «History Workshop Journal», 9, 1980, pp. 5-36; M. Rohrwasser, *Freuds Lektüren: Von Arthur Conan Doyle bis zu Arthur Schnitzler*, Imago Psychosozial Verlag, Gießen 2005, pp. 17-86.

Va comunque rilevato che non ci si è mai davvero soffermati sulle concomitanze biografiche tra gli stessi Freud e Doyle, i quali ebbero entrambi formazione medica. Ci troviamo pertanto di fronte a una sorta di circolarità esistenziale, scientifica e artistica. Il fondatore della psicoanalisi era un clinico con esperienze di fisiologo e neurologo, che impiegava una metodologia prossima a quella del personaggio uscito dalla penna di un narratore che aveva inventato una tecnica investigativa fondata sull'osservazione più scrupolosa, a sua volta elaborazione – come vedremo – del modello diagnostico della propria prima professione di medico; professione che poi era anche la medesima del sodale holmesiano, il bonario e rotondo dottor John Watson, benché spogliata di sagacia.

Naturalmente, le somiglianze e le affinità non si esauriscono qui. Ne scoveremo altre, ma intanto vale ribadire come Freud e Conan Doyle fossero accomunati tanto dalla matrice scientifica quanto dalla vena creativa. Sia il «romanziere» dell'investigazione mentale sia lo «scienziato» dell'investigazione criminale erano personalità dalle doti molteplici, cui ben si adattano le parole che Doyle stesso adoperò per celebrare le figure di spicco della cultura: «Certi uomini sono più grandi del loro lavoro. Il loro lavoro rappresenta solo un aspetto del loro carattere, invece ce ne potrebbe essere un'altra dozzina, tutti degni di nota, che si riuniscono per formare una creatura complessa e unica»[3].

2. *La lezione dei maestri*

Sigismund Schlomo (poi abbreviato in Sigmund) Freud era nato in Moravia, suddito dell'imperatore Francesco Giuseppe, il 6 maggio 1856, lo stesso anno del commediografo G.B. Shaw, del pittore John S. Sargent e della scoperta dell'Uomo di Neanderthal. Arthur Conan Doyle era nato a Edimburgo, suddito della regina Vittoria, il 22 maggio 1859, lo stesso anno del filosofo Bergson e dello psicologo

[3] A.C. Doyle, *Oltre la porta magica* (1907), tr. it. Piano B, Prato 2010, p. 49.

Pierre Janet, del capitano Dreyfus, del fuorilegge Billy the Kid e della pubblicazione dell'*Origine delle Specie* di Charles Darwin.
Il padre di Sigmund, Jacob, era un modesto commerciante di stoffe ebreo che non riuscì mai ad affermarsi. La sua terza moglie, Amalia, spirito vivace e quasi coetanea dei figli di primo letto del consorte, ebbe a sua volta due maschi e cinque femmine: Sigmund fu il prediletto[4]. Arthur Conan apparteneva a una famiglia irlandese trapiantata in Scozia, e percorsa da una vena artistica. Il nonno, John Doyle, era stato ritenuto il principe dei caricaturisti politici (si firmava HB), mentre gli zii Richard e Henry Edward ricoprivano rispettivamente le qualifiche di primo vignettista del popolare «Punch» e di direttore della National Gallery d'Irlanda; il terzo, James, era storico e anche lui illustratore. Il padre Charles, architetto spiantato, era il meno dotato, e il figlio stesso probabilmente lo superava nel disegno[5]. La madre Mary era una ragazzina quando aveva preso marito, e si rivelò uno spirito esuberante e un intelletto brillante. Anche in questo caso, il primo figlio maschio, Arthur, fu il prediletto. Il quale fu affidato ai padri gesuiti a Stonyhurst, e poi a Feldkirch, in Austria, dove prese un'infarinatura di tedesco[6].
Nel ricostruire le esistenze parallele di Freud e di Doyle, non indulgeremo a riferire di come l'uno, a diciannove anni, si fosse recato Oltremanica, affetto da quel che chiamava «mal d'Inghilterra», per recare visita ai fratellastri Emanuel e Philipp, i quali si erano trapiantati a Manchester, avevano messo su famiglia e avevano avviato fiorenti attività commerciali, né indulgeremo a riportare di come fosse rimasto incantato dal mare e dalle spiagge della vicina Blackpool, il suo clima e il cibo. Era

[4] Vedi P. Gay, *Freud. Una vita per i nostri tempi* (1988), tr. it. CDE, Milano 1988; P.-A. Alt, *Sigmund Freud* (2016), tr. it. Hoepli, Milano 2022.
[5] Arthur si dilettava appunto nel disegno; il suo schizzo più noto è *The Old Horse* (1929), rappresentazione dell'autore sotto forma di ronzino, che arranca tirandosi dietro il fardello dei familiari e delle imprese di un settantennio.
[6] Vedi J.D. Carr, *La vita di Sir Arthur Conan Doyle* (1949), tr. it. Rizzoli, Milano 1956; A. Lycett, *Conan Doyle* (2007), tr. it. Exceltior 1881, Milano 2011.

la *dear old England*, assurta a modello culturale e politico, patria di adozione dei due congiunti, tanto più vecchi di lui, al punto che Sigmund bambino li considerava alla stregua di zii[7]. Analogamente, non ci soffermeremo a evidenziare come il piccolo Arthur nutrisse ammirazione per gli zii artisti trasferiti a Londra, ove lo accoglievano volentieri e gli stimolavano entusiasmo per la capitale dell'impero britannico[8]. Non va invece dimenticato che per quanto in seguito assumesse convinzioni laiche, Conan Doyle apparteneva a una famiglia di religione cattolica, in Inghilterra assoluta minoranza, come quella ebraica di Freud in Austria (per quanto anch'egli presto smettesse di praticarla). Né va trascurato che sia Freud sia Doyle scelsero lo studio della medicina in quanto al tempo era ritenuta la più «umanistica» tra le facoltà scientifiche, e perché «è un piacere così grande fare un po' di bene che» un dottore dovrebbe «pagare per averne il privilegio»[9].

Se Conan Doyle diverrà il narratore universalmente conosciuto, anche il «gemello» viennese, confesserà di essersi dilettato a ideare trame per racconti, alcune ambientate in Oriente[10]. Per quanto si fosse ripromesso di «seppellire la bacchetta magica» della creazione artistica, in tarda età Freud arriverà a considerare la sua carriera scientifica una diversione: «sono diventato medico essendo stato costretto a distogliermi dai miei originari propositi, e il trionfo della mia esistenza consiste nell'aver ritrovato dopo una deviazione tortuosa e lunghissima, l'orientamento dei miei esordi»[11].

---

[7] P. Gay, *Freud: percorsi di lettura* (1990), tr. it. Il Pensiero Scientifico, Roma 1994, pp. 39 e 43. Cfr. S. Freud, *L'interpretazione dei sogni*, in *Opere*, Bollati Boringhieri, Torino 1989, III, pp. 246 e 253.

[8] Cfr. J.D. Carr, *La vita di Sir Arthur Conan Doyle* cit., pp. 15, 26-27.

[9] A.C. Doyle, *I discorsi del chirurgo*, in *Un medico d'altri tempi* (1894), tr. it. Passigli, Firenze 2012, p. 118.

[10] S. Freud, lettera a Martha Bernays (1° aprile 1884), in M. Lavagetto, *Freud, la letteratura e altro*, Einaudi, Torino 2001, p. 9.

[11] S. Freud, *Poscritto del 1927* a *Il problema dell'analisi condotta da non medici* (1926), tr. it. in *Opere*, cit., X, p. 418.

In ogni caso, Freud era sempre stato di ferme convinzioni darwiniane[12]. Charles Darwin non solo aveva ipotizzato l'origine delle specie e dell'uomo per selezione naturale, ma aveva anche avuto un ruolo precorritore nell'occuparsi dell'espressione delle emozioni. In ciò si era avvalso degli studi di Duchenne, neurologo alquanto singolare, nonché di Charles Bell, illustre anatomista[13]. Conan Doyle aveva avuto a sua volta la massima considerazione per Darwin – dotato di «una mente onnicomprensiva» senza precedenti – così come di Spencer e di Stuart Mill[14]. All'Università di Edimburgo aveva eletto a proprio mentore Joseph Bell, nipote del precedente, dotato di straordinarie doti di diagnosta e di osservatore. Non ancora ventenne, Doyle nel 1876 era rimasto strabiliato di fronte alle sue esibizioni: allampanato, bruna capigliatura setolosa, Bell in aula ordinava allo studente (Arthur) di far entrare i pazienti uno a uno e inesorabilmente ne coglieva con un solo sguardo il malanno e anche il mestiere e le abitudini[15].

In modo sorprendentemente affine, Jean-Martin Charcot, seguace di Duchenne, la chioma candida e le orbite buie, possedeva altrettanto occhio clinico, la capacità di catturare istantaneamente la fisionomia di una persona e di schizzarla a mente sulla carta, inoppugnabile. Nel corso delle sue lezioni, Charcot, grazie all'ipnosi di cui faceva uso nei casi d'isteria, aveva svelato a Freud, borsista presso l'ospedale parigino della Salpêtrière (1885), l'azione della parte sommersa della psiche, decifrandola con l'intuizione creativa[16].

[12] Cfr. P. Gay, *Freud*, cit., p. 33.
[13] C. Darwin, *L'espressione delle emozioni nell'uomo e negli animali* (1872), tr. it. Bollati Boringhieri, Torino 2012. Cfr. G. Scarpelli, *L'eredità della scimmia*, in «B@abelonline.print», V, 2008, pp. 159-167. Darwin apprezzava il Bell osservatore delle espressioni facciali umane, ma ne criticava la convinzione che gli animali non ne avessero.
[14] A.C. Doyle, *Oltre la porta magica*, cit., p. 143.
[15] J.D. Carr, *La vita di Sir Arthur Conan Doyle*, cit., pp. 38-39.
[16] S. Freud, *Charcot* (1893), tr. it. in *Opere*, cit., II, pp. 105-116; *Autobiografia* (1925), *ibid.*, X, pp. 75-141. Freud tradusse in tedesco un'opera

Sarà Bell a ispirare un giorno in Doyle la figura del suo campione dell'investigazione criminale, Sherlock Holmes (Watson era semmai la proiezione di Doyle stesso), ma sarà Charcot a svelare a Freud l'esistenza dell'inconscio.

3. *La lampada rossa*

Va riconosciuto che le vite dei due personaggi per un periodo si erano fatte divergenti, ma era stato di breve durata. Freud aveva impiegato otto anni per laurearsi all'Università di Vienna, nel 1881. Aveva trovato impiego all'Istituto Zoologico, quindi all'Istituto di Fisiologia e poi all'Ospedale Generale della capitale austriaca. Conan Doyle aveva invece pubblicato il suo primo racconto nel 1879 (*The Mystery of Sasassa Valley*), e si era imbarcato per sei mesi in qualità di apprendista medico di bordo sulla baleniera *Hope* (1880), a caccia nel Circolo Polare Artico; sarebbe seguita un'altra esperienza marinaresca, sulla *Mayumba*, lungo le coste africane (1882). Il dato significante, comunque, resta che anche Arthur si era addottorato in medicina in quel 1881, a Edimburgo, con una tesi sulla atassia locomotoria.

Seguono gli anni di iniziazione alla pratica professionale. Doyle si è stabilito nel 1882 a Southsea, al numero 1 di Bush Villas, e ha installato davanti al portone una targa lustra e un globo rosso luminescente, quale vistosa insegna di medico generico. La sala di consultazione l'ha allestita al di là delle sue possibilità economiche, con tappeto, tavolo di quercia, quadri e uno strumentario fornito. Né lo stratagemma della lampada da caserma dei pompieri, né l'ambulatorio elegante sortiscono però l'effetto desiderato da Arthur, il quale trascorre le giornate in attesa di pazienti che non si presentano (di momenti altrettanto vuoti patirà anche il suo Sherlock). Narrerà nella raccolta intitolata *La lampada rossa* (1894):

---

di Charcot (*Neue Vorlesungen über die Krankheiten des Nervensystems*, Toeplitz und Deuticke, Leipzig und Wien 1886).

> Uomini, donne e bambini… durante una giornata ne passavano a migliaia, e tuttavia tutti sembravano trascorrere in fretta, occupati nei loro affari, gettando appena un'occhiata alla piccola placca di ottone o evitando di sprecare un pensiero riguardo a quell'uomo in attesa nel suo studio. Eppure, quanti di loro sicuramente, naturalmente, avrebbero potuto trovare giovamento nella sua assistenza professionale. Uomini dispeptici, donne anemiche, volti pustolosi, carnagioni biliari…[17].

E Freud? Non se la passava molto meglio. Aveva pubblicato sulla «Neue Freie Presse» l'annuncio che «il dottor Sigmund Freud» era tornato da «un soggiorno di sei mesi da Parigi», e che risiedeva al numero 7 della Rathausstraße[18]. Al portone era apposta una targa di vetro a lettere d'oro. Il quartierino era spazioso e ben ammobiliato, con tanto di saletta oftalmoscopica. Qualche paziente si presentava, ma quasi nessuno era in grado di pagare la visita. Un minimo per sostentarsi proveniva da ex degenti del dottor Joseph Breuer, protettore del giovane Sigmund, il quale scriveva alla fidanzata Martha Bernays: «in poche settimane il mio denaro, che non ho ancora riscosso, sarà finito»[19]. In quelle ore nell'ambulatorio deserto, nacque in Freud l'esigenza di elaborare un sistema di pensiero atto all'esplorazione dell'inconscio, sulla scorta di ricerche sulla funzionalità del sistema nervoso e sui fenomeni isterici (in collaborazione col più anziano Breuer), ma anche dello studio dell'ipnosi e delle opere di pionieri della psichiatria quali Charcot e Bernheim, e di un filosofo quale Stuart Mill, per giungere alla analisi e all'interpretazione del sogno e del rapporto tra nevrosi e sessualità[20].

Parallelamente, in Conan Doyle alle prese con la propria ristagnante carriera medica, si andava rafforzando e consolidando la vocazione del narratore, da lui accolta con una specie di «gioia selvaggia», che finalmente lo avrebbe reso padrone di se

[17] A.C. Doyle, *Falsa partenza*, in *La lampada rossa* (1894), tr. it. Passigli, Firenze 2011, p. 34. Vedi A. Lycett, *Conan Doyle*, cit., pp. 83-93.
[18] E. Jones, *Vita e opere di Freud* (1953), tr. it. Il Saggiatore, Milano 1966, I, p. 186.
[19] S. Freud, *Lettere 1873-1939* (1960), tr. it. Boringhieri, Torino 1960, p. 196.
[20] Gli scritti di Freud sugli argomenti elencati occupano i primi 3 volumi delle *Opere*, cit.

stesso. Lentamente, ma inesorabilmente, andrà affermandosi come autore dapprima di racconti – anche sulle sue disavventure professionali – in seguito con l'edificazione della figura del *detective* che gli avrebbe assicurato il successo planetario, anche se lo avrebbe condannato a non più liberarsene[21].

Si trattò dunque da parte sia di Freud sia di Doyle di «fiaschi felici» (preferibile il titolo di una novella di Melville all'abusato *serendipity*)[22], cioè di insuccessi che generarono la fortuna dei due giovani medici squattrinati.

[21] Holmes fa la sua prima apparizione in *A Study in Scarlet* (1887).
[22] H. Melville, *The Happy Failure* (1854), tradotto da Enzo Giachino: *Il fiasco felice*, in *Racconti*, Einaudi, Torino 1954.

4. *Cattive amicizie ed esperimenti azzardosi*

Freud e Doyle diverranno l'uno padre della psicoanalisi, l'altro di Sherlock Holmes, ma i loro rispettivi cammini continuano a risultare punteggiati da esperienze simili, anche se non sempre luminose.

È nota la storia dello stretto rapporto di amicizia tra Freud e Wilhelm Fliess, otorinolaringoiatra tedesco dalle teorie ingegnose e bislacche sulla corrispondenza riflessa tra mucosa nasale e apparato genitale, e sui cicli biologici femminili e maschili, nonché chirurgo spericolato: rischiò di uccidere con un intervento sbagliato una paziente di Freud stesso, Emma Eckstein. Ciò nonostante, il sodalizio con Fleiss fu necessario per Freud non solo perché gli fornì un appoggio morale negli anni cruciali della sua vita, ma anche perché gli favorì l'indagine introspettiva, finalizzata a individuare i barlumi della teoria psicoanalitica, tra cui la bisessualità e la sublimazione[23].

D'altronde, Conan Doyle, nel 1882, subito prima dell'apprendistato a Southsea, si era lasciato irretire da un esuberante collega dall'etica professionale non proprio adamantina, George T. Budd. Il quale si era procurato a Plymouth una variopinta clientela degna del circo Barnum: curava con metodi ciarlataneschi, che gli fruttavano denaro sonante. Aveva fatto leva su una certa ingenuità di Arthur, sfruttandolo e poi accusandolo di avergli allontanato i pazienti a causa del suo esercizio della medicina tradizionale e superato. Si era giunti a una rottura del sodalizio quando Budd aveva sfasciato a martellate la placca «Dr. Arthur Conan Doyle», posta all'ingresso dello studio in comune. Va però ammesso che la breve attività insieme a Budd aveva costituito per Doyle un'esperienza che andava oltre quella scientifica. La conoscenza ravvicinata con le più svariate tipologie di persone gli sarebbe tornata utilissima nella creazione dei personaggi per le proprie opere[24].

[23] Cfr. S. Freud, *Lettere a Wilhelm Fliess, 1887-1904* (1985), tr. it. Bollati Boringhieri, Torino 1990; F. Sulloway, *Freud biologo della psiche* (1979), tr. it. Feltrinelli, Milano 1982.

[24] Nel romanzo epistolare autobiografico *The Stark Munro Letters* (1895), Doyle raccontò le sue esperienze con Budd, modificandone il nome in James Collingworth. Vedi D.N. Pearce, *The Illness of George Tumavine Budd and Its Influence in the Literary Career of Sir Arthur Conan Doyle*, in «Journal of Medical Biography», III, 1995, pp. 236-238.

L'altra cantonata in cui era incorso il giovane Freud era consistita nello sbandierare sconsiderata fiducia sulle proprietà benefiche della cocaina. Dopo averla provata su di sé in dosi blande, nel 1884 aveva pubblicato il saggio in cui vantava le doti farmacologiche dell'alcaloide, che poteva essere impiegato nella cura della nevrastenia, della cachessia, dei disturbi di stomaco, come euforizzante, analgesico e afrodisiaco, insomma una panacea. Impiegò la cocaina per alleviare la condizione dell'amico Ernst Von Fleischl-Marxow, autore di ricerche sull'attività elettrica del sistema nervoso, il quale pativa per le complicazioni dell'amputazione di un pollice ed era diventato morfinomane. La prescrizione di Freud, benché elargita con le migliori intenzioni, causò ulteriore dipendenza di Fleischl e lo condusse alla morte[25]. L'elogio malriposto della cocaina e l'esito infausto della terapia provocarono in Freud sensi di colpa e rischiarono di offuscargli la carriera. E però, chi sa, lasciarono una qualche traccia Oltremanica: il ricorso che Holmes fa della cocaina in soluzione sette per cento, quando si trova a corto di casi criminosi e cade nella depressione[26].

A onor del vero, cinque anni prima di Freud, Conan Doyle aveva pubblicato il resoconto dell'esperimento sugli effetti di un altro alcaloide, il *Gelseminum* (o *Gelsemium*), affine alla stricnina e impiegato come miorilassante, per trattare nevralgie e neuropatie. Era andata così. Durante il tirocinio a Birmingham, presso il paterno dottor Reginald Hoare (che si potrebbe ritenere un corrispettivo di Breuer), tormentato dalla cefalea, Arthur si era autosomministrato la tintura di gelsomino in quantità sempre crescenti, allo scopo di stabilire la soglia del sovraddosaggio. L'azzardata sperimentazione, segnata dall'insorgere progressivo di vertigini e paralisi facciali, si era arrestata al raggiungimento

[25] S. Freud, *Sulla cocaina* (1884), tr. it. Marsilio, Padova 1976; P. Gay, *Freud*, cit., pp. 40-41.
[26] Nicholas Meyer in *La soluzione sette per cento* (1974), tr. it. Rizzoli, Milano 1976, fa incontrare Freud con Holmes, allo scopo di disintossicarlo. Dal romanzo è tratto un film del 1976, regia di Herbert Ross.

di una dose di duecento *minime* (equivalente ad altrettanti granuli faramacopeici), la quale provocava una forma estrema di diarrea[27]. Superare quel *limen* significava la morte. Probabilmente in questa occasione Doyle si dimostrò più cauto di Freud.

5. *Psicoanalisi e ricerca psichica*

Il giorno dell'Epifania del 1891 Conan Doyle giunse a Vienna insieme alla moglie Louisa. Nevicava a larghe falde. Vi rimase tre mesi, per seguire le lezioni di oftalmologia al Krankenhaus e migliorare la conoscenza del tedesco. Si manteneva scrivendo racconti nel tempo libero. Freud si era sposato con Martha, era agli inizi del suo sodalizio con Fliess e andava pubblicando saggi sulle afasie e sulla paralisi infantile. Arthur e Sigmund non si incontrarono personalmente, possiamo immaginare che si sfiorarono soltanto, nell'edificio universitario, lungo l'Alserstraße, oppure in un caffè. E tuttavia questa non fu l'unica occasione.

Freud erigeva l'edificio della psicoanalisi sulla base della teoria sessuale e del centrale complesso di Edipo. Doyle invece prenderà a essere attratto sempre più fortemente dallo spiritismo. Questa disposizione si trasformerà quasi in un'ossessione, anche se fondata su un presupposto – almeno terminologico – affine all'indagine freudiana: lo studio del mondo psichico. La Society for Psychical Research, per l'appunto, fondata nel 1882 e di cui facevano o faranno parte, oltre a Conan Doyle, i fisici Crookes, Lodge, Flammarion, i coniugi Curie, gli psicologi Janet, Liébeault, Bernheim, i filosofi Sidgwick, Bergson e William James, l'ideatore della teoria della selezione naturale Alfred R. Wallace, e persino Cesare Lombroso, aveva l'intento di svolgere uno studio scientifico dei fenomeni cosiddetti paranormali. Era una compagine internazionale che poche accademie potevano vantare.

[27] A.C. Doyle, *Gelseminum as a Poison*, in «The British Medical Journal», 20 settembre 1879, p. 483. Vedi anche K. Jones, I. Kelvin, *The Sherlock Holmes Pharmacopoeia: Being an Examination into the Uses and Effects of Drugs and Poisons in the Sherlock Holmes Stories*, Oakmagic Books, Weston-super-Mare 2002.

Col tempo, tuttavia, si venne a creare nell'associazione uno spartiacque tra coloro che giudicavano saggio continuare a indagare solamente laddove la determinazione razionale offriva una certa attendibilità – prevalentemente riguardo alla supposta trasmissione del pensiero – e coloro che invece ritenevano la ricerca psichica atta ad affrontare ogni aspetto parapsicologico, compresa la possibilità di comunicazione con i defunti. Doyle, assieme al suo nuovo mentore Wallace, apparteneva a questa frangia: finirà per diventare acritico sostenitore dell'attività dei medium[28].

Anche Freud era stato cooptato nella Society for Psychical Research. Comunicava nel 1911 al discepolo Jung che l'istituzione «mi ha invitato a porre la mia candidatura a socio corrispondente […] Primo segno di interesse di questa *dear old England*. La *List of members* è veramente formidabile»[29]. L'orgoglio che trapelava dalle sue parole era motivato dalla constatazione che la psicoanalisi, vista in patria come una dottrina «scandalosa», otteneva riconoscimento all'estero. Più tardi però, colto dal timore di compromettersi e di essere accomunato agli spiritisti, Freud aveva puntato i piedi, da un lato cercando di spiegare la telepatia con la psicoanalisi stessa, dall'altro dichiarando che l'associazione era «riuscita a estorcermi un articolo inglese sull'inconscio», ma esso, «naturalmente,» non diceva «nulla di nuovo»[30].

Nemmeno in questa circostanza Freud e Doyle entrarono dunque in contatto diretto, nondimeno possediamo ormai abbastanza elementi per provare a tirare le somme. Se Hillman ha affermato che Freud era uno scrittore con la maschera del medico, abile autore di romanzi terapeutici[31], potremmo a nostra volta dire che Conan

[28] Tale argomento è trattato nel mio *Il cranio di cristallo. Evoluzione della specie e spiritualismo*, Bollati Boringhieri, Torino 1993, pp. 134-175.
[29] *Lettere tra Freud e Jung* (1974), tr. it. Boringhieri, Torino 1974, p. 426.
[30] *Op. cit.*, p. 522; S. Freud, *Nota sull'inconscio in psicoanalisi* (1912), in *Opere*, cit., VI, pp. 573-581.
[31] J. Hillman, *Le storie che curano*, cit., p. 8.

Doyle era un medico con la maschera di scrittore? Sarebbe eccessivo, forse, e semmai potrebbe valere per il Doyle precedente la sua fascinazione per i fantasmi. Ma sarebbe altrettanto generico sostenere semplicemente che ambedue appartenevano alla categoria di medici che furono anche artisti, da Čechov a Bulgakov, da Cronin a Mario Tobino. Tra Freud e Doyle esisteva un particolare, inconsapevole affratellamento metodologico, esistenziale e di formazione, che piuttosto ci autorizza a riunirli sotto il motto che fu di Cartesio: *larvatus prodeo*. I due autori, in definitiva, «avanzavano mascherati», per scrutare la realtà senza condizionamenti, per costruire liberamente altri mondi, forse non meno veri[32].

[32] R. Descartes, *Cogitationes privatae* (1619), 1.

S&F_n. 27_2022

ALTERAZIONI

GIOVANNI ALTADONNA

LA CRITICA ALL'ANTROPOLOGIA RAZZIALE IN STEPHEN JAY GOULD

1. Osservazioni *di uno scienziato sulla dismisura umana*
2. *Un precedente importante:* Ontogenesi e filogenesi
3. Intelligenza e pregiudizio: *contesto storico-epistemologico*
4. *Gould contro Morton: pregiudizio?*

ABSTRACT: *THE CRITIQUE OF RACIAL ANTHROPOLOGY IN STEPHEN JAY GOULD*
In this article I aim to draw a summary of Stephen Jay Gould's critique of racial anthropology, part of the more general contestation of biological determinism and the ideological use of biology. It is highlighted how this theme has represented a constant commitment, both intellectual and civil, for the American paleontologist, with particolar reference to Ontogeny and phylogeny and to The Mismeasure of Man. Finally, it is briefly described the controversy surrounding Gould's critique of Samuel G. Morton, which shows the long range of Gould's work and the importance of the epistemological problem of the relationship between data and expectation.

1. *Osservazioni di uno scienziato sulla dismisura umana*

«Alla memoria di mia nonna e di mio nonno, che qui vennero, lottarono e prosperarono, nonostante Goddard»[1]. È introdotto da questa epigrafe *Intelligenza e pregiudizio* – titolo originale *The Mismeasure of Man* – fondamentale saggio di Stephen Jay Gould, una delle figure più importanti della storia della

[1] S.J. Gould, *Intelligenza e pregiudizio. Contro i fondamenti scientifici del razzismo* (1996), tr. it. Il Saggiatore, Milano 2016.

scienza del XX secolo[2]. Egli, paleontologo di professione, era convinto che la storia della scienza avesse molto da insegnare agli stessi scienziati circa la natura della loro impresa. In particolare, Gould non perse mai occasione per criticare le teorie deterministiche e riduzioniste in biologia; accompagnando sempre alle precise ragioni epistemologiche della contestazione il monito verso le implicazioni sociali e politiche della scienza. Non senza ragioni personali e biografiche.

Infatti, Stephen Jay Gould era nipote di immigrati *yiddish*, ebrei ungheresi, negli Stati Uniti. Egli non fa mistero del fatto che una delle ragioni della stesura di *Intelligenza e pregiudizio* fosse proprio il ricordo dei propri nonni e di quanti, come loro, affrontarono l'emigrazione verso gli Stati Uniti in un periodo storico (i primi decenni del Novecento) in cui la società civile e la politica rinunciarono *de facto* ai principi di uguaglianza e libertà che *de jure* fondano la nazione americana. Ciò avvenne nel nome di supposte verità scientifiche, le quali miravano a preservare la "razza" e lo stile di vita *wasp* contro tutte le categorie giudicate "inferiori": dai neri alle donne, dagli immigrati dell'Europa mediterranea e orientale ai (presunti) malati di mente[3].

La critica all'antropologia razziale che Gould conduce in *Intelligenza e pregiudizio* rappresenta solo il momento culminante di una riflessione che attraversa il suo intero itinerario intellettuale: la critica al determinismo biologico e al rapporto fra scienza e ideologia. Questo articolo intende ripercorrere brevemente soltanto alcuni aspetti di questo complesso argomento. Saranno esposte: le prime istanze contro il determinismo biologico in *Ontogenesi e filogenesi*; le implicazioni epistemologiche del

[2] Cfr. R. York, B. Clark, *The science and humanism of Stephen Jay Gould*, Monthly Review Press, New York 2011; A. Ottaviani, *Stephen Jay Gould*, Ediesse, Roma 2012; F. Civile, B. Danesi, A.M. Rossi (a cura di), *Grazie brontosauro! Per Stephen Jay Gould*, ETS, Pisa 2012.
[3] Cfr. S.J. Gould, *op. cit.*, pp. 30-34.

tema in *Intelligenza e pregiudizio*; la controversia circa la critica di Gould a Samuel G. Morton.

2. *Un precedente importante:* Ontogenesi e filogenesi

La prima opera importante in cui Gould valuta le implicazioni ideologiche della biologia è *Ontogenesi e filogenesi*, il suo primo trattato specialistico. Esso è dedicato allo studio storico e teorico della teoria della ricapitolazione (o teoria biogenetica) di Ernst Haeckel, secondo cui l'ontogenesi (ovvero lo sviluppo embrionale) ripercorre pedissequamente le tappe della filogenesi (cioè della storia evolutiva), e di conseguenza i tratti giovanili di una certa specie corrisponderebbero a quelli adulti di un progenitore. In estrema sintesi, la tesi di Gould è che, pur ammettendo la confutazione della teoria di Haeckel, di essa sia opportuno conservare e sviluppare due nozioni fondamentali per le scienze biologiche: l'*eterocronia* – «le mutazioni nei tempi relativi di apparizione e nel ritmo dello sviluppo dei caratteri già presenti nei progenitori»[4] – e la *pedomorfosi*, ovvero la conservazioni di tratti fenotipici infantili nell'adulto, per ritardo (*neotenia*) o accelerazione (*progenesi*) dello sviluppo somatico.

Ontogenesi e filogenesi contiene un intero capitolo (il quinto) espressamente dedicato agli usi della teoria della ricapitolazione come supporto "scientifico" alla dottrina del determinismo biologico. Esso mira infatti a evidenziare «che le arcane ricerche degli accademici di clausura possono essere pregne di rilevanza sociale, e che il distacco scientifico e l'oggettività assoluta sono solo un mito»[5]. Il titolo, *Influenza dilagante*, è più che appropriato per definire la pervasività della teoria di Haeckel in campi diversi da quello della biologia evoluzionistica: «sospetto che la sua influenza, in quanto esportata dalla teoria evolutiva

[4] Id., *Ontogenesi e filogenesi* (1977), tr. it. Mimesis, Milano-Udine 2013, p. 16.
[5] *Ibid.*, p. 19.

ad altri campi, fu inferiore soltanto a quella della selezione naturale nel diciannovesimo secolo»[6]. Gould estrapola degli esempi da cinque ambiti in cui la teoria della ricapitolazione ebbe un peso determinante: l'antropologia criminale, il razzismo, lo sviluppo infantile, l'educazione primaria e la psicoanalisi. Ai fini della nostra analisi è interessante ripercorrere brevemente i primi due.

L'antropologia criminale della scuola di Lombroso fondò l'origine ereditaria e innata della criminalità in certi individui proprio sulla teoria della ricapitolazione. Il delinquente è, secondo questa dottrina, l'esponente di un passato evolutivo distante e violento, che si manifesta in un mondo civile e progredito a partire da certe caratteristiche tipiche: gli atavismi. In conformità alla teoria di Haeckel, il criminale non è però soltanto un primitivo in una società di uomini moderni (filogenesi), ma possiede anche, in maniera permanente, la violenza tipica dello stadio infantile che gli individui normali superano durante lo sviluppo (ontogenesi)[7].

Gli esempi più significativi dell'applicazione ideologica della teoria della ricapitolazione sono però relativi, secondo Gould, al razzismo scientifico: «Pochi legami sono più intimi e pervasivi di quelli tra il razzismo e le affermazioni degli scienziati (non le chiamerei affermazioni scientifiche) sulla diversità umana prima della Seconda Guerra Mondiale. Di conseguenza, non c'è da stupirsi del fatto che il primo argomento per la ricapitolazione in morfologia fu formulato secondo un'impronta razzista»[8]. La teoria della ricapitolazione fornisce, nelle sue conseguenze politiche, una giustificazione razionale del colonialismo e delle discriminazioni razziali in generale:

> Per chiunque voglia affermare l'ineguaglianza innata delle razze, pochi argomenti biologici possono essere più attraenti della ricapitolazione, con la sua insistenza sul fatto che i bambini delle

[6] *Ibid.*, p. 113.
[7] Cfr. *ibid.*, pp. 117-118.
[8] *Ibid.*, p. 121.

> razze superiori (invariabilmente i propri) passano attraverso e oltre le condizioni permanenti degli adulti nelle razze inferiori. Se gli adulti delle razze inferiori sono come i bambini bianchi, allora devono essere trattati come tali – sottomessi, disciplinati, guidati (oppure, nella tradizione paternalistica, educati ma ugualmente sottomessi). L'argomento del "primitivo-bambino" fu uno dei più importanti nell'arsenale degli argomenti razzisti forniti dalla scienza per giustificare la schiavitù e l'imperialismo[9].

Gould non manca di sottolineare come l'imporsi delle pretese scientifiche del razzismo mostri quanto sia frequente, nell'affermazione ideologica di una teoria scientifica, il ricorso all'argomento positivistico dell'oggettività del metodo induttivo come garanzia di ottenere una teoria *libera-da-valori*. Sennonché l'ironia tragica di questa illusione è che gli stessi dati che avrebbero dovuto mostrare una gerarchia di facoltà mentali innate fra le popolazioni umane sono *privi di valore*[10].

Dopo aver esposto una carrellata di esempi relativi a questo tema, Gould precisa: «non sto selezionando le affermazioni pazzesche di un periodo passato. Sto citando le opere principali di leader riconosciuti. L'influsso del determinismo sociale, la mancanza di sensibilità nei confronti dell'influenza dell'ambiente, e il desiderio manifesto di incoronare il proprio gruppo come biologicamente superiore sono caratteristiche del tempo – e ancora oggi non sono estinte»[11]. La conformità della prassi politica agli assunti razziali tratti dalla teoria della ricapitolazione, specifica Gould, non fu accidentale: ciò «è testimoniato da due tipi di evidenze. Innanzitutto, molti scienziati trassero esplicitamente conclusioni politiche nei loro libri di amplia diffusione. […] In secondo luogo, molti politici e uomini di stato utilizzarono l'argomento del primitivo-bambino facendo un inchino esplicito alla scienza e alla ricapitolazione»[12].

