STUDIUM

CONTEMPORARY HUMANISM
OPEN ACCESS ANNALS

2024

STUDIUM
Rivista trimestrale

Gli articoli della Rivista sono sottoposti a doppio referaggio cieco. La documentazione resta agli atti. Per consulenze specifiche ci si avvarrà anche di professori esterni al Consiglio scientifico. Agli autori è richiesto di inviare, insieme all'articolo, un breve sunto in italiano e in inglese.

Abbonamento 2024 € 72,00 / Europa € 120,00 / extra Europa € 130,00 / sostenitore € 156,00
Un fascicolo € 19,00. L'abbonamento decorre dal 1° gennaio.
e-mail: rivista@edizionistudium.it
www.edizionistudium.it

EDIZIONI STUDIUM S.R.L.

REDAZIONE E AMMINISTRAZIONE
Edizioni Studium s.r.l., via Giuseppe Gioachino Belli, 86 - 00193 Roma
Tel. 06.6865846 / 6875456, c.c. post. 834010

Autorizzazione del Trib. di Roma n. 255 del 24.3.1949
Direttore responsabile: Giuseppe Bertagna

STUDIUM - **Rivista trimestrale**

ISSN 0039-4130

www.edizionistudium.it

Contemporary Humanism
Open Access Annals

2024

Studium
edizioni

TABLE OF CONTENTS

HUMAN FREEDOM AT THE TEST OF AI AND NEUROSCIENCE
Stefano Biancu, Mathieu Guillermin, Fabio Macioce (eds.)

ISSN 0039-4130

ISSN 0039-4130

FIDUCIA: ANALISI E PROSPETTIVE

ISSN 0039-4130

SOMMARI/SUMMARY

Mario De Caro, *Does Imputability Require Free Will? The Discussion in the Civil Law Tradition*

It is a highly debated issue, in both civil law and common law traditions, whether the concept of imputability presupposes that of free will. This article critically examines a classic position within Italian jurisprudence (which is firmly rooted in the civil law tradition) according to which metaphysical discussions about free will – being too abstruse and abstract – should have no place in the legal definition of imputability. Four arguments are presented to show that this view is inadequate.

Fiorella Battaglia, *Democracy and Education at the Time of AI*

New technologies present some risks to the functional relationship between democracy and education. Scholarship which values the common ground of educational and political themes has often been influenced by Dewey's pragmatism (Democracy and Education 1916). I argue that democracy and education mutually strengthen each other and analyse how their relationship is complicated by the epistemic impact of AI systems on democracy and education.

Angelo Tumminelli, *AI and Democratic Citizenship. The Consequences of Infodemics for Human Freedom*

This paper aims at investigating the political risks and implications for personal freedom due to the circulation of GANs (Generative Adversarial Network) technology by presenting, in particular, an interdisciplinary mapping of the consequences related to infodemics, both at a geopolitical and at an individual level. The theoretical gains to be presented in the paper are aimed at questioning a manipulative use of these technologies aimed at the spread of infodemics and the assertion of authoritarian powers, in order to promote a "humanisation" and an ethical circulation that knows how to use these tools in a fair and democratic way.

Laura Palazzani, *Health in the Age of AI and Neuroscience: Ethical Challenges to Autonomy and Freedom*

We are faced with a "new biotechnological wave" that includes neuroscience and AI, which have a potentially "disruptive" impact on health. New possibilities are emerging, driven by the speed of innovation: the objective of the ethical discussion is to identify and discuss challenges, analyze any regulatory gaps to provide governance and/or specific regulation solutions and inform society. The paper discusses the applications of Neuro-AI in the field of medicine and the transformations in the concept of health with specific reference to the theme of human autonomy/freedom. The focus is on neuroscience and neurotechnologies at the convergence with AI, distinguishing the applications for therapy, research, enhancement, and neural data, outlining the implication on liberty, autonomy, mental privacy and cognitive liberty. There is also a reference to the use of neurotechnologies and AI outside of the medical field and the discussion on the so called neurorights.

Martina Properzi, *The Human Body and the Challenges of Augmentative Technology*

The advent of augmentation technology presents a transformative challenge: the metamorphosis of the human user into a cyber-human, that is to say, a human/machine hybrid. This article maintains that a significant portion of the ongoing discourse concerning the ethical implications of augmentative technology is shaped by a "neurocentric" bias. We propose a reorganization of the current debate that prioritizes the embodied subject.

Helga Martins, Joana Romeiro & Sílvia Caldeira, Helga Martins, *Unlocking the Soul: AI and Neuroscience Insights into Spirituality*

This paper examines the relationship between artificial intelligence (AI), neuroscience, and spirituality, beginning with the etymological roots of "spirituality", derived from the Latin spiritus, which signifies the essence of life. Spirituality encompasses various dimensions, including intrapersonal, interpersonal, and transpersonal aspects, all integral to holistic healthcare approaches. Advances in neuroscience, particularly through neuroimaging technologies, have illuminated the brain's role in spiritual experiences, involving regions like the temporal lobes and prefrontal cortex. The paper also critiques AI's potential to simplify human complexity into algorithms, raising questions about its ability to address profound spiritual needs. While AI

ISSN 0039-4130

and neuroscience offer promising tools for understanding spirituality, ethical challenges arise regarding the subjective nature of spiritual beliefs. The work advocates for a holistic perspective in healthcare that values spiritual dimensions alongside biomedical approaches, echoing Lewis-Williams' notion that not all experiences require explanation to hold significance.