Al fine di evidenziare ulteriormente il peso decisivo dei pregiudizi razziali nei primi decenni del Novecento, e più in

[9] *Ibid.*, p. 122.
[10] Cfr. *ibid.*, p. 123.
[11] *Ibid.*, p. 125.
[12] *Ibid.*, pp. 125-126.

generale l'inconsistenza del paradigma empirista, Gould evidenzia come la diffusione dell'idea di pedomorfosi in luogo della ricapitolazione avrebbe dovuto condurre gli scienziati a conclusioni diametralmente opposte circa la supremazia biologica dei bianchi nei confronti delle altre "razze". Se, in accordo alla pedomorfosi, gli individui evolutivamente più "progrediti" sono quelli che da adulti mantengono caratteristiche somatiche infantili, bisognava concludere che gli adulti neri, assomigliando ai bambini bianchi, dovevano essere superiori agli adulti caucasici. Osserva Gould:

> Inutile a dirsi, non accadde niente del genere. Ho presentato questa proposta di ciò che sarebbe potuto accadere non come un'ipotesi plausibile, ma semplicemente per illustrare con un po' di retorica l'assurdità di ogni affermazione secondo cui gli scienziati agiscono "oggettivamente" su questioni così vitali per gli interessi dei loro padroni (e per la loro stessa posizione privilegiata). Infatti, i fondatori della pedomorfosi dimenticarono tranquillamente tutti i vecchi dati sugli "adulti neri simili ai bambini bianchi" e cominciarono a cercare delle informazioni opposte che avrebbero riaffermato il razzismo sulla base dell'opposto modello pedomorfico[13].

Alla fine del capitolo, Gould osserva che la teoria della ricapitolazione implica il determinismo biologico in tutti i contesti in cui è stata applicata. Inoltre, esso può essere declinato ideologicamente sia in senso "liberale" (nel caso della psicologia dello sviluppo e della pedagogia) che in senso "conservatore" (da parte dell'antropologia criminale e razziale):

> Il tema comune è il determinismo biologico. Tutti i sostenitori della ricapitolazione l'hanno usato per fare affermazioni biologiche sulla natura umana e per difendere la nozione di inevitabilità per alcuni aspetti del comportamento e dello status sociale. […] Considerare l'innato equivalente all'inevitabile conduce quasi sempre a giustificare lo *status quo* – poiché questo non sarebbe sorto se fosse stato in opposizione con la nostra natura.[14]

È interessante osservare che il capitolo IV di *Intelligenza e pregiudizio, La misurazione dei corpi. Due casi della natura scimmiesca degli indesiderabili*[15] è dedicato proprio alla teoria della ricapitolazione e all'antropologia criminale di Lombroso

[13] *Ibid.*, p. 128.
[14] *Ibid.*, pp. 153-154.
[15] Cfr. Id., *Intelligenza e pregiudizio*, cit., pp. 145-179.

quali teorie quantitative di ispirazione evoluzionistica volte a discriminare su base biologica i popoli "inferiori", le donne, i poveri, i criminali. Ciò mostra chiaramente la portata di *Ontogenesi e filogenesi* per il successivo itinerario intellettuale di Gould, non solo per quanto pertiene al tema generale del determinismo biologico, ma anche riguardo all'impiego di particolari argomenti e concetti. Infatti, anche l'importanza della neotenia quale fattore fondamentale dell'evoluzione biologica e culturale umana sarà oggetto di particolare attenzione nel capitolo conclusivo di *Intelligenza e pregiudizio* (*Una conclusione positiva*)[16] dove Gould scrive: «Non conserviamo soltanto lo stampo anatomico dell'infanzia, ma pure la sua flessibilità mentale. L'idea che la selezione naturale dovrebbe aver lavorato per la flessibilità nell'evoluzione animale non è un concetto *ad hoc* nato per la speranza ma un'implicazione della neotenia come processo fondamentale della nostra evoluzione. Gli uomini sono animali che apprendono»[17]. Quanto al determinismo biologico come dottrina volta a consolidare lo *status quo*, la quale condiziona le possibilità di realizzazione individuale di molte persone, sempre in *Intelligenza e pregiudizio* Gould afferma: «Il determinismo biologico è, nella sua essenza, una *teoria dei limiti*. Essa considera lo stato corrente di gruppi come una misura di dove essi dovrebbero e devono stare (anche se permette ad alcuni rari individui di elevarsi in conseguenza della loro fortunata biologia)»[18].

Gould conclude il capitolo quinto di *Ontogenesi e filogenesi*, che come abbiamo cercato di mostrare può essere a buon diritto considerato un illustre precedente di *Intelligenza e pregiudizio*, con una partecipata e decisa presa di posizione contro il determinismo biologico, dal tono quasi autobiografico:

[16] Cfr. *ibid.*, pp. 371-384.
[17] *Ibid.*, p. 384.
[18] *Ibid.*, pp. 56-57.

> Sappiamo oggi che tutti gli argomenti citati in questo capitolo sui vincoli basati sulla ricapitolazione sono sostanzialmente errati. Eppure essi ebbero grande importanza all'epoca, perché la scienza si esprimeva in loro favore e l'asserzione liberale di una potenzialità non vincolata puzzava di sentimentalismo. Gli antideterministi hanno sempre questa croce da portare, ma il collasso degli argomenti ricapitolazionisti sulle potenzialità umane suggerisce che la storia potrebbe essere dalla loro parte[19].

3. Intelligenza e pregiudizio: *contesto storico-epistemologico*

Intelligenza e pregiudizio può essere considerato il culmine dell'attenzione di Gould verso i rapporti complessi fra l'obiettività in cui si impegna la scienza e le sue componenti extra-scientifiche. In questo libro il vasto tema del determinismo biologico, ovvero la dottrina per cui «le norme comportamentali comuni e le differenze sociali ed economiche tra i gruppi umani – in primo luogo razze, classi e sessi – derivano da distinzioni innate ereditate, e che la società, in questo senso, è un esatto riflesso della biologia»[20], è circoscritto attraverso "tre cornici teoriche" a un'indagine sulla giustificazione delle differenze mentali tra i gruppi umani operata dagli scienziati nel corso dell'Ottocento e ancora nel Novecento[21]. Il libro non affronta in generale il tema vastissimo del rapporto fra scienza e pregiudizi sociali, bensì «tratta *una forma particolare* di rivendicazione *quantificata* sulla classificazione dei gruppi umani: la tesi secondo cui l'intelligenza può essere correttamente rappresentata da un solo numero, in grado di classificare gli uomini secondo la scala lineare di un valore mentale connaturato e inalterabile»[22]. Gould ritiene che la critica al determinismo biologico sia «senza tempo e insieme legata al momento»[23]. Essa ha validità universale in quanto gli argomenti deterministi si fondano su fallacie epistemologiche quali:

> il *riduzionismo*, ovvero il desiderio di spiegare vasti fenomeni irriducibilmente complessi e in parte casuali con il comportamento

[19] Id., *Ontogenesi e filogenesi*, cit., p. 154.
[20] Id., *Intelligenza e pregiudizio*, cit., p. 48.
[21] Cfr. *ibid.*, pp. 13-19.
[22] *Ibid.*, p. 12.
[23] *Ibid.*, p. 19.

> deterministico delle singole e più piccole parti costituenti […]; la *materializzazione*, ovvero la propensione a convertire un concetto astratto (come l'intelligenza) in un'entità concreta […]; la tendenza a *ragionare per dicotomie*, ossia la nostra inclinazione ad analizzare la realtà complessa e ininterrotta attraverso coppie di opposti […]; e la *gerarchia*, ossia la nostra inclinazione a ordinare gli elementi classificandoli secondi una scala lineare di valore crescente […][24].

Ma essa è anche congiunturale nella misura in cui tali motivi riaffiorano in corrispondenza di certe condizioni sociopolitiche che li favoriscono[25]. La scelta di approntare una edizione riveduta del libro è legata proprio a una ragione occasionale: la pubblicazione, nel 1994, di *The Bell Curve* di Herrnstein e Murray, la più rilevante difesa del determinismo biologico dell'intelligenza apparsa in tempi recenti.

Un aspetto della questione che Gould tiene a sottolineare in quanto più volte frainteso riguarda il dualismo natura-cultura. Egli sostiene: «Di tutte le false e fuorvianti dicotomie che ostacolano la nostra comprensione della complessità del mondo, "natura contro cultura" deve classificarsi fra le prime due o tre (ma è un'opposizione fasulla, enfatizzata solo dall'eufonia dei sostantivi)»[26]. Ebbene, una strategia retorica usata sovente dai deterministi per difendersi dalle accuse è precisare la percentuale di componente innata (biologica) e componente appresa (culturale) dell'intelligenza. Gould osserva in proposito che trincerarsi dietro tale argomento significa non cogliere il nucleo della questione:

> I problemi davvero rilevanti sono la flessibilità e la malleabilità, non le fallaci ripartizioni in percentuali. Un tratto può essere per il 90 per cento ereditario, eppure del tutto malleabile. Un paio di occhiali da venti dollari acquistato nella farmacia locale può correggere completamente un difetto della vista ereditario al 100 per cento. Un biodeterminista "al 60 per cento" non è un moderato interazionista, ma un determinista di quelli "piuttosto innocui"[27].

[24] *Ibid.*, pp. 19-20.
[25] Cfr. *ibid.*, p. 21.
[26] *Ibid.*, p. 27.
[27] *Ibid.*, p. 28.

Nel primo capitolo[28], Gould afferma chiaramente che, piuttosto che postulare un'illusoria dicotomia fra cattivi deterministi e scienziati obbiettivi, *Intelligenza e pregiudizio* ha come scopo la critica contro «il mito che la scienza stessa sia un'impresa oggettiva, fatta correttamente solo quando gli scienziati possono uscire dal guscio della loro cultura e osservare il mondo com'è realmente»[29]. Questo perché «la scienza, dal momento che viene praticata dall'uomo, è un'attività socialmente inserita»[30] e pertanto il rilevamento di pregiudizi inconsci nell'attività dei deterministi biologici deve essere intesa come particolare aspetto dell'assunto più generale per cui «la scienza debba essere compresa come un fenomeno sociale, come una coraggiosa impresa umana, non come il lavoro di automi programmati per raccogliere informazione pura»[31]. Tale convinzione, precisa Gould, non equivale ad abbracciare una visione relativistica della scienza: «ritengo che una realtà fattuale esista e che la scienza, sebbene in maniera spesso ottusa ed errata, possa imparare da essa»[32]. Eppure, quando la scienza si concentra su temi di particolare rilevanza sociale e politica, essa è quasi inevitabilmente influenzata dal contesto culturale e da assunzioni concettuali errate ma persistenti nella tradizione filosofica occidentale, quali la *reificazione* della coscienza e la *classificazione gerarchica*. Entrambe queste nozioni compongono un terzo errore epistemologico, ovvero l'idea che l'intelligenza sia *un'entità misurabile*, secondo i metodi della craniometria nel XIX secolo e mediante l'uso dei test mentali nel XX secolo.

> Questo libro riguarda, quindi, l'astrazione dell'intelligenza come entità singola, la sua collocazione dentro il cervello, la sua quantificazione in un numero per ogni individuo e l'uso di questi numeri per classificare le persone in una singola serie di valore, per trovare invariabilmente che i gruppi oppressi e svantaggiati – razze, classi o sessi – sono innatamente inferiori e meritano il loro stato.

[28] Cfr. *ibid.*, pp. 47-58.
[29] *Ibid.*, p. 49.
[30] *Ibid.*, p. 50.
[31] *Ibid.*
[32] *Ibid.*

> In breve, questo libro riguarda l'erronea misurazione dell'uomo, così come è stata condotta fino ad oggi[33].

Gould precisa poi il suo metodo di lavoro interdisciplinare: l'analisi dei calcoli degli illustri scienziati dei secoli scorsi per denunciare la falsa pretesa di oggettività dei dati numerici. «Se – come credo di aver dimostrato – i dati quantitativi sono soggetti alla costrizione culturale come ogni altro aspetto della scienza, allora non devono avere nessuna speciale pretesa sulla verità finale»[34].

Fra i vari capitoli che compongono *Intelligenza e pregiudizio*, in questa sede ci occupiamo in particolare del secondo: *La poligenesi e la craniometria in America prima di Darwin. Neri e indiani considerati come specie separate e inferiori*[35]. Esso ha per oggetto l'antropologia razziale fra Settecento e Ottocento, caratterizzata dal dibattito fra monogenismo e poligenismo: teorie che sostengono, rispettivamente, un'origine comune o separata dei gruppi umani. La poligenesi postulava l'esistenza delle razze umane come il risultato di atti di creazione separati: essa era la tesi "dura" del razzismo pre-darwiniano. Sebbene meno diffusa rispetto alla teoria monogenista (in quanto implicava un'interpretazione non letterale della Genesi), la poligenesi riscosse sostenitori di spicco: nel XVIII secolo, David Hume e Charles White nell'ambiente anglosassone, e soprattutto Louis Agassiz e Samuel George Morton, a metà Ottocento, in quello americano.

Proprio il medico di Philadelphia Samuel G. Morton era ben noto ai suoi contemporanei per aver radunato un'imponente collezione di crani umani allo scopo di misurare la capacità cerebrale delle razze, con l'assunto di verificare la gerarchia dell'intelligenza fra quelle che egli riteneva specie umane separate. «Morton iniziò la sua collezione di crani umani negli anni venti; ne aveva più di

[33] *Ibid.*, p. 53.
[34] *Ibid.*, p. 55.
[35] Cfr. *ibid.*, pp. 59-103.

mille quando morì nel 1851. Amici (e nemici) si riferivano al suo grande ossario come al "Golgota americano"»[36]. Il suo metodo, che dava per implicita un'equivalenza fra dimensioni del cervello e intelligenza individuale, si basava sul riempimento della cavità cranica con semi di senape e, in seguito, con pallini di piombo (che restituivano risultati meno incerti). Le sue conclusioni, riassunte in tabelle indicanti la capacità cranica media per ogni razza, sono esposte in tre volumi: *Crania Americana* (1839) sugli indiani d'America, *Crania Aegyptiaca* (1844) sui crani delle tombe egizie, e infine la tabulazione complessiva dei dati tratti dalla sua collezione (1849). Queste tavole

> rappresentano il maggior contributo della poligenesi americana ai dibattiti sull'ordinamento razziale. Esse sono sopravvissute alla teoria delle creazioni separate e sono state ristampate ripetutamente nel corso del XIX secolo come dati "rigorosi" e irrefutabili sul valore intellettuale delle razze umane. Senza bisogno di dirlo, esse si accordavano con i pregiudizi di ogni buon Yankee: i bianchi in cima, gli indiani nel mezzo e i neri in fondo; e, tra i bianchi, teutoni e anglosassoni in cima, ebrei nel mezzo e indù in fondo[37].

Dal momento che Morton non omise di segnalare le misurazioni "grezze" nei suoi resoconti, Gould ha potuto rianalizzarle, scoprendo una manipolazione dei dati significativa (ed evidentemente involontaria, altrimenti non si spiegherebbe perché Morton non nascose i dati che avrebbero potuto smentirlo). Questa osservazione ci riconduce direttamente alla cifra epistemologica del testo di Gould:

> La frode cosciente è probabilmente rara nella scienza. Non è neanche molto interessante, perché essa ci dice poco sulla natura dell'attività scientifica. I mentitori, se scoperti, vengono scomunicati; gli scienziati dichiarano che la loro professione si è vigilata appropriatamente e ritornano al lavoro con la mitologia intatta e obiettivamente vendicata. La prevalenza di mistificazioni *inconsce*, d'altro canto, suggerisce una conclusione generale sul contesto sociale della scienza. Perché, se gli scienziati possono essere onestamente autoingannati sino al livello di Morton, allora il pregiudizio sottostante può essere trovato ovunque, anche nei fondamenti del misurare ossa e fare somme[38].

[36] *Ibid.*, p. 81.
[37] *Ibid.*, p. 84.
[38] *Ibid.*, p. 85.

Gould esamina dettagliatamente le fallacie e le omissioni contenute nei calcoli di Morton nelle tre opere sopra citate, allo scopo di far emergere non solo gli errori personali di Morton, ma delle essenziali e impreviste fragilità epistemiche, potenzialmente presenti in ogni impresa scientifica[39].

In particolare, Gould evidenzia che la discrepanza fra le misurazioni con i semi di senape e con i pallini di piombo (i quali riducono significativamente il margine di errore), appare prevedibilmente maggiore per i gruppi umani considerati inferiori, mentre è minore per i bianchi:

> Scenari plausibili sono facili da costruire. Morton, misurando con i semi, prende un cranio di nero terribilmente grande, lo riempie senza pigiare e lo scuote appena. Prende poi un cranio di bianco penosamente piccolo, lo scuote energicamente e pressa con forza con il pollice nel *foramen magnum*. È facilmente fatto, senza motivazione cosciente; le aspettative sono una potente guida all'azione[40].

Dopo aver riscontrato analoghe distorsioni dei dati anche nella tabulazione finale del 1849, Gould conclude: «In breve la mia correzione dell'ordinamento convenzionale di Morton *non* rivela differenze significative tra le razze in base agli stessi dati forniti da Morton»[41]. In definitiva, egli individua quattro categorie di errori inconsapevoli in Samuel G. Morton: 1. «inconsistenze propizie e criteri mutevoli»; 2. «soggettività diretta verso un pregiudizio aprioristico»; 3. «omissioni procedurali apparentemente ovvie»; 4. «calcoli errati e opportune omissioni»[42].

4. *Gould contro Morton: pregiudizio?*

La critica di Gould all'opera di Morton, sviluppata nel secondo capitolo di *Intelligenza e pregiudizio* a partire da un articolo

[39] Cfr. *ibid.*, pp. 87-100.
[40] *Ibid.*, p. 96.
[41] *Ibid.*, p. 98.
[42] *Ibid.*, pp. 98-99.

dello stesso Gould pubblicato su "Science" nel 1978[43], è stata oggetto di una diatriba iniziata nel 2011 e continuata fino a tempi recentissimi.

Essa è stata innescata da un articolo a firma di un gruppo di antropologi americani coordinati da Jason E. Lewis, intitolato (in evidente parodia rispetto al titolo originale del libro di Gould): *The Mismeasure of Science: Stephen Jay Gould versus Samuel George Morton on Skulls and Bias*[44]. Tale studio è rivolto a rispondere alla domanda: «did Morton allow his a priori views on human variation to impact the data and analyses he published, as Gould argues?»[45]. La risposta degli autori è negativa: «Our results resolve this historical controversy, demonstrating that Morton did not manipulate data to support his preconceptions, contra Gould. In fact, the Morton case provides an example of how the scientific method can shield results from cultural biases»[46]. Essi asseriscono ciò sulla base di una revisione delle analisi di Gould sulle misurazioni di Morton e, soprattutto, di una nuova misurazione di «almost half of the skulls that Morton had originally measured»[47]. Gli autori concludono che, «Ironically, Gould's own analysis of Morton is likely the stronger example of a bias influencing results»[48]. Essi considerano la loro analisi in pieno accordo con i presupposti metodologici di Gould, tanto da affermare che, se egli avesse potuto leggerla, l'avrebbe accolta come un appropriato esercizio epistemologico, sebbene contrario alle proprie conclusioni. Lewis e colleghi sostengono che Morton, nonostante Gould, può essere un buon esempio di come, sebbene gli scienziati possano essere soggetti al pregiudizio, i metodi della

[43] Cfr. Id., *Morton's ranking of races by cranial capacity: Unconscious manipulation of data may be a scientific norm*, in «Science», 200, 1978, pp. 503-509.
[44] Cfr. J.E. Lewis, D. DeGusta, M.R. Meyer, J.M. Monge, A.E. Mann, R.L. Holloway, *The Mismeasure of Science: Stephen Jay Gould versus Samuel George Morton on Skulls and Bias*, in «PLoSBiol», 9(6), 2011: e1001071.
[45] *Ibid.*, p. 1.
[46] *Ibid.*
[47] *Ibid.*
[48] *Ibid.*, p. 5.

scienza in quanto tali possono costituire una garanzia contro il rischio di deformazioni dettate dalle aspettative: «The Morton case, rather than illustrating the ubiquity of bias, instead shows the ability of science to escape the bounds and blinders of cultural contexts»[49].

Il *paper* di Lewis e colleghi ha sollevato diverse reazioni. Il giornalista Nicholas Wade, in un articolo del 2011 per il *New York Times*, ha posto l'accento sulla nuova misurazione dei crani della collezione di Morton come prova del fatto che lo studio dello scienziato ottocentesco non fosse condizionato dal pregiudizio. Nello stesso articolo sono riportati alcuni giudizi di Lewis e colleghi su Gould. Se il primo autore, l'antropologo di Stanford Jason E. Lewis, non si sbilancia nel giudicare deliberato o meno il (presunto) fraintendimento di Gould su Morton, il collega Ralph L. Hollolay della Columbia University è molto meno accomodante, definendo Gould «a charlatan», viziato dall'ideologia[50]. D'altra parte, come è stato evidenziato da Rickards e Biondi nel 2011, Lewis e colleghi hanno effettuato le misurazioni su solo 308 dei 670 crani studiati da Morton, un campione insufficiente per verificare o smentire alcunché[51].

La prima risposta all'articolo di Lewis e colleghi apparsa in sede accademica risale al 2014 e si deve a Michael Weisberg il quale, pur evidenziando alcune incongruenze nella critica di Gould a Morton, ritiene la sua tesi della distorsione inconsapevole dei dati nell'opera dello scienziato ottocentesco del tutto valida:

> […] Gould's charge that Morton's analyses exhibit racial bias seems well-justified. Gould made some analytical errors which were uncovered by Lewis et al., but his two most important claims – that there is evidence that Morton's seed-based measurements exhibit racial bias and

[49] *Ibid.*, p. 6.
[50] N. Wade, *Scientists Measure the Accuracy of a Racism Claim*, in «New York Times», June 14, 2011, section D, p. 4.
[51] Cfr. O. Rickards, G. Biondi, *Gould si è sbagliato su Morton?*, pubblicato il 07/07/2011 su https://www.scienzainrete.it/articolo/gould-si-%C3%A8-sbagliato-su-morton/olga-rickards-gianfranco-biondi/2011-07-07 (ultimo accesso il 25/02/2021).

> that there are no significant differences in mean cranial capacities across races in Morton's collection – are sound[52].

Un'altra replica a Lewis e colleghi è stata avanzata nel 2015 da Jonathan M. Kaplan, Massimo Pigliucci e Joshua A. Banta[53]. La tesi da loro sostenuta è che, da un lato, Lewis e colleghi siano nel torto nel riabilitare Morton e dunque la critica di Gould a quest'ultimo sia pertinente, ma dall'altro essi non ritengono che questa offra un'alternativa convincente. Sia Lewis e colleghi sia Gould condividerebbero la convinzione che i dati di Morton possano essere interpretati in qualche modo, ora in modo corretto, ora distorti da un pregiudizio inconsapevole. Kaplan e colleghi, piuttosto, ritengono che, quand'anche il campione esaminato da Morton e revisionato da Gould e da Lewis e colleghi sia esaustivo (e non lo è), da un suo riesame non si potrebbe ricavare alcuna informazione utile dal momento che non risulterebbe chiaro in che termini classificare le eventuali variazioni riscontrate (dal momento che i confini fra "popolazioni" umane non sono ben definite e la stessa classificazione che Morton diede ai suoi crani presenta delle ambiguità). Essi scrivono:

> […] while we agree with Lewis et al. That Gould's statistical analysis of Morton's data is in many ways no better than Morton's own, we believe that Lewis et al.'s work is at least equally problematic. Gould was, in ourview, right to recognize that there was something very wrong with Morton's analysis; but he went wrong himself in trying to find a "better" analysis. Lewis et al. are right that Gould's analysis isn't better, but wrong to think that Morton's is appropriate[54].

Kaplan *et al.* riscontrano diverse criticità metodologiche nel lavoro di Lewis e colleghi: essi (così come Michael Weisberg e Diane B. Paul)[55] ritengono che la nuova misurazione dei crani di Morton da parte del team di Lewis, sebbene abbia ricevuto ampia

[52] M. Weisberg, *Remeasuring man*, in «Evolution&Development», XVI, 3, 2014, pp. 166-178. Il testo citato si trova a p. 177.
[53] J.M. Kaplan, M. Pigliucci, J.A. Banta, *Gould on Morton, Redux: What can the debate reveal about the limits of data?*, in «Studies in History and Philosophy of Biological and BiomedicalSciences», 30, 2015, pp. 1-10.
[54] *Ibid.*, p. 1.
[55] Cfr. M. Weisberg, *Remeasuring man*, cit.; M. Weisberg, D.B. Paul, *Morton, Gould, and Bias: A Comment on "The Mismeasure of Science"*, in «PLoSBiol» 14(4), 2016: e1002444.

visibilità dai media come prova dell'attendibilità delle loro conclusioni (cfr. *supra* l'articolo di Wade), non costituisca l'aspetto più importante del loro lavoro; anzi, affermano che esso sia del tutto irrilevante rispetto alla critica a Gould. Come ricordano Kaplan e colleghi, infatti, lo stesso Gould aveva ammesso che «le ricalibrature non variavano mai più di un pollice cubo e possiamo accettare il giudizio di Morton secondo cui le misure con i pallini erano obiettive, accurate e ripetibili, mentre le misure precedenti ottenute con i semi erano altamente soggettive e irregolari»[56]. Dunque la tesi che Lewis e colleghi attribuiscono a Gould, per cui Morton avrebbe erroneamente misurato i crani (anche) col metodo dei pallini di piombo, è inconsistente: proprio perché i dati di Morton ricavati col secondo metodo risultano corretti, Gould ha potuto rintracciare la discrasia fra i valori ottenuti con le due misurazioni e, incidentalmente, verificare che il divario maggiore si riscontra nel caso degli africani e ipotizzare così l'inconsapevole sottostima della capacità craniale delle "razze" non-caucasiche[57]. Kaplan e colleghi si soffermano altresì su rappresentazioni distorte e assurdità riportate da Lewis e colleghi a proposito delle tesi di Gould; ad esempio quella per cui egli avrebbe riconosciuto in Morton uno scienziato autore di consapevoli alterazioni dei dati e di misurazioni erronee; tesi, questa, insostenibile, dal momento che lo stesso Gould scrive:

> In breve, e per dirla esplicitamente, i suoi compendi [di Morton, *scil.*] sono un mosaico di fandonie e mistificazioni nel chiaro interesse di verificare convinzioni aprioristiche. Tuttavia – e questo è l'aspetto più affascinante del caso – non ho trovato alcuna prova di frode cosciente; invero, se Morton fosse stato un truffatore, non avrebbe pubblicato i suoi dati così apertamente[58].

[56] S.J. Gould, *Intelligenza e pregiudizio*, cit., p. 95.
[57] Cfr. *ibid.*, pp. 84-85; cfr. anche M. Weisberg, *Remeasuring man*, cit.; M. Weisberg, D.B. Paul, *Morton, Gould, and Bias…*, cit.; P.W. Mitchell, *The fault in his seeds: Lost notes to the case of bias in Samuel George Morton's cranial race science*, in «PLoSBiol», 16(10), 2018: e2007008.
[58] S.J. Gould, *Intelligenza e pregiudizio*, cit., pp. 84-85 e cfr. *supra*, § 3.

Dopodiché, Kaplan e colleghi si concentrano sulla tesi più impegnativa (e, a giudizio di chi scrive, più discutibile) del loro articolo: essi sostengono che vada posta una distinzione fra ciò che intendeva dimostrare Morton (da loro giudicata una questione non chiara) e ciò che secondo Gould egli intendeva mostrare, ovvero la classificazione gerarchica delle razze umane sulla base del criterio dell'intelligenza, di cui la capacità cranica è considerata un riflesso esatto[59]. Essi scrivono: «It is possible that Morton had no particolar purpose in mind»[60]. Mi sembra sia abbastanza chiaro, a partire dalle citazioni di Morton in merito riportate da Gould[61], che non possano sussistere dubbi sul fatto che la collezione di Morton e gli studi relativi avessero come esplicito intento quello di provare, se non esclusivamente, quantomeno *anche* le tesi dell'antropologia razziale nella versione poligenista[62].

L'ultimo intervento circa tale controversia è quello pubblicato da Paul W. Mitchell nel 2018[63]. Egli, tramite lo studio di fonti archivistiche, e in particolare annotazioni manoscritte inedite da parte di Morton sulle sue opere, è giunto alla conclusione che sebbene la critica di Gould alle misurazioni effettuate da Morton con i semi debba essere considerata errata, resta pienamente valido il monito di Gould circa la pervasività del pregiudizio nella scienza. La tesi di Mitchell è che, sebbene le misurazioni di Morton (comprese, *contra* Gould, quelle effettuate con il metodo dei semi) possano ritenersi oggettive, non lo sono le teorie che Morton ricava da quei dati. L'autore evidenzia a tal proposito come un contemporaneo di Morton, Friedrich Tiedemann, sostenesse l'uguaglianza delle razze umane e l'abolizione della schiavitù sulla base di metodi analoghi e di dati craniometrici simili a

[59] Cfr. J.M. Kaplan *et al.*, *Gould on Morton, Redux…*, cit., p. 5.
[60] *Ibid.*, p. 8.
[61] Cfr. S.J. Gould, *Intelligenza e pregiudizio*, cit., pp. 81-87 (*passim*).
[62] Cfr. anche M. Weisberg, *Remeasuring Man*, cit.; P.W. Mitchell, *The fault in his seeds…*, cit., p. 3.
[63] Cfr. P.W. Mitchell, *The fault in his seeds…*, cit.

quelli di Morton, negli stessi anni in cui lo scienziato di Philadelphia usava tali dati e tali metodi per affermare la poligenesi delle razze umane, la gerarchia intellettuale delle razze e l'ineluttabilità naturale della schiavitù dei neri. Mitchell a tal proposito sottolinea i limiti epistemici dei dati (tema già evidenziato da Kaplan *et al.*) e il carattere insidioso del pregiudizio all'interno del contesto psicologico, storico e culturale di elaborazione delle teorie scientifiche.

Va peraltro notato che lo stesso Gould, dedicando a Tiedemann un saggio[64], aveva additato in lui un esempio di come l'aspettativa possa condizionare l'esame dei dati non solo nel caso di scienziati razzisti e deterministi, ma anche nel caso di scienziati con ideologie diametralmente opposte. Sotto questo aspetto, la "questione" Morton rimane decisamente istruttiva: non si tratta di una diatriba fra scienziati "razzisti" viziati dall'ideologia e scienziati "liberali" fedeli all'obbiettività, quanto piuttosto un monito sull'insidiosa pervasività del pregiudizio nella scienza. A tal proposito, ritengo sia opportuno lasciare la parola allo stesso Gould il quale, in *Intelligenza e pregiudizio*, aveva precisato:

> Questo libro cerca di dimostrare le debolezze scientifiche e i contesti politici delle tesi deterministiche. Però, con ciò non intendo contrapporre cattivi deterministi che smarriscono la strada dell'obiettività scientifica e illuminati antideterministi che si accostano ai dati con mente aperta e vedono quindi la verità. Critico, piuttosto, il mito che la scienza stessa sia un'impresa oggettiva, fatta correttamente solo quando gli scienziati possono uscire dal guscio della loro cultura e osservare il mondo com'è realmente[65].

GIOVANNI ALTADONNA ha conseguito la Laurea Magistrale in Scienze Filosofiche presso l'Università degli Studi di Catania

altadonnagiovanni415@gmail.com

[64] Cfr. S.J. Gould, *Il grande fisiologo di Heidelberg*, in Id., *I Have Landed. Le storie, la Storia* (2002), tr. it. Codice, Torino 2009, pp. 403-422.
[65] Id., *Intelligenza e pregiudizio*, cit., p. 49; cfr. *supra*, § 3.

S&F_n. 27_2022

COMUNICAZIONE

DAVID BENATAR

PERCHÉ È MEGLIO NON VENIRE MAI AL MONDO[1]

ABSTRACT: ***WHY IT IS BETTER NEVER TO COME INTO EXISTENCE***
Not coming into existence would be good, since existence has no net advantage over non-existence. Indeed, the presence of pain is bad, and the presence of pleasure is good for every existent. However, an equivalent symmetrical consideration does not apply to the absence of pain and pleasure. Indeed, the absence of pain is good even if there is no one to experience that absence, while the absence of pleasure is not bad if there is nobody to be deprived of that pleasure. The argument just presented forms the core of the article, along with other considerations about the preferability of not bringing children into the world, the quality of human life, the unreliability of subjective judgments about one's existential condition, the preferability to phase human extinction

I

Nella letteratura sulle potenziali persone future c'è un presupposto condiviso per cui, a parità di condizioni, non si fa nulla di sbagliato portando al mondo persone le cui vite nel

[1] N.*d*.T. L'articolo qui presentato per la prima volta in traduzione italiana, e pubblicato nella sua edizione originale nel 1996 sulla prestigiosa Rivista *American philosophical quarterly*, rappresenta, in un certo senso, la genesi della riflessione di David Benatar sull'antinatalismo e il primo *locus philosophicus* in cui viene formulato in modo sistematico l'*asymmetry argument*, in base al quale non è possibile stabilire una simmetria, dal punto di vista *assiologico*, tra presenza di piacere e dolore e assenza di piacere e dolore. Questo argomento è stato poi ripreso e sviluppato in maniera più approfondita e articolata nel famoso testo *Better never to have been. The harm of coming into existence*, pubblicato nel 2006. In *Why it is better never to come into existence* sono però presenti *in nuce* tutte le principali linee del pessimismo benatariano: la scarsa qualità della vita umana e l'inaffidabilità dei giudizi soggettivi sulla qualità dell'esistenza, la differenza tra vite degne di cominciare e vite degne di continuare, l'estinzione umana graduale.
Ringrazio *APQ* per aver autorizzato questa traduzione e David Benatar per aver accolto con entusiasmo l'idea.

complesso saranno buone. Questa convinzione, a sua volta, si basa su un'altra idea, ossia che essere portati al mondo (con delle prospettive di vita dignitose) costituisca un beneficio (anche se non essere nati non rappresenta un danno). Tutto ciò viene presupposto senza fornire argomentazioni. Io intendo sostenere che il presupposto di fondo è sbagliato. Essere portati al mondo non è un vantaggio ma sempre un male. Molti penseranno che questa affermazione profondamente inquietante sia contro-intuitiva e cercheranno di rigettarla. Per questa ragione, propongo non solo di fornire delle argomentazioni a sostegno di questa affermazione, ma anche delle indicazioni sul perché le persone potrebbero manifestare delle resistenze verso di essa.

II

Che le cose negative accadano a tutti è un dato di fatto. Nessuna vita è priva di difficoltà. È facile pensare ai milioni di persone che vivono nella povertà o a coloro i quali vivono con delle disabilità per la maggior parte della loro vita. Alcuni di noi sono abbastanza fortunati da essere risparmiati da questi destini ma, nonostante ciò, la maggior parte di noi soffrirà, in qualche fase della propria vita, per la presenza di condizioni di salute avverse. Spesso la sofferenza è straziante, anche se ci colpisce solo negli ultimi giorni di vita. Alcuni sono condannati, a causa della loro costituzione, a vivere per anni in una condizione di fragilità. *Tutti* prima o poi ci troveremo di fronte alla morte[2]. Raramente prendiamo in considerazione le sofferenze che attendono i neonati: dolore, delusione, ansia, lutti e poi la morte. Non siamo in grado di prevedere per ciascun neonato quale forma specifica assumeranno tali sofferenze o quanto saranno gravi, ma possiamo essere certi che almeno alcune

[2] Avviene solo in casi estremamente rari, se mai avviene, che la morte sia un bene, nonostante essa sia il minore tra due mali quando la continuazione dell'esistenza risulta intollerabile.

di esse saranno presenti. (Soltanto a quelli che muoiono prematuramente alcune di esse verranno risparmiate, a parte la morte). I non-esistenti non andranno incontro a nulla di tutto ciò. Soltanto gli esistenti patiscono sofferenze. Naturalmente non ho raccontato l'intera storia. Non solo le brutte cose ma anche le belle cose accadono soltanto a coloro i quali esistono. Soltanto coloro i quali esistono possono fare esperienza di piaceri, gioie, e soddisfazioni. Pertanto, questo direbbe probabilmente un ottimista, dobbiamo soppesare i piaceri della vita rispetto alle sofferenze e dal momento che i primi superano le seconde la vita è degna di essere vissuta. Venire al mondo a queste condizioni, sulla base di tale prospettiva, rappresenta un beneficio. Tuttavia, non è affatto questa la conclusione alla quale bisogna pervenire. La ragione è che vi è una differenza decisiva tra dolori e piaceri che rende i vantaggi dell'esistenza rispetto alla non esistenza[3] vacui, mentre gli svantaggi concreti. È incontrovertibile affermare che:

1) la presenza di dolore è male

e

2) la presenza di piacere è bene

Tuttavia, non è possibile applicare la stessa valutazione simmetrica all'*assenza* di dolore e piacere dal momento che

3) L'assenza di dolore è bene anche se nessuno farà esperienza di quel bene

mentre

4) l'assenza di piacere non è male a meno che non vi sia qualcuno per il quale tale assenza rappresenti una privazione.

[3] Il termine "non esistenza" risulta ambiguo da diversi punti di vista. Si può applicare a quelli che non esisteranno mai e a quelli che attualmente non esistono. Quest'ultimo caso può essere a sua volta suddiviso tra quelli che non esistono ancora e quelli che non esistono più. Nel presente contesto sto utilizzando "non esistenza" per riferirmi a quelli che non esisteranno mai. Joel Feinberg ha sostenuto che coloro i quali non esistono ancora e coloro i quali non esistono più possono essere danneggiati. Io sono d'accordo con questa prospettiva. Quello che ho da dire qui si applica però solo a coloro i quali non esisteranno mai.

La mia prospettiva sull'asimmetria tra 3) e 4) è ampiamente condivisa. Per supportare questa affermazione possono essere fornite diverse ragioni.

Innanzitutto, questa prospettiva rappresenta la migliore spiegazione dell'idea ampiamente sostenuta per cui, mentre c'è un dovere di evitare di portare al mondo persone sofferenti, non c'è alcun dovere di portare al mondo persone felici. In altre parole, la ragione per cui pensiamo che vi sia un dovere di non portare al mondo persone sofferenti è che la presenza di quella sofferenza rappresenterebbe un male (per coloro i quali soffriranno) e l'assenza della sofferenza un bene (anche se non vi fosse nessuno a fare esperienza dell'assenza di sofferenza). Al contrario, pensiamo che non vi sia alcun dovere di portare persone felici al mondo perché mentre il loro piacere sarebbe un bene, la sua assenza non sarebbe un male (dato che non ci sarebbe nessuno a subire la privazione).

Si potrebbe obiettare che c'è una spiegazione alternativa a supporto dell'idea che esistono dei doveri procreativi, una spiegazione che non fa riferimento alla mia affermazione sull'asimmetria tra 3) e 4). Si potrebbe suggerire che la ragione per cui noi abbiamo un dovere di evitare di portare persone sofferenti al mondo ma non un dovere di portare persone felici al mondo è che abbiamo dei doveri negativi di evitare il dolore ma non dei corrispondenti doveri positivi di produrre felicità. I giudizi sui nostri doveri procreativi sono pertanto come i giudizi su tutti gli altri doveri. Ora per quelli che negano che noi abbiamo doveri positivi, questa potrebbe in effetti rappresentare una spiegazione alternativa rispetto a quella che ho fornito io. Tuttavia, anche tra quelli che pensano che abbiamo effettivamente dei doveri positivi, soltanto pochi pensano che tra questi vi sia un dovere di portare persone felici al mondo. Per questa ragione la mia spiegazione risulta preferibile rispetto a quella alternativa. Una seconda ragione a sostegno

della mia affermazione sull'asimmetria tra 3) e 4) è che se da un lato risulta strano fornire tra le ragioni per avere un bambino che per quel bambino venire al mondo rappresenterà un bene, dall'altro in alcuni casi evitiamo di portare un bambino al mondo a causa dei suoi possibili interessi. In tal senso, se i figli si facessero con l'obiettivo di recar loro beneficio ci sarebbe una ragione morale robusta, almeno per molte persone, per averne di più. Al contrario, si ritiene che la nostra preoccupazione per il benessere di potenziali figli che soffriranno sia una ragione valida per decidere di *non* averne. Se l'assenza di piaceri fosse un male a prescindere dal fatto che lo *sia* per qualcuno, allora avere figli per il loro bene non risulterebbe qualcosa di strano. E se non fosse vero che l'assenza di dolori è un bene anche nel caso in cui non vi sia qualcuno a farne esperienza allora non potremmo dire che sarebbe un bene evitare di portare al mondo bambini sofferenti.

Infine, possono essere forniti argomenti a supporto della mia affermazione a partire da un'asimmetria correlata, questa volta con riferimento ai nostri giudizi retrospettivi. Portare persone al mondo così come non riuscire a portarle può essere motivo di dispiacere. Tuttavia, ci si può dispiacere *per il bene di una persona*, la cui esistenza dipendeva dalle nostre decisioni, soltanto per aver portato una persona al mondo. Si può essere addolorati per il fatto di non aver avuto figli ma non perché i figli che uno avrebbe potuto avere sono stati privati della propria esistenza. Il rimorso per non aver avuto figli è un rimorso per noi stessi, un dolore per non essere riusciti a vivere l'esperienza di fare un figlio e crescerlo. Tuttavia, noi ci dispiacciamo di aver portato al mondo un figlio con una vita infelice e ci dispiacciamo per quel bambino oltre che per noi. La ragione per la quale non ci lamentiamo per non essere riusciti a portare qualcuno al mondo è perché l'assenza di piaceri non costituisce un male. So che i giudizi alla base di questa

asimmetria non sono universalmente condivisi. Per esempio, gli utilitaristi positivi che sono interessati non soltanto alla minimizzazione del dolore ma anche alla massimizzazione del piacere sarebbero spinti a lamentarsi per l'assenza di possibile piacere aggiuntivo, anche se non ci fosse nessuno privato di quel piacere.

In base alla loro prospettiva, c'è un dovere di portare al mondo delle persone se ciò incrementerà l'utilità. Solitamente tale dovere si manifesterà come dovere di portare al mondo persone felici. Tuttavia, in certe situazioni il dovere potrebbe essere di portare una persona sofferente al mondo se ciò determinerà un incremento netto della felicità, recando beneficio ad altri. Dico questo non per sostenere che tutti gli utilitaristi positivi *sono tenuti* a rifiutare la prospettiva sull'asimmetria di 3) e 4). Gli utilitaristi positivi che guardano con favore all'asimmetria potrebbero stabilire una distinzione tra (i) promuovere la felicità delle persone (che esistono o esisteranno indipendentemente dalle scelte di un individuo) e (ii) incrementare la felicità creando persone. Essi potrebbero a quel punto, coerentemente con l'utilitarismo positivo, ritenere che soltanto (i) sia un requisito di moralità. Questa rappresenta la versione preferibile dell'utilitarismo positivo. Se si ritenesse che anche (ii) costituisca un requisito per la moralità le persone dovrebbero essere considerate come semplici mezzi per la produzione della felicità. Se gli argomenti che ho sviluppato finora sono validi allora la prospettiva sull'asimmetria tra piacere e dolore risulta diffusa e i suoi oppositori pochi.

Il mio argomento procederà mostrando come, data questa prospettiva comune, ne consegue che sia meglio non venire al mondo.

Per mostrare questo è necessario confrontare due scenari, uno (A) nel quale *X esiste* e un altro (B) nel quale *X non è mai esistito*.

Questo insieme alle prospettive già menzionate può essere rappresentato attraverso una matrice.

Scenario A (X esiste)	Scenario B (X non è mai esistito)
1) Presenza di dolore (Male)	3)Assenza di dolore (Bene)
2) Presenza di piacere (Bene)	4)Assenza di piacere (Non male)

È indubbio che 1) sia male e 2) sia bene. Tuttavia, in linea con le intuizioni precedentemente menzionate 3) è bene anche se non c'è nessuno a fare esperienza di quel bene ma 4) non è male perché non c'è nessuno che venga privato dei piaceri assenti. Richiamandomi alla precedente difesa dell'asimmetria dovremmo notare che i modi alternativi di valutare 3) e 4), in base ai quali viene mantenuta una simmetria tra dolore e piacere, sono destinati a fallire, quantomeno se dobbiamo preservare alcuni importanti giudizi comuni. La prima opzione è

1) male	3) bene
2) bene	4) male

Qui per preservare la simmetria l'assenza di piacere è stata definita "male". Questo giudizio è troppo forte perché se l'assenza di piacere nello scenario B è "male" piuttosto che "non male" allora noi dovremmo provare dei rimpianti per il fatto che

X non è venuto al mondo. Ma noi non pensiamo che le cose stiano in questi termini. Il secondo modo per avere una valutazione simmetrica di piacere e dolore è:

1) male	3) non male
2) bene	4) non bene

In questo caso per preservare la simmetria l'assenza di dolore 3) è stata definita “non male” piuttosto che “bene” e l'assenza di piacere è stata definita “non bene” piuttosto che “non male”. Sulla base di una certa interpretazione “non male” è equivalente a “bene” e “non bene” è equivalente a “male” ma questa non è l'interpretazione utilizzata in questa matrice perché se lo fosse non differirebbe da e presenterebbe le stesse criticità della precedente matrice. “Non male” significa “non male ma neanche bene”. Questo però è troppo debole. Non portare un bambino sofferente al mondo è più che semplicemente non male. È un bene. Anche ritenere che l'assenza di piacere sia “non bene” è troppo debole dal momento che non dice abbastanza. Naturalmente l'assenza di piacere non è ciò che chiameremmo bene. Tuttavia, la vera questione, quando l'assenza di piacere non comporta alcuna privazione per uno specifico individuo, è se tale assenza sia anche “non male” o se essa sia un “male”. La risposta è che essa “non è bene ma neanche male” piuttosto che “non bene ma male”. Poiché “non male” rappresenta una valutazione più completa rispetto a “non bene” io la preferisco. Tuttavia, anche coloro che desiderano mantenere "non bene" non riusciranno in questo modo a ripristinare una simmetria. Se il dolore è male e il piacere è bene, ma l'assenza di dolore è bene e l'assenza di piacere è non bene allora non c'è alcuna simmetria tra piacere e dolore.

Avendo rifiutato le valutazioni alternative ritorno al mio schema iniziale. Nel primo confronto vediamo che la non esistenza è preferibile all'esistenza. Il vantaggio è reale. Nel secondo confronto, invece, i piaceri dell'esistenza, sebbene siano un bene, non costituiscono un vantaggio reale rispetto alla non esistenza, perché l'assenza di piacere non è male.

Affinché il bene costituisca un vantaggio concreto rispetto alla non esistenza, dovrebbe verificarsi che la sua assenza sia un male.

Per illustrare questo concetto, consideriamo un'analogia che, essendo costruita intorno al confronto tra due persone esistenti, in questo risulta differente dal confronto tra esistenza e non esistenza, tuttavia può essere istruttiva.

S è soggetto a regolari episodi di malattia. Fortunatamente per lui, ha anche una costituzione che gli consente di guarire rapidamente. H non ha la capacità di guarire rapidamente, ma non si ammala mai. Per S ammalarsi è male ed è bene che guarisca rapidamente.

È bene che H non si ammali mai, ma non è male che non abbia la capacità di guarire rapidamente. La capacità di guarire rapidamente, sebbene sia bene per S, non costituisce un vantaggio concreto rispetto a H.

Questo perché l'assenza di tale capacità non è negativa per H (e H non sta peggio di quello che sarebbe stato se avesse avuto le capacità di recupero di S). S non sta meglio di H in nessun modo, anche se S sta meglio di quanto lui stesso sarebbe stato se non avesse avuto la capacità di guarire rapidamente.

Possiamo accertare i vantaggi e gli svantaggi relativi dell'esistenza e della non esistenza in un altro modo, sempre nella mia matrice originale, ma confrontando (2) con (3) e (4) con (1). Ci sono vantaggi sia nell'esistere che nel non esistere. È un bene che gli esistenti godano dei loro piaceri. È altrettanto positivo che i dolori siano evitati grazie alla non

esistenza. Tuttavia, questa è solo una parte del quadro. Poiché non c'è nulla di male nel non venire mai al mondo ma c'è qualcosa di male nel venire al mondo, tutto sommato la non esistenza risulta preferibile.

III

Una delle conclusioni che emergono da alcune delle riflessioni condotte finora è che l'analisi costi benefici dell'ottimista - nella quale i piaceri della vita si bilanciano rispetto ai mali - non è convincente come confronto tra la desiderabilità dell'esistenza e della non esistenza. L'analisi dell'ottimista è sbagliata perché comporta la formulazione di un confronto scorretto. Se vogliamo determinare se la non esistenza sia preferibile rispetto all'esistenza o viceversa allora dobbiamo confrontare la parte sinistra con quella destra dello schema, che rappresentano gli scenari alternativi in cui *X* esiste e *X* non esiste. Confrontare i riquadri superiore e inferiore di sinistra ci dice qualcosa di abbastanza differente; ossia quanto la vita di X sia buona o cattiva. Comprendere questa differenza rende chiaro che nonostante l'esistenza non esibisca alcun vantaggio rispetto alla non esistenza alcune vite presentano dei vantaggi rispetto ad altre. Non ogni venuta al mondo è svantaggiosa allo stesso modo. Più le caratteristiche positive di una vita superano quelle negative migliore è la vita e meno svantaggiosa risulta l'esistenza. Ma nella misura in cui ci sono degli aspetti negativi la vita non è preferibile rispetto al non essere mai venuti al mondo.

A partire da questa conclusione, c'è una differenza tra dire che è meglio non venire al mondo e stabilire quanto sia grande il male di essere venuti al mondo. Finora ho formulato argomentazioni solo a supporto della prima affermazione.

Quanto sia grande il male dell'esistenza varia da persona a persona ma io intendo suggerire ora che il male risulta in definitiva rilevante per chiunque. Tuttavia, deve essere sottolineato che la prospettiva per cui sarebbe meglio non venire mai al mondo è logicamente distinta dalla mia opinione su quanto sia grande il male dell'esistenza.

Si può abbracciare la prima prospettiva e tuttavia negare che il male sia grande. Similmente, se uno ritiene che il male dell'esistenza non sia grande non può inferire da ciò che l'esistenza sia preferibile rispetto alla non esistenza. Tendiamo a dimenticare quanto siano grandi le sofferenze che tutti noi patiamo. C'è una forte tendenza a considerare quanto vadano bene le nostre vite in relazione a quelle degli altri. Se noi viviamo più a lungo e con meno malattie e maggiore comfort rispetto ad altri ci consideriamo fortunati. E dovremmo farlo. Tuttavia, allo stesso tempo non dovremmo perdere di vista quanto siano rilevanti le sofferenze che patiamo. Che le persone tendano a perdere di vista questo rappresenta un'importante ragione psicologica per cui molti avvertono delle resistenze rispetto alla mia conclusione che venire al mondo non costituisce un beneficio. Molti hanno poche difficoltà a comprendere perché delle vite di qualità relativamente scarsa potrebbero non costituire un vantaggio. Avrebbero molte meno difficoltà a estendere questo giudizio a tutte le vite, se vedessero davvero quanto grandi sono le sofferenze che tutte le persone patiscono.

Consideriamo ad esempio la morte, dal momento che si tratta di qualcosa che tutti affronteremo. Consideriamo tragica la morte di una persona di quarant'anni, ma tendiamo a essere abbastanza disinteressati rispetto alla morte di un novantenne. Chiaramente, la morte di quest'ultimo è di gran lunga preferibile rispetto alla morte del quarantenne (a parità di condizioni), ma questo non toglie nulla al male intrinseco della morte di un novantenne. Immaginiamo quanto sarebbe diverso il nostro giudizio sulla morte

di un novantenne se le persone vivessero normalmente centoventi anni. Al contrario, c'è stato un tempo in cui le persone raramente vivevano fino a cinquant'anni. Immagino che a quell'epoca vivere fino a quarant'anni non fosse considerato una tale tragedia[4]. È chiaro che i nostri giudizi comuni su quali morti siano effettivamente un male sono flessibili. La mia opinione è che tutte le morti, *ceteris paribus*, costituiscono un grave male. Quanto sia grande il male può variare in relazione agli altri o alle attuali norme (che a loro volta sono determinate dall'aspettativa di vita degli altri), ma ogni morte rappresenta di per sé un'enorme tragedia. Che siamo nati già destinati alla morte costituisce un grave male. Quest'idea della morte non è condivisa. Una prospettiva opposta vede infatti la morte come qualcosa di equivalente alla non esistenza del pre-concepimento. Quelli che abbracciano questa prospettiva negheranno che la morte costituisca un male. Potrebbero persino provare a suggerire che la mia prospettiva presenta delle contraddizioni dal momento che io penso che la non-esistenza sia preferibile rispetto all'esistenza, ma poi ritengo che la cessazione dell'esistenza costituisca un male. Se venire al mondo è un male, come può anche la cessazione dell'esistenza essere un male? La risposta è questa.

Mentre la non esistenza del pre-concepimento o la non esistenza di potenziali persone che non diventeranno mai effettive è qualcosa che non accade a nessuno, la morte (la *cessazione* dell'esistenza) è qualcosa che accade a qualcuno. Accade alle persone che muoiono. Mentre Epicuro ha ragione nel dire che se c'è la morte io non ci sono e se io ci sono la morte non c'è, non consegue da ciò che io non abbia alcuna ragione per giudicare la mia morte un male. Dopotutto, si tratta della cessazione di me e

[4] Oggi, nelle parti più povere del mondo, l'aspettativa di vita corrisponde a quella dei Paesi sviluppati nei secoli passati. Si noti che consideriamo la brevità della vita delle persone in questi Paesi più poveri (e a volte anche delle persone in epoche precedenti) come tragica, ma proprio perché stiamo confrontando la loro durata di vita con quella a cui siamo abituati.

questa prospettiva è qualcosa di cui ci si può rammaricare intensamente.

IV

Un'importante obiezione al confronto che ho stabilito tra il venire all'esistenza di X e il non venire all'esistenza di X è che non è possibile confrontare l'esistenza e la non esistenza. Si dice che la non esistenza non è uno stato nel quale qualcuno possa trovarsi e per questa ragione non è possibile che esso risulti migliore o peggiore dell'esistenza.

Altri hanno già risposto a questa obiezione. Per esempio, Joel Feinberg ha messo in evidenza[5] che confrontare l'esistenza di X con la non esistenza di X non equivale a confrontare due possibili condizioni o stati di X. Piuttosto equivale a confrontare l'esistenza di X con uno stato di cose alternativo (lo scenario B nel mio schema) nel quale X non esiste. Un confronto del genere è possibile. Si noti che quando io dico che la non esistenza è migliore, preferibile, o che presenta un vantaggio rispetto all'esistenza non mi sto impegnando nel dire che sia migliore, preferibile o vantaggiosa per il non esistente. I non esistenti non esistono e per questa ragione le cose non possono letteralmente essere migliori per loro o a loro vantaggio. Quando io dico che la non esistenza è preferibile, quel giudizio è formulato nei termini degli interessi della persona che sarebbe altrimenti venuta all'esistenza. L'argomento è che per qualsiasi persona (sia potenziale sia effettiva) lo scenario alternativo della non esistenza risulta migliore[6].

[5] J. Feinberg, *Wrongful lives and the counterfactual element in harming*, in «Freedom and Fulfillment», Princeton University Press 1992, p. 19.

[6] Un certo sostegno per questi confronti può essere ricavato considerando la differenza tra la vita infelice di X e l'inesistenza di X. Molti trovano problematico anche questo paragone, ma altri comprenderanno l'idea che la non esistenza sia preferibile per X che altrimenti esisterebbe. Per loro, questo tipo di confronto potrebbe essere la punta di diamante, che porterà agli altri scenari comparativi che sto suggerendo.

È proprio perché la valutazione viene formulata sempre nei termini della persona che esisterebbe o che effettivamente esiste (vale a dire la persona nello scenario A) che la mia prospettiva non è impersonalista[7] anche se il confronto viene istituito con uno stato di cose (lo scenario B) e non con lo stato di una persona[8].

V

Che l'esistenza costituisca un male potrebbe essere una conclusione difficile da digerire. La maggior parte delle persone non è dispiaciuta della propria esistenza. Molti sono felici di essere venuti al mondo perché apprezzano le loro vite. Ma queste valutazioni sono sbagliate proprio per le ragioni che ho sottolineato. Il fatto che una persona apprezzi la propria vita non rende l'esistenza di una persona migliore della non esistenza perché se un individuo non fosse venuto al mondo non ci sarebbe stato nessuno ad essere privato della gioia di condurre quella vita e in tal modo l'assenza della gioia non sarebbe stata un male. Si noti al contrario che ha senso dispiacersi di essere venuti al mondo se non si apprezza la propria vita. In questo caso se uno non fosse venuto al mondo non ci sarebbe stata alcuna sofferenza da patire. Questo sarebbe un bene anche se non ci fosse nessuno a fare esperienza di quel bene.

Ora si potrebbe obiettare che non ci si può sbagliare per nessuna ragione rispetto a se l'esistenza sia preferibile alla non esistenza. Si potrebbe dire che così come uno non può sbagliarsi sul fatto di provare dolore non può sbagliarsi neanche rispetto a se è felice di essere nato. Dunque, se “sono felice di essere nato”, una proposizione rispetto alla quale molte persone sarebbero d'accordo, fosse equivalente a “è meglio che io venga

[7] Cfr. D. Parfit, *Reasons and persons*, OUP, Oxford 1984, in part. Capitolo 17.
[8] Difendo quest’affermazione in maniera più dettagliata in altri lavori.

al mondo" allora non ci si potrebbe sbagliare rispetto a se l'esistenza sia migliore della non esistenza. Il problema con questa tipologia di ragionamento è che queste due proposizioni non sono equivalenti. Anche se un individuo non può sbagliarsi rispetto a se attualmente è felice di essere venuto al mondo non ne consegue che uno non possa sbagliarsi rispetto a se sia meglio venire al mondo. Possiamo immaginare che qualcuno sia felice, in un certo momento della sua vita, di esistere e successivamente (o in precedenza), magari nel pieno dell'agonia, rimpianga di essere venuto al mondo. Ora non può essere vero che (tutto sommato) sia meglio essere venuti al mondo e allo stesso tempo non essere venuti al mondo. Ma ciò è esattamente quello che dovremmo dire in un caso del genere se fosse vero che essere felici o infelici di essere venuti al mondo fosse equivalente al fatto di essere effettivamente meglio o peggio che un individuo venga al mondo. Ciò risulta vero persino in quei casi in cui le persone non cambiano idea rispetto a se sono felici di essere nate.

VI

Se ciò che ho detto è corretto allora non può esserci nessun dovere di portare le persone al mondo. Ciò dimostra anche che è effettivamente sbagliato avere figli oppure la procreazione non è né obbligatoria né proibita? È forse vero che il nostro dovere di non portare persone al mondo si applica non solo a quelli che rispetto ad altri soffrono ma a tutte le potenziali persone? Una risposta affermativa sarebbe nettamente contraria ad alcune delle tendenze umane più profondamente radicate e potenti, quelle riproduttive. Nel valutare se sia sbagliato avere figli noi dobbiamo stare molto attenti e in guardia rispetto a queste caratteristiche della nostra natura, dal momento che possiedono l'enorme potere di condizionarci a loro favore. Allo stesso tempo, abbracciare l'opinione che la procreazione sia sbagliata, senza aver prima valutato la rilevanza morale di queste pulsioni

sarebbe avventato. I bambini non possono essere portati al mondo per il loro bene. Le persone fanno i bambini per altre ragioni, la maggior parte delle quali è in funzione dei propri interessi. I genitori soddisfano desideri biologici di procreazione. E trovano soddisfazione nell'educare e crescere i bambini. I bambini costituiscono spesso una polizza assicurativa per la vecchiaia.

I discendenti forniscono ai genitori una sorta di immortalità attraverso il materiale genetico, i valori e le idee che i genitori trasmettono ai loro figli e che sopravvivono nei loro figli e nipoti dopo che i genitori stessi sono morti. Queste sono tutte buone ragioni affinché le persone desiderino avere figli ma nessuna di esse mostra perché avere figli non sia sbagliato. Soddisfare i propri interessi non è sempre un male. Spesso è un bene ma quando farlo comporta sofferenze significative per altri solitamente non è giustificato. A questo punto, un modo per difendere l'avere dei bambini anche se si accetta la mia prospettiva per cui l'esistenza è un male consiste nel negare che il male sia grande. Si potrebbe pertanto sostenere che il male sia bilanciato dai benefici per i genitori. Tuttavia, c'è qualche ragione per pensare che, anche se uno compie il passaggio ulteriore e concorda sul fatto che l'esistenza sia un *grande* male, *potrebbe* ancora non essere immorale avere figli.

Mi permetto di aggiungere per ragioni che renderò più chiare nel prosieguo che non sono convinto da questi argomenti. Formulo la possibile difesa dell'avere dei figli non perché ritengo che questa attività *deve* essere accettabile – probabilmente l'esistenza è così negativa che *è* sbagliato avere figli – ma perché qualcosa di così apprezzato come la procreazione non va condannato a cuor leggero. È moralmente rilevante che la maggior parte delle persone le cui vite vanno relativamente bene non consideri le proprie vite come un male. Esse non rimpiangono di essere venute al mondo. I miei argomenti suggeriscono che queste

prospettive potrebbero essere meno che razionali ma questo non le priva di tutto il loro significato morale. Poiché la maggior parte delle persone che vive vite confortevoli è felice di essere venuta al mondo i futuri genitori all'interno di questa tipologia di persone sono giustificati nel presumere che se avessero bambini anche i loro bambini si sentirebbero in questo modo. Dal momento che non è possibile ottenere il consenso delle persone di venire al mondo prima della loro esistenza questo presupposto potrebbe giocare un ruolo chiave nella giustificazione di avere dei figli. Quando possiamo presumere che quelli che porteremo al mondo non saranno dispiaciuti di ciò, abbiamo il diritto, così potrebbe svilupparsi l'argomento, di dare espressione ai nostri interessi procreativi. Se questi interessi possono essere soddisfatti avendo o un figlio con una vita relativamente buona o un figlio con una vita relativamente cattiva sarebbe sbagliato se i genitori portassero quest'ultimo al mondo anche se quel figlio non sarebbe dispiaciuto della sua esistenza. Ciò avviene perché se i futuri genitori devono soddisfare i loro interessi procreativi essi devono farlo con i minori costi possibili. Quanto meno la vita che essi porteranno al mondo sarà negativa tanto minori saranno i costi. Alcuni costi (come quelli di una discendenza che condurrà una vita con un livello di decenza inferiore al minimo) sono così grandi che supereranno sempre gli interessi dei genitori. Quei casi in cui la discendenza finirà per rimpiangere la propria esistenza sono straordinariamente tragici ma se i genitori non possono ragionevolmente prevederlo noi non possiamo dire, suggerirebbe l'argomento, che essi commettono uno sbaglio nel seguire i loro rilevanti interessi nell'avere dei figli. Immaginiamo quanto sarebbero diverse le cose se la maggioranza o quantomeno una minoranza non trascurabile di persone rimpiangesse di essere venuta al mondo. A queste condizioni, questa possibile giustificazione per avere bambini certamente sarebbe condannata. L'argomento per cui

potrebbe non essere immorale avere bambini è in qualche misura sbagliato. Per esempio, la sua forma paternalistica è stata ampiamente criticata in altri contesti a causa della sua incapacità di escludere quelle interferenze dannose nelle vite delle persone (come ad esempio l'indottrinamento) che comportano una conseguente adesione alle interferenze stesse. Non sono così sicuro che questa obiezione abbia forza nel contesto dell'avere bambini. Ciò avviene perché l'azione sbagliata di portare persone al mondo è diversa dai fattori che causano la seguente approvazione di quell'azione da parte della progenie. In tal senso essa appare differente dal danno da indottrinamento. Tuttavia, rimangono altre simili preoccupazioni. Arrivare ad abbracciare le prospettive che siamo stati spinti a sostenere è una forma di preferenza adattiva, in cui un'interferenza paternalistica finisce per essere abbracciata. Tuttavia, ci sono altre tipologie di preferenza adattiva rispetto alle quali manifestiamo dei sospetti. I beni desiderati che si rivelano irraggiungibili possono smettere di essere desiderati (l'uva acerba). È vero anche l'inverso. Non è inusuale che le persone si ritrovino in circostanze sfortunate[9] e che adattino le loro preferenze per gestire la propria difficile condizione[10]. Se venire al mondo costituisce un danno così grande come ho suggerito e se ciò rappresenta un fardello psicologico difficile da sostenere allora è verosimile che noi potremmo essere impegnati in una forma di autoinganno di massa su quanto magnifiche siano per noi le cose. Alcuni potrebbero ritenere non plausibile questa indicazione. Essi dovrebbero considerare alcuni dettagli. Innanzitutto, esiste il fenomeno di come le valutazioni sulla qualità della vita delle persone differisce e cambia. Tra le persone prive di una qualche grave malattia o disabilità si

[9] Spesso, anche se non sempre, ciò avrà inizio come un tentativo di salvare la faccia ma anche in tale caso può essere interiorizzato.

[10] *N.d.T* Benatar qui usa un'espressione che in lingua italiana risulta intraducibile. Sweet lemons si potrebbe, in ogni caso, rendere con "indorare la pillola".

ritiene spesso che queste condizioni costituiscano danni sufficientemente gravi da far sì che non essere mai venuti al mondo sarebbe preferibile rispetto all'esistenza con questi danni. In alcune circostanze la tesi è persino più forte, ossia che cessare di esistere sia preferibile rispetto alla continuazione dell'esistenza con questo tipo di malattie o disabilità. Molto spesso, tuttavia, le persone che hanno o finiscono con il ritrovarsi nelle stesse condizioni terribili si adattano a queste ultime e preferiscono vivere in queste condizioni piuttosto che non essere mai esistite o cessare di esistere. Questo potrebbe suggerire, come alcuni sostenitori dei diritti dei disabili hanno sostenuto, che la soglia nei giudizi circa che cosa costituisca una qualità di vita minimamente dignitosa è fissata troppo in alto. Tuttavia, il fenomeno è ugualmente compatibile con la tesi per cui la soglia ordinaria è posta troppo in basso così che almeno alcuni di noi dovrebbero superarla. Quest'ultima idea esprime esattamente il giudizio che noi possiamo immaginare verrebbe formulato da un extraterrestre con una vita felice e priva di qualsiasi sofferenza o difficoltà. Egli guarderebbe con pietà alla nostra specie e vedrebbe la delusione, l'angoscia, i lutti, il dolore e la sofferenza che scandiscono ogni vita umana e giudicherebbe la nostra esistenza come noi (relativamente in salute e con un corpo in forma) giudicheremmo l'esistenza di individui quadriplegici costretti a letto come peggiore dell'alternativa della non esistenza.

I nostri giudizi su che cosa costituisca un limite accettabile alla sofferenza sono profondamente radicati nello stato del nostro benessere. Come possiamo essere così sicuri di non essere colpevoli di auto-inganno?

Ma perché una simile forma di auto-inganno dovrebbe essere così pervasiva? Una spiegazione potrebbe essere rintracciata nelle forti ragioni evolutive per cui potremmo essere disposti a giudicare le nostre vite come vantaggiose. Una prospettiva del

genere facilita la sopravvivenza, sia dell'individuo sia della specie.

Questi temi meritano un trattamento più strutturato di quello che io qui sia in grado di offrire. Non sono sicuro, pertanto, che l'argomento indicato per la permissibilità (in alcuni casi) di avere dei figli sia valido. Tuttavia, l'idea che le preferenze adattive potrebbero svolgere un ruolo fornisce in effetti una risposta ad un'obiezione sollevata da alcuni critici, che il fatto che la maggior parte delle persone non rimpiange di essere venuta al mondo fornisce una ragione convincente per ritenere che le loro vite costituiscano un vantaggio e pertanto che la mia conclusione opposta deve essere falsa. Quello che mostra la preoccupazione per le preferenze adattive è che la mera credenza che un individuo sia stato avvantaggiato non è sufficiente a mostrare che quell'individuo sia stato avvantaggiato o che la propria valutazione sia razionale. Non considereremmo l'accettazione della condizione di schiavitù da parte di uno schiavo come una prova *definitiva* che la schiavitù sia nel suo interesse. Di fronte a un'argomentazione per cui egli non è stato avvantaggiato dalla sua condizione di schiavitù noi giudicheremmo con sospetto il suo entusiasmo per quella condizione. Dovremmo fare la stessa cosa rispetto all'entusiasmo che le persone manifestano per essere venute al mondo.