Justin Nnaemeka Onyeukaziri, *Artificial Intelligence and the Question on Ethico-Moral Algorithmic Representation*

This paper is a discourse on the question of ethical and moral algorithmic representation in artificial intelligence (AI) systems. It raises questions that border on moral metaphysics and ethical epistemology, such as free agency and ethical determinism on one hand and moral apprehension and ethical cognition on the other. It argues that considering the metaphysical nature of free agency in the intrinsic relations between reason and desire in moral cognitive operations, at the root of ethical and moral actions, the question of the algorithmic representation of the human capacity for ethical and moral operations in AI systems is a possibility that cannot be automatized in AI systems.

Fernand Doridot, *About some Characteristics of Contemporary Discourses on Converging Technologies*

Some twenty years ago, Roco and Bainbridge's NBIC report on converging technologies seemed to some observers to augur a profoundly anti-humanist and reductive "metaphysical research programme", from which it was necessary to distance oneself. This article uses the joint questioning of developments in AI and neuroscience from the point of view of humanism as an occasion for assessing the topicality and development of this concept of technological convergence, as it is expressed and received today in different parts of the world. It is shown that this concept has had real scientific efficiency and has spread to most regions of the world, although it has consistently been accompanied by renewed criticism, both regarding its underlying assumptions and its mode of governance, which are expressed differently depending on scientific and cultural contexts.

Sylvain Lavelle, *What a Human is, could be and should be. The Anthropology of the Human and the Philosophy of Humanism*

The anthropological question "What is a human being?" is at the heart of

ISSN 0039-4130

Kant's philosophy, which also distinguishes the field of theory from that of practice. In the wake of ancient and modern questioning on the essence of man, philosophical anthropology is supposed to provide an understanding of human nature based on a factual-descriptive approach. But given its equally normative-prescriptive tendency in Kant, it is not without ambiguities and is exposed to the risk of a violation of Hume's Law which distinguishes the Is and the Ought. Its development into a scientific anthropology, then into a moral anthropology, which is only a branch of the previous one, leaves the question of the articulation between fact and norm unresolved. The common image of the human is today challenged by the rise and power of certain anthropo-techniques, encouraged by the stream of trans/post-humanism and its perspective of an "enhanced human". The profound modification of the human being that it envisages implies that the question of philosophical anthropology is changing. It challenges the new humanism in its capacity to propose an articulation between what a human is, what a human can be and what a human should be. This is the program of an anthropological philosophy, to be distinguished from a philosophical anthropology, that, at the foundation of a critical humanism, strives to constitute an onto-deontology of the limits of humanity.

Dominique Lambert, *AI and Freedom: some Ideas for a Debate*

In this paper, we give first a brief definition of AI and Freedom. We then describe the two main facets of intelligence and freedom. Rejecting any technophobia or technolatry, we show, on the one hand, that AI can be at the service of humans, their intelligence and their willpower. But, on the other hand, we emphasize the fact that AI can also really and deeply hinder freedom. In order to study the relationships between AI and freedom, and the ethical questions they address, we consider three regulatory models: collectivist, individualistic, personalistic. And finally, we suggest that the way to respect freedom implies a regulation in tension between the collective and the personal human dimensions. This leads us to a "common good personalism" proposition.

Sara Fernandes, Leonor Almeida and Alexandre Castro Caldas, *Free Will, Neurosciences & Robotics*

This essay explores the neuroethical problem of free will through philosophy, neurosciences, and robotics. While studies by Libet, Haggard, and Haynes reveal unconscious brain processes preceding conscious intentions, they do not negate free will. Drawing on Kant, P.Ricœur, and C.Taylor, the essay distinguishes it from freedom and defends the importance of self-determination, ali-

ISSN 0039-4130

gning actions with personal values and identity. Alan Winfield's work highlights AI's limits in replicating human adaptability and moral reasoning. These insights emphasize that true freedom and ethical decision-making rely on human relationships, empathy, and contextual understanding, especially in healthcare.

Elad Magomedov, *Existentialism as a Humanism in the Techno-scientific Era*

From a Sartrean perspective, the neurosciences and artificial intelligence do not challenge human freedom, but merely situate it. The true challenge consists in taking up the responsibility for our choices, when our responsibility seems to be compromised by something beyond our control. The more our age makes it seem that we are unfree, the more responsible we become to act in accordance with our freedom.

Zsolt Almási, *Human Agency Reloaded in our Techno-social Ecosystem*

This paper examines the decentralisation of the human subject in posthumanist philosophy, focusing on the interplay between human agency and technological indeterminacy. Using Artificial Narrow Intelligence as a case study, it explores how generative/predictive/invocational AI challenges traditional conceptions of freedom and intentionality. Drawing on Alasdair MacIntyre's notion of "human practice," the paper argues that human agency persists within these technologies through an intentional commitment to excellence and self-transcendence. By situating the human within a technosocial ecosystem, this study contends that collaborative engagement with technology involves a moral imperative to refine one's capabilities and achieve a unique, authentic voice.

Yves Poullet, *Human Dignity at an AI and Neurosciences Age*

The questions we face might be summarized as follows: what does our right to dignity mean in an age of ubiquitous technology, present in our pockets, our walls, our supermarkets, ... in our bodies? What does this right imply at a time when science can modify our genetic make-up and decipher the functioning of our brains to read our thoughts or guide our actions? Dignity is undoubtedly the first of the fundamental rights that underpin everyone's right to self-determination and equality. At the time of emerging technologies, both digital and neuroscientific, we need to imagine new facets of this right to dignity. The aim of this contribution is to go beyond regulatory texts and envisage what a fourth generation of human rights could be, including the right to participate in decisions concerning the future of our information society, the redefinition

ISSN 0039-4130

of universal service access in the information society, and the protection of our mental integrity, … In view of the risks raised by the potential uses of these technologies not only to our individual liberties but beyond social justice, environment, rule of law and democracy, the duty of precaution imposes not an a priori distrust of these innovations but, at least, the duty to reflect on and, where necessary, to reduce these risks. "AI for good" requires it.

Marco Russo, *Right Thing at the Right Time: A Phronetic Look at AI*

The ethical discussion of AI is dominated by the normative approach, which identifies a set of rules for production and use of the technology. Although important, this top-down approach is insufficient because the behaviours of humans, but in part also of AI, are unpredictable: notwithstanding any control procedure, interaction with AI accentuates the contingency of praxis. Virtue ethics in the Aristotelian tradition is based precisely on this contingency and the need to develop phronesis (practical wisdom) to act righteously under conditions of uncertainty. The article aims to highlight the importance of such a bottom-up phronetic approach, at least as a supplement to the standard approach. After outlining the basic features of phronesis, we ask whether and why only humans can be wise.

Thierry Magnin, *Christian Thought and Humanism at the time of Artificial Intelligence and Neurosciences*

Humanity has to learn to live with so-called intelligent machines, whose certain capacities increasingly surpass those of humans in many fields, particularly when they pilot nanotechnologies, biotechnologies and neurotechnologies. This paper aims to explore how Christian social thought can offer meaningful reference points for a digital world that serves humanity. Thus human dignity, common good, solidarity, subsidiarity and participation,the universal destination of goods and the preferential option for the poors, guide the reflection, through an integral ecology. In the face of intelligent machines, the human dignity is explored in terms of differences between human intelligence and AI (which is a bodiless intelligence), through the tripartite anthropology of body-soul-spirit and its resonance with life sciences and neurosciences today. It is shown how the adaptability to a transitioning world precisely calls upon humanity's own unique qualities: its personal and collective resilience, which is rooted in a harmonious interaction between the bodily, psycho-social and spiritual dimensions.

ISSN 0039-4130

Heup Young KIM, *New Humanism in the Age of Artificial Intelligence: A Theo-Daoian Reflection*

This paper explores humanism in the age of AI through the lens of Theo-Dao, an East Asian theology integrating Daoism and Confucianism. It critiques the technocratic paradigm's focus on technological advancement at the expense of ethical and environmental considerations. The paper also examines posthumanism and transhumanism, highlighting their limitations. It proposes an inclusive humanism that emphasizes Confucian virtues, cosmogonic relationality (Taiji), and theo-anthropo-cosmic wholeness (Dao). This approach seeks to recalibrate humanism for the AI age, advocating for a future where technology enhances human virtues and reconnects humanity with the Earth and its inhabitants.

Michael D. Prendergast, *Religious Bias Benchmarks for ChatGPT*

The objectives of this study are to estimate the frequency of six types of biases in ChatGPT's responses to faith-specific morality and ethics questions, assess how those biases vary by both faith and ChatGPT model version and determine how model engineering techniques affect the assessments. Five of six biases examined were widespread in the responses, and no engineering technique successfully reduced all of them.

Jan Juhani Steinmann, *The coming God. Soteriological figures in Kierkegaard, Nietzsche and Heidegger*

Based on Martin Heidegger's famous statement that only a god can save us, this article thematises three soteriological figures as they appear in the thinking of Kierkegaard, Nietzsche and Heidegger: Christ, Dionysus and the last God. The respective crises and transitional movements that underlie the necessity and possibility of their appearance, as well as their specific appearance itself, are explained. At the end, the convergence of all three figures in the event of the coming God is briefly reflected upon.

Gael Trottmann-Calame, *An all-too-modern modernity. A genealogical investigation*

Nietzsche is famous for having asserted that man is something to be overcome. So, wouldn't we have reason to think that the philosopher from Sils-maria would be like the prophet of this very contemporary craze for augmented man (transhumanism) ? In truth, far from embodying the overcoming of which Nietzsche championed (Selbstüberwindung) – calling for a new type (Übermensch) –, the other human or the new intelligence to which some aspire

ISSN 0039-4130

would rather be a fall, a negation, in short the assumption of the "fragment man", the "last man". Poison rather than cure, the vain quest for a different humanity rather than a metamorphosed one, invites us to ask, with and following Nietzsche: "What is it here that hates so much"?