Anche se avere figli non è immorale (dato il presupposto che noi potremmo essere in diritto di farlo) il mio argomento suggerisce quantomeno che ciò non è moralmente desiderabile. Sebbene la nostra potenziale discendenza potrebbe non rimpiangere di essere venuta al mondo certamente non rimpiangerà di non essere venuta al mondo. Dal momento che non è effettivamente nel suo interesse venire al mondo il corso moralmente desiderabile di azione è di assicurare che ciò non avvenga. Un'implicazione della mia prospettiva è che sarebbe preferibile per la nostra specie estinguersi. Sarebbe meglio se non ci fossero più persone. Molte

persone ma non io trovano una prospettiva del genere intrinsecamente intollerabile. Immaginiamo che tutti decidano di intraprendere un percorso di non procreazione o che anche in assenza di un accordo agissero sulla base di un ideale non procreativo. Non ci sarebbero più bambini che vengono al mondo e la popolazione umana invecchierebbe e infine si estinguerebbe. Non c'è alcuna possibilità che questo accada. Se la nostra specie si estinguerà non sarà perché avremo liberamente scelto questo evento (sebbene potrebbe essere l'effetto di altre azioni liberamente scelte). Nonostante ciò, la possibilità deve essere considerata perché è un'implicazione teorica della mia prospettiva.

Sono d'accordo sul fatto che ci sarebbero alcuni aspetti della scomparsa dell'umanità che risulterebbero tragici. L'ultima generazione prima dell'estinzione vivrebbe in un mondo nel quale le strutture della società collasserebbero. Non ci sarebbero più generazioni giovani a lavorare nei campi, a preservare l'ordine, a far funzionare gli ospedali e le case di cura per anziani e a seppellire i morti. La situazione è effettivamente desolante. È difficile sapere se le sofferenze delle ultime persone sarebbero in qualche modo maggiori di quelle di molte persone di ciascuna generazione. Dubito che lo sarebbero ma immaginiamo per un momento l'opposto. Ho suggerito (con un po' di trepidazione) che avere bambini potrebbe non sempre essere immorale. Assumiamo in prima battuta che la mia prospettiva sia corretta. Che cosa accadrebbe se, nonostante avere dei figli fosse ammissibile, le persone agissero in base all'ideale, rinunciassero ad avere dei figli e come risultato soffrissero tremendamente? Come potrebbe ciò essere accettabile come ideale morale? La prima cosa da notare è che sarebbe un risultato che una generazione volenterosa (anche se con timore) accetterebbe su di sé in nome dell'ideale morale. Sarebbe una decisione supererogatoria o eroica da compiere per le persone (specialmente se sapessero che tutti gli

altri stanno compiendo la stessa scelta). Esse accetterebbero ulteriore sofferenza su di sé per risparmiare alle potenziali persone future i dolori dell'esistenza. Ciò sarebbe qualcosa da ammirare anche se le conseguenze per gli eroi sarebbero estremamente spiacevoli. Se non muoviamo obiezioni ai sacrifici eroici in altri contesti perché dovremmo muovere delle obiezioni quando tali gesti preverrebbero qualsiasi futura sofferenza? Ma che cosa accadrebbe se la tesi che avere bambini è ammissibile si rivelasse sbagliata? Persino in quel caso noi dovremmo considerare che se c'è qualcosa di tragico in relazione alla scomparsa dell'umanità non è la scomparsa in se stessa ma la sofferenza che ne scaturisce. Io ritengo che le persone che pensano che la scomparsa in se stessa sarebbe tragica sarebbero fortemente spinte a fornire una spiegazione di questo nei termini degli interessi di quelli che sarebbero potuti venire al mondo. Chi ci sarebbe a soffrire per la scomparsa di Homo sapiens? Una possibile idea è che impatterebbe sulle persone che sanno che sta per accadere. Tuttavia, questa sarebbe semplicemente un'altra caratteristica della sofferenza che prefigura la fine della vita umana[11].

(Traduzione italiana di Luca Lo Sapio)

DAVID BENATAR è Direttore del Dipartimento di Filosofia e del Bioethics Centre della Cape Town University in Sudafrica

Liz.gubb@uct.ac.za

[11] Sono grato ai revisori di *APQ* per i numerosi e acuti commenti che mi hanno aiutato a effettuare dei rilevanti miglioramenti all'articolo.

S&F_n. 27_2022

ARTE

ARMANDO MINOPOLI

L'ESSENZIALE NON È INVISIBILE AGLI OCCHI.

DIVAGAZIONI SUL CONCETTO DI ESSENZA IN ARCHITETTURA

1. *Chiave di (s)volta* 2. *Abitare* 3. *Doppia linea*
4. *Casa* 5. *Archè* 6. *(Es)Senza fine*

ABSTRACT: ***WHAT IS ESSENTIAL IS NOT INVISIBLE TO THE EYE. REFLECTIONS ON THE CONCEPT OF "ESSENTIAL" IN ARCHITECTURE***
This work wonders about the concept of "essential" in architecture. In particular, it refers to the architecture of the house conceived as the founding place of the act of living by man. The concept of "essential" is explored both in a strictly architectural key – the space that is constituted only through the fundamental type-morphological elements – and in a conceptual and metaphorical key, trying to identify those components – even immaterial, invisible – which however represent the true (essential) meaning of people's existence.
In the end, the text argues that if the ideal of the essential, conceived by subtraction, corresponds to the shape of an empty space, of a bare space, it is precisely that void – as a trace full of memory – that represents a new beginning every time.

Idea di una residenza, alla quale remoto è l'esterno.
Un recinto imbiancato e uniforme: il primo guscio non offre occasioni.
L'accesso unico è in posizione marginale: varcata la soglia, l'esilio è rifugio.
Il grande atrio prepara al piacere del recesso: l'una nell'altra le valve della segregazione.
L'esilio è un singolare modo di osservare il mondo: guardare all'interno come altri le stelle
(F. Venezia, *Scritti brevi. 1975 – 1989*)

Trenta raggi convergono nel mezzo ma è il vuoto del centro che fa avanzare il carro.
Si modella l'argilla per fare vasi ma è dal vuoto interno che dipende il suo uso.
Una casa è scandita da porte e finestre, è ancora il vuoto che la fa diventare abitabile.
L'essere dà possibilità ma è per il non essere che la utilizziamo.
(Lao Tze, *Tao Te King*)

Il mio consiglio più importante è: quando fate architettura fate il meno rumore possibile […] Regola principale per chi si mette a progettare, fare silenzio attorno per essere più attenti, e capaci di vedere piccolo: tra le cose.
(V. Gregotti, *Dieci buoni consigli*)

1. *Chiave di (s)volta*

L'uomo primitivo, un bel giorno, che fosse soleggiato oppure piovoso poco importa, si fermò. Quel giorno, l'uomo primitivo si ammutinò dalla sua condizione di acquisitivo e si fece stanziale, vivendo questo naturale trapasso geologico che gli scienziati avrebbero definito dal Paleolitico al Neolitico, con assoluta naturalezza senza la ben che minima consapevolezza della svolta rivoluzionaria che aveva innescato nell'evoluzione naturale della propria specie e delle conseguenze che avrebbe determinato circa il destino del mondo. Non aveva gli strumenti culturali per comprendere ciò che avrebbe causato, non aveva neanche le parole per definire quelle circostanze, figurarsi la struttura mentale analitica e logico-cognitiva per ragionare e prevedere.

L'uomo primitivo, quel giorno, uscì dalla caverna e decise che era giunto il momento di abbandonare i panni del nomade e allo stesso tempo decise che non gli andava più di vivere in una caverna ma che meritava condizioni più confortevoli per proteggersi dalle

avversità climatiche, per difendersi dalle aggressioni animali, per riposarsi dopo le fatiche dell'approvvigionamento, e infine per ristorarsi e procreare. Perché limitarsi alla sola attività del procreare? Non aveva forse già nozione del sentimento dell'amore? Non si sa. Di certo nessun cuore trafitto da Cupido arciere, per diventar degno di essere citato nell'arte rupestre, è stato mai trovato tra le immagini di tori, leoni, uri e mammut disegnati sulle pareti delle caverne dove l'uomo primitivo trovava riparo. A quei tempi, dalle parti di Lascaux, Chauvet, Altamira, c'era da pensare alla pancia e a salvare la pelle, altro che mielose moine sentimentali. Eppure l'uomo primitivo, privo di vocabolario scritto e orale, proprio per poter comunicare, imparò a disegnare molto bene, talmente bene che Picasso avrebbe rubato la potenza espressiva dell'essenzialità primordiale e spontanea di quelle figure. Fatto sta che il prototipo primigenio dell'uomo contemporaneo, sebbene privo di televisore e aria condizionata, decise di porre intorno a sé una protezione che si fece volume per ospitare uno spazio cavo interno. Era il passo epocale che, nel liberarlo dalle tenebre caliginose del mondo della sostruzione, lo aprì alla luce solare del mondo della costruzione. Sarà forse per questo che l'uomo, unico tra gli animali del pianeta, scava una fossa per seppellire i suoi morti ed erige un sepolcro per onorarli.

Esorcizzare la paura della morte sacralizzandola attraverso un atto creativo[1]: tutto questo avvenne a un certo punto della

[1] «Il primo atto di architettura è forse collegato al lutto. Di fronte il lutto l'uomo risponde con un atto creativo. La sepoltura è la prima architettura progettata da un uomo come grembo fecondo. Che si tratti di una buca o uno scavo, di un rogo o di una pira, non ha importanza, l'atto di progettazione è stato compiuto di fronte l'evidenza di una perdita che ci riguarda in prima persona in quanto viventi e mortali. Questa evidenza muove l'uomo a compassione: è una identificazione traumatica, una mimesi immobilizzante che cerca una via simbolica di fuga, un moto contrario, un tentativo di riparazione con cui affrontare e dominare la morte. Una via creativa (la gestazione di una partoriente, la semina), immaginaria, (*super-naturae*, metafisica), in grado di opporsi a quella distruttiva della morte. Sia in una sepoltura per un morto, sia in una capanna o un dolmen per un vivo, sia in una pelliccia per un malato, l'architettura è direttamente connessa alla "protezione", una "tecnica del

Preistoria, in un periodo molto lontano che solerti e infaticabili studiosi delle scienze naturali, sociali e umane hanno comunque determinato con incredibile precisione cronologica e geografica, fermando sull'orologio l'istante X del Tempo e cerchiando su carta geografica il punto P di coordinate (x; y) dello Spazio in cui tutto ciò avvenne. Era tempo che l'uomo la smettesse di girovagare in lungo e largo per territori impervi e rischiosi alla ricerca vorace e ossessiva degli elementi-alimenti che soddisfacessero il suo primario bisogno di nutrirsi per sopravvivere. Esausto e stremato per aver camminato per centinaia di migliaia e migliaia di anni trascinandosi a spalla i propri beni e averi – la figura geometrica del cerchio non era ancora stata promossa a mezzo tecnico di ruota – l'uomo primitivo quel giorno si fermò e sentì necessità di trovare una propria stabilità, individuare un punto di riferimento certo e fermo da cui partire e a cui tornare. Non sapeva ancora che questa sconvolgente fase sociopsicologica sarebbe stata prodromica all'introduzione nella sua sfera psico-emotiva del sentimento della nostalgia che in età moderna si sarebbe sublimata in *Saudade* per i portoghesi e *Pucundria* per i napoletani.

L'uomo primitivo non prese coscienza che la sua scelta avrebbe rappresentato una vera e propria svolta epocale nella metamorfosi della propria identità (da *homo faber* a ben più titolato *homo costruens* e un istante dopo *homo destruens*) e un cambiamento radicale nel suo modo di rapportarsi alla Madre Natura inaugurando anche, a sua insaputa, quel fenomeno che sarebbe stato poi definito antropizzazione dell'ambiente. Cementificazione del suolo, speculazione edilizia, contesti naturali violentati, inquinamento e surriscaldamento terrestre, buco dell'ozono, sarebbero arrivati ma questa è un'altra storia di cui l'ignaro uomo primitivo non merita la responsabilità a meno di non

sentimento" e non della ragione con cui farà i conti soltanto dopo», in L. Curatolo, *Riflessioni sparse*, testo non editato.

rischiare di trasformarlo da intrepido eroe rivoluzionario ad artefice di efferati crimini ambientali.

L'ignaro non sapeva quanto la sua scelta esistenziale fosse dirompente nel potenziale costruttivo (e anche distruttivo) e quale valenza escatologica si nascondesse nel semplice gesto di definire un ambito sul terreno e assegnargli un riscontro altimetrico: si stavano definendo i concetti di spazio, di spazio incluso in un volume racchiuso da proteggere ulteriormente con una successiva linea perimetrale di delimitazione. Linea che avrebbe poi differenziato il senso del dentro da quello del fuori, che sarebbe stata in seguito nominata recinto ma anche limite, confine, demarcazione, barriera, steccato, muro, muraglia, fortificazione, andando a definire l'enorme differenza semantica, culturale, politica e ideologica tra la *res propria* e la *res publica*.

Fatto sta che il primo recinto fu eretto e in quello stesso istante si poté cominciare a sognare un mondo senza separazioni. D'altronde non può esserci un *pensare senza* se non si è affermato dapprima un *pensare con* che ha prodotto ciò che poi si sognerà di eliminare. John Lennon che canta *Imagine* qualcosa al riguardo avrebbe detto ma siamo troppo troppo avanti rispetto all'istante decisivo in cui l'uomo primitivo, ancora molto bestiale, si disse che il suo posto era *qui*: qui dove abitare, qui dove tornare.

All'inizio di tutto, all'inizio della qui descritta Genesi 2.0, rinnovata versione della ben più celebre precedente, non fu pertanto la luce ma fu la necessità di un ricovero, di un riparo, di un *demorare*, termine derivato da *morari* e cioè indugiare, aspettare e pazientare, impiegare tempo (traduzione precisa del termine *demorar* che accomuna, per lemma e significato, nelle lingue iberiche, portoghesi e spagnoli). Il riparo è prima di tutto un tetto o meglio la forma rispondente alla funzione tetto sotto cui stare per proteggersi da pioggia battente e sole cocente.

Il primo riparo-tetto non fu disegnato, oggi diremmo progettato. Non si fece forma attraverso il di-segno, costituendosi costruzione nel partire da immagine pensata corrispondente a un lemma denotativo che si fa segno per dare figura al pensiero astratto. Fu atto istintivo e fu atto intuitivo comprendendo la logica costruttiva e funzionale a prescindere dalle regole statiche e fisiche che la tecnologia e la scienza delle costruzioni avrebbero poi codificato.

Al sesto giorno – per tornare alla versione 1.0 – ultimo della creazione e vigilia di quello del legittimo riposo, li avevamo lasciati nudi e immersi nella beatitudine del Giardino dell'Eden, sotto la meraviglia del cielo nelle sue varie e principali manifestazioni cosmiche e metereologiche. Null'altro sarebbe servito alla prima coppia di umani, se non una foglia a tutela del minimo senso del pudore, per vivere in eterno in armonia con la natura che mai si sarebbe svelata matrigna. Eppure, la disobbedienza in nome del libero arbitrio sotto le mentite spoglie della tentazione, ebbe il suo costo: l'espulsione dal Paradiso Terrestre, il ripudio del Sommo Creatore, la mancanza di difesa dal gelido inverno, dalla torrida estate e dall'aggressione delle belve.

Come si deduce, qualunque fosse la provenienza del nostro eroe – disceso, per antropogenesi policentrica, da un progenitore ominide lungo un processo durato sei milioni di anni o derivato, per tradizione biblica e monogenetica, da una coppia maschio-femmina, al secolo Adamo ed Eva, creata di sana pianta nel tempo minimo di un fatidico sesto giorno – urgeva indossare un abito a protezione del proprio corpo e a rappresentazione del proprio ruolo.

Un abito che proprio per proteggere e rappresentare si fece abitazione, affinché il posto scelto dal nostro eroe lo si potesse

cominciare ad abitare e anche poeticamente[2], così come Hölderlin nella sua immaginazione visionaria e romantica disse.

2. *Abitare*

Abitare: studiare, omaggiare, commemorare, governare, celebrare, pregare, curare, svagarsi, rilassarsi. Case, e non solo case. Pure: scuole, biblioteche, musei, monumenti, chiese, ospedali, teatri, stadi, ristoranti, piazze, colorando «tutti i muri, strade, vicoli e palazzi perché Margherita ama i colori»[3].

Abitare è l'inizio di tutto e comincia dove si decide di restare. Abitare è possibile solo scegliendo un luogo e successivamente uno spazio in cui risiedere. In questo primo atto (in questa prima azione) c'è in embrione tutto il potenziale che si sprigionerà negli infiniti modi possibili attraverso cui diventerà forma costruita[4].

Abitare è termine omnicomprensivo e include tutte le complementari forme possibili dello "stare al mondo" che seguono al "venire al mondo", fatidico e dirompente istante altrimenti detto nascita: evento che comincia in un pianto immediatamente confortato dal morbido calore del seno materno e dal tepore protettivo di una culla, futura casa che i greci chiameranno *oikos*, i romani *domus*, i tedeschi *Heimat*. Luoghi e spazi in cui risiedere, riconoscersi, proteggersi, cambiare – anche di trasloco in trasloco, tra matrimoni che si compiono e lutti che si abbattono – e infine ritornare a conclusione di una lunga giornata o di un lungo girovagare esistenziale.

Solcare quel varco di ingresso vuol dire puntualmente riassaporare la rassicurazione di quella culla, prolungamento esterno dell'utero materno, rimasto come ricordo nei più remoti meandri

[2] «Poeticamente abita l'uomo su questa terra», F. Hölderlin citato da M. Heidegger in *Saggi e discorsi* (1936-1953), Mursia 1976, p. 130.
[3] Riccardo Cocciante, *Margherita*, 1976.
[4] Cfr. M. Fukuoka, *La rivoluzione del filo di paglia* (1975), tr. it. Libreria Editrice Fiorentina, Firenze 1980; M. Sahlins, *L'economia dell'età della pietra* (1980), tr. it. Elèuthera, Milano 2020; R. Sennett, *L'uomo artigiano*, tr. it. Feltrinelli, Milano 2008.

della nostra memoria primordiale, a cui torniamo con il conforto di una spiegazione psicoanalitica e biologica che ci chiarisce i motivi di questo strano fenomeno di vivere la macchina mentale del tempo a ritroso.

3. *Doppia linea*

A ritroso nello Spazio-Tempo remoto, fu il principio della capanna costruita con sottili e leggeri ed elastici rami di legno tenuti insieme da un cordame derivato dalla sfibratura degli stessi tronchi che fungevano da struttura portante e tutto racchiuso da un involucro composto da pellame oppure paglia mista a fango secco. E fu anche il principio del trilite, due piedritti lapidei tozzi e massicci a sostenere una grossa e spessa lastra di pietra. La doppia linea costruttiva fu l'archetipo da cui tutto si originò generando il modo di pensare le forme e costruire spazi e volumi, che sarebbero stati definiti con l'erudito termine di "architettura" nell'istante in cui l'atto istintivo del proteggersi dentro o sotto qualcosa sarebbe diventato l'intenzione consapevole di reificare la sintesi perfetta delle tre categorie denominate *utilitas*, *firmitas*, *venustas*, tenute insieme dai principi di misura, equilibrio, proporzione e armonia[5].

Volendo esporci su quale delle due linee preferire, il bambino che è in noi opta per la capanna perché troppo vivido è ancora il ricordo dei film western, dei soldatini Atlantic e del Manuale delle Giovani Marmotte.

Poi, a un'analisi più erudita, tralasciando l'aspetto nostalgico legato alle memorie della preadolescenza, scopriamo che la nostra preferenza è culturalmente comprovata perché il cerchio su cui si fonda la capanna-cono è la disposizione geometrica (al netto di un π del tutto sconosciuto e quindi assolutamente innata) di un gruppo di persone che nel disporsi attorno a un fuoco si trasforma in comunità trascorrendo il tempo dapprima solo mangiando poi,

[5] Vitruvio, *De Architectura* (I sec. a.C), Einaudi, Torino 1997.

evolvendosi, cianciando tra storie vere di vita quotidiana e racconti "romanzati" di vita magica, inaugurando la fervida e affascinante epopea del Mito - suggestivo, magnetico, immaginifico – attraverso cui chiedersi i primi perché della Preistoria. Ah, la cosmogonia (della prima ora)!

Le strade imboccate furono due, avendo come "portali" di ingresso la capanna da un lato e il dolmen dall'altra. A partire da questi si sarebbero sviluppati i due principali filoni tecnico-costruttivi (tettonica) e tipo-morfologici (architettura) classificati come "opere di carpenteria" (dalla capanna madre) e "opere di pietra" (dal dolmen padre)[6] lungo i quali oscilla da sempre la produzione architettonica divenuta oggi sperimentazione spinta della tecnologia, da un lato, oppure ritorno a linguaggi semantici più legati alla tradizione costruttiva, dall'altro.

Nel primo gruppo: la capanna, lo *yurt* mongolo, il *tipì* degli Indiani nomadi delle praterie nordamericane, la tradizione costruttiva anglosassone, il *balloon frame* fino alle opere di Glenn Murcutt, l'ingegneria dell'Ottocento dei padiglioni espositivi e anche industriali in ferro e vetro, le *Case Study Houses*, il Padiglione di Barcellona e Casa Farnsworth di Mies van der Rohe fino al minimalismo odierno, il Beaubourg di Renzo Piano e l'architettura High Tech di Norman Foster (*archistar* e pure *Sir*), tutta l'architettura del surmodernismo, tutto il pensiero architettonico contemporaneo che concepisce l'edificio come una macchina ad alta complessità e ad alta efficienza energetica (struttura in legno o acciaio tamponata con elementi leggeri o racchiusa da grandi superfici vetrate schermate da lamelle fisse o mobili, con conseguente riduzione della consistenza muraria, rarefazione dell'involucro che si alleggerisce e perde l'usuale immagine di gravità).

[6] K. Frampton, *Costruzioni pesanti e leggere. Riflessioni sul futuro della forma architettonica*, in «Lotus», (99), 1998, pp. 24-31.

Nel secondo gruppo: il dolmen, che si fa trilite erudito nell'architettura greca prima di modificarsi in arco in quella romana, l'invenzione del mattone che riduce a una misura ergonomicamente gestibile il componente primario della costruzione di mura altrimenti megalitiche, tutta la tradizione tecnico costruttiva del Mediterraneo che ne deriva, la casa greca (*oikos*), la casa romana (*domus*), le città compatte, la monumentalità del grande impero romano, le basiliche romaniche, i castelli medioevali, i palazzi rinascimentali, l'architettura che prosegue la propria valenza massiva col dinamismo del Barocco, l'imponenza delle maestose residenze reali e imperiali che si susseguono, l'articolazione volumetrica a forte valenza plastica del modernismo, «il gioco sapiente rigoroso e magnifico dei volumi assemblati nella luce»[7], da Le Corbusier ad Àlvaro Siza.
In poche parole, l'antico e reiterato e seguitante oscillare tra i valori di leggerezza, trasparenza, involucro, doppia pelle, vetro, metallo e così via, e massa, gravità, muro, spessore, bucature, intonaco, tornando sempre all'archetipo di mamma capanna e papà dolmen e da loro ricominciare per proseguire.

4. *Casa*

Quando si chiede a un bambino di disegnare una casa, o meglio elaborare un disegno che rappresenti ciò che la sua immaginazione ha elaborato-prefigurato-composto-proporzionato, il disegno che ne scaturisce prevede, con percentuale statisticamente molto elevata, una rappresentazione bidimensionale (piatta) di una figura composta (a insaputa del bambino) da forme geometriche elementari (quelle "antiche" o primordiali della cosiddetta "geometria piana") che sono generalmente un quadrilatero sormontato da un triangolo (rigorosamente isoscele e talvolta equilatero) corredato su uno dei lati obliqui da un rettangolo molto snello.

[7] Le Corbusier, *Verso una architettura* (1973), Longanesi, Milano 1989, p. 16.

La composizione prevedrà una impostazione geometrica rigorosamente (e a quanto pare, per statistica, anche irrimediabilmente) simmetrica. A completare la parte definibile "facciata" o più propriamente "prospetto" e per istanza simbolica e rappresentativa più precisamente "prospetto principale", ci sarà un rettangolo verticale posto in asse (spesso coronato da un segno semicircolare) e due quadratini ulteriormente suddivisi ognuno in quattro quadratini inscritti istintivamente dalle parti di quel punto che in geometria si definisce baricentro. Ad adornare il tutto, poste nella parte inferiore del foglio, due linee sinuose si affiancano convergendo e culminando nel rettangolo centrale suddetto che a questo punto potrà rivendicare il diritto di chiamarsi "porta", che a sua volta indurrà l'osservatore ad assegnare alle due sinusoidi la definizione di "sentiero" o "via di casa".

Il bambino protagonista di questo articolato *incipit*, cosa ha inventato? Cosa di questo vasto repertorio segnico-geometrico, a cui associare un corposo lessico architettonico, è frutto della propria (uno) fantasia, (due) creatività, (tre) invenzione, (quattro) immaginazione[8]?

Niente: sulla *tabula rasa* del bambino, a sua insaputa, presuntuosi e prepotenti adulti ritenutisi degni di capacità e competenza e autorità hanno inciso la propria immagine di casa ideale, scivolando a ritroso nel tempo passato, probabilmente perduto e magari ritrovato con piena soddisfazione di Marcel Proust se lo venisse a sapere.

Tutto questo nel caso di un bambino educato in nome e per conto dei "sacri valori" rassicuranti della cultura cattolica occidentale ma lo stesso varrebbe se il bambino fosse di altra appartenenza geografica e sociale.

Va da sé che se chiedessimo di disegnare il proprio ideale di casa a un bambino inuit, apache, mogul, islamico, cinese del Fujan,

[8] Cfr. B. Munari, *Fantasia*, Editori Laterza, Roma-Bari 1977.

rom, il bambino (etnicamente individuato) disegnerebbe un *igloo*, un *tipì*, uno *yurt*, una casa a tarma, una *Tou Lou*, una roulotte.
Nel tramutare queste immagini bidimensionali che rappresentano una vista frontale che gli architetti chiamerebbero "prospetto", in disegno planimetrico denominato "pianta", parleremmo di tipologia il cui etimo affonda la sua radice nella parola greca *týpos* che vuol dire "impronta". E questa parola fa una gran bella differenza in architettura, determinando la chiave di volta nella progettazione se è vero che fa allo stesso modo una bella differenza se l'impronta che vediamo lasciata sul terreno è quella di un gabbiano, di una tigre, di un elefante, oppure quella delle ruote della macchina con cui è scappato un assassino. Un buon architetto si vede da come concepisce il disegno di pianta[9], affermava il padre fondatore dell'architettura organica, Frank Lloyd Wright. Alla tipologia si associa la morfologia e la magica triade si completa con la topologia e così pianta, alzato e luogo si combinano in una sintesi che si auspica sempre essere giusta ed equilibrata.
Per tornare al bambino, il fatto che egli abbia disegnato il proprio ideale di casa non come naturalmente avrebbe potuto fare se avesse seguito lo Spazio candido e immacolato della propria fantasia, vuol dire che se gli si fornissero altri modelli di riferimento iconografico potrebbe comporre – per emulazione o più criticamente per rielaborazione – la propria casa ideale alla maniera di Ville Savoye (1928 – 1931) di Le Corbusier, di Casa Kaufmann (1936–1939) di Frank Lloyd Wright, di Casa Farnsworth (1945 - 1951) di Mies van der Rohe, di Casa Duarte (1981-1985) di Àlvaro Siza[10], giusto per fare degli esempi, tra i più significativi.

[9] Cfr. F. L. Wright, *Per la causa dell'architettura* (1975), tr. it. Gangemi Editore S.p.A., Roma 1996, p. 93.
[10] In certi casi pare che non ci sia relazione dialettica tra le fonti culturali e il risultato, con il rischio che l'immagine venga vissuta come simulacro da ripresentare in quanto tale, anziché iniziatrice di un'elaborazione dove «la materia della coscienza immaginativa si allontana

L'interrogativo che ci si pone trasla il *focus* dell'attenzione dal concetto di *ideale* – che per definizione è inteso come il risultato di un processo di astrazione logica oppure di elaborazione immaginativa senza rispondenza nella realtà – al concetto di *essenziale* che di fatto può essere considerato il suo contrappunto, perché riferito alla realtà immutabile, sostanziale delle cose.

Di fatto la casa ideale, non solo quella immaginata dal bambino, è piena di cose. Una casa essenziale è al contrario immaginata come un posto ripulito di tutto quello che si ritiene essere non necessario.

5. *Archè*

La fatidica e fatale domanda su cosa possa ritenersi essenziale in architettura, sul se esistano elementi costitutivi e imprescindibili al punto da poter rappresentare una essenza architettonica, porta con sé una molteplicità di risposte.

E in effetti, le componenti che concorrono alla definizione del termine "architettura" abbracciano uno spettro che va dalle discipline umanistiche a quelle scientifiche, passando per la soggettività dell'emozione e della reazione poetica. Constatato che la risposta alla domanda fondamentale non potrebbe essere quella corrispondente a 42[11], espressa da Pensiero Profondo (che si riservò, tra l'altro, di pensarci per sette milioni e mezzo di anni) non resta che seguire i principi basici dell'antropologia culturale e indagare sul campo: chiedere alle persone.

---

dalla materia della percezione, si compenetra di conoscenza, e ne attenua la somiglianza con l'oggetto dell'immagine» (J.-P. Sartre, *Immagine e coscienza* (1940), tr. it. Einaudi Editore, Torino 1964). «Il pericolo è perdere di vista la forza immaginante che può sorgere nella progettazione architettonica e che riguarda la capacità di reinterpretare le eredità disciplinari grazie all'uso degli strumenti concreti del mestiere e all'importanza che gli stessi assumono nel percorso progettuale» (P. Mestriner, *Immaginare Spazi,* testo non editato).

[11] «42 è la risposta alla domanda fondamentale sulla vita, l'universo e tutto quanto». D. Adams, *Guida galattica per autostoppisti*, Mondadori, Milano 2008, p. 178.

La risposta più comune è “il tetto”. Sì, il tetto perché “senza tetto non c’è inizio”. E in effetti, per sostenere un tetto, c’è bisogno di una struttura che quindi dovrebbe guadagnare nella gerarchia degli elementi imprescindibili il primo posto tra quelli costitutivi. Ciò equivale a dire che un conto sono le componenti strumentali alla realizzazione di una cosa, un conto è la “realizzanda” cosa che diventa l’idea, intesa come entità essenzialmente mentale e pure spirituale, astratta, posizionata nel trascendente, assurgendo pertanto a categoria di ordine superiore.

Nell’andare avanti con l’indagine, capita spesso di perdersi irrimediabilmente nella fantasmagoria dei ricchi premi e *cotillos* di tendaggi arabescati, cuscini fiorati, vasche idromassaggio e caminetti *soap opera* americana del pieno edonismo reaganiano anni ‘80.

Insomma, del concetto di essenza da intravedersi anche nel più remoto angolo perduto nei meandri della precisione semantica, manco a pagarlo a peso d’oro.

In principio si pensava che il “principio” di tutto fosse l’acqua, ma anche l’aria, ma anche l’*ápeiron*. Queste furono le risposte che i tre pionieri della filosofia diedero interrogandosi sull’origine di tutto. Poi le cose cambiarono un po’, e oggi le risposte sono diverse, univoche e determinate, e soprattutto scientificamente dimostrabili e comprovabili. Però è un’altra categoria del pensiero, la scienza.

E il principio costitutivo dell’architettura? L’origine di tutto, l’elemento archetipico, l’atto fondativo, il punto uno (e non quello zero, attenzione) da cui tutto è scaturito e che pertanto possiamo definire determinante e rispetto alla domanda fondamentale che ci siamo posti “essenziale”?

Nella struttura del termine “essenza” c’è la parola “senza” che implicitamente suggerisce l’operazione indispensabile da attuare, e cioè la rimozione del non necessario (o superfluo), affinché a

restare siano solo quegli elementi (in architettura, nell'analisi tipo-morfologica della tradizione storica oppure di un contesto urbano o organismo architettonico preesistente, definiti anche "caratteri" o "invarianti") ritenuti indispensabili, necessari e sufficienti, per garantire l'esistenza stessa di un corpo tangibile del mondo fisico oppure di entità astratta.

In architettura questo lavoro di depurazione del superfluo, di eliminazione dell'orpello, di rimozione del posticcio rimandabile alla definizione di "decoro" si sviluppò in contrapposizione al ridondante formalismo caratterizzato dal recupero degli stilemi del passato (gli *ismi* ottocenteschi) culminato poi nella corrente dell'*Art Nouveau* (denominato in differenti modi per area geografica) a cui però va riconosciuta una legittima dignità culturale in quanto movimento artistico e filosofico ideologicamente fondato. Il processo di semplificazione – per sottrazione – della forma architettonica vede in Adolf Loos uno degli esponenti più rappresentativi di questo periodo "rivoluzionario" anti-formalista. Di esemplare importanza culturale è soprattutto il suo lavoro di saggista[12], che ha apportato un contributo critico-intellettuale che sarà ereditato dal Movimento Moderno, evolvendosi in chiave purista con Le Corbusier e minimalista con Mies van der Rohe.

Álvaro Siza e Campo Baeza si inseriscono nel solco tracciato dal maestro francese, Eduardo Souto de Moura e John Pawson in quello delineato dal maestro tedesco, solo per citare alcuni esponenti rappresentativi e soprassedendo sulle precisazioni tassonomiche e critiche delle numerose varianti con cui l'architettura successiva alla rivoluzione ideologica del Movimento Moderno si è andata sviluppando.

[12] A. Loos, *Parole nel vuoto* (1921), tr. it. Adelphi Edizioni, Milano 1972.

6. *(Es)Senza fine*

Eliminare, eliminare il superfluo, sottrarre, semplificare[13]. «*Less is more*» esclamò un giorno l'architetto Mies van der Rohe. L'affermazione divenne il nuovo paradigma attraverso cui pensare un'architettura fondata sui valori fondamentali di razionalità, ordine, semplicità, rifuggendo il fantasioso e l'arbitrario, per procedere invece dalla complessità alla essenzialità fino al punto in cui ogni parte del sistema è funzionale al sistema nella sua interezza e nulla può essere aggiunto o tolto. Il pensiero di Mies van der Rohe determina una riflessione sull'architettura dell'epoca e propone una modalità del progettare che non tenga conto della forma come entità pensata *a priori* ma che focalizzi l'attenzione sul processo da cui poi far derivare la forma come conseguenza. Per parafrasare Mies van der Rohe, «Non siamo contrari alla forma, ma siamo contrari alla forma come fine». Si depurano, così, lo spazio e il volume dagli elementi ritenuti accessori e si valorizza, in forma rigorosamente essenziale, solo quanto considerato indispensabile per significato e funzione. Il *modus operandi* dell'architetto tedesco avrebbe poi aperto la strada all'architettura minimalista.

L'operazione, sebbene finalizzata alla realizzazione di una architettura essenziale, è prima di tutto una operazione intellettuale (e filosofica) di cui l'architettura diventa il riflesso. Il ragionamento viene perseguito fino ai limiti dell'effettiva possibilità realizzativa e abitativa, fino al punto in cui ciò che resta è lo spazio o meglio il vuoto in cui la stessa presenza dell'uomo parrebbe essere ritenuta episodica e accessoria. Lo spazio divenuto spoglio diventa vuoto, il vuoto evidenzia il nulla, il nulla manifesta l'assenza, l'assenza è silenzio. Le persone non ci sono più, le persone non abitano più le case. La definizione di "essenziale", tutta riferita alla

[13] Si veda, sul punto, C. Martí Aríz, *Silenzi eloquenti* (1999), tr. it. Christian Marinotti Edizioni, Milano 2006.

individuazione degli elementi fondativi e imprescindibili, in architettura, non può prescindere dalla presenza del vero elemento sostanziale che è l'essere umano.