Tomaso Pignocchi, *Language and Soteriology: Desire, Illusion and Liberation in Wittgenstein's and Buddhist Philosophies*

This text proposes a soteriological reading of Wittgenstein's later philosophy, in which "salvation" is understood as a form of "liberation" aimed at overcoming compulsive thought patterns that generate suffering. According to the Austrian philosopher, our misunderstanding of the functioning of language creates metaphysical expectations which, when disappointed by reality, lead to a state of existential distress. A parallel will thus be drawn between Wittgenstein's philosophy and certain principles found in Buddhist philosophies, to highlight how Wittgensteinian thought can likewise be interpreted as a practice of psychological liberation, based on the idea that suffering does not arise from the external world but rather from the confused relationship that, through language, we establish between "the world" and "the way we see the world".

Marco Tassella, *Moral Luck: an Accessible Exploration*

The moral luck paradox poses a challenge to the principle that moral responsibility should depend uniquely on factors within an agent's control. In their papers, Thomas Nagel and Bernard Williams highlighted how uncontrollable elements—such as consequences, circumstances, character, and causal history—may impact moral judgments, creating a tension between our theoretical image of moral assessment and everyday practical evaluations. This paradox raises critical questions in ethics and law, both fields where judgments may be swayed by lucky factors. By distinguishing between causal and consequential luck, we could reframe the debate in a healthier way, in order to reveal the complexities of responsibility to get a more complete understanding of moral agency, justice, and moral evaluation.

Federico Rudari, *Duchamp, Materiality, and Intersubjectivity: From Phenomenology to Aesthetics*

Starting from a material approach to Marcel Duchamp's Foutain (1917) and his ready-mades, the paper looks at the contemporary experience of artistic objects and exhibition architectures as a complex act that is less and less grounded on historical and technical value but rather on the manifold act of "spectating".

ISSN 0039-4130

This concept, practice, and phenomenon is thus understood as the interrelation of three fundamental elements: the visitor's body, architecture as a phenomenal and semio-narrative tool, and artworks as physical objects.

Costanza Vizzani, *The Theoretical Foundations of the Feminist Debate on Reproductive Technologies*

With this contribution I intend to highlight some of the theoretical assumptions underlying the feminist debate on reproductive technologies. In particular, I will discuss the comparison between feminism of difference and feminism that is critical of binarism, pointing out the different theoretical positions on the concept of motherhood and reproduction. I have chosen, as a key access, the thought of Simone de Beauvoir and Luce Irigaray as representatives of feminism of difference, and Monique Wittig and Judith Butler as critical representatives of heteronormativity and binarism. Indeed, the works of these authors can be counted among those that have most contributed to outlining some fundamental conceptual junctures taken as a point of reference in the contemporary feminist debate on reproductive technologies.

Sarah Horton, *Alienation and Self-Knowledge in Maine de Biran*

Maine de Biran's philosophy of effort teaches us that the limits to self-knowledge and even to experience are constitutive of human being. There is neither awareness of oneself nor self-knowledge without the strangeness that resists them, for humans are constituted by a non-assimilable exteriority, and one can know oneself, to the degree that such a thing is possible—and it is indeed an important task—only on the basis of this resistance, even this alienation. The one who knows himself or herself is the one who admits the impossibility of knowing oneself—but this impossibility founds all possible knowledge.

Cecilia Benassi, *Il metodo e l'intero. Nota sull'eredità di Pavel Florenskij*

Il presente studio esplora l'eredità intellettuale di Pavel Florenskij, soffermandosi sulla ricostruzione del suo pensiero e del suo metodo creativo in un contesto limitato dal difficile accesso agli archivi e da incongruenze nelle informazioni disponibili sui suoi lasciti. A partire dalle lettere dal gulag e dalle testimonianze dei figli, emerge un quadro della metodologia di Florenskij, che appare basata sull'integrazione di discipline diverse e sull'idea di un intero che precede e genera organicamente le sue parti. Lo studio propone un'analisi del pensiero dell'autore ponendo al centro la sua visione integrale, intesa anche come chiave del suo metodo conoscitivo-creativo e della sua eredità.

ISSN 0039-4130

This study explores the intellectual legacy of Pavel Florenskij, focusing on the reconstruction of his thought and creative method within a context constrained by limited access to archives and inconsistencies in the available information on his intellectual estate. Drawing on letters from the gulag and testimonies from his children, a picture of Florenskij's methodology emerges, revealing a *system* grounded in the integration of diverse disciplines and the notion of a whole that precedes and organically generates its parts. The study offers an analysis of the author's thought, emphasizing his integral vision, which is also understood as the key to his epistemological-creative method and his legacy.

Flavia Chieffi, *The role of «symbolic consciousness» in Virgilio Melchiorre's philosophy*

This paper analyzes the role of «symbolic consciousness» in Virgilio Melchiorre's philosophy. Starting from the tension in perspective consciousness to transcend itself, it highlights the intentional duplicity of consciousness, both situated and capable of desituation. This duplicity is due to the constitutive relationality of every reality, arising from an intentional movement converging on the Being, everywhere participating and transcending itself. Symbolic consciousness proves to be the most adequate way to express the relationship between Being and human being.

Francesca Fioretti, *Civic and Citizenship Education in Italy. From School Organization to Teaching Practices*

This paper aims to provide an overview of civic and citizenship education in Italy, tracing its evolution from its introduction in the 1950s to the current legislation that reinforces its cross-curricular teaching. The main findings of the survey conducted in four lower secondary schools in Rome are then presented, exploring the organizational and teaching practices adopted to implement civic and citizenship education and highlighting the main differences and similarities among the schools.

Marco Valerio, *Learning to Teach Civic and Citizenship Education and Education for Sustainable Development During Pre-service Teacher Training*

This study investigates the integration of Civic and Citizenship Education (CCE) and Education for Sustainable Development (ESD) in pre-service teacher training programs for pre-primary and primary school teachers in Italy

ISSN 0039-4130

and Portugal. Using a multiple-case study approach, it examines how CCE and ESD are embedded within teacher education programs, focusing on curriculum content, pedagogical approaches, and the perceived preparedness of future teachers to teach CCE and ESD. Findings aim to shed light on how CCE and ESD are integrated into four selected university contexts and to propose a model for integrating CCE and ESD into teacher training curricula.

Francesco Marcelli, *Catholic University Students in the 1940s and 1950s. The Importance of a Professional, Human and Religious Formation.*

This article highlights the importance of the professional, human and religious education received by young Italian Catholics in the university environment in the 40s and 50s of the last century. The Fuci and the Movimento Laureati, as well as Pax Romana at an international level, formed the future ruling class. In fact, many of their members, partly as a result of the education received, rose to positions of primary importance in professional, social and political spheres.

Giammarco Basile, *Flaminio Piccoli, the DC and Centrist Democrat International (CDI): Methodology and Goals*

This essay aims to outline the research lines concerning the political personality of Flaminio Piccoli, a member of the Christian Democracy (DC), in which he was active from his political debut in 1945, in Trento, until the conclusion of the Christian Democrat experience in 1994. This contribution seeks to define the areas of investigation and the sources that could help highlight on the main aspects of Piccoli's political career and, consequently, the political evolution of the contemporary Italy over three analysis' level: local, national, and international.

Enrico Di Meo, *Mechanism and Free Will: A possible Convergence Hypothesis*

This paper explores Charles Taylor's convergence hypothesis, introduced in a 1971 paper. Taylor proposes a multi-layered ontology that accommodates mechanistic explanations without reducing human behaviour to mere causal determinism. This framework seeks to reconcile the insights of various sciences, particularly neuroscience, with our intuitive understanding of free will and self-determination. To further illuminate this hypothesis, we will examine the perspective of Ian McGilchrist, a British neuroscientist and psychiatrist, who offers insights into the nature of consciousness and human agency that can be seen as complementary to Taylor's ones.

ISSN 0039-4130

Alessia Cadelo, *The Power of Algorithms to Redefine Human Autonomy*

Recommender systems, since they select the contents being displayed to the users, help users navigate online, but they raise a number of ethical issues. Here, the focus is on the impact the recommender systems may have on personal autonomy. Firstly, the concept of autonomy will be considered philosophically, from two different perspectives, procedural and relational. On these grounds, it will be illustrated that recommender systems are a form of digital nudging and therefore may undermine human autonomy. They could interfere especially with authenticity and reshape personal identity.

Folco Cimagalli, Giuseppina Signorello, *Posso fidarmi? La fiducia nelle relazioni del "Dopo di noi"*

Il presente lavoro indaga il ruolo della fiducia nelle relazioni di aiuto, con particolare attenzione al contesto del "Dopo di noi", espressione utilizzata per descrivere il periodo in cui i familiari delle persone con disabilità non sono più in grado di prendersi cura di loro. La fiducia è un elemento fondamentale nelle dinamiche sociali e riveste un'importanza cruciale nel processo decisionale delle famiglie che pianificano il futuro del proprio familiare con disabilità. Nello studio, si sostiene che la fiducia particolaristica, quella generalista e quella politico-istituzionale si intrecciano all'interno delle relazioni di aiuto; nel contesto del "Dopo di noi", essa rappresenta un elemento fondamentale tanto per il buon esito del processo che per il benessere "emotivo" delle persone coinvolte. Appare dunque fondamentale che le politiche sociali si orientino verso modelli che favoriscano la costruzione e il mantenimento della fiducia tra famiglie, persone con disabilità, operatori e organizzazioni. La creazione di relazioni fiduciarie solide è determinante per affrontare le sfide del "Dopo di noi", assicurando soluzioni sostenibili e adeguate, che rispondano alle esigenze delle persone con disabilità e delle loro famiglie nel lungo periodo. Solo in questo modo sarà possibile garantire un sistema di supporto che rispetti la dignità e le aspirazioni di tutti gli attori coinvolti, promuovendo un futuro più inclusivo e partecipato.