Lo spazio acquisisce senso e significato solo in quanto abitato dall'uomo che in tal modo ne diventa l'essenza costitutiva. Se non abitato, e in generale esperito, lo spazio pur definito "essenziale" non ha motivo di esistere.

È così che viene rappresentata l'architettura in fotografia: un fotogramma che congela un istante in cui non c'è presenza umana, gli ambienti sono ordinati e lindi, non c'è strofinaccio sul tavolo, non c'è posacenere sporco, le sedie sono allineate, sulle mensole non sono previste le bomboniere, i souvenir, le foto ricordo. La rappresentazione trasmette un insolito senso di interruzione e sospensione del tempo.

Ma il tempo scorre e onora il Dio Kairos, danzando tra la memoria del passato e la prefigurazione del futuro, e le due dimensioni temporali danzando confondendosi (e questo è certamente il labirinto di Jorge Luis Borges), inciampando nell'oblio, e tornando a riscattare il passato, animando il presente segnato dall'ora esatta di due lancette da cui desumere nello stesso momento quanto manca a domani o quanto all'appuntamento a cui sperare di arrivare puntuale.

Lo spazio nudo piano piano si veste, "indossa" i panni della nostra esistenza attraverso i nostri oggetti che diventano gli "abiti" dello spazio stesso in cui si svolge il nostro abitare una casa. Gli oggetti testimoniano della nostra presenza, talvolta più straziante assenza, diventando elementi di una esistenza in perenne transito tra le variegate sfumature di una realtà dondolante tra gli estremi della felicità e della tragedia.

Ambienti e oggetti che raccontano di noi e del nostro stare al mondo e che ci sopravvivranno poi quando si faranno tracce del nostro passaggio[14].

Oggetti che coabitano con noi, animando la nostra casa e trasformandosi essi stessi in soggetti[15] nell'istante esatto in cui gli si attribuisce un valore affettivo, parte integrante della casa che erroneamente si ritiene essere una "scena fissa" e che al contrario è cerchio magico le cui stanze travalicano i limiti fisici della sola estensione planivolumetrica[16].

Case in cui offrire riparo e riposo a «…noi irresponsabilmente poco attenti agli attimi di felicità, dimentichi delle nostre angosce di barbari nomadi»[17], così lontane da quelle «abitazioni per esteti anoressici che si possono guardare ma non abitare»[18] della recente architettura-manifesto di Rem Koolhaas o di Herzog e de Meuron come lo era già stata quella di Mies e di Philip Johnson. «*Là tout n'est qu'ordre et beauté*»[19].

---

[14] Come scrive Georges Perec: «Il tempo che passa (la mia Storia) deposita residui che si accumulano: foto, disegni, feltri di pennarello da tempo asciugati, cartelline, vuoti a perdere e vuoti a rendere, imballaggi di sigari, scatole, gomme, cartoline, libri, polvere e soprammobili: ed è quel che io chiamo la mia fortuna» (G. Perec, *Specie di spazi* (1974), tr. it. Bollati Boringhieri, Torino 1989, p. 34).

[15] Si vedano, sul punto, E. Coccia, *Filosofia della casa. Lo spazio domestico e la felicità*, Torino, Einaudi 2021; E. Coccia, *Metamorfosi. Siamo un'unica, sola vita*, Torino, Einaudi 2022.

[16] «Abitare una camera, che cos'è? Abitare un luogo, vuol dire impossessarsene? Che significa impossessarsi di un luogo? A partire da quando un luogo diventa veramente vostro? Quando si mette in ammollo tre paia di calzini in un catino di plastica rosa? Quando si fanno riscaldare degli spaghetti su un camping-gas? Quando sono state utilizzate tutte le grucce spaiate del guardaroba? Quando si è fissata alla parete con delle puntine una vecchia cartolina che raffigura il Sogno di Sant'Orsola del Carpaccio? Quando vi si sono provati i tormenti dell'attesa, o le esaltazioni della passione, o i supplizi del mal di denti? Quando si appende alle finestre le tende del proprio gusto, e tappezzati i muri e levigati i parquet?» (G. Perec, *Specie di spazi*, cit., p. 34).

[17] Á. Siza, *Vivere una casa*, in «Domus», (771), 1995, p. 77.

[18] L. Prestinenza Puglisi, *Rem Koolhaas. Trasparenze metropolitane*, testo e immagine, Torino 1997, p. 14.

[19] «Grosso modo, abbiamo tratteggiato fin qui lo svolgimento morfologico degli *inabitabili*, dense e rinfrescanti raffiche d'arte che non piegano il capo al menomo utilitarismo: nessuno vi entra, nessuno vi stende, nessuno vi si siede sulle calcagna; nessuno si insinua nelle concavità, nessuno saluta con la mano dall'impraticabile balcone, nessuno agita il fazzoletto, nessuno vi si getta di sotto. *Là tout n'est qu'ordre et beauté*», J. L. Borges-A. Bioy Casares, *Sboccia*

Ninnoli accumulati alla buona in ceste di fortuna, scatoli imballati colmi di cose opportunamente raggruppate e classificate con precisione tassonomica esemplare, fili elettrici a centro stanza penduli che alimentavano lampadari adorni di gemme vitree a forma di goccia sfaccettata e scintillante e magnetica che il bambino che siamo stati raggiungeva allungando la manina ogni sera per svolgere il rito della buonanotte stando seduto a cavalcioni sulle spalle del proprio papà, un (mobile) tavolo immobile a centro stanza vuota, il rimbombo di parole appena pronunciate nella stessa stanza vuota, il letto spoglio sotto cui ci si nascondeva, pareti di colore sbiadito su cui leggere per contrasto chiaroscurale la composizione delle riquadrature a misura varia lasciate da quadri e fotografie e poster spesso irrimediabilmente inclinati. È proprio questo disegno involontariamente astratto a tramutarsi nell'emblema conclusivo e inappellabile del senso dell'assenza in una casa definitivamente vuota di vita vissuta e del senso della presenza in una casa indefinitivamente pregna di vita vissuta.

Alla fine resta questo.

Anzi no, non alla fine bensì tutte le volte al re-inizio.

ARMANDO MINOPOLI architetto, insegna Design all'Istituto Suor Orsola Benincasa di Napoli

arminopoli@hotmail.com

un'arte, in *Cronache di Bustos Domecq* (1967), tr. it. Einaudi, Torino 1999, p. 57.

S&F_n. 27_2022

RECENSIONI&REPORTS

VALERIO SPECCHIO

IL CORPO INAGGIRABILE:

TESSUTO E SOSPENSIONE TRA IO E MONDO

Riflessioni sul Convegno Internazionale:
"Merleau-Ponty: ritornare alla percezione /
Merleau-Ponty:revenir à la perception"
Università degli Studi di Napoli Federico II, 8-9-10 Giugno 2022

1. ***Introduzione*** **2.** ***Corpo e linguaggio*** **3.** ***Spazio e movimento***
4.*Intersoggettività tra evento e memoria:* **Polemos** ***ed*** **Eros**

ABSTRACT: ***INESCAPABLE BODY: TEXTURE AND SUSPENSION BETWEEN THE SELF AND THE WORLD***
This paper tries to present the discussion that took place in the International Conference "Merleau-Ponty: Revenir à la Perception". For this purpose, I have divided the different questions that made up the debate in three macro-thematic areas: body and language, space and movement, intersubjectivity between event and memory. Every macro-area is naturally linked with the others and tries to deliver the actuality of Merleau-Ponty's thought starting from the analyses proposed during the conference.

1. *Introduzione*

Per tentare di restituire, sottoforma di testimonianza, l'esperienza del convegno internazionale "Merleau-Ponty: Ritornare alla percezione", tenutosi a Napoli, in Aula Aliotta, dall'otto al dieci giugno, non posso che partire dall'enorme piacere che mi ha accompagnato. E questo per diversi motivi.

Per iniziare, la possibilità di essere *fisicamente* presente in uno spazio di discussione, tra una costellazione di ricercatori che nel loro vicendevole rapportarsi hanno *incarnato* quel che per me

risulta essere il fondamento e il fine dell'università: incontro, dialogo, crescita nel confronto continuo di saperi.

In secondo luogo – e questo sarà anche motivo di scelta metodologica per la traduzione dell'esperienza – il convegno, strutturato in quattro panel, ha disegnato traiettorie teorico-speculative che nel loro intrecciarsi rendono praticamente impossibile restituire una completa ed esauriente tessitura dei contenuti proposti. Pertanto, scartando l'ipotesi di una cronistoria che nel migliore dei casi risulterebbe forma metodologica inappropriata per l'oggetto in questione, tenterò di *inseguire* le tematiche affrontate attraverso tre macro-aree, confessando l'impossibilità materiale di una completa restituzione dell'esperienza. Tutto ciò, a mo' di conferma del tema stesso del convegno: la percezione che, nel suo esser vissuta da un corpo proprio, non troverà mai pace in una definitiva traduzione positivistica bensì, aprendo alla relazione con l'*altro*, non potrà che *ritornare* a "bussare" in vista di nuova immersione esistenziale.

Infine, non posso negare la gioia dell'aver incontrato e conosciuto nuove persone che ancora una volta confermano l'impossibilità di far ricerca se non a partire dalla relazione umana stessa in quanto motivo fondante e fondativo di ogni possibile circuito di confronto.

2. *Corpo e Linguaggio*

Aderendo al metodo fenomenologico, in quanto approccio filosofico che ricolloca le essenze nella fatticità dell'esistenza, partirò dalla prima macro-area tematica: l'intreccio tra il concetto di corpo e il concetto di linguaggio in Merleau-Ponty.

Difatti, all'interno dell'architettonica merleau-pontyana, tale intreccio risulta essere il più fecondo per inquadrare il peculiare approccio fenomenologico del pensatore francese e al contempo il versante più delicato.

In vista di un superamento del dualismo cartesiano in cui l'atteggiamento riflessivo vorrebbe separare il soggetto dall'oggetto e il corpo da se stesso, neutralizzando l'ambiguità dell'esistenza, Merleau-Ponty ci dice che «l'oscurità si propaga per l'intero mondo percepito», poiché l'unico modo di conoscere il mondo è vivere il dramma unico che l'io in quanto corpo è.
Il corpo è sempre *altro* da ciò che è, mai chiuso e mai superato nell'intreccio tra natura e cultura, sessualità e libertà. L'esperienza del corpo proprio è ciò che tiene l'uomo nel tessuto del mondo in quanto non vi è distinzione ultima ma reciproca interrelazione: «il mio corpo proprio è abbozzo provvisorio del mio essere totale».
Nella comprensione dell'uomo *nel* mondo (che è sempre già là e al contempo messo tra parentesi attraverso il momento dell'epochè fenomenologica) nel ritorno alle *cose stesse*, vi è la sconfessione di ogni scienza riduzionistico-deterministica. Nel ritorno al mondo anteriore alla conoscenza, la percezione è il corpo vissuto attraverso il quale dallo sfondo si staccano figure, in cui la *situazione* è l'ambiente proprio dell'umano.
La riflessione, non ritirandosi dal mondo ma attraversandolo, non aspira all'unità della coscienza come fondamento originario in vista di una manomissione degli enti. La presa di distanza dalle cose permette la trascendenza nel rilevare il mondo nella sua paradossalità. Da qui lo stupore, il *thaumazein* in quanto colpo originario dato dalla percezione del problema.
La percezione, quindi, è possibilità d'accesso al tessuto del mondo di cui il soggetto stesso è parte, tutt'uno con l'oggetto. Lo sguardo, quindi, a cui apre la fenomenologia è l'intenzionalità fungente in quanto unità naturale e ante-predicativa del mondo. Il mondo fenomenologico non è essere puro ma *senso* che traspare nell'interazione tra i corpi vissuti. Su questo punto è possibile menzionare l'intervento di Antonino Bondi – in quanto espressione di una rilettura semiotica di Merleau-Ponty tra strutturalismo e

fenomenologia – che propone una ricostruzione del modello di strutturazione di senso secondo l'ipotesi semiogenetica.

L'uomo è immerso e, al contempo, tessuto del mondo che lo circonda, in quanto la costruzione di significato è intimamente vissuta, i segni-oggetto sono una pulsazione vibrante nella dinamicità soggetto-mondo. Vi è una tensione costante tra la *determinatezza* degli oggetti e l'*indeterminatezza* del contesto umano. Il concetto centrale adoperato da Bondi è stato il concetto di *motivo* e la relazione di *motivazione* fra corpo, espressione e mondo. Nella continua ritrascrizione ontologico-artificiale tra uomo e mondo vi sono delle trame intenzionali e dei progetti anticipativi che caratterizzano la *prassi*. La motivazione governa le relazioni percettive e semiotiche fra soggetto e mondo, consentendo di pensare insieme i flussi esperienziali e le forme che s'inanellano in concatenamenti di senso multipli. In una tale ottica, la motivazione descrive una varietà di rapporti che governano i processi di formazione delle esperienze di senso e focalizza le tensioni fra l'indeterminazione delle formazioni allo stato nascente e la loro partecipazione alla costruzione di unità di formato superiore. I motivi costituiscono il modo in cui le forme sorgono da uno sfondo, iscrivendosi in sistemi di valori e significazioni. La motivazione, allora, altro non è che la storia delle *riprese* che riposa su alcune identità rese temporaneamente stabili, vale a dire i *segni* in quanto traccia d'iscrizione (mitica, immaginaria, etc.), che danno vita a forme di identità narrativa.

Non a caso il concetto stesso di motivazione, soprattutto nell'incompiuto *Le visible et l'invisible,* trova riflesso nello scomparire, nell'andare verso l'altro nella trasfigurazione di senso della metafora, nell'iconologia e musicologia.

Percezione-visione-intenzione conformano e alimentano quella dimensione caotica, quel *tourbillon* nella spazialità avvolgente da cui fluisce il fronte ecosemiotico umano nell'intreccio tra cose

del mondo e forme soggettive, o meglio, quell'oscurità che inerisce l'intero mondo percepito di cui parla Merleau-Ponty.
Il fenomeno si dispiega nella relazione con il *motivante*, la realtà non è causa originaria ma flusso, pluralità di senso, presa del corpo in un campo motivato. Il linguaggio non è referenziale ma intelaiatura socio-emotiva, la percezione è originariamente semiotica.
Su tale versante si ricollega l'intervento di Paolo Missiroli sulla riflessione antropologica contenuta nella *Fenomenologia della percezione* attraverso le nozioni di *apertura* ed *espressione*. Secondo tale prospettiva, Merleau-Ponty supererebbe la concezione inizio-novecentesca dell'uomo in quanto essere manchevole, consegnato alla "sprovvedutezza morfologica" nella polarizzazione del negativo psico-fisico. Il riferimento principale per questa rilettura merleau-pontyana risulta essere la psicologia della *Gestalt*: l'uomo si risolve nell'espressione e al contempo è espressione di un qualcosa, di un silenzio primordiale che viene spezzato dalla voce, dal gesto espressivo che non può esser ricondotto a una sommatoria facente capo a un totale, ma la forma stessa del gesto sonoro è più del totale traducibile in segni. Da questo punto di vista, la questione del negativo e della distanza può trovare la propria declinazione in una concezione d'esistenza che trova nel paesaggio del mondo quell'inoggettivabile di cui il corpo stesso è espressione ultima e mai ultimata nel movimento di "sfuggimento" del soggetto da se stesso.
L'alterità, quindi, tanto nel/del corpo, tanto nel/del linguaggio, non viene sottomessa in quanto negativo fondamento della *risoluzione* ma riporta alla riflessione in quanto mistero del mondo, in un "movimento di meditazione infinita" per riprendere le parole dello stesso Merleau-Ponty.
La percezione, pertanto, non è somma di sensazioni discrete ma struttura. Il testo del mondo esterno è squarciato e ricostituito di volta in volta nella coscienza intenzionale, nelle forme, nei

gesti e nei segni. L'esperienza sensibile, quindi, «è un processo vitale tanto quanto la procreazione», così come scritto nella *Phénoménologie de la perception*.

L'unità del mondo non è costruita per associazione ma è condizione di percezione. Ogni sentire è comunicazione vitale con il mondo, tessuto intenzionale. Il primo atto filosofico, allora, consiste nel ritornare al mondo vissuto al di qua dell'oggetto: io-altro-mondo, ecco il campo fenomenico. L'immediato è la struttura in cui soggetto e oggetto sono tutt'uno. Il corpo è veicolo primo di rapporto al mondo, al contempo ciò che più è proprio e ciò che incarna l'irriducibilità del patto costitutivo tra il visibile e l'invisibile, le zone d'ombra da cui emerge ogni determinazione mai definitiva.

3. *Spazio e Movimento*

A questo punto è possibile affrontare la seconda macro-area tematica.

La spazialità, così come concepita da Merleau-Ponty, è reciproca flessione dello schema corporeo nell'unità spazio-temporale del *corpo proprio* come termine terzo sospeso tra figura e sfondo. La significazione è data dal corpo nel rimando tra sensibilità e significato, struttura e forma, fisiologia e psicologia. Il corpo è nucleo significante, nodo tra esistenza ed essenza, tra io e mondo.

Il filosofo francese, sulla scia della prospettiva inaugurata da Kant nella correlazione tra fenomeno e condizioni di manifestazione, dà primato alla relazione per arrivare all'immediato, in una circolarità che nel partire dalla fatticità vi ritorna decostruendo i due approcci apparentemente antitetici di razionalismo ed empirismo. In una tale ottica, la conoscenza non è altro che emergenza e oblio tra sfondi e figure: è l'orizzonte indeterminato ad assicurare le identità.

Nella struttura oggetto-orizzonte la temporalità è flusso, rete, melodia, in cui vi è un unico momento: la durata. Non c'è successione riorganizzativa di causa ed effetto secondo linearità. Il problema sta nell'unità relazionale, vi è un avvilupparsi reciproco tra io e mondo, la significazione è immanente al sensibile che va a significare, così come lo spazio è intreccio continuo *al* movimento.

L'analisi della parola e dell'espressione, che ci fa riconoscere la natura enigmatica del corpo proprio in cui l'umano è connaturato all'artificio originario della potenza originaria da cui deriva, è essa stessa movimento fonetico. Strutturazione simultanea di corpo, emozione ed espressione, spazio e tempo. Il mondo linguistico e intersoggettivo non è distinto dal mondo, è il mondo stesso, ciò che lo tiene insieme è il *movimento*. L'intenzione significante, come il corpo, è una mancanza che cerca di colmarsi al pari del gesto e del movimento che cerca l'equilibrio perduto del silenzio primordiale.

Lo schema corporeo, quindi, è l'unità spaziale e temporale, l'unità senso-motoria del corpo. Non è il semplice risultato delle associazioni stabilite nell'esperienza ma una presa di coscienza della postura nel mondo intersensoriale, una "forma" ancora una volta nel senso della *Gestaltpsychologie*. Lo schema corporeo è dinamico poiché sempre in situazione. Il corpo è il terzo termine sottinteso tra figura e sfondo attraverso cui il movimento fonda lo spazio, così come indicato da Prisca Amoroso. La simultaneità mondo-io trova nell'infanzia quella correlazione confusa tra sfondo e figure, quel contatto ingenuo, in cui il corpo risulta inaggirabile. Come a dire che ogni successiva fase socio-pedagogico-culturale interposta tra l'indistinto mondo-io e la significazione non potrà che portar con sé quell'irriducibile che il corpo stesso incarna.

Su questo punto è possibile riprendere le parole di Andrea Pace Giannotta in merito alla co-emergenza di corpo e mondo

nell'ontologia della carne concepita da Merleau-Ponty. Difatti, ne *Il visibile e l'invisibile*, il filosofo francese approda a una concezione del carattere fenomenico della percezione al di là dell'alternativa tra internalismo ed esternalismo. Secondo tale analisi, al cuore del processo percettivo vi è una dimensione qualitativa che precede la distinzione tra soggettività e oggettività e che, piuttosto, si colloca alla base della loro genesi comune. Merleau-Ponty individua la *carne* come dimensione ontologica fondamentale che è all'origine dell'intreccio chiasmatico tra soggetto e oggetto, interno ed esterno, mente e mondo, spazio e movimento. Ogni metafisica della sostanza viene messa in scacco da una metafisica della corporeità, o meglio da una filosofia del processo, in cui la percezione rifluisce in un rimando di eventi e flussi di qualità non discrete, avvicinandosi alla teoria del monismo neutrale.

4. *Intersoggettività tra Evento e Memoria:* Polemos *ed* Eros

In quest'ultima sezione il focus verterà sul confluire intersoggetivo tra corpo proprio e corpo altrui, o meglio, tra l'originario tessuto che inerisce il corpo al mondo, l'io al non-io, l'individuo alla società e in ultimo, l'amante all'amato. I concetti di *polemos* ed *eros*, riflettendosi nella questione di evento e memoria, riemergono, nel loro inscindibile legame, in quanto punto di raccordo delle riflessioni di cui è andato componendosi il convegno. Su questo versante, non è un caso che il paragrafo conclusivo della *Fenomenologia della percezione* sia proprio intitolato "la libertà".

Per Merleau-Ponty ogni uomo, immergendosi nel mondo, assume la propria situazione in cui il concetto stesso di libertà è impigliato "nelle contraddizioni dell'impegno".

Ed è proprio in questa sovrapposizione tra impegno e libertà che va a inserirsi l'intervento di Chiara Palermo su «la memoria del corpo: percezione e storia in Merleau-Ponty e Kader Attia».

La dimensione sociale e politica del corpo vissuto trova nella *ferita* e nella deformazione fisica proposta da alcune opere di Kader Attia la propria materializzazione. L'artista, infatti, sembra adoperare i materiali e le tecniche produttive in vista di una terapia circa la memoria collettiva della coscienza occidentale, non volendo tradire il rimosso psichico che contraddistingue l'individuo cosiddetto civilizzato. In Kader Attia la riparazione è fuoriuscita dal regno animale. Il corpo vissuto del sociale è la storia in quanto attestato di reciproca fallibilità, tanto del corpo proprio quanto dell'*impegno*. Il divenire che si manifesta nel turbinio della storia, per riprendere un'espressione benjaminiana, non può accettare la propria condizione al pari dei soggetti affetti dalla sindrome dell'arto fantasma. In tal senso, Merleau-Ponty, rifiutando una visione idealistica della storia, non può che toccare con mano l'esperienza della clinica, in quanto storia materiale e resoconto dei traumi scanditi dall'esistenza.

La storia clinica allora, in quanto riflesso della storia occidentale, risulta nient'altro che confessione di un passato che non decide a farsi tale, ma a ripetersi e proliferare in quanto eterno presente.

Su questo tema si riallaccia l'intervento di Gael Caignard circa le riflessioni di Merleau-Ponty sulla storia e "le piège de l'événement".

Il filosofo francese, difatti, prendendo le distanze dall'approccio *engagé* di stampo sartriano, tentò di disinnescare la trappola dell'evento alimentata dall'estrema vicinanza tra la percezione dell'evento storico e il dibattito conseguente. Il nesso tra funzione percettiva e riflessione, nell'ottica merleau-pontiniana, è intimamente legato. L'estrema visibilità dell'evento ne nasconde la fatticità, l'essenza. In tal senso, in continuità con le analisi di Merleau-Ponty, eventi come l'11 Settembre o la pandemia di SARS-CoV-2 ci restituiscono la problematicità del

riflettere la storia nell'eccesso di prossimità all'evento, seppur nell'ossimoro di una distanza incolmabile: eco ultima del corpo vissuto nella propria sospensione tra evento e memoria.

Proprio a partire da questa sospensione, l'intervento di Salvatore Prinzi tenta di restituire Merleau-Ponty al proprio vissuto biografico, scandito dalle due Guerre mondiali, la resistenza francese, la Guerra fredda e la decolonizzazione. La guerra, difatti, ha occupato un posto dirimente nello sviluppo delle riflessioni del pensatore francese che, come accennato, spingono in direzione di una filosofia della *coesistenza* che faccia i conti con le contraddizioni dell'impegno. La libertà, al pari del corpo vissuto, non può essere una sommatoria di organi/azioni ma una struttura di inerenza al mondo, così come la società non può essere spartizione di campi dualisticamente intesi così come vorrebbe ogni propaganda bellica.

Infine, sempre in un'ottica di superamento del dualismo cartesiano, che tutt'oggi si ripercuote tanto sul campo di guerra tanto negli ambiti di ricerca universitari, mi trovo "obbligato" a concludere su quello che è il più scivoloso dei temi, oltre che il più singolarmente vissuto: l'*eros*. Accentratore di energie sociali e sublimazione in senso freudiano, il rapporto amoroso risulta ciò che più *interessa* la dimensione sociale nel proprio legame tra singolarità. L'individuo, in vista di un riconoscimento mai sazio, poiché senza fondamento, sembrerebbe vivere la propria esperienza amorosa in un intraducibile che egli stesso riconosce.

O meglio, un inoggettivabile che spererebbe di non oggettivare mai fino in fondo.

Riprendendo l'intervento di Andrea Nicolini, quindi, è possibile ritrovare nella letteratura, e in special modo in Marcel Proust, la forma estetica che più si avvicina al concetto di "carne viva nell'invisibile" così come concepito da Merleau-Ponty.

L'esistenza individuale è rifrazione singolare del mondo verso se stesso. L'altro, l'amato, è l'invisibile in quanto piega del

visibile enigmatico. Ed è proprio questo volto, il volto dell'amato, in cui la letteratura esplora la prossimità vertiginosa con l'essere secondo il senso merleau-pontyano: un volto, al contempo fatto di carne e di forme, che nel proprio ritornare *sempre* non torna mai.

Con la letteratura si tratta di trafiggere la carne in un *logos* selvaggio: contingenza e scontro. L'amato è immanenza della pelle e, al contempo, trascendenza dell'amore.

Il volto continua a sfuggire nella menzogna della memoria in quanto riflesso dell'amore. Finché il corpo proprio è vissuto nell'altro, l'amato in quanto oggetto non può che corrispondersi al soggetto, in un'evanescenza ontologica che permette l'abbandono dell'evento, e quindi dello spazio, e della memoria, e quindi del tempo. È sulla *carne* del pensiero, ossia sulla co-implicazione di pensiero ed emozione, che la filosofia troppo spesso ha disertato rispetto alla letteratura. Le *cose*, quindi, agli occhi di Merleau-Ponty si presentano *irrecusabili*. Tanto la persona amata quanto gli schiavi della catena di montaggio occidentale, tanto l'*altro* dal corpo vissuto quanto il corpo vissuto stesso nella propria irriducibile alterità. La libertà non può che volersi nell'uscita dalla singolarità.

La filosofia non ha altra funzione che riportarci alle cose stesse, nella loro fatticità. *"L'uomo non è che un nodo di relazioni, solo le relazioni contano per l'uomo"*. Commozione, molteplicità intersoggettiva, divenire altro del corpo: riparare le ferite nella contingenza della storia singolare e collettiva, carne e dispersione.Il corpo è un'assenza, una dimensione di fuga, movimento di coesistenza e mai sommatoria di funzioni chiamate organi. La percezione, allora, consegnandoci alla pluralità irriducibile, ci *conferma* nella resistenza dell'oggetto a ogni determinazione tramite l'assoluto inaggirabile: il corpo.

specchiovalerio@gmail.com

Yuk Hui
Cosmotecnica. La questione della tecnologia in Cina
tr. it. di Sara Baranzoni, Nero, Roma 2021, pp. 288, € 20

Gli ultimi anni hanno visto un fiorire di pubblicazioni dedicate alla crisi climatica che, da un punto di vista filosofico, spesso giungono a mettere in questione il paradigma occidentale moderno della diade oppositiva uomo-natura. Il dibattito sull'Antropocene si è sempre più polarizzato tra il rifiuto cieco della tecnologia dovuto alla "scoperta" dei danni che l'ideologia dell'illimitato progresso ha comportato sulla natura, e la glorificazione delle sempre maggiori conquiste della specie umana in ogni campo, dal microcosmo all'esplorazione dell'universo.

Il testo di Hui, pubblicato in originale in inglese nel 2016, rientra perfettamente in questo ambito di interesse e il suo punto di partenza è l'origine del concetto di tecnologia in Cina e quindi la filosofia orientale. Per quanto infatti la tecnica sia da sempre presente sulla Terra come attività umana, il concetto filosofico di tecnica non può essere considerato universale. Tale questione non può continuare a essere ritenuta marginale dal pensiero occidentale: alla luce delle complesse conseguenze della globalizzazione tecnologica e del dibattito sull'Antropocene, le relazioni tra il pensiero tecnologico orientale e quello

occidentale risultano infatti determinanti e ci aiutano a allargare finalmente lo spazio di riflessione. Per questo, uno dei grandi contributi di questo saggio è di mettere in evidenza lo strettissimo legame che corre tra lo sviluppo di una filosofia della tecnologia cinese e la contemporanea lotta alla crisi climatica.

L'obiettivo di Hui sembra particolarmente ambizioso e si rivela addirittura più complesso di quanto sembri in prima battuta poiché l'autore inizia il suo lavoro dichiarando che il pensiero cinese, in realtà, non ha mai sviluppato un concetto di tecnica come lo abbiamo pensato noi occidentali e non possiede alcuna coscienza tecnologica. Tale impostazione del lavoro può risultare contraddittoria con le premesse dell'autore che inizialmente propone di mettere in campo un confronto tra due modi di vivere la tecnologia, per dichiarare subito dopo la presenza di una forte asimmetria concettuale tra questi, per cui, il modello orientale non sarebbe in alcun modo paragonabile al nostro. Ciò conduce allora alla necessità di ridefinire il termine da indagare, che non è più quello di tecnica/tecnologia.

La proposta di Hui è quindi di concepire la tecnica non solo come categoria filosofica ma come categoria ontologica per interrogarla su una configurazione più estesa, cioè la cosmologia propria della cultura dalla quale è emersa. Da qui la scelta del termine *cosmotecnica* che sana la tradizionale opposizione natura/tecnica ma anche la progressione da magia/mito a scienza: «La cosmotecnica propone di riprendere la questione della modernità reinventando il sé e la tecnologia allo stesso tempo, dando la precedenza a morale ed etica» (p. 236). Se l'*episteme* globale dominata dalla mentalità capitalistica occidentale venisse rimpiazzata da una varietà di saperi e conoscenze si potrebbe produrre un nuovo modo di esistenza.

La prima delle due parti che compongono il saggio, dedicata allo studio storico di un pensiero tecnologico in Cina, incontra subito

una difficoltà anche di tipo linguistico nel proporre una comparazione tra il pensiero cinese e quello europeo: non è possibile trovare nel pensiero cinese né l'equivalente della *téchne* greca né del paradigma prometeico su cui si è fondata la filosofia della tecnica occidentale. È proprio il mito di Prometeo, infatti, caratterizzato dalla profonda frattura tra uomini e dei e da una concezione dell'esercizio della *téchne* come violenza e dominio sulla *physis*, che, assunto a paradigma del pensiero tecnico in Occidente, segna una distanza fondamentale con il pensiero mitico orientale. Il pensiero cinese presenta piuttosto una visione olistica in cui *Qi* e *Dao*, le due categorie filosofiche di cui Hui sceglie di seguire lo sviluppo, sono inseparabili e testimoniano di un'originaria armonia tra produzione umana e principio cosmologico: *Dao* ha bisogno di *Qi* per manifestarsi in forme sensibili e *Qi* ha bisogno di *Dao* per diventare perfetto.

Questa differenza tra la mitologia occidentale e quella cinese si riflette poi in configurazioni della tecnica molto diverse; perciò solo con il superamento dell'opposizione tra natura e tecnica, con l'unificazione tra ordine cosmico e ordine morale attraverso la ricerca di una cosmotecnica e situando storicamente la relazione tra *Qi* e *Dao*, si può rendere giustizia a una filosofia della tecnologia cinese.

Per cogliere al meglio l'operazione di Hui in tutta la sua complessità è necessario calarsi quindi in una nuova dimensione che non richiede soltanto un ripensamento della tecnica ma che dimostra quanto questo ripensamento sia inscindibilmente legato a altre, forse più fondamentali, questioni filosofiche. Non si può produrre una filosofia della tecnica cinese se non partendo dalla tematizzazione della questione ontologica: «Piuttosto che l'"Essere", ciò che sta a cuore tanto all'insegnamento confuciano quanto a quello daoista è il tema del "Vivente", inteso come la possibilità di condurre una vita morale o buona» (p. 65).

La relazione tra ontologia e tecnica nella filosofia cinese, così evidenziata, permette quindi di portare allo scoperto anche l'ontologia sottostante la filosofia della tecnica occidentale ovvero quel legame originario e fondativo di cui sembrano essersi perse le tracce. In tal senso, mentre in Occidente la tecnica è imitazione e perfezionamento della natura, in Cina è piuttosto legata alla realizzazione del bene morale del mondo ed è subordinata all'ordine cosmologico.

È solo a causa di una contingenza storica che la Cina assume per sé la condizione tecnologica occidentale: all'indomani delle Guerre dell'Oppio, la Cina capisce di avere la necessità di sviluppare la propria tecnologia e procede a un adattamento forzato al sistema occidentale. In questo processo i due termini fondativi di *Qi* e *Dao* vengono stravolti e l'originaria visione olistica si trasforma in dualismo cartesiano. Si è messo in atto «un movimento di modernizzazione inteso come la possibilità di utilizzare il *Qi* per realizzare il *Dao*» (p.132). Questa enfasi strumentale sull'"uso" rovescia il *Dao* e il *Qi* e la cosmotecnica olistica viene rimpiazzata dal meccanicismo occidentale. Tutto ciò che risultava incompatibile con quella modernità, e intralciava quindi le forze di sviluppo, viene relegato alla categoria di "tradizione".

La seconda parte del saggio che pone al centro la relazione tra la tecnica e il tempo, può risultare quindi come l'applicazione delle categorie individuate nella prima parte e dei risultati di quella riflessione come possibili risposte alle sfide della globalizzazione. La globalizzazione tecnologica e il dibattito sull'Antropocene stanno creando un'*impasse*: la Cina si trova nelle condizioni di dovere rifiutare un modo di concepire la scienza e la tecnica che è frutto di una modernità che non hai mai vissuto. Come chiarisce Hui, infatti, la fine della modernità in Europa non ha affatto significato la sua fine in generale, anzi, la dinamica accelerativa che ha coinvolto l'Occidente e che si è riproposta in

Oriente, ha prima condotto la Cina a sincronizzarsi con un pensiero tecnologico che non le è originariamente proprio, e in seguito a riproporlo, facendosi paradossalmente essa stessa portatrice di quella dinamica "moderna" di sfruttamento, nei paesi del Terzo mondo. La globalizzazione ha reso impossibile distinguere il concetto cinese di tecnica da quello occidentale poiché l'accelerazione promossa da Xiaoping ha sviluppato un universalismo che è assimilazione e che conduce all'oblio ogni specificità culturale.