This work explores the role of trust in helping relationships, with particular attention to the context of the "After Us" period—an expression used to describe the time when the family members of people with disabilities are no longer able to care for them. Trust is a fundamental element in social dynamics and plays a crucial role in the decision-making process of families planning for the future of their relative with a disability. The study argues that particularistic trust, generalized trust, and political-institutional trust inter-

ISSN 0039-4130

twine within helping relationships; in the context of the "After Us" period, trust represents a key factor both for the success of the process and for the emotional well-being of those involved. It is therefore essential that social policies be oriented towards models that foster the building and maintenance of trust among families, people with disabilities, professionals, and organizations. The creation of solid trust-based relationships is crucial for facing the challenges of the "After Us" period, ensuring sustainable and appropriate solutions that meet the long-term needs of people with disabilities and their families. Only in this way will it be possible to guarantee a support system that respects the dignity and aspirations of all parties involved, promoting a more inclusive and participatory future.

Simone Mulargia, *Mi fido, quindi fai tu. La fiducia come chiave di lettura nella comunicazione dagli anni '50 a oggi*

In questo intervento analizzerò la dimensione della fiducia in tre figure chiave della comunicazione: l'opinion leader, la celebrità e l'influencer. Partendo dal modello del "flusso a due fasi della comunicazione" di Lazarsfeld, esaminerò il ruolo degli opinion leader come mediatori fiduciari tra i media e il pubblico. Successivamente, analizzerò la figura della celebrità, approfondendo il concetto di pseudo-evento di Boorstin e il ruolo delle star come testi culturali, secondo Dyer, fino ad arrivare al fenomeno della "celanthropy". Infine, esplorerò l'emergere degli influencer e il concetto di autenticità come risorsa strategica, evidenziando il passaggio dalla fiducia interpersonale alla fiducia mediatizzata. La riflessione si chiuderà con una discussione sul rapporto tra fiducia e partecipazione, problematizzando la delega decisionale e le nuove forme di coinvolgimento attivo nell'ecosistema digitale.

This paper examines the role of trust in three key figures of communication: the opinion leader, the celebrity, and the influencer. Starting from Lazarsfeld's "two-step flow of communication" model, I will explore how opinion leaders act as trust mediators between media and audiences. Then, I will analyze the celebrity phenomenon, addressing Boorstin's concept of the pseudo-event and Dyer's view of stars as cultural texts, leading to the notion of "celanthropy." Finally, I will investigate the rise of influencers and the notion of authenticity as a strategic resource, highlighting the shift from interpersonal trust to mediatized trust. The discussion will conclude by questioning the relationship between trust and participation, considering decision-making delegation and new forms of active engagement in the digital ecosystem.

ISSN 0039-4130

Michele Ciancimino, *La fiducia nell'esperienza giuridica contemporanea. Brevi note introduttive*

Il contributo propone una riflessione introduttiva sul tema della fiducia nell'esperienza giuridica contemporanea. Muovendo dalla constatazione di una crisi di fiducia nelle istituzioni pubbliche e nel sistema giuridico, si analizza la fiducia come elemento strutturale e relazionale dell'ordinamento, riscoprendone la funzione tanto di presupposto dell'efficacia del diritto, quanto di obiettivo dell'azione giuridica. Viene, quindi, evidenziata la necessità di un ripensamento culturale del diritto, volto a promuovere una cultura giuridica capace di integrare valori condivisi e di favorire la coesione sociale. A partire da alcune ipotesi di rilevanza della fiducia in diversi ambiti giuridici, infine, si valorizza la possibilità di considerare la fiducia non solo come categoria giuridica, ma anche come strumento di rigenerazione culturale e sociale.

This contribution offers an introductory reflection on trust in contemporary legal experience. Observing a widespread crisis of trust in public institutions and the legal system, it analyses trust as both a structural and relational element of the legal order, rediscovering its function as a prerequisite for the effectiveness of law and as a goal of legal action. Thus, the need for a cultural rethinking of law is highlighted, aiming to promote a legal culture capable of integrating shared values and fostering social cohesion. Drawing from selected examples of the significance of trust in various legal contexts, the article ultimately emphasises the possibility of understanding trust not only as a legal category, but also as a tool for cultural and social regeneration.

Marco Valerio, *La fiducia dei giovani studenti nel proprio futuro e nelle istituzioni. Uno sguardo ai risultati ICCS 2022*

Questo intervento analizza la fiducia degli studenti nelle istituzioni e le loro aspettative per il futuro, basandosi sui dati dell'indagine comparativa ICCS 2022 (International Civic and Citizenship Education Study) condotta dalla IEA (International Association for the Evaluation of Educational Achievement). In particolare, viene esaminato il caso italiano in relazione ai risultati degli altri paesi, evidenziando tendenze e variazioni rispetto al ciclo precedente. I risultati mostrano un generale calo della fiducia nelle istituzioni, mentre le aspettative per il futuro rimangono positive. Infine, si discuterà l'importanza dei fattori di contesto – storici, economici e culturali – nell'interpretare adeguatamente tali risultati, nonché il ruolo attivo della scuola nel promuovere la fiducia degli studenti nelle istituzioni.