Nasce l'esigenza di costruzione di una *episteme* nuova non basata più sull'equivalenza o la somiglianza tra il linguaggio della metafisica e della tecnica occidentali e il pensiero tradizionale cinese, bensì sulla differenza, «una *traduzione* che permetta a una *transduzione* di aver luogo» (p. 252). Allo stesso modo la sincronizzazione egemonica dell'Occidente, dovrebbe essere sostituita da una nuova costituzione del tempo; non solo in una nuova narrazione, ma soprattutto nella creazione di una nuova pratica e nuovi saperi che si confrontino con la modernità e attraverso questo confronto la superino. L'Antropocene è strettamente legato al concetto di modernità e a una cosmotecnica che possiamo identificare con il capitalismo. Il superamento della rischiosa condizione globale in cui ci troviamo oggi è possibile allora sostituendo l'episteme capitalistica con altri saperi e conoscenze. Ciò non implica però un ritorno ai significati tradizionali di categorie come *Qi* e *Dao*, ma richiede una loro risignificazione appropriata alla nuova epoca.

La riappropriazione della tecnologia moderna dal punto di vista della cosmotecnica deve avvenire attraverso due passaggi: «il primo, tentato qui, richiede una riconfigurazione di categorie metafisiche fondamentali come *Qi* e *Dao* in quanto base e fondamento; il secondo, esige che su questa base si ricostituisca un'*episteme* capace a sua volta di condizionare l'invenzione, lo

sviluppo e l'innovazione tecnica, cosicché quest'ultima non sia più ridotta a mera imitazione o ripetizione» (p. 250).

I riferimenti presenti nel testo di Hui sono moltissimi e spaziano chiaramente da un paese all'altro, così come permette la formazione dell'autore stesso: originario di Honk Kong, dove si laurea, Hui prosegue la sua attività in Francia conseguendo il suo *PhD* con Bernard Stiegler, per poi insegnare in Germania, in Cina e a Hong Kong. Da Heidegger a Nishitani e la Scuola di Kyoto, Stiegler e Mou Zongsan, ogni tassello di questa ricerca è funzionale all'arricchimento della prospettiva e alla comprensione di tradizioni diverse attraverso confronti e influenze che l'autore non manca di sottolineare puntualmente. La prima parte, molto affascinante e a volte quasi esoterica, ci avvicina a un mondo o meglio, a una visione del mondo, quasi incomprensibile, ma le citazioni dai più grandi classici della filosofia cinese, per quanto caratterizzate da un linguaggio complesso, si rivelano uno degli elementi di grande forza di questa ricerca.

La sensazione che lascia la lettura del testo di Hui è quella di essere travolti e colpiti dall'estrema diversità di coordinate rispetto a tutto ciò con cui abbiamo familiarità, tuttavia, per quanto sia un testo sulla cultura filosofica e tecnologica cinese, *Cosmotecnica* ci stimola a riflettere anche sulla cultura occidentale, sul pensiero mitico e sulla cosmologia che l'hanno originata e di cui ancora mantiene i tratti. Leggere della perdita nella cultura cinese di un legame fondamentale tra ordine morale e ordine del mondo, può servire all'Occidente come specchio in cui riconoscere le proprie ferite, come quella frattura tra ontologia e tecnica e che ha reso entrambe, nella loro parzialità, incapaci di rispondere agli interrogativi posti dai propri oggetti di indagine e che oggi va certamente ripensata verso una integrazione capace di affrontare adeguatamente gli squilibri della contemporaneità. Negli ultimi anni la filosofia postcoloniale si sta imponendo nel dibattito sfidandoci a pensare fuori dalla

Weltanschauung occidentale e della narrazione della Storia universale come storia europea. Il libro di Hui va anche oltre questa intenzione, alza l'asticella e si impone come un vero e proprio rompicapo, certamente molto stimolante e necessario.

FRANCESCA DE SIMONE

Luciano Floridi
Etica dell'intelligenza artificiale.
Sviluppi, opportunità, sfide
Raffaello Cortina Editore, Milano 2022, pp. 384, € 26

La rivoluzione digitale può essere annoverata tra i grandi *turning point* della nostra contemporaneità, al pari ad esempio della rivoluzione agraria o industriale, per via della sua enorme e innegabile portata trasformativa.

Oggigiorno sembra quasi banale sottolineare che le intelligenze artificiali, le IA, vivano al nostro fianco in ogni possibile ambito della vita umana, dallo shopping al marketing, dalle interazioni sociali all'intrattenimento, dai servizi bancari ai viaggi, dalla logistica al mondo del lavoro, e così via. Eppure, lo stesso termine "intelligenza artificiale" sembra ancora evocare, nella percezione comune così come nell'alveo del giornalismo sensazionalistico – e non ultimo anche entro alcuni settori dell'ambito accademico – una recondita paura di essere, come esseri umani intelligenti, lasciati indietro, superati e "detronizzati" – nella migliore delle ipotesi – dai nostri stessi artefatti. Complice, probabilmente, quasi un secolo di produzione fantascientifica: durante il secolo scorso, infatti, le intelligenze artificiali e le loro possibilità potevano essere contemplate o esplorate solo con l'immaginazione

della fiction, impiegando talvolta dei narrativi passaggi obbligati verso una frattura tra organico e inorganico, culminante con la futura e inevitabile sconfitta dell'umano. Lo stesso termine "robot" viene coniato negli anni '20 del Novecento da Karel Čapek nell'ambito della sua ormai celeberrima opera teatrale *R.U.R.*: tratto dal ceco *robota*, che significa lavoro forzato, o schiavitù, il robot di Čapek, il *primo* robot, era l'operaio artificiale deputato a liberare l'essere umano dalla fatica del lavoro. E, come si è ormai abituati a aspettarsi, lo schiavo umanoide diventerà man mano sempre più senziente, e si ribellerà in maniera violenta al suo padrone/creatore, determinando così la fine dell'umanità. Questo nodo narrativo rappresenta un *topos* letterario di grande successo, che dalla pagina o dal palcoscenico teatrale si riverserà sempre più, all'approssimarsi del nuovo millennio, nell'immaginario collettivo, e quindi nella cultura di massa. Hollywood lo ha declinato in mille modi: basti pensare a Skynet in *Terminator* o all'agente Smith in *Matrix*, a Samantha nel film *Her*, o a Ava in *Ex Machina*.

Le propaggini accademiche di questa visione estremamente competitiva tra intelligenze organiche e artificiali porteranno a una rinnovata paura semi-luddista, basata sulla previsione dell'avvento della Singolarità: un preciso momento nel tempo in cui l'intelligenza artificiale supererà di gran lunga l'intelletto umano, condannando quest'ultimo a languire per sempre nell'obsolescenza. Questa ipotesi eccessivamente catastrofica, e probabilmente, come si diceva, fortemente influenzata dalle narrazioni fantascientifiche, troverebbe un suo fondamento matematico nella celebre legge di Moore: nello sviluppo dei computer, il numero dei transistor sui circuiti integrati, e dunque la potenza di calcolo, raddoppierebbe ogni diciotto mesi. Questa affermazione di carattere empirico è stata coniata negli anni '60, agli albori dunque dell'era digitale, e poteva sembrare assolutamente valida in quel preciso momento storico di esplosione

delle tecnologie digitali; essa è tuttavia decaduta in tempi recenti – la crescita esponenziale di un fattore per un dato lasso di tempo non autorizza a pensare che essa debba procedere *ad infinitum*.

Nessun futuro catastrofico è dunque alle porte, e, soprattutto, l'intelligenza artificiale è già qui; invece di minacciare l'esistenza della specie umana, o meglio, surclassare il suo posto di dominio entro l'ordine naturale – questa la reale paura – l'IA e l'avvento del digitale hanno profondamente modificato il paesaggio umano, silenziosamente ma inevitabilmente. Sviluppi, opportunità e sfide, dunque, una volta sgomberato doverosamente il campo da suggestioni e paure cinematografiche, sono, nel testo di Floridi, ben vagliate e soppesate con rigore filosofico, allo scopo di delineare una possibile via da seguire per sfruttare al meglio il digitale e le sue promesse, e, attraverso di esso, costruire un mondo migliore.

L'assunto di partenza del volume, anche in vista di fornire una prima definizione di IA, consiste nell'attestare che le intelligenze artificiali non siano semplicemente super-intelligenze, poiché, nei fatti, non lo sono, ma che siano un qualcosa, anzitutto, che trasforma profondamente la realtà. Questi due punti, *cosa* effettivamente sia l'IA, e *come* essa operi nel mondo, sono assolutamente rilevanti in vista di una problematizzazione etica del rapporto tra umano e artificiale nella contemporaneità. Per quanto riguarda il primo punto, Floridi riconosce che l'IA sia difficilmente definibile in termini chiari, essendo (e funzionando come) un insieme di tecnologie, pratiche e artefatti matematici dalle più vaste applicazioni; l'autore si appoggia perciò alla definizione classica fornita da McCarthy, Minsky, Rochester e Shannon, nella loro *Proposta per il progetto estivo di ricerca sull'intelligenza artificiale di Darthmouth*, ovvero quello che è considerato il documento seminale per i primi studi sulle IA, datato 1955: «Il problema dell'intelligenza

artificiale è quello di far sì che una macchina agisca con modalità che sarebbero definite intelligenti se un essere umano si comportasse allo stesso modo» (citazione riportata dall'autore a p. 35). Questa definizione per Floridi resta assolutamente valida anche nel presente, e sulla scorta di tale definizione l'autore procede a rilevare, lucidamente, che il *proprium* dell'IA sia quello di avere una sorta di *potere di scissione*: lungi dall'essere un felice matrimonio tra il regno biologico e quello artificiale, l'IA comporta invece un definitivo *divorzio* tra l'agire in modo intelligente e la necessità di essere intelligenti per farlo. Le attuali tecnologie cosiddette intelligenti mostrano, per Floridi, solo comportamenti intelligenti, tali perché definibili sulla scorta della performance umana: non possiedono realmente intelligenza, ma sono in grado di svolgere compiti che richiederebbero a un umano intelligenza per portarli a termine. Una calcolatrice ad esempio può svolgere calcoli anche molto complessi, e anche meglio e più velocemente di un essere umano, ma senza per questo essere intelligente (né tantomeno super-intelligente) *stricto sensu*. Ciò che conta è il risultato, seguendo il ragionamento dell'autore; dunque l'IA inaugura una nuova forma dell'*agire*, e poco avrebbe a che far col *pensare*: l'IA non è in grado di pensare in modo intelligente, ma può assolutamente agire in modo intelligente. Ciò che resta da rilevare è dunque che l'IA consista in una riserva di agire intelligente, seppur dalle capacità non illimitate, ma alla portata di tutti.

Il potere di scollamento tra agire e intelligenza è caratteristica fondante dell'agire artificiale, ma esistono altri scollamenti operati direttamente dal digitale sull'umano, il che ci porta al secondo punto: *come* l'intelligenza artificiale operi nel mondo. Effettivamente, le generazioni precedenti hanno conosciuto esclusivamente una dimensione del vivere "analogica", entro la quale, ad esempio, *presenza* e *posizione* in senso spaziale sono

aspetti coincidenti e assolutamente inscindibili, che descrivevano l'essere situati nel qui e nell'ora. Il digitale ha, secondo Floridi, provocato invece una frattura anche tra questi ambiti, considerati assolutamente immutabili prima e senza il digitale: posso essere fisicamente in un luogo, ma connesso attraverso la tecnologia digitale a un portale online; posso cioè trovarmi in due luoghi nello stesso momento, essere posizionato in un luogo fisico ma presente contemporaneamente in un luogo virtuale. Per le generazioni dei nativi digitali, e soprattutto per chiunque abbia vissuto la transizione digitale pur ricordando la vita analogica, la telepresenza non è più fantascienza, ma una realtà ormai diffusa e quotidiana, quasi banale, un tratto comune a tutta l'esperienza umana nelle società dell'informazione della contemporaneità. In questo modo, si comprende come tale potere di creazione di nuovi spazi e ambienti che abitiamo e nuove forme di agire con cui interagiamo, proprio delle intelligenze artificiali, possa essere descritto come una *re-ontologizzazione* del mondo: non si tratta semplicemente, o soltanto, di una re-ingegnerizzazione del mondo, una ristrutturazione nuova o inedita, o di una semplice ricombinazione, ma più propriamente di una modifica radicale della sua natura intrinseca, della sua ontologia. Ovviamente, il passaggio a questo nuovo statuto ontologico comporta anche una *ri-epistemologizzazione* della modernità, vale a dire una riconfigurazione, operata direttamente dall'influsso del digitale nelle pratiche quotidiane, di conoscenze e idee prima consolidate e considerate stabili – come la coincidenza tra presenza e posizione.

È in questo senso che l'infosfera, o meglio l'insieme degli spazi virtuali aperti dal digitale e abitati da esseri umani materiali nell'epoca di Internet, sta compenetrandosi sempre di più con il cosiddetto "mondo reale": per Floridi non viviamo esattamente *online*, ma *onlife*. Lungi dall'essere una banale constatazione di costume contemporaneo, il potere di scollamento dell'IA diventa

ora un potere di *incollamento*; per usare la terminologia dell'autore, è il *taglia e incolla* della realtà che comporta anche una ri-significazione e una ri-strutturazione degli spazi reali – delle città o dei trasporti, ad esempio, o delle nostre abitazioni – in modo che meglio si adattino alle delimitate funzioni e possibilità delle IA di farle funzionare come e/o meglio di intelligenze umane. Per tale ragione, non solo il digitale ha re-ontologizzato l'esistenza umana, comportando anche una modifica radicale della sua epistemologia (taglia), ma ha realmente modificato anche lo spazio "analogico" in vista di un accrescimento di funzionalità dell'esperienza *onlife*, e di una migliore integrazione dell'agire artificiale nella realtà fisica (incolla). O in altre parole, e per dirla con Floridi, siamo noi umani a aver modificato il nostro spazio per adattarlo, o "avvolgerlo", attorno alle tecnologie smart. La progettazione degli ambienti fisici, siano essi un *hub* di Amazon, una catena di montaggio industriale robotizzata, o un'abitazione con elettrodomestici smart, dipende oggi interamente dalle possibilità dell'IA: se un agente in tali ambienti può essere artificiale, bisogna progettare, ristrutturare e dunque "avvolgere" tali ambienti intorno all'agente, in modo tale che sia in condizioni di svolgere con successo i suoi compiti. Il potere trasformativo delle IA si concreta anche in questo: è in atto un progressivo adattamento dell'ambiente all'IA piuttosto che dell'IA all'ambiente.

Sono molti i problemi etici sollevati dal prepotente ingresso del digitale nella vita quotidiana, si pensi ai concetti di progettazione e responsabilità; in questo importante contributo, Luciano Floridi guida il lettore attraverso una corretta concettualizzazione delle IA, per spianare la strada a una tematizzazione etica sulle intelligenze artificiali a partire da ciò che esse sono, da ciò che possono e non possono fare, per arrivare a considerare anzitutto i possibili usi (benevoli o

malevoli) che se ne possano fare. Floridi ritiene che il problema non risieda nell'IA in sé – l'apparato tecnoscientifico che ne permette e ne ospita l'esistenza non è in sé buono o cattivo – ma, piuttosto, nel grado di intenzionalità che vorremo imprimere a essa nel guidarne le applicazioni, e verso quali scopi. Siamo ancora in pieno potere di direzionare queste tecnologie, potenzialmente verso ogni obiettivo; l'imperativo che abbiamo, dunque, è solo quello di comprenderne le opportunità e coglierne l'impatto in prospettiva futura, facendo lavorare l'IA verso il miglioramento delle società umane.

Il vero intento del volume è quello di proporre dunque tutta una serie di riflessioni strettamente etiche, raccolte in quattro macroaree concettuali: rendere l'IA più adatta all'ambiente umano, per permetterle di affrontare problemi concreti come i disastri ambientali, o la povertà, o la disuguaglianza sociale; dovremmo rendere l'IA più adatta all'umano, e, per dirla con Kant, assicurarci che le applicazioni di essa facciano in modo che gli esseri umani vengano trattati sempre come fini, e mai come mezzi; dovremmo mettere la "stupidità" dell'IA al servizio dell'intelligenza umana, facendo dunque in modo che i benefici trasformativi vengano condivisi da tutti, e non solo da alcuni; in ultimo, dovremmo fare in modo che l'IA ci renda più umani, scongiurando il rischio di usi potenzialmente o intenzionalmente "maligni", che danneggerebbero la *polis* umana e l'intero pianeta. Attorno a queste riflessioni si sviluppano raccomandazioni o avvertenze *pratiche*, basate su anni di esperienza e lavoro sul campo, di carattere etico, politico e economico, per condurre questa nuova tecnologia, così promettente e così potente, verso il bene comune: per rendere le nostre società contemporanee più eque; per aiutare a risanare il pianeta dall'enorme minaccia – stavolta davvero catastrofica – del cambiamento climatico causato direttamente da interventi umani; per riequilibrare il mondo del lavoro alla luce delle possibilità dell'IA, per migliorare i

rapporti sociali, personali e internazionali, e così via. Si comprende dunque che quello dell'autore sia un progetto umano, più che tecnologico, non ancora completo ma certamente da compiersi: come egli stesso ricorda, «il mondo ha bisogno di una buona filosofia» (p. 277), che traghetti con sicurezza verso il bene l'immenso potere tecnologico a disposizione delle generazioni presenti e future.

STEFANO PALUMBO

Angela Balzano, Elisa Bosisio, Ilaria Santoemma
(a cura di)
Conchiglie, pinguini, staminali. Verso futuri transpecie
DeriveApprodi, Roma 2022, pp. 288, € 18

La storia, anche quella biologica, è stata scritta e raccontata dai dominatori; da qui deriva l'urgenza di narrare discorsi sulla vita a partire da condizioni di marginalità per improvvisare contro-manovre e resistenze. Questa antologia, *Conchiglie, pinguini, staminali*, si configura come una cartografia transfemminista in cui il "com-pensare", come modalità relazionale del pensiero, è messo alla prova nel tentativo di smantellare teorie e prassi nelle quali è l'Uomo a darsi come forma d'espressione dominante. L'obiettivo è quello, dunque, di indagare i rischi perpetuati dalla visione antropocentrica, che si impone mistificatoriamente come neutrale, nei confronti dell'alterità umana e non-umana. Qualsiasi pensiero *altro* e qualsiasi strategia di resistenza passano attraverso lo smantellamento del discorso egemone, che si è insinuato in qualsiasi ambito dell'organizzazione sociale costringendo a guardare il mondo attraverso gli occhi dei dominatori. Un primo passo è quello di operare una sostituzione dello sguardo attraverso un processo gioioso di distruzione-creazione in cui la destrutturazione del

pensiero dicotomico (primo fra tutti: il binomio natura-cultura) avvia la proliferazione di punti di vista alternativi e che mettono in crisi il potere di appropriazione come configurazione primaria della/nella relazione con l'altro-da-sé. Così il punto di vista transfemminista si esprime attraverso un posizionamento inedito nei confronti delle popolazioni umane e non-umane, liberando l'alterità dalla sua riduzione alle logiche di coercizione identitaria e di assoggettamento strumentale. L'ontologia classica, che impone un soggetto come figura simbolica del possesso in tutte le sue forme, fa spazio a una "promiscuità" ontologica che rivela la centralità della relazione, nella forma dell'incontro/scontro, nei processi di soggettivazione.

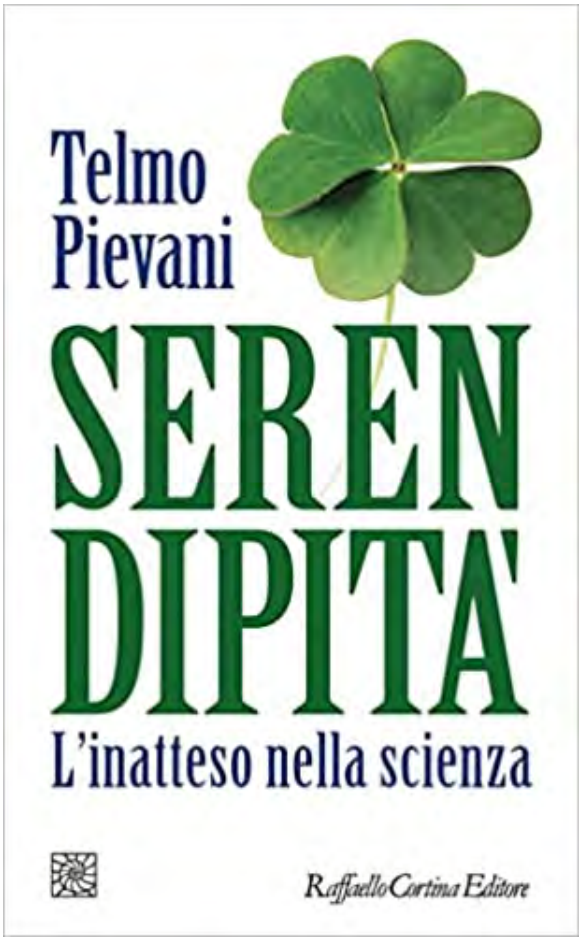
Quali relazioni inedite e quali sperimentazioni innesca questo pensiero femminista rispetto alle "promesse" delle bio-infotecnologie? Fatti fuori i dualismi propri di un pensiero normato, questo posizionamento non può mai risolversi in una postura tecnofila o tecnofoba, ma tende a immaginare un utilizzo piegato a vantaggio del vivente tutto per farla finita con lo sviluppo di tecnologie egoiste e egocentrate. Alla replicazione biotecnologica del sapiens nell'ordine della famiglia eterosessuale si oppone una «bio-ingegneria più bizzarra e transpecie, più frocia e fuori luogo» (pp. 10-11) che agisce, per contrasto, dissolvendo il modello familiare dominante: *Per farla finita con la famiglia* è proprio il titolo di un testo della curatrice Angela Balzano. Il *fil rouge* è il superamento della (ri)produzione familista bianca e capitalista come apertura dell'orizzonte di sopravvivenza delle alterità minacciate dalla crisi ecosistemica indotta dalla superiorità secolare del sapiens nella sua declinazione di *homo oeconomicus*.

A questo armamentario teorico si aggiunge una concezione della "cura" in netto contrasto con il paradigma umanistico-moralistico: dare rilevanza alla cura vuol dire, in questa prospettiva, abbracciare le relazionalità sostenibili e generative e disfare

quelle prettamente strumentali che si danno attraverso violente verticalizzazioni. Una relazionalità "orizzontale" è il motore di connessioni trasformative che rinnovano lo schema dei legami ereditari e identitari. Come ha suggerito Donna Haraway, prendersi cura non equivale a occupare la posizione gerarchicamente più alta, ma lasciarsi coinvolgere in una relazione e pensare tra le pieghe di se stessi per cogliere gli infiniti mondi *simpoietici* che ci costituiscono. Stimando come insufficienti le proposte di una *green economy* che reitera la fissazione del punto di vista dei dominatori come punto di vista assoluto e esclusivo sul mondo, si tratta di aprire prospettive "mostruose", incomprensibili allo sguardo convenzionale e per questo cariche di potenzialità, non sussumibili ai linguaggi di Stato, per avanzare riflessioni/azioni etiche nel riguardo effettivo delle minoranze.

Melania Moltelo

Telmo Pievani
Serendipità. L'inatteso nella scienza
Raffaello Cortina Editore, Milano 2021, pp. 256, € 15



Gli ultimi due anni che abbiamo vissuto testimoniano meglio di qualunque discorso quanto importante sia la comunicazione della scienza, al fine di restituirne una percezione corretta – ugualmente distante sia da un anacronistico positivismo sia dal relativismo post-moderno – tale da educare la collettività circa l'importanza e i limiti di questa titanica impresa umana, nella quale più si sa e più si scopre di non sapere. Ogni nuovo contributo in tal senso assume, pertanto, una valenza non solo accademica, ma anche civile.

Tale è il caso del saggio di Telmo Pievani, *Serendipità. L'inatteso nella scienza*. In questo libro, l'autore fa sapiente uso di una prospettiva plurale (che comprende l'etimologia, la storia delle idee, la sociologia della scienza, l'epistemologia) per bonificare questo termine dalla caligine degli aneddoti, talvolta impropri, usati comunemente per spiegarlo. Piuttosto, in ossequio all'ammonimento di Socrate, per cui la definizione di qualcosa è cosa diversa dall'enumerazione di esempi (*Tht*. 146c-147d), l'autore ci conduce in un itinerario, insieme narrativo e analitico, volto a far chiarezza su quel fenomeno curioso per cui, mentre si è alla ricerca di qualcosa, si trova tutt'altro.

E la prima sorpresa è che già solo questa definizione, radicata nel senso comune, cela una storia tortuosa, che affonda le radici nella novellistica medievale dell'Oriente. Nel racconto dei tre principi di Sarandib, la settima delle *Otto novelle del paradiso* del poeta indiano Amir Khusrau, il termine è codificato nel suo significato etimologico: «Sarandib era il nome persiano antico dello Sri Lanka, dell'isola di Ceylon, o Silan, prima ancora Serendip o Sarendip» (p. 26). Il re di Sarandib esilia i suoi figli affinché conoscano il mondo e acquisiscano saggezza al fine di governare il regno, secondo il classico modello del rito di iniziazione. I tre fratelli incarnano «il piacere della libera investigazione: affrontare le incertezze del mondo e provare a interpretarle, risalendo da indizi apparentemente trascurabili a realtà nascoste non immediatamente esperibili con i sensi» (*ibid.*). Essi non hanno un obiettivo preliminare e non lo trovano casualmente; non trovano per vie impreviste ciò che cercano né scovano qualcosa di inaspettato; per il semplice motivo che non cercano alcunché: i tre principi di Sarandib, nella novella di Khusrau (e ancor più nella trasposizione italiana rinascimentale di Cristoforo Armeno) sono piuttosto «l'archetipo del detective» (pp. 29-30) e maestri di *abduzione*.

La quale è una componente della serendipità, ma non la serendipità di per sé. Voltaire si ispira alla figura dell'investigatore che fa uso del ragionamento abduttivo per il personaggio di Zadig, saggio babilonese del IX secolo, protagonista dell'omonimo racconto filosofico. «Per Voltaire, ciò che noi oggi chiamiamo abduzione significa apprezzare le piccole differenze nei fenomeni osservati al fine di arrivare a un risultato preciso […]: descrivere un oggetto mai visto, prevedere l'esistenza di qualcosa che non è mai direttamente finito sotto i nostri sensi, o anche indovinare una serie di eventi del passato che, retrospettivamente, spiegano quegli indizi» (pp. 49-50). Tale è il mestiere del paleontologo, come osservò lo stesso Georges Cuvier,

il quale scrisse «che il suo mestiere di investigatore del tempo profondo è proprio quello dello Zadig di Voltaire» (p. 51): è la «profezia retrospettiva», come la chiamò Thomas H. Huxley (p. 170).

Da dove nasce dunque il significato di serendipità noto ai più, quello per cui si trova qualcosa di inaspettato mentre si cerca tutt'altro? Il malinteso si deve a Horace Walpole, che nell'Inghilterra del Settecento conia la parola *serendipity* ancorandola all'universo semantico della ricerca antiquaria, del collezionismo eclettico, e enfatizzando le componenti di caso e sagacia della ricerca indiziaria dei principi di Sarandib (i quali sono sì sagaci, ma la cui ricerca si svolge all'insegna dell'abduzione più che del caso). Ciò che colpisce in questa storia dell'idea è che lo stesso termine serendipità nasce quindi da una deviazione, da un errore generativo, in modo serendipitoso: Walpole inventa la parola in maniera non del tutto fedele al senso della favola persiana, per un vezzo linguistico (pp. 40-42).

Tale narrazione è propedeutica al nucleo fondamentale del testo: la proposta di una «*scienza dell'inatteso*» (p. 83) che indaghi – al di là dell'ambito letterario in cui è sorta e dell'aneddotica che ne occulta il senso – l'importanza della serendipità nella scienza. E come ogni impresa in cui è necessario porre ordine nella complessità del molteplice, è necessaria una tassonomia dei significati storicamente codificati di questo concetto. Pievani propone una classificazione della serendipità, dalla più alla meno accidentale: 1. *scoperta casuale* (quella dei principi di Serendippo, i quali si imbattono accidentalmente in misteri da risolvere, senza cercarli di per sé); 2. *serendipità in senso forte* (che è quella di Walpole), ovvero «trovare qualcosa che nemmeno sapevi di star cercando, grazie al caso e alla sagacia, senza alcuna intenzionalità» (p. 75); 3. *serendipità in senso debole*, quando imprevedibile non è la scoperta (che è invece cercata con impegno), bensì la modalità con cui si perviene a

essa; 4. *processo indiziario*, «quando l'indagatore sa benissimo qual è il problema e grazie alle sue capacità abduttive risolve il mistero come un detective» (pp. 75-76), o come un paleontologo, o come un medico, o come Zadig.

La serendipità in senso stretto, quella epistemologicamente più interessante, è la seconda. La serendipità in senso forte, osserva Pievani (p. 77), può infatti istruirci circa le modalità della scoperta scientifica. «Se capiamo le circostanze in cui la serendipità fiorisce, potremmo promuoverla con misure e finanziamenti adeguati. Il tema non è di intrattenimento. Dall'importanza quantitativa della serendipità dipende ad esempio la giustificazione o meno dei fondi dati alla ricerca di base, cioè quella non finalizzata a priori a un risultato applicativo bensì guidata dalla sola curiosità di conoscere la natura, un tipo di indagine che a prima vista sembrerebbe più consona alle gioie della serendipità» (p. 129). Ma se la serendipità è per definizione «l'inatteso che irrompe» (p. 130), appare paradossale cercare di ridurla a schemi che la rendano predicibile.

Se non possiamo pianificare a tavolino le ricerche serendipitose, nondimeno possiamo indagare quali sono le condizioni che le favoriscono: in altre parole, possiamo studiare l'*ecologia della serendipità*. Ingredienti fondamentali in tal senso pare siano: incoraggiare la ricerca libera, su argomenti scelti dal ricercatore, con scadenze dilatate, in gruppi di lavoro interdisciplinari e magari con membri diversi a livello di cultura, genere e religione (così come l'antico regno di Sarandib era un crocevia di civiltà diverse) (pp. 153-156).

"Libertà" è parola chiave della serendipità almeno quanto "sagacia" e "incertezza". «Il piacere della serendipità sta nell'incertezza che per una volta non porta sciagure» (p. 136). Di fronte all'incertezza dell'esito e del percorso, la mente umana è sovente abituata a pensare in modo dicotomico: ciò vale in modo particolare per la serendipità, la quale è stata interpretata ora

come massima fortuna (rimarcando la componente casuale e "democratica" del processo) ora come massima applicazione metodica (enfatizzando le capacità dello scienziato e sfociando talora nel mito dell'"eureka!"). «Serve una via mediana, quella indicata da Pasteur: il successo che arride agli audaci; il caso che premia i sagaci. Non è solo fortuna, ma non è nemmeno solo metodo» (*ibid.*). Bisogna che l'occasione imprevista, l'osservazione inaspettata venga colta e ponderata da una mente preparata, competente, capace di fare le giuste associazioni. La serendipità in tal senso è democratica (in quanto ci pone tutti ugualmente ignoranti di fronte all'ignoto) ma non qualunquista (le grandi scoperte non sono alla portata di chiunque) (pp. 137-138).

Spesso, l'abuso dell'argomento "fortuna" è sostenuto da una visione trionfalistica della scienza e teleologica della storia. L'antidoto intellettuale in questi casi è imparare a ragionare in modo controfattuale, a dare il giusto peso alle potenzialità espresse a fronte di quelle (molto più numerose) non realizzatesi. «Non vi è nulla di più pericoloso e fuorviante del senno di poi che addomestica le ricostruzioni a posteriori» (p. 140). Questo è un problema che deve indurci a assumere con cautela le testimonianze degli stessi scienziati, dalle autobiografie agli articoli scientifici, nei quali i requisiti di forma del saggio accademico impediscono il più delle volte di riferire il contesto psicologico e ambientale della scoperta, riducendo quest'ultima esclusivamente alla sua giustificazione logica e dando così l'impressione (falsa) che la sua genesi sia altrettanto lineare e razionale. «Insomma, non possiamo fidarci più di tanto dei racconti a posteriori degli scienziati! Ancor meno delle loro autobiografie celebrative. Sono tutte viziate dal senno di poi, dalle semplificazioni retrospettive, dal fatto inesorabile di sapere come è andata a finire la storia» (p. 143).

Ma se il caso deve accompagnarsi al talento, qual è la "valuta" per candidarsi alla serendipità? Oltre a una smaniosa curiosità,

occorrono ampie conoscenze pregresse, contatti giusti con altri professionisti, una spiccata attenzione ai dettagli apparentemente secondari e una propensione ai campi di ricerca poco esplorati. «A una scoperta scientifica serendipitosa bisogna candidarsi, come si compra un biglietto della lotteria candidandosi alla vincita, non trovando per caso un biglietto vincente per terra. Un accidente resta un accidente, finché non succede alla persona giusta: a quel punto diventa una scoperta» (p. 148). Bisogna, poi, saper tollerare l'errore: invece di scartare le anomalie come imprecisioni fastidiose, avere l'onestà e il coraggio di approfondirle per comprenderne la causa; stabilire cioè se si tratta davvero di refusi o piuttosto di incongruenze promettenti, che svelano scenari di ricerca inediti. In tal senso, bisognerebbe proprio adottare un atteggiamento di «trascuratezza controllata» (p. 149), come scrisse il biologo italiano, naturalizzato statunitense, Salvador Luria.

«Secondo il filosofo della scienza Aharon Kantorovich, la scoperta serendipitosa, essenziale per comprendere la creatività scientifica soprattutto nelle fasi di forte cambiamento, è equiparabile alle mutazioni casuali nell'evoluzione biologica: errori e deviazioni involontarie perturbano le procedure di routine introducendo variazioni, sottoposte poi ai filtri selettivi della mente dello scienziato stesso e della sua comunità di riferimento» (pp. 151-152). Anche se tale analogia, come precisa Pievani, non va presa alla lettera, essa ci aiuta a comprendere il valore generativo dell'errore nella pratica scientifica. «Le vie dell'errore sono infinite e nella scienza si fa bricolage, anche resuscitando idee dormienti. Quale che sia, l'errore nella scienza non è necessariamente un allontanamento dalla verità, ma anche un modo per acquisire conoscenza. Evitare gli errori è meschino, sosteneva Karl R. Popper. La scienza deve affrontare problemi che siano sufficientemente difficili da rendere l'errore inevitabile, e generativo» (p. 153). Senza

contare che, per certi versi, le teorie errate sono più interessanti di quelle corrette, non solo perché può capitare che dietro l'errore si nascondano scenari insospettabili forieri di teorie corrette, ma anche in quanto studiare a posteriori la genesi di una teoria errata mostra i percorsi tortuosi della scienza in atto (che solitamente non si conservano nei *paper* pubblicati). In tale senso, Pievani scrive icasticamente: «Le teorie giuste sono giuste nello stesso modo. Le teorie sbagliate sono sbagliate ognuna a modo suo» (p. 219).

Dietro un errore scientifico può celarsi una soluzione per un problema diverso da quello inizialmente considerato, tramite un riuso creativo delle risorse disponibili. Il bricolage è una delle analogie fra evoluzione biologica e evoluzione della scienza (pp. 227-228). Come l'evoluzione biologica è un'esplorazione di possibilità, che spesso opera mediante rabberciamenti, riutilizzo creativo di strutture preesistenti, sviluppatesi per un adattamento diverso o per nessuna ragione adattativa, così anche la storia della scienza è costellata da scoperte cooptate da una disciplina all'altra, secondo percorsi tortuosi e imprevedibili. «Esiste insomma un opportunismo serendipitoso sia nell'evoluzione sia nella cognizione» (p. 229), che costituisce l'onda lunga della nozione di *exaptation*.

Un'altra analogia fra evoluzione biologica e storia della scienza è il fenomeno della convergenza. Di fronte a problemi adattativi simili, specie distanti filogeneticamente finiscono per sviluppare i medesimi adattamenti. Accade talvolta che anche scienziati alle prese con una stessa domanda di ricerca raggiungano, indipendentemente fra loro, soluzioni coincidenti. In tal senso, l'argomento "la scoperta era nell'aria" è attendibile: non sempre le scoperte scientifiche sono serendipitose. Ma ciò non significa che esse sarebbero giunte necessariamente, prescindendo dagli scenari congiunturali in cui ogni scienziato opera. Darwin e Wallace svilupparono indipendentemente la teoria dell'evoluzione

per selezione naturale; ma, come ricorda Pievani, «cosa avrebbero scoperto Darwin e Wallace senza le occasioni altamente serendipitose dei loro viaggi?» (p. 226). Come osservava Gould: «Dove sarebbe stato Darwin nel 1837 senza Gould, Owen e la dinamica vita scientifica di Londra e di Cambridge?» [S.J. Gould, *Il sorriso del fenicottero* (1985), tr. it. Feltrinelli, Milano 2009[2], p. 287].

Pievani sottolinea il paradosso per cui Internet, con le sue enormi potenzialità in termini di ricerca e condivisione delle informazioni, sia diventato uno strumento che consolida le vie già percorse, preclude l'inaspettato, rafforza in maniera retroattiva le idee preconcette: il nemico per eccellenza della serendipità. «Nel cosmo digitale è sempre più raro che si trovi qualcosa che non si stava cercando, tanto più che a ogni passo l'algoritmo ci propone connessioni prevedibili, scontate, ovvie, fastidiosamente fondate sul pregresso cumulativo delle nostre scelte e navigazioni» (p. 157). La storia della serendipità, in tutte le sue forme, ci mostra nella maniera più lampante che le scoperte più grandi si ottengono invece se si rinuncia alla logica dell'*hic et nunc*. «Serve una scienza più lenta, liberata dalla pressione di ottenere rapidamente risultati certi a breve termine, per rispondere a competizioni interne ai settori, a sollecitazioni di interessi economici e di mercato, o a ansie sociali emergenziali. La serendipità fiorisce sul terreno di una scienza più autonoma e indisciplinata, una scienza che rallenta un po', per dare più importanza alle domande che alle risposte» (p. 159).

Ancora nei primi decenni dell'Ottocento, i naturalisti accompagnavano l'epiteto "filosofico" alle loro ricerche come sinonimo di "rigoroso". Oggi potremmo ben parlare di scienza "filosofica" come scienza "serendipitosa", se è vero che la filosofia, così come la serendipità, non solo ammette ma esige lentezza, ponderazione, il disinteressato piacere dell'indagine prima ancora che della scoperta.

Ma perché la serendipità è così diffusa? Essa intride talvolta non solo la storia della scienza, ma anche la storia degli individui: potenzialità, congiunture, fortuna e talento si mescolano nella vicenda umana e scientifica del biologo giapponese Osamu Shimomura (pp. 196-202). Essa mostra che la serendipità «è un atto perturbativo, proattivo, di ricerca, che porta a conseguenze non intenzionali. Come tutte le nostre azioni, una volta compiute entrano a far parte di un'ecologia più ampia di relazioni e interpretazioni, foriera di potenziali conseguenze inattese, e suggeriscono magari contributi involontari in altri contesti. La serendipità è un piccolo sassetto nello stagno che fa emergere l'inatteso. Succede quando la natura risponde in maniera imprevedibile alle nostre domande» (pp. 201-202). La frequenza della serendipità, specie quando si studiano i sistemi complessi, è dovuta all'abisso che intercorre fra conoscenza e realtà. «Da Zadig a Galileo, tutti cercano di decifrare il libro della natura per svelarne aspetti nascosti. Il problema affascinante è che nessuno sa quante pagine restano da leggere» (p. 203). Ciò vale non solo per le scienze fisico-matematiche, ma anche per la biologia, e in particolare per la classificazione biologica: «E fa grande tristezza pensare che noi stiamo estinguendo a spron battuto quella cornucopia di forme di vita che ancora non abbiamo nemmeno scoperto e studiato» (p. 214). Che il divario fra conoscenza e realtà sia abissale è ragione di sgomento e di meraviglia: ci ricorda la nostra finitudine e la fatica di Sisifo che lo scienziato sperimenta nello sfidarla (p. 217; cfr. anche T. Pievani, *Finitudine. Un romanzo filosofico su fragilità e libertà*, Raffaello Cortina Editore, Milano 2020).

La suggestiva immagine di Pierre Duhem ripresa da Jacques Hadaman – per cui lo scienziato è come un pittore che prima dipinge un paesaggio, e poi apre la finestra per verificare il grado di approssimazione fra la sua opera e la realtà là fuori – è il pretesto usato da Pievani per introdurre l'ultimo ma fondamentale

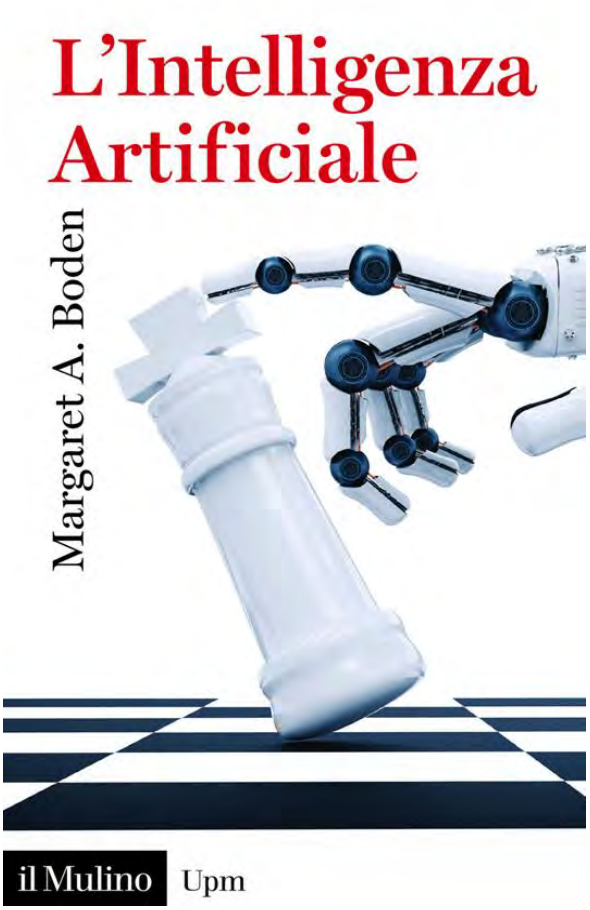
problema legato alla serendipità: spiegare come sia possibile che «concetti matematici nati per gioco, come esercizi formali ispirati dall'amore per bellezza e eleganza, senza alcuna applicazione immaginabile, si rivelano a posteriori come descrizioni potenti di nuove realtà fisiche» (p. 235). Il fisico Eugene Wigner chiamò questo tema "l'irragionevole efficacia della matematica nelle scienze naturali". A tale questione sono state proposte diverse risposte. L'ipotesi di Pievani «è che un contributo alla soluzione di questo classico problema di filosofia della scienza (il rapporto fra matematica e realtà, fra mente e mondo) possa venire proprio dalla serendipità, e da un'interpretazione evoluzionistica della scienza» (pp. 223-224). «La serendipità, così importante tanto nell'evoluzione quanto nella cognizione, sembra suggerire che questa propizia consonanza tra mente e mondo, questa connessione tra enti mentali e forme naturali, non sia un dono miracoloso da accettare e basta, ma possa avere un'origine naturale. Dopo tutto, il nostro cervello è parte integrante della realtà che indaga e si è evoluto dentro quella realtà, adattandosi alle sue pressioni selettive e ai suoi vincoli. Non è stupefacente, quindi, che le sue costruzioni matematiche siano (talvolta, non sempre) in sintonia con la realtà» (p. 246). «In tal senso, la (non più?) irragionevole efficacia della matematica sarebbe una sintonizzazione evolutiva tra la naturalità della nostra mente e il resto della natura di cui fa parte» (p. 248).

Pievani propone dunque una soluzione naturalistica ma non riduzionistica alla misteriosa sintonia fra matematica e realtà fisica la quale, mentre sottolinea la pervasività della serendipità a livello storico e epistemologico, evidenzia d'altra parte che la mente umana stessa è prodotto dell'evoluzione. «La scienza dell'inatteso, in ultimo, ci parla insomma della natura umana. Ci parla niente meno che della natura della mente umana, o

meglio: di una parte della natura, cioè la mente umana, alle prese con il resto della natura di cui fa parte» (p. 249).

GIOVANNI ALTADONNA

Margaret A. Boden
L'Intelligenza Artificiale
a cura di Diego Marconi - tr. it. di Fabrizio Calzavarini
Il Mulino, Bologna 2019, pp. 188, € 16



Nel 1833, in piena epoca positivista, le tavole matematiche deputate al calcolo complesso costituivano l'unico metodo, peraltro incredibilmente prono all'errore umano, per assicurare il corretto funzionamento dell'industria e dell'infrastruttura inglese, che vedeva all'epoca la sua esplosione. In quell'anno, il matematico Charles Babbage incontra a una festa Ada Byron, unica figlia legittima del poeta Lord Byron, che diventerà per matrimonio contessa di Lovelace. Ada, seppur giovanissima, è istruita in matematica e ha ottime capacità logiche, e soprattutto ha una mente curiosa; Babbage sta perfezionando il progetto di una nuova macchina, dopo il fallimento di un primo progetto preparatorio, *the Analytical Engine*. L'incontro tra i due, e i successivi studi e progetti di entrambi sul *motore analitico* getteranno le basi per le moderne scienze su computer e intelligenze artificiali. L'apporto di Ada Lovelace, in particolare, consiste nell'aver intravisto le potenzialità della macchina di Babbage: seppure si trattasse di un "mostro" in metallo e ottone dal peso previsto di quattro tonnellate, un ammasso di ingranaggi e ruote dentate in

numero, ironicamente, quasi incalcolabile – come se fosse uscito dalle pagine di un racconto di fantascienza *steampunk* – Lovelace comprende presto che la sua architettura potesse sia automatizzare il calcolo delle tavole matematiche, espungendo così la possibilità dell'errore – anatema di ogni matematico – sia devolversi interamente a compiti nuovi e inaspettati. L'intuizione di Ada Lovelace consiste nella creazione di "schede punzonate", antenate dell'algoritmo, da inserire nella macchina per *programmare* il suo comportamento e ampliare così enormemente, e potenzialmente all'infinito, il raggio di applicazione del *motore analitico* concepito da Babbage. Lovelace prevede che tale macchina, rimasta solo allo stato progettuale nei disegni di Babbage, e oggi osservabile solo grazie a riproduzioni su piccola scala, può essere *programmata* per passare dal calcolo a «comporre elaborati e scientifici brani musicali di qualsiasi grado di complessità e lunghezza», aprendo «un'epoca gloriosa nella storia delle scienze» (p. 12). Ada Lovelace riconosce dunque le potenzialità di una macchina funzionalmente progettata per essere dedicata al calcolo algebrico, rendendola programmabile: le funzionalità logico-simboliche dell'apparato permettono di espandere il suo dominio grazie alla programmazione, in modo da poter rappresentare anche tutti gli oggetti dell'universo e giungere a ricombinazioni e applicazioni inedite. Gli elementi basilari della *programmazione* odierna vengono descritti e pioneristicamente individuati da Ada Lovelace: i programmi immagazzinati, la gerarchia di sottoprogrammi, il reindirizzamento, la microprogrammazione, il *looping* e perfino i *bugs*. Per quanto i suoi metodi non siano mai stati effettivamente implementati, e per quanto non ci sia verifica sperimentale di come integrare la creazione musicale o il ragionamento scientifico sul *motore analitico* di Babbage, Ada Lovelace viene spesso individuata come la prima programmatrice della storia. L'intelligenza artificiale aveva una prima architettura, e

disponeva anche di un progetto di funzionalità possibili, ma era ancora lontana dal diventare una realtà.

Ci vorrà più di un intero secolo, e tutta l'esperienza crittografica di un altro geniale matematico inglese, Alan Turing, per arrivare a una concettualizzazione pienamente "moderna" delle macchine calcolanti, i cosiddetti computer, e delle loro possibilità in merito a intelligenza, pensiero e comportamenti intelligenti. La prima macchina immaginata da Turing nel 1936, la *macchina di Turing universale*, è un sistema simbolico immaginario in cui ogni tipo di computazione può, almeno in linea di principio, essere operata attraverso un motore matematico in grado di ridurre e ricombinare simboli binari, rappresentati da *0* e *1*, *vero* e *falso*, *on* e *off*. Turing spenderà quasi due decenni a riflettere su come questa nozione astratta avrebbe potuto tradursi in una macchina fisica – il primo computer moderno, completato a Manchester nel 1948, non sarebbe infatti stato costruito senza l'aiuto progettuale di Turing – ma appare già evidente che l'impostazione del matematico, riguardo a cosa potesse essere l'intelligenza artificiale e l'automazione del pensiero, e quale ordine di problemi essi potessero sollevare, fosse pioneristica sin dall'inizio: gli scopi dell'IA di Turing erano già pienamente "moderni". Il celebre *test di Turing*, infatti, rappresenta un vero e proprio manifesto fondativo degli albori dell'era dell'IA, e ancora oggi non perde di validità: ogni intelligenza artificiale che si proponga come tale vi sarà sottoposta, senza successo nella maggior parte dei casi. Pubblicato in un articolo nel 1950 sulla rivista filosofica *Mind*, il test propone che un dato operatore umano sia assolutamente in grado di riconoscere, almeno il 30% delle volte, se l'agente con cui sta interagendo sia un altro umano, o un computer; nel caso in cui non sia possibile distinguere l'umano dall'artificiale, il test è superato dal computer, e in tal caso, suggerisce Turing, non ci sarebbe ragione di negare che una macchina matematica possa *pensare*. E è proprio

questo il punto. Turing identifica tutta una serie di questioni fondamentali riguardanti l'intelligenza e il funzionamento della mente (come il gioco, le percezioni, l'apprendimento o il linguaggio): è necessario scomporre i "processi computazionali" che avvengono all'interno del cervello, e che gli permettono effettivamente di *pensare*, se si vuole, in qualche modo, replicare l'attività pensante in macchine (eventualmente) intelligenti.

Da un punto di vista pragmatico, seppure il test di Turing possa suscitare un forte interesse filosofico, esso è di scarsa utilità alle intelligenze artificiali esistenti: la maggior parte di esse mira a svolgere compiti utili, esibire comportamenti che chiameremmo intelligenti, ma non aspira a imitare l'intelligenza umana. Eppure, i rapidi sviluppi in tale campo scientifico mettono in campo l'esigenza di riconsiderare le intelligenze artificiali a partire dalle prime, geniali formulazioni, dal momento che hanno guidato la loro venuta all'essere. A differenza di Turing, ci troviamo già a coesistere con le IA in molteplici ambiti della vita quotidiana; le riflessioni del matematico inglese resistono però al trascorrere del tempo proprio perché hanno il grande merito di approcciare, per la prima volta, la questione dell'intelligenza a partire dal supporto biologico che ne permette l'esistenza, cioè la mente.

Margaret A. Boden in questo saggio raccoglie una varietà di questioni, temi e problemi, recenti e passati, sul macrotema dell'intelligenza artificiale; il testo rappresenta un'agile ma comprensiva introduzione alla questione, che viene osservata da molteplici punti di vista nei suoi caratteri fondamentali. Esperta di scienze cognitive, l'autrice procede a definire l'IA nella storia della scienza e del pensiero, cercando di presentare i vari tipi di intelligenza artificiale, la loro implementazione, i loro scopi e le loro promesse, accompagnando il lettore nelle molteplici declinazioni, tecnologiche ma anche filosofiche, delle intelligenze artificiali.

La portata metodologica dell'IA recente è vastissima, e se è vero che l'applicazione di sistemi artificiali costruiti a arte è potenzialmente infinita, e, oggi, persino di grande successo nello svolgimento di compiti intelligenti in determinati settori specialistici, è anche vero che agli albori il termine intelligenza artificiale veniva concepito dai suoi pionieri come il raggiungimento di macchine dotate di un'intelligenza *generale*: una macchina virtuale può svolgere, anche benissimo, uno specifico compito grazie a una programmazione di finalità specifica, ma quello che Boden definisce il *Santo Graal* dell'intelligenza artificiale (p. 25) è il raggiungimento di una AGI – *Artificial General Intelligence* – e cioè una macchina virtuale dotata di ragionamento e cognizione *generali*, i cui campi di applicabilità sarebbero immensi, e non più unicamente specifici. Possiamo, in altre parole, programmare una macchina virtuale in grado di giocare a scacchi, e anche di battere il campione mondiale in carica la maggior parte delle volte – come nel caso del match tra il supercomputer *Deep Blue* e il campione di scacchi Kasparov nel 1996-97; ma questa macchina sarà in grado di svolgere solo e unicamente quel compito. Seppur immaginata sin dagli anni '70, l'AGI rappresenta una sfida per il XXI secolo, e è ancora tale proprio perché non basterebbe l'incremento esponenziale delle capacità di calcolo dei computer, a lasciare, banalmente, "emergere" un'intelligenza artificiale di tipo generale; si tratta di un problema complesso che coinvolge il concetto stesso di intelligenza, e chiama in causa tutta una serie di astrazioni preliminari (come l'implemento della logica, l'euristica, la pianificazione, la semplificazione matematica, e la rappresentazione della conoscenza). In sostanza, per quanto il campo dell'IA abbia fatto passi da gigante dal secondo dopoguerra alla contemporaneità, l'intelligenza artificiale generale, il sogno di Turing, resta ancora un progetto incompiuto, per la difficoltà di replicare una reale intelligenza in grado, ad

esempio, di comprendere e utilizzare in modo proprio il linguaggio, di ricevere percezioni di tipo fenomenico e non solo funzionale, di mostrare reale creatività sulla base dei dati noti, di includere nelle sue capacità l'influsso neurologico delle emozioni, di mostrare capacità di deliberazione e auto-riflessione, e così via. Si comprende dunque che in prima battuta si possa, in tal senso, impiegare nello studio dell'IA un approccio di tipo *cognitivista*, in grado di isolare e eventualmente riprodurre i meccanismi fondamentali in grado di permettere il pensiero, conscio o subconscio – compito non semplice già all'interno delle sole discipline neuroscientifiche.
Il vero problema filosofico posto dall'IA, e in particolare dall'ipotesi del raggiungimento di una IA dall'intelligenza generale, è il problema *mente-corpo*: nella sua declinazione specifica, quella relativa a forme di intelligenza non biologiche ma artificiali, il dibattito si concentra anche sulla possibilità che una visione "eccessivamente" cognitivista (che considera cioè il contenuto della scatola cranica e le sue funzioni come proprietà necessarie ma *sufficienti* per l'emergenza del pensiero) sia eccessivamente sbilanciata, e che non permetta, in ultima analisi, di generare intelligenze artificiali generali. Tutto sommato, il pensiero, nelle forme di vita umane, emerge dall'interazione, inscindibile e dinamica, tra il corpo – sempre *situato* nel qui e nell'ora, e esposto a stimoli e percezioni ininterrotti – *e* la mente. In mancanza di esperienze "incarnate" non potrebbe esserci reale *cognizione*, e nessuna AGI instanziata solo "su schermo", e dunque completamente "scorporata" potrebbe mostrare reale intelligenza senza i caratteri generali propri del *vivente*: auto-organizzazione, autonomia, emergenza, sviluppo, adattamento, evoluzione e metabolismo. Se è vero che alcune di queste caratteristiche possono essere simulate "in astratto", senza "carne", e cioè rappresentate attraverso l'elaborazione di informazioni, altre, come ad esempio il metabolismo, sono

proprietà strettamente e irriducibilmente fisiche. L'*A-life*, la vita artificiale che instanzia l'IA in robot fisici, è sicuramente più vicina a questo approccio *fenomenologico*, secondo il quale la *vita* (e specificamente la vita corporea) è necessaria per la mente, e dunque per il pensiero; per quanto le IA possano mostrare comportamenti intelligenti e performance impensabili ai tempi di Turing, la *reale* intelligenza rimarrebbe ancora fuori portata proprio per via di quest'enfasi così marcata sulla cognizione.
In ultima analisi, Boden sostiene che la frattura tra l'approccio cognitivista e quello fenomenologico sia insanabile, e ciò sarebbe in parte dovuto all'impossibilità di definire univocamente la natura della mente e dell'intelligenza, e di trovare eventualmente un consenso nella formulazione (o ri-formulazione) di questi concetti. In fondo, una riflessione sull'IA, sin dalle sue prime e ipotetiche concezioni, fino alla sua recente realizzazione nell'alveo delle tecnoscienze post-digitali, chiama in causa una riconsiderazione più attenta dei concetti fondamentali di vita, corpo, mente, autocoscienza, cognizione e emergenza del pensiero; la sfida concettuale posta dall'intelligenza artificiale rappresenta un'arena filosofica ancora aperta, dunque, e le sue possibili risoluzioni sono ancora in via di definizione.

STEFANO PALUMBO

Maria Teresa Catena
Breve storia del corpo
Mimesis, Milano-Udine 2020, pp. 201, € 18

È vero: «un certo desidero di *assenza*» del corpo (p. 12) continua a permeare le società e i singoli. Un corpo percepito come peso, rischio, lordura. Il peso del tempo che avanza e della impossibilità di azioni e comportamenti che la sua struttura anatomo-fisiologica rende infattibili. Il rischio sempre incombente della malattia e del morire. La lordura rispetto ai luminosi e rarefatti orizzonti di ciò che di volta in volta è stato chiamato anima, spirito, razionalità, mente, virtuale.

Espressione paradigmatica di una perplessità verso il corpo che diventa spesso diffidenza e persino rifiuto è il cartesianesimo. Infatti, il corpo «quest'evidenza silenziosa che mi accompagna, una volta che si cerchi di afferrarla e comprenderla, rivela una sua insondabile opacità; e forse nessuno meglio di René Descartes ha saputo descrivere quanto facile sia cogliersi immediatamente come esseri pensanti e quanto difficile, invece, afferrare quest'estraneità che pure il corpo è» (p. 15). Espressione paradossale e ambigua, dato che Descartes è all'origine anche delle tradizioni che rifiutano qualunque cosa si possa ricondurre all'*anima*, a favore del puro meccanismo del mondo. Un meccanismo che però diventa espressione ancora una volta di una razionalità disincarnata, puramente formale, algoritmica, computazionale, virtuale. Lo diventa non solo nella tradizione della scienza

galileiana ma anche e soprattutto nelle correnti transumaniste del presente.

Complessa, come si intuisce, è la vicenda del corpo nella storia. Di essa Maria Teresa Catena formula una sintesi illuminante a partire dai concetti e dalle pratiche del corpo *presente* e del corpo *assente*. Due paradigmi che nel corso del tempo si susseguono, si intrecciano, si separano, si trasformano.

Momento e forma fondamentale è il corpo pagano:

> Composto della materia eterna e informe che, in quanto tale, condivide l'incrollabile perpetuità dell'universo, nel quale giocano dèi esuberanti, il corpo è visto dai pagani come un oggetto del mondo naturale, qualcosa che occupa il suo legittimo posto nella grande catena dell'essere che lega l'uomo agli dei da una parte, alle bestie dall'altra. […] Più che portatrice di una forza sorda, duttile e temibile, di cui sospettare continuamente, la corporeità è, nell'Antichità, espressione dello spirito ardente che attraversa la natura; è composta, scrive Cicerone, dello stesso calore che irradia dalle stelle (p. 44).

Anche Platone, il troppo frainteso Platone, pur privilegiando in alcuni dialoghi e momenti la ψυχή non ha mail ritenuto il σῶμα un elemento di perdizione e anzi nel *Timeo* lo pone al centro dell'armonia individuale e cosmica. Il 'dualismo' platonico è in gran parte presunto ed è invece assai più reale il dualismo cartesiano e post-cartesiano che mostra ancor oggi tutta la sua forza ermeneutica e prassica nel corpo assente delle correnti e posizioni estropiane e transumanistiche che pervadono anche alcuni sviluppi dell'Intelligenza Artificiale e del variegato arcipelago delle tecnologie virtuali. In esse opera infatti la tendenza

> a far *assentare* il corpo, a mettere da parte la lentezza della sua evoluzione biologica per tentare di realizzare quel che si crede essere la natura più propria dell'uomo, cioè quell'intelligenza che in tutta evidenza - come già recitava il vecchio adagio del *corpo assente* - trova il suo senso indipendentemente e contro la *prigione* di una carne che, in fondo, continua a restare il vero limite al pieno dispiegamento della sua dimensione spirituale (p. 156).

Ritornano le tendenze, quasi le implorazioni, orfiche contro la carne, contro il corpo, a vantaggio di un «materialismo escatologico e/o platonismo tecnofilo» (p. 173) che sembra

volersi sbarazzare di insegnamenti diversi tra loro ma fondamentali quali il naturalismo darwiniano, la fedeltà nietzscheana ai corpi e alla terra, la fenomenologia della *chair* di Merleau-Ponty, tutte espressioni del *corpo immanente* che siamo, che l'animalità è.

La riflessione di Darwin «fa del corpo […] una presenza imprenscindibile che, di *antropos*, racconta insieme, la radicale immanenza, la sua decentrata continuità con gli altri esseri e la radice materiale delle sue cosiddette facoltà superiori» (p. 81). Le manifestazioni politiche della centralità del corpo sono innumerevoli e permeano tutta la cultura umana: dai riti alle torture, dall'abbigliamento alla medicina. Anche nel presente epidemico stiamo ben osservando come il controllo delle 'anime' e delle 'menti' passi sempre e si compia nel controllo della corporeità umana. Lo ha ben indicato, tra gli altri, Foucault, che ha individuato

> un micromotore, capace di agire sui corpi e sulla vita disciplinandoli attraverso delle dinamiche estremamente complesse e diversificate. Più nello specifico: il dispositivo disciplinare si attua grazie all'organizzazione di situazioni e luoghi, in particolare istituzioni chiuse, in cui poter esercitare su gruppi relativamente ristretti o inquadrati di individui un controllo che permette, al contempo, di avere a disposizione un sapere completo su questi stessi individui (p. 104).

Plurale è quindi il corpo. Fatto «di significati, figure e nature» (p. 15) diverse, ricche, simboliche, materiche, storiche, psichiche, sanitarie, politiche; «un che di dinamico ed eterogeneo» (p. 61) che ha la caratteristica inoltrepassabile di rimanere sempre infisso al centro della storia e della vita. Perché la corporeità umana *è* la storia, il corpo *è* la vita. Storia e vita che nascono dal «*corpo presente*, disseminato in

modi e figure diverse» (p. 184) e del quale questo libro ci racconta con rigorosa passione la vicenda.

Alberto Giovanni Biuso

Federica Porcheddu
Ripensare il terzo a partire da Lévinas
Trascendenza e reciprocità
Mimesis, Milano-Udine 2021, pp. 288, € 24

C'è una questione che occupa parti rilevanti della discussione filosofica contemporanea e che può servire per collocare l'interesse di questo libro. Essa è descritta bene in un denso volume di Roberto Esposito uscito qualche tempo fa e dedicato al *Pensiero istituente*. Detto in modo sintetico, si tratta di riconoscere un'impazienza rispetto agli effetti ontologici e politici del pensiero destituente (e anche di quello costituente). Come una sorta di consunzione filosofica. Un *J'accuse* che si rivolge a un intero spirito del tempo, potremmo dire. Piuttosto, l'urgenza filosofica è ora quella di rilanciare un pensiero istituente. Perché parto proprio da qui, nel recensire il bel volume di Federica Porcheddu dal titolo *Ripensare il terzo a partire da Lévinas. Trascendenza e reciprocità*? Probabilmente perché, a leggerlo con sguardo grandangolare, questa inquietudine – ma anche una certa forma di ammirazione - per una certa epoca filosofica sta alla base dell'interpretazione di Lévinas che viene qui proposta. Si tratta, in termini più semplici, di affrontare Lévinas radicalmente, per comprendere se alla fine dei conti la sua etica prima si possa leggere esclusivamente come un pensiero destituente l'ontologia o, piuttosto, in essa non vi sia una pretesa a istituire, ancorché non chiara o implicita. Tra le

tante parole chiave che si succedono in queste pagine, molto dense ma anche molto chiare e mosse da un'attenzione al lettore che non è per nulla scontata di questi tempi, è forse la categoria di "istituzione" quella che mi colpisce di più.

Già nel primo capitolo, dedicato al debito di Lévinas nei confronti di Husserl e Heidegger, emerge infatti una sorta di insoddisfazione, conseguenza dell'irrigidimento ontologico cui la fenomenologia husserliana sembra condannarsi. Al di là dell'attendibilità storico-filosofica dell'interpretazione levinassiana di Husserl, queste pagine hanno il pregio di far capire con evidenza che la lettura di Husserl resta subordinata al primato dell'ontologia: «La posta in gioco in tutta l'opera consiste quindi nel rilevare come la teoria dell'intuizione di Husserl derivi dalla nuova impostazione ontologica della sua filosofia. La tesi centrale che emerge è che la fenomenologia husserliana, pur partendo da problemi strettamente gnoseologici, offra invero una nuova concezione dell'essere» (p. 29). Eppure risulta insoddisfacente anche il tentativo heideggeriano di «radicalizzare, completandola, la nozione husserliana di intenzionalità evitandone allo stesso tempo le aporie» (p. 39). Heidegger avrebbe radicalizzato la postura ontologica che è già tratto fondamentale di Husserl.

Perché questa radicalizzazione ontologica non sarebbe sufficiente? La risposta a questa domanda, che viene delineata estesamente nel secondo e nel terzo capitolo, non è affatto scontata.

Non si tratta infatti semplicemente di ricostruire l'evasione levinassiana dall'ontologia, ma di riconoscere piuttosto che tramite essa è la stessa ontologia a essere rimessa in questione: «Il bisogno di evasione – pieno di speranze chimeriche o no, poco importa – ci conduce al cuore della filosofia. Permette di rinnovare l'antico problema dell'essere in quanto essere» (p. 45).

C'è dunque un'ambiguità per certi versi deliberata, nel progetto del filosofo. Che l'ontologia non sia più fondamentale è il modo per prendere sul serio l'antico problema dell'essere. L'autrice sottolinea molto bene quanto tutto il percorso di rovesciamento del primato metafisico, dall'ontologia all'etica, avvenga saturando le stesse armi dell'essere, se così possiamo dire. Nota di merito particolare è, a mio parere, aver individuato la centralità teoretica della questione della *separazione*. Essa appare come un vero e proprio nodo gordiano, tramite il quale osservare tutta la tensione del progetto levinassiano. È tramite la separazione che comprendiamo una irriducibilità della pretesa istituente nel cuore stesso di un tentativo di destituzione ontologica. Una tesi fondamentale che emerge è precisamente che vi sia uno spazio istituente i cui confini sono segnati rispettivamente dalla separazione e dall'irruzione del terzo. In questo senso il secondo e il terzo capitolo vanno letti in piena continuità, perché il loro oggetto di lavoro non è la separazione e il terzo, ma questo unico spazio segnato da entrambi.

Infatti, il soggetto di godimento non si annichilisce nella messa in questione dell'altro, ma permane (si rilegga questa pagina di *Totalità e Infinito* riportata a p. 77: «La rottura della totalità che si attua con il godimento della solitudine – o con la solitudine del godimento – è radicale. Quando la presenza critica di Altri metterà in questione questo egoismo, non distruggerà questa solitudine»). Solitudine ed esteriorità avvengono *nello stesso tempo*. Questa *simultaneità* (che vedremo a breve) di pluralità e alterità è ciò che comporta la necessità di un ripensamento radicale del terzo e dell'ontologia (nei termini che sto proponendo, della "pretesa istituente"): «ciò che tento di mostrare muove dall'ipotesi di fondo di questo lavoro, ovvero dal sospetto che il concetto di soggetto e di alterità in Lévinas possa configurarsi come realmente effettivo solo a patto di conservare, da un lato, la dimensione ontologica […] e dall'altro

di rovesciare la relazione asimmetrica per far posto a una certa reciprocità» (p. 83).

I capitoli che seguono saranno dunque dedicati all'approfondimento di questa ipotesi di fondo. A cominciare dal terzo capitolo in cui, attraverso una puntuale ricostruzione dei riferimenti testuali alla questione del terzo, vengono sottolineati alcuni nuclei concettuali fondamentali come la fraternità, l'illeità, la giustizia. Voglio però soffermarmi in particolare su una distinzione, quella tra contemporaneità e simultaneità. Il terzo è infatti semplicemente contemporaneo alla relazione etica? La risposta dell'autrice è molto interessante: «la pluralità è sì contemporanea rispetto alla relazione duale, ma proprio il suo essere contemporanea è ciò che impedisce alla giustizia di porsi effettivamente come ciò che potrebbe garantire l'inclusione di tutti i volti che non fanno parte della relazione duale. L'interposizione del terzo come contemporaneo fa precipitare la relazione in un movimento per il quale la chiamata dell'alterità dell'altro non può far altro che richiamare un'alleanza esclusiva. Affinché la relazione non si chiuda su se stessa, ma possa configurarsi come apertura e garanzia della dimensione della pluralità, è necessario che il terzo sia simultaneo e non contemporaneo» (p. 137).

A me pare che i capitoli successivi, dedicati al confronto con Ricoeur e Nancy, siano un tentativo di dimostrare precisamente la necessità della simultaneità e la sua irriducibilità rispetto alla contemporaneità.