ISSN 0039-4130

This paper analyses students trust in institutions and expectations for their own future, based on data from the ICCS 2022 (International Civic and Citizenship Education Study) comparative survey conducted by the IEA (International Association for the Evaluation of Educational Achievement). Specifically, it examines the Italian case in relation to the results of other countries, highlighting trends and variations compared to the previous cycle. The findings reveal a general decline in trust in institutions, while expectations for the future remain positive. Finally, the discussion will address the importance of contextual factors – historical, economic, and cultural – in adequately interpreting these results, as well as the active role of schools in fostering students' trust in institutions.

Francesco Luigi Reina, *La fiducia come cura della patologia sociale nel sistema penale*

Il contributo analizza l'evoluzione della rilevanza della fiducia all'interno del sistema penale italiano, sia nel rapporto verticale tra Stato e consociato, sia nei rapporti orizzontali, cioè tra i consociati che dal reato siano stati lesi direttamente o indirettamente. Grazie alle recenti innovazioni legislative, alla tradizionale giustizia punitiva, connotata da sfiducia verso il reo, si sono affiancati istituti di giustizia riparativa in cui la fiducia nel pentimento del colpevole incide direttamente sugli esiti della vicenda giudiziaria.

The article analyzes the evolution of the significance of trust within the Italian penal system, both in the vertical relationship between the State and the individual, and in the horizontal relationships among individuals who have been directly or indirectly harmed by a crime. Due to recent legislative innovations, restorative justice mechanisms have been introduced alongside traditional punitive justice, which is characterized by distrust towards the offender. In restorative justice, the trust in the remorse of the guilty party directly influences the outcomes of judicial proceedings.

Vincenzo Mignano, *La fiducia nel diritto internazionale e dell'Unione Europea: inquadramento e strumenti*

La definizione del concetto di fiducia nell'ambito del diritto internazionale e del diritto dell'Unione Europea (UE) costituisce un tema complesso e di non facile sintetizzazione. Esso, di fatto, implica un rapporto dalle sfumature più generali che attiene alla relazione che sussiste tra la fiducia stessa e il diritto latamente considerato, nonché agli effetti che da tale relazione discendono. Muovendo da tale prospettiva, il presente contributo mira ad esaminare il concetto di fiducia sotto un duplice profilo di indagine. Il primo attiene all'inquadramento della

ISSN 0039-4130

natura che definisce la fiducia negli ordinamenti considerati. Il secondo, strettamente connesso al precedente, concerne alcuni strumenti normativi attraverso cui tale concetto trova attuazione. L'analisi delle questioni sopra elencate consentirà di addivenire alla conclusione secondo cui la fiducia negli ordinamenti giuridici oggetto di esame appare di difficile definizione, soprattutto in relazione alla concezione sia "fiduciaria" che "non fiduciaria" che alcuni degli strumenti che verranno considerati hanno assunto negli anni.

Defining the concept of trust in the International and European Union (EU) law framework constitutes a challenging and not easily summarised issue. As a matter of fact, it implies a relationship with more general shades concerning the connection that exists between trust itself and law as well as the effects stemming from this relationship. From this perspective, this article aims to examine the concept of trust from a twofold viewpoint. The first concerns the framing of the nature qualifying trust in the legal systems under consideration. The second, closely linked to the previous one, deals with specific regulatory instruments by means of which this concept is implemented. The analysis of the issues listed above will lead to the conclusion that trust in the legal systems under examination appears to be difficult to define, especially in relation to both the "fiduciary" and "non-fiduciary" conceptions that some of the instruments that will be analysed have assumed over the years.

Mael Bombaci, *Gli influencer tra credibilità, identificazione e trasparenza: linee di ricerca sulla negoziazione della fiducia*

L'articolo analizza il rapporto tra influencer e audiences attraverso tre dimensioni chiave della fiducia: credibilità, identificazione e autenticità. Attraverso i risultai di tre ricerche empiriche, si tenta di fare emergere la complessità della figura dell'influencer e l'impatto delle sue pratiche sulla partecipazione collettiva. I risultati mostrano come la fiducia possa, in alcuni casi, rimanere confinata a una dimensione individuale, modellare desideri e percezioni sociali o rafforzare dinamiche di spettatorialità passiva. Questo solleva interrogativi sulle implicazioni della fiducia in rapporto agli influencer, che potrebbe rappresentare tanto una leva per nuove forme di partecipazione quanto un elemento che ne rafforza la passività.

The article analyses the relationship between influencers and audiences through three key dimensions of trust: credibility, identification and authenticity. Three empirical studies reveal the complexity of the influencer figure and the impact of their practices on collective participation. The results show that trust can, in some cases, remain limited to an individual dimension, shape social desires and perceptions, or reinforce the dynamics of passive specta-

ISSN 0039-4130

torship. This raises questions about the implications of trust in relation to influencers, which can be both a lever for new forms of participation and an element that reinforces their passivity.

Giovanna Arigliani, *Il tempo della permanenza. Riflessioni pedagogiche sulla fiducia come "sustanza di cose sperate"*

Il contributo si prefigge di affrontare il tema della fiducia con l'intenzione di mettere in luce le caratteristiche e le spinte che la fiducia mette in gioco nell'ambito della relazionalità. Ponendo alcune domande che ruotano attorno all'imprevedibilità insita nella fiducia e intesa come parola contenitrice, seguiranno alcune riflessioni propedeutiche ad una ri-lettura della fiducia fino a che questa possa riappropriarsi del suo tempo che è il tempo della permanenza, una promessa di reciproca appartenenza e riconoscimento.