Il capitolo dedicato a Ricoeur è molto denso. Ciò che conviene ricordare è quanto in Ricoeur la questione iniziale – il rapporto tra pensiero istituente e destituente – sia essenziale. Si potrebbe dire che la particolarità dell'interpretazione ricoeuriana di Lévinas sia precisamente quella di sollevare incessantemente l'obiezione istituente (in questo senso potremmo dire che la lettura di Ricoeur è agli antipodi di tutte le

interpretazioni destituenti, a partire forse da quella derridiana). C'è infatti *simultaneamente* a *Autrui* un altro che emerge tramite lo spazio delle istituzioni. L'orizzonte dell'etica come filosofia prima conterrebbe sia il prossimo sia il *socius*. Certo, l'ipotesi ricoeuriana finirebbe per risultare doppiamente ingenerosa, nei confronti del soggetto separato e nei confronti dell'altro in quanto altro. Un'ingenerosità dovuta – mi pare questa la tesi dell'autrice – fondamentalmente a una cautela metodologica circa l'ontologia: «vorrei sottolineare come in Ricoeur vi sia una sorta di persistente esitazione tra l'ontologia classica e l'ontologia moderna, a cui si somma una certa ambivalenza tra fede e filosofia, che sembrerebbe frenare una piena assunzione del paradigma triadico» (p. 189). Insomma, si potrebbe con una battuta riconoscere che ciò che davvero accomuna i due autori è, più che un'attenzione estesa ai temi etici, una specie di inibizione ontologica che, nel caso di Lévinas, prende la via breve quanto radicale dell'altrimenti che essere, mentre nel caso di Ricoeur prende la via lunga e controllata dell'istituzionalizzazione *in vista di* un'ontologia (l'ontologia come terra promessa).

Se è questa inibizione ontologica a non permettere fino in fondo una presa di responsabilità filosofica nei confronti del terzo, il capitolo su Nancy rappresenta un vero e proprio rovesciamento. Il motivo è intuibile: perché nel filosofo di Strasburgo non vi è alcuna inibizione, anzi la questione ontologica viene immediatamente posta come la questione prima dentro cui collocare il confronto con l'etica di Lévinas (ne sarebbe indizio determinante il fatto che questo confronto sia sempre mediato da Heidegger). Certo, il prezzo da pagare è notevole: una metamorfosi dell'etica della trascendenza in etica dell'immanenza. Ma il guadagno appare certo: istituzione e ontologia non sono soltanto simultanee, ma sono la stessa cosa.

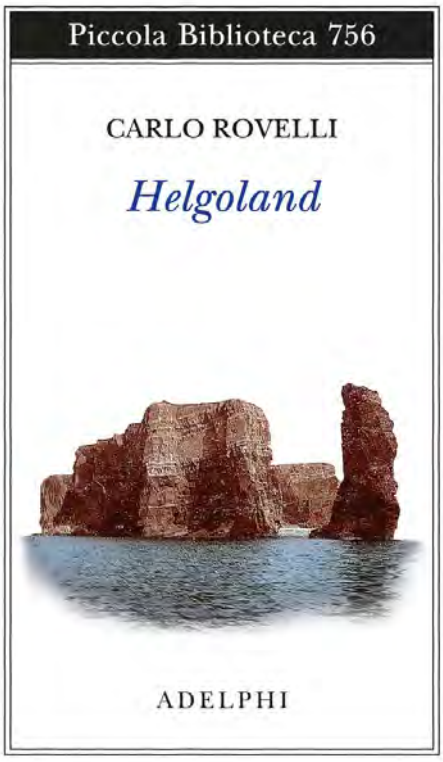
È per questo motivo che gli ultimi due capitoli sono forse i più riusciti del volume: il nodo gordiano non si scioglie ma si allenta risolutamente e, finalmente, è possibile far emergere la tesi originale dell'autrice. La possibilità di «individuare uno spazio in cui condivisione e inclusione non implichino due dimensioni irriducibili […] dipende in maniera non secondaria da una diversa interpretazione della prossimità e della relazionalità che trova – a mio avviso – una via percorribile in quel nesso che lega insieme il terzo, la reciprocità e la trascendenza» (p. 256).

Insomma, il nostro punto di partenza sembra confermarsi: è lecito pensare Lévinas all'interno di un paradigma istituente? Questo volume, pur non volendo e potendo dare una risposta definitiva, ci mette davanti degli indizi filosofici che incoraggiano un supplemento di indagini.

Sergio Labate

Carlo Rovelli
Helgoland
Adelphi Edizioni Digitale, Milano 2020, pp. 161, € 8
Adelphi Edizioni, Milano 2020, pp. 219, € 15

Telmo Pievani
Finitudine. Un romanzo filosofico su fragilità e libertà
Raffaello Cortina Editore, Milano 2020, pp. 280, € 16



Nel corso dei secoli, l'uomo ha imparato così bene ad addomesticare tutto ciò che è presente sulla terra tanto da convincersi che la Natura gli appartenga, mentre è vero l'esatto contrario. Assumendo sempre di più le sembianze di un parassita, egli vive convinto che ogni cosa presente sulla superficie terrestre sia lì per lui e, se non è sfruttabile come si vorrebbe, va adattata alle sue esigenze, solitamente rovinata. Usurpatore di un trono che non gli era assolutamente destinato, l'uomo si è convinto che un abisso minimo, ma invalicabile, lo separa dal regno animale, riconosce solo una respirazione elementare alle piante e il pensiero che possa avere delle qualità comuni ai

minerali non lo sfiora nemmeno. Ha le sue debolezze e ne è ben consapevole: teme le calamità naturali e la morte, ma si considera un essere superiore perché è in grado di spiegarsi questi eventi. L'esser consapevole dell'uomo è il suo tratto distintivo, ciò che lo distingue dalle altre specie viventi innalzandolo, eppure sembra che con il passare del tempo l'uomo abbia smesso di essere consapevole della cosa più importante: la sua precarietà. Quello che non bisognerebbe mai dimenticare è che ciò che l'uomo è riuscito a diventare è frutto di un miracolo ma il miracolo può avvenire anche in senso inverso: tutto ciò che l'uomo ha costruito può essere distrutto, tutto è soggetto a usura e alla morte. In qualsiasi momento la Natura può porre fine alla vita dell'uomo, spazzando via lui, quello che crede essere il suo mondo e la sua memoria. La vera sovrana imbattibile è la Natura solo che, in quanto tale, non ha bisogno di rimarcare la sua superiorità in ogni momento. Lo ricorda all'improvviso con cataclismi, pandemie e tutto ciò che per noi può essere una tragedia, e lo è proprio perché ci ricorda ciò che noi vogliamo costantemente dimenticare: noi non siamo i padroni del mondo e non abbiamo tutto sotto controllo. In quest'ottica, come ricorda perfettamente Rovelli nel suo ultimo libro, *Helgoland*, l'avvento della fisica dei quanti, la scoperta che la fisica classica non fosse sufficiente per spiegare il funzionamento del mondo tutto, ha rappresento un duro colpo per l'orgoglio umano, l'onnipotenza e la sua smania di controllo. La fisica classica offriva infatti un'immagine nitida della realtà, letta come un vasto spazio dove corrono particelle che costituiscono ogni materia, e che sono spinte e tirate da forze e il cui moto seguiva leggi deterministiche (p. 54). Grazie alla meccanica classica si era diffusa la convinzione che in ogni istante noi avremmo potuto, almeno in teoria, conoscere la posizione, la velocità, l'accelerazione di un corpo, inteso come insieme di particelle: niente di più bello per un maniaco del controllo. Tuttavia, la chiarezza concettuale della fisica

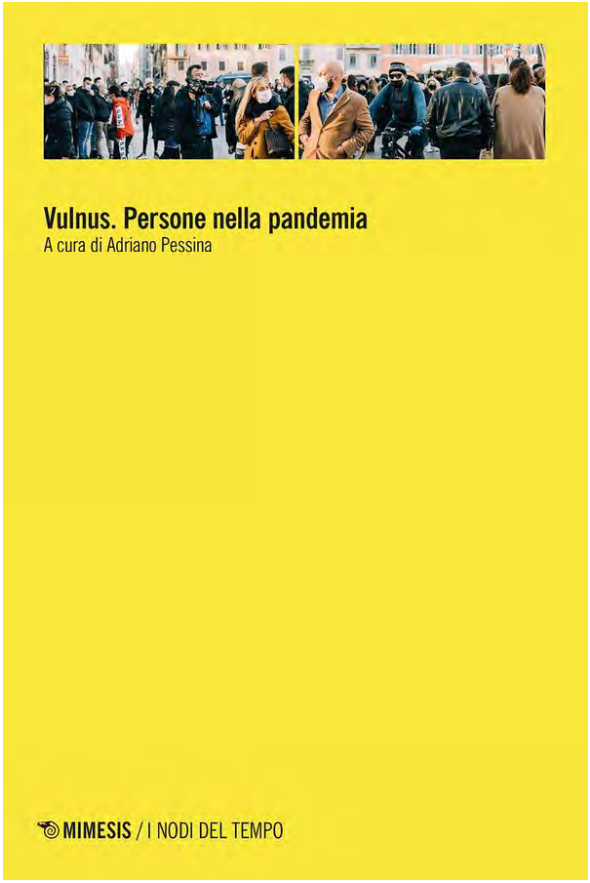
classica è stata spazzata via dai quanti. La meccanica quantistica è la teoria più significativa della scienza nella descrizione della natura microscopica ma, allo stesso tempo, è inquietante e misteriosa, e questo perché ci chiede di rinunciare a tutto ciò che ci sembrava solido e inattaccabile, ci chiede di accettare che la nostra realtà è diversa da quanto immaginavamo o credevamo di sapere. La solidità della visione classica del mondo non è che nostra miopia. Le certezze della fisica classica sono sostituite da probabilità. L'immagine del mondo nitida e solida della vecchia fisica classica è un'illusione (p. 79). Il cambio di prospettiva è radicale: una descrizione efficace della natura deve basarsi solo sulle grandezze osservabili mediante le leggi fisiche dedotte fino a quel momento. Si può descrivere soltanto ciò che è osservabile, l'inosservabile in quanto tale non può essere noto. Occhio che non vede, scienza che non sa. Questo è il punto di partenza della teoria dei quanti, questa è l'intuizione di Heisenberg, alla cui scoperta, e precisamente al luogo in cui essa avviene, è dedicato il titolo del libro di Rovelli. Heisenberg sollevò un velo e da quel momento non sarebbe stato più possibile tornare indietro. Alla sua intuizione seguì un movimento che avrebbe per sempre cambiato la visione del mondo, un movimento composto da tutti futuri premi Nobel: la fisica dei ragazzi. La teoria dei quanti è infatti innanzitutto gioco di squadra, gioco di menti che non hanno avuto paura di abbandonare la strada vecchia, conosciuta, confortevole, che non hanno avuto il coraggio di ignorare ciò che era palese ai loro occhi: non è possibile determinare tutto e il mondo microscopico ne è la dimostrazione. La teoria dei quanti è frutto di un gioco di squadra, una squadra eterogenea composta da filosofi e scienziati che con i loro contributi hanno giocato un ruolo fondamentale. Non bisogna aver paura di abbandonare le certezze che si hanno e di certo non possono farlo dei pensatori degni di questo nome, che si interrogano sul senso della vita e sul funzionamento del mondo, e ancor di più non possono farlo gli

scienziati. A più riprese nel suo testo Rovelli sottolinea come, infatti, la scienza sia questo: pensiero in movimento continuo che rimette in discussione ogni cosa, che non ha paura di sovvertire l'ordine del mondo o pensarne uno migliore (p. 55). La scienza è esplorazione di nuovi modi di pensare il mondo, capacità di rimettere costantemente in discussione i nostri concetti. Posta dunque l'intuizione di Heisenberg, qual è questa nuova realtà che ci viene disvelata? Una realtà fatta di relazioni più che di singoli individui in grado di osservare e non influenzare ciò che li circonda. Per comprendere la Natura dobbiamo guardare alle relazioni tra i vari oggetti fisici, e questo perché la realtà è una rete di relazioni. Un oggetto che non interagisce con un altro non esiste. Questo è il grande apporto della fisica quantistica: salta l'idea che il mondo debba essere costituito da una sostanza che ha attributi e ci forza a pensare tutto in termini di relazioni (p. 96). Nulla esiste come entità autonoma, nulla quindi neanche l'uomo. Noi siamo parte della Natura, non siamo osservatori esterni del mondo ma siamo nel mondo. Ogni cosa è contemporaneamente osservata, osservabile e osservante. Il punto di vista dall'esterno è un punto di vista che non esiste: ogni descrizione del mondo avviene dal suo interno (p. 120). L'uomo è parte della realtà, il nostro discorso sulla realtà è esso stesso parte della realtà. Di relazioni è fatto il nostro io, le nostre società, la nostra vita culturale, spirituale e politica. L'intuizione di Heisenberg e le sue conseguenze ci invitano ad abbandonare ogni antropocentrismo e ad accettare il da farsi: l'uomo fa parte della Natura e bisogna che a essa ritorni. Bisogna restituire l'uomo alla Natura. Se abbandoniamo il nostro egocentrismo, l'amore per il nostro Io, l'ossessione per la nostra individualità, il miraggio persistente di essere gli unici a poter comprendere il mondo, anche il nostro essere temporanei, il nostro essere mortali, diventa digeribile. Questo il messaggio del testo di Telmo Pievani, *Finitudine*, il quale riprende la tematica

camusiana dell'assurdità della vita, dell'insensatezza del vivere se tutto poi è destinato a finire. Immaginando che sia proprio un sopravvissuto Camus a scrivere, in collaborazione con il suo grande amico biologo Jacques Monod, e dimostrando come tutte le sfide tentate alla finitudine siano destinate a fallire, Pievani vuole aprire un varco, vuole provare a rendere più accettabile la morte: noi, intesi come unicità, siamo destinati a perire ma la nostra esistenza non è per questo inutile: la vita è un inestimabile e casuale regalo, esistiamo perché è uscito il nostro numero alla roulette russa, e il nostro obiettivo deve essere quello di vivere il più possibile e di farlo al meglio. E come Sisifo, ogni giorno, anche noi non dobbiamo stancarci dei nostri eterni ritorni. Nulla è privo di senso e, anzi, è come se la finitudine stessa desse un senso alla nostra vita: la sua accettazione ci rende liberi, non illusi, non nichilisti, ma liberi (p. 220). È una libertà a termine, d'accordo, ma se l'unica fatalità della vita è la morte, tutto il resto è un mondo libero in cui l'uomo è padrone dei suoi giorni. Dei suoi giorni, non del mondo (p. 260). La nostra esistenza non ha nessun senso, non siamo alla nascita destinati a compiere grandi gesta, ma nulla vieta che possiamo dare noi un senso alla nostra vita. Dobbiamo lottare per diminuire le sofferenze e rendere questo mondo desiderabile oltre che vivibile, per renderlo un posto buono per la vita per noi e per i nostri simili, presenti e futuri. Non siamo parenti degli angeli ma siamo parenti delle scimmie, facciamo parte di questo mondo e proteggerlo, rallentarne l'inevitabile fine, potrebbe essere il nostro compito, il senso del nostro essere. La partita contro la morte è già persa, ma solo se l'obiettivo è l'immortalità. Se la posta in gioco diventa l'esser soddisfatti di come si è vissuto, l'assenza di rimpianti, allora l'esito della sfida è tutto da vedere.

GIORDANA COLICCHIO

Adriano Pessina (a cura di)
Vulnus. Persone nella pandemia
Mimesis, Milano-Udine 2022, pp. 171, € 15



Di fronte alla crisi – ambientale, sociale, bellica – a quale compito è chiamata la filosofia?

Il compito che le è probabilmente più proprio: ripensare, indagare, "curare" la dimensione individuale, i suoi turbamenti e trasformazioni, nel suo rapporto con quella collettiva e viceversa.

Questo vale in senso forse ancora più forte se pensiamo alla crisi pandemica che abbiamo appena attraversato – e che ancora sembra di là dall'essere superata. Nell'intreccio tra queste due dimensioni, quella dell'evento collettivo e del dramma che si consuma in una sfera più intima, è forse possibile ricercare e individuare, se non la "verità" di quanto è accaduto, almeno delle tracce e degli spunti per riflettere a 360° sul presente e sulle sue contraddizioni.

Il Covid-19 è stato (ed è) un *trauma collettivo massivo* (p. 59), dato il suo carattere complesso e proteiforme, sono molteplici le curvature che la riflessione su questo evento può assumere – riflessione sui dispositivi di controllo, sul rapporto tra individuo e comunità, sulla fragilità umana e sul nostro rapporto con il dolore, sulla "gestione" (pubblica e privata) dell'isolamento e della *solitudine*. Nel libro curato da Adriano

Pessina, questa pluralità di spunti e di interpretazioni vengono messi in rilievo attraverso una serie di saggi che, sebbene mantengano il profilo dell'indagine sulla stretta attualità, travalicano la logica del testo d'occasione, offrendo riflessioni che, pur essendo legate all'oggi, lo superano, mostrando un più ampio respiro.

Urgenza e necessità di non fermarsi al fatto presente sono infatti ciò che deve necessariamente essere conciliato perché questi tempi *non sfuggano*, non fluiscano via impensati e dunque muti – incapaci di parlarci – o, per un movimento uguale e contrario, ci costringano a fissarci nel trauma, a non essere, a non pensare fuori dal suo perimetro, come se non esistesse un prima e un dopo: «per il filosofo di Jena tutto accade secondo un movimento cieco della storia e i grandi cambiamenti sono sempre un'opera notturna, difficilmente interpretabile nell'immediato. Rimane, perciò, come compito, il "dovere" di una ripar-*azione* terapeutica e virgola soprattutto, con l'aiuto della filosofia, *medicina doloris* dell'umano, di "comprendere" il lavorio carsico dello Spirito» (p. 75).

Il punto di partenza comune dei contributori sembra essere dunque quello della malattia che diviene potenzialmente, pur nella sua tragicità, *occasione*; come sostiene Virginia Woolf in *Ammalarsi* – testo ripreso da Papa nel suo saggio – «la malattia con i suoi simboli del dolore (…) è da sempre socialmente disturbante, anche perché *veracem*: espressione di verità» (p. 67). La *crisi* pandemica, come la radice greca stessa del termine suggerisce, non è solo momento complesso e drammatico, passaggio, spartiacque, ma anche occasione per il pensiero, per distinguere e separare, per fare chiarezza. Ci costringe a fare i conti con una mutata esperienza della morte e della solitudine. Mutata, ma non del tutto nuova.

La forza della riflessione su questo evento sta infatti, e in questo consiste la promessa mantenuta dal testo qui in

discussione, nel tenere assieme il tema dell'eccezionalità/eccezione – ciò che è accaduto è da considerarsi fuori dal comune – e della continuità: la malattia, il modo, collettivo e individuale, che abbiamo di affrontarla, sono rivelatori del fatto che parte di ciò che abbiamo vissuto già è, già era, nel bene come nel male, in preparazione, presente in potenza; «ragionare sui tratti di continuità del tempo del Covid-19 con quello pre-pandemico significa prendere consapevolezza del fatto che il primo compito dell'etica della sovranità che stiamo cercando di delineare è quello di andare a vedere in che misura la nostra vita anteriore al virus, che abbiamo imparato a considerare come la nostra *normalità*, possa essere considerata effettivamente come *normale* (nel senso di buona)» (p. 49).

Anche il ritorno alla centralità del *mero biologico* (cfr. p. 60), del corpo, che può apparirci come ripresa del vecchio, della società disciplinare che, secondo la felice intuizione di Han più volte ripresa nel testo, sembrava essere stata surclassata una volta per tutte da società della prestazione, del controllo sull'auto-controllo – instillato dal dilagare del meccanismo della performance e dell'auto-realizzazione – si configura come una riproposizione non meccanica, ma anzi del tutto nuova di questi dispositivi, di un'ibridazione tra i due modelli.

Come i cittadini di Orano raccontati ne *La Peste*, gli uomini e le donne di quest'epoca hanno vissuto il quotidiano con un misto di insofferenza, rabbia, smarrimento, incredulità. Ma la solitudine forzata, cui sono, siamo, stati costretti dalle norme di contenimento del contagio non è e non deve essere necessariamente un male. Rispunta il tema dell'occasione, sia pure, forse, dell'*occasione mancata*: il tempo del dolore e del lutto può aprire lo spazio alla solidarietà e all'introspezione, alla valorizzazione dei legami comunitari.

La stessa solitudine, parola chiave, se si vogliono descrivere e comprendere questi ultimi due anni, se ripoliticizzata, nel senso

più ampio e migliore del termine, può rovesciarsi in una risorsa salvifica e attivante. Se così intesa «la solitudine diventa un bene da conservare, così da sottrarre se stessi e la propria condizione allo sguardo di chi non può "cum-patire", ma solo commentare. Nei giorni della pandemia (…) si è assistito alla narrazione di una solitudine di un isolamento segnati drammaticamente non dal pudore, ma dallo strazio di una forzata estraneazione delle relazioni parentali e amicali: l'isolamento, dettato dalle norme sanitarie, ha trasformato gli ospedali le case di cura in luoghi inospitali perché ha impedito ogni forma di contatto fisico con chiunque non fosse "addetto" alle concitate prassi di cure e assistenza» (p. 32). Questo isolamento, questo abbandono, è stato spesso materiale, e ancor più spesso anche psicologico ed esistenziale. Lo si è visto reparti di rianimazione affollati in cui si sono sperimentate la contiguità e la separazione assolute (*ibid.*), nella quasi totale assenza di riti di passaggio significativi e da vivere collettivamente per accompagnare la dipartita dei propri congiunti, nei discorsi inquietanti sulla gerarchia delle vite da salvare/abbandonare in assenza di risorse, sulle esistenze "utili" o "gravose" (o meglio di cui sgravarsi). Eppure, di fronte al consolidarsi di una cultura sempre più individualistica, la retorica della "responsabilità collettiva", se non è semplice chiacchiera ma autentico richiamo a un *noi* (cfr. p. 34), può farci immaginare una "buona solitudine": non isolamento, competizione, egoismo, ma ripiegamento finalizzato all'apertura, alla più stretta connessione con l'Altro; «c'è infatti una solitudine originaria, non sofferente, non disperata, non di cui non si fa mai memoria e di cui, invece, si sente la nostalgia. È la solitudine che appartiene all'esperienza dell'interiorità e che richiede di essere praticata in un certo isolamento» (p. 37) laddove «non può darsi alcuna forma di solitudine e di isolamento se non sullo sfondo di una socievolezza persa» (p. 30).

L'esperienza di questa solitudine, che è ricerca e sentimento della mancanza, avrebbe potuto mettere in evidenza una volta per tutte l'impraticabilità e l'infondatezza dell'idea di autonomia assoluta dell'individuo, del mito del "self made man", dell'assoluta autosufficienza che mistifica l'origine della comunità: né gli individui, né le comunità si producono separatamente e autonomamente, la loro vitalità e possibilità di equilibrio e evoluzione è frutto della collaborazione e della cura reciproca.

Se la malattia difficilmente può essere interpretata come un dono (cfr. pp. 77 e sgg.), in questo caso, sarebbe almeno potuta essere rivelatrice di quel fondo comune dell'umano, di quella fragilità costitutiva, del vincolo che inevitabilmente ci lega; mettere in evidenza il fatto semplice, eppure troppo spesso cancellato e mistificato, su cui si fonda ogni gruppo, consesso umano: ovvero che la dipendenza non implica necessariamente una relazione gerarchica, uno stato di inferiorità, non comporta un giudizio morale o di valore. Non è e non deve essere considerata una condizione stigmatizzata e deviante. Che siamo tutti dipendenti, anzi interdipendenti. Che vulnerabilità e interdipendenza sono caratteri costitutivi dell'umano.

È riuscita la pandemia in questa demistificazione? A trasformare la nostra solitudine da *passione triste* a forza attivante?

Gli autori del volume non sembrano esserne del tutto convinti, e forse a ragione. Ma c'è ancora tempo, per riflettere e per cambiare, non disperiamo.

Viola Carofalo

NORME REDAZIONALI

I testi vanno inviati esclusivamente via email a
redazione@scienzaefilosofia.it
in formato Word con le seguenti modalità:

Abstract in inglese (max. 200 parole)

Testo	Le note vanno inserite a fine testo con:
Carattere: Calibri o Times o Times New Roman	Carattere: Calibri o Times o Times New Roman
Corpo: 12 Interlinea: 1,5	Corpo: 10 Interlinea: singola

Per favorire la fruibilità telematica della rivista, i contributi devono aggirarsi tra le 15.000 – 20.000 battute, tranne rare eccezioni, e gli articoli vanno sempre divisi per paragrafi. Anche le note devono essere essenziali, limitate all'indicazione dei riferimenti della citazione e/o del riferimento bibliografico e non dovrebbero contenere argomentazioni o ulteriori approfondimenti critici rispetto al testo. È indispensabile un abstract in lingua inglese (max. 200 parole). A esclusione delle figure connesse e parti integranti di un articolo, le immagini che accompagnano i singoli articoli sono selezionate secondo il gusto (e il capriccio) della Redazione e non pretendono, almeno nell'intenzione – per l'inconscio ci stiamo attrezzando – alcun rinvio didascalico.

Note

Norme generali
a) Autore: nome puntato e cognome in Maiuscolo/minuscolo tondo seguito da una virgola. Se si tratta di due o più autori, citarli tutti di seguito inframmezzati da virgole o trattino. Evitare l'uso di Aa.Vv. e inserire il curatore o i curatori come Autori seguito da "(a cura di)"
b) Titolo: Maiuscolo/minuscolo corsivo sempre, seguito da virgola.
c) Editore: occorre inserire la Casa Editrice.
d) Città e data: Maiuscolo/minuscolo tondo, non inframezzate da virgola. Le città straniere vanno in lingua originale.
e) L'anno di edizione. Nel caso in cui non si cita dalla prima edizione a stampa, occorre specificare l'edizione con un apice.
Esempio:
[1] G. Agamben, *L'aperto. L'uomo e l'animale*, Bollati Boringhieri, Torino 2002.
[2] A. Caronia, *Il Cyborg. Saggio sull'uomo artificiale* (1984), Shake, Milano 2008.
[3] E. Morin, *Il paradigma perduto. Che cos'è la natura umana?* (1973), tr. it. Feltrinelli, Milano 2001.
[4] G. Hottois, *Species Technica*, Vrin, Paris 2002.
[5] P. Amodio, R. De Maio, G. Lissa (a cura di), *La Sho'ah tra interpretazione e memoria*, Vivarium, Napoli 1998.
[6] G. Macchia, *Il paradiso della ragione*, Laterza, Roma-Bari 1961², p. 12. ["²" sta per seconda edizione].

Nel caso in cui si tratti di uno scritto già precedentemente citato, le indicazioni circa l'opera possono essere abbreviate con le seguenti diciture: "cit." (in tondo), "*op. cit.*" (in *corsivo*), "*ibid.*" o "*Ibid.*" (in *corsivo*).
Dopo la prima citazione per esteso si accetta il richiamo abbreviato costituito da: Autore, Prime parole del titolo seguite da puntini di sospensione e dall'indicazione "cit." (invariata anche nel caso di articoli di riviste).
Esempio:
[12] A. Caronia, *Il Cyborg*..., cit.

Casi in cui si usa "cit.":
Quando si tratta di opera citata in precedenza ma non nella Nota immediatamente precedente (per quest'ultimo caso si veda più avanti).
Esempio:
[1] E. Morin, *Il paradigma perduto. Che cos'è la natura umana?*, cit.

- Casi in cui si usa "*op. cit.*" (in *corsivo*):
Quando si tratta di un Autore di cui fino a quel punto si è citata un'unica opera.
Esempio:
[1] B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari 1942, pp. 232- 233.
[2] G. Hottois, *Species Technica*, Vrin, Paris 2002.
[3] B. Croce, *op. cit.*, p. 230. [Il riferimento è qui chiaramente a *Discorsi di varia filosofia*, poiché nessun'altra opera di Croce era stata precedentemente citata].

Nel caso in cui, invece, siano già state citate due o più opere dello stesso Autore, o nel caso in cui in seguito si citeranno altre opere dello stesso autore, *op. cit.* va usato solo la prima volta, poi si utilizzerà "cit.".
Esempio:
[1] B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari 1942, pp. 232- 233.
[2] G. Hottois, *Species Technica*, Vrin, Paris 2002.
[3] B. Croce, *op. cit.*, p. 230.
[4] Id., *Saggio sullo Hegel*, Laterza, Roma-Bari 1913, p. 44.
[5] P. Piovani, *Conoscenza storica e coscienza morale*, Morano, Napoli 1966, p. 120.
[Se a questo punto si dovesse citare nuovamente B. Croce, *Discorsi di varia filosofia*, per non creare confusione con *Saggio sullo Hegel*, si è costretti a ripetere almeno il titolo seguito da "cit."; la Nota "[6]" sarà dunque]:
[6] B. Croce, *Discorsi di varia filosofia*, cit., pp. 234-235.
In sostanza, "*op. cit.*" sostituisce il titolo dell'opera (è questo il motivo per cui va in *corsivo*) e comprende anche le indicazioni tipografiche; cit. sostituisce solo le indicazioni tipografiche (è questo il motivo per cui non va mai in *corsivo*).

- Casi in cui si usa "*ibid.*" o "*Ibid.*" (in *corsivo*):
a) Quando si tratta di un riferimento identico alla Nota precedente.
Esempio:
[1] B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari, 1942, pp. 232- 233.
[2] *Ibid.* [Ciò significa che ci riferisce ancora una volta a B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari 1942, pp. 232- 233].
[N.B.: *Ibid.* vale anche quando si tratta della stessa opera, ma il riferimento è ad altra pagina e/o volume o tomo (che vanno specificati)]:
[3] *Ibid.*, p. 240.
[4] *Ibid.*, vol. I, p. 12.
b) Quando ci si riferisce a uno scritto diverso, ma dello stesso autore (ad esempio nelle raccolte moderne di opere classiche. In tal caso, inoltre, la data della prima pubblicazione va tra parentesi).
Esempio:
[1] F. Galiani, *Della moneta* (1750), in Id., *Opere*, a cura di F. Diaz e L. Guerci, in *Illuministi italiani*, Ricciardi, Milano-Napoli 1975, t. VI, pp. 1-314.
[2] Id., *Dialogues sur le commerce des bleds* (1770), *ibid.*, pp. 345-612. [*ibid.* in tal caso sotituisce: F. Galiani, *Opere*, a cura di F. Diaz e L. Guerci, in *Illuministi italiani*, Ricciardi, Milano-Napoli 1975, t. VI].
c) Quando ci si riferisce a uno scritto contenuto in opera generale (l'esempio classico sono i volumi collettanei) citata nella Nota immediatamente precedente:
Esempio:
[1] G. Spini, *Alcuni appunti sui libertini italiani*, in *Il libertinismo in Europa*, a cura di S. Bertelli, Ricciardi, Milano-Napoli 1980, pp. 117-124.
[2] P. Rossi, *Discussioni sulle tesi libertine su linguaggio e barbarie*, *ibid.*, pp. 319-350. [*ibid.* in tal caso sostituisce: *Il libertinismo in Europa*, a cura di S. Bertelli, Ricciardi, Milano-Napoli 1980].

Tutte queste indicazioni valgono non solo quando si tratta di Note diverse, ma anche quando, nella stessa Nota, si cita più di un'opera.
Esempio:
[1] Cfr. G. Spini, *Alcuni appunti sui libertini italiani*, in *Il libertinismo in Europa*, a cura di S. Bertelli, Milano-Napoli, 1980, pp. 117-124; ma si veda anche P. Rossi, *Discussioni sulle tesi libertine su linguaggio e barbarie*, *ibid.*, pp. 319-350.

Nel caso in cui si tratta dell'edizione moderna di un classico, è indispensabile specificare tra parentesi l'anno di pubblicazione e quindi il curatore, in particolare se si tratta di edizioni critiche.
Esempio:
[1] G. Galilei, *Dialogo sopra i due massimi sistemi* (1632), a cura di L. Sosio, Einaudi, Torino 1970, pp. 34-35.

Opere in traduzione

Quando si cita dalle traduzioni è consentito omettere il titolo originale, ma occorre sempre specificare la data dell'edizione originale tra parentesi, e l'editore della traduzione preceduto dall'abbreviazione "tr. it.", "tr. fr." ecc.
Esempio:
[1] M. Heidegger, *Essere e tempo* (1927), tr. it. Utet, Torino 1969, p. 124.
[2] Id., *Les problèmes fondamentaux de la phénoménologie* (1927), tr. fr. Gallimard, Paris 1985.

Articoli di riviste

La citazione completa è così composta:
Autore, Titolo del saggio, indicazione "in" seguita dal titolo della riviste tra *virgolette basse*, annata in numeri romani, numero del fascicolo in numeri arabi (sempre preferito all'indicazione del mese), numeri delle pagine.
Esempio:
[1] D. Ferin, *Profilo di Tranquillo Marangoni*, in «Grafica d'arte», XV, 57, 2004, pp. 22-25

Citazioni

Le citazioni nel testo possono essere introdotte in due modi:

1) se si tratta di brani molto lunghi o di particolare rilevanza possono essere trascritti con corpo più piccolo rispetto al resto del testo, preceduti e seguiti da una riga vuota e senza virgolette.
2) se si tratta di citazioni più brevi o interrotte e spezzettate da interventi del redattore dell'articolo vanno messe nel corpo del testo principale, introdotte da caporali: «Xxxxxxx»
   Nel caso 2) un'eventuale citazione nelle citazione va posta tra virgolette inglesi semplici: «Xxxx "Xxxxxxx"»

Segno di nota al termine di una citazione

Quando la citazione rimanda a una nota, il richiamo di nota deve venire subito dopo l'ultima parola nel caso 1, subito dopo le virgolette nel caso 2: solo dopo va introdotto il segno di punteggiatura che conclude la frase.
Esempio:
«Conobbi il tremolar della marina»[2].

Congiunzioni ("d" eufonica)

Si preferisce limitare l'uso della "d" eufonica ai soli casi in cui essa serva a staccare due vocali uguali.
Esempio:
"a essi" e non "ad essi"; "ad anticipare" e non "a anticipare".
È consentito "ad esempio", ma: "a esempio", in frasi del tipo "venire citato a esempio".

Bibliografie

Evitare le bibliografie, i testi di riferimento vanno in nota.

Avvertenza sulle note

Sempre per garantire una più immediata fruibilità di lettura, le note devono essere essenziali e non introdurre nuovi elementi di analisi critica. Questi ultimi vanno solo ed esclusivamente nel testo.

Titoli e Paragrafi

Sempre per garantire una più immediata fruibilità di lettura, gli articoli vanno titolati e suddivisi in paragrafi. Qualora l'autore non provvedesse, il redattore che cura l'editing dell'articolo è tenuto a dare il titolo all'articolo e a suddividere l'articolo in diversi e brevi paragrafi.

S&F_scienzaefilosofia.it

ISSN 2036 _ 2927

www.scienzaefilosofia.it