This paper aims to address the theme of trust with the intention of highlighting the characteristics and dynamics it involves in relationships. By raising questions concerning the unpredictability inherent in trust as a container word, the discussion will then move toward preliminary reflections that lead to a reinterpretation of trust. This will ultimately allow trust to reclaim its own time—the time of permanence, a promise of reciprocal belonging and recognition.

Giulia Anselmo, Pierfrancesco Minicangeli, *La fiducia nel diritto civile*

La fiducia ha sempre rivestito un ruolo centrale e complesso nel diritto privato, rappresentando un elemento di fondamentale importanza per la regolamentazione delle relazioni private. Piuttosto che basarsi su sanzioni, il diritto privato è plasmato da principi che promuovono la cooperazione e la prevedibilità, con la fiducia che funge da elemento fondamentale per le interazioni giuridiche. Questo contributo esplora il modo in cui la fiducia si riflette in concetti chiave quali la buona fede, le legittime aspettative e l'equità, e il modo in cui essa informa vari istituti giuridici, dal diritto contrattuale al diritto di famiglia e alle tutele per gli individui più vulnerabili. Particolare attenzione viene data al rapporto fiduciario, che storicamente esemplifica la dimensione personale della fiducia nel diritto privato. Negli ultimi anni, tuttavia, la crescente complessità degli scambi economici e il declino della fiducia negli intermediari istituzionali hanno spinto verso nuovi modelli tecnologici di fiducia. La tecnologia *blockchain* e gli *smart contract* rappresentano una trasformazione significativa in questo senso, proponendo

ISSN 0039-4130

sistemi decentralizzati che sostituiscono la fiducia "personale" o istituzionale con processi automatizzati che promettono trasparenza.

Trust has always played a central and complex role in private law, acting as an underlying element in the regulation of private relationships. Rather than relying on sanctions, private law is shaped by principles that promote cooperation and predictability, with trust serving as a fundamental basis for legal interactions. This contribution explores how trust is reflected in key concepts such as good faith, legitimate expectations, and equity, and how it informs various legal institutions from contract law to family law and protections for vulnerable individuals. Particular attention is given to the fiduciary relationship, which historically exemplifies the personal dimension of trust in private law. In recent years, however, the increasing complexity of economic exchanges and the decline in confidence in institutional intermediaries have prompted a shift toward new technological models of trust. Blockchain technology and smart contracts represent a significant transformation in this regard, proposing decentralized systems that replace personal or institutional trust with automated and transparent processes.

Lucia Battistel, *Like Hermes «the ox-thief» or a child «with jam on his hands»: Notes on Trust from Piero Bigongiari's Metapoetic Reflections*

Drawing upon reflections by the Italian poet Piero Bigongiari (1914-1997), this paper will examine the theme of trust and its metapoetic function. The focus will be, in particular, on the image of Hermes «the ox-thief» and that of the child «with jam on his hands», which Bigongiari uses to describe his own activity as a poet and which both, albeit in different ways, offer an interesting key to renegotiating the relationship of trust within the "literary relationship".

ISSN 0039-4130

Preface

Stefano Biancu, Mathieu Guillermin, Fabio Macioce

To think how human freedom is being challenged today by new technologies (especially AI) and new knowledge (especially neuroscience) is to think about how to be human in an age like this. The relationship between human freedom and new knowledge and technologies is particularly paradoxical. Hans Jonas' analysis is illuminating in this respect. He observes that the scientific and technological revolutions "started in freedom and as an exercise of human freedom". But, "while the revolution was started by revolutionaries, it is now continued, although still a revolution, by the orthodox. What began in acts of supreme and daring freedom has set up its own necessity and proceeds on its course like a second, determinate nature – no less deterministic for being man-made". Somehow, human freedom can undermine itself. Scientific and technological development can acquire so much momentum that it begins "carrying its carriers along as its appointed instruments" (Jonas 1974, p. 48).

This paradoxical relationship highlights the pressing need for a deep exploration of humanism, of what it means to be human, if we wish to defuse the various threats that AI and neuroscience represent to human freedom, as well as to benefit from the wonderful opportunities they reveal.

In this respect, the philosophical tradition can be a key component of our reflection. The question of freedom represents the fundamental question of humanism, at least since that sort of "Manifesto" that is the "Oratio De Hominis Dignitate", which Giovanni Pico della Mirandola wrote in 1487.

In Pico's "Oratio" God addresses these very famous words to the human being, "I have placed you at the very centre of the world, so that from that vantage point you may with greater ease glance round about you on all that the world contains. We have made you a creature neither of heaven nor of earth, neither mortal nor immortal, in order that you may, as the free and proud shaper of your own being, fashion yourself in the form you may prefer." (Pico della Mirandola, 1958, p. 7).

The human being is placed by God at the centre of the world. Unique among all creatures, humans have the power to freely shape themselves as they see fit. Humanity is therefore engaged in a free project, called on to freely and creatively fulfil themselves. This is indeed the core of humanism: that humanity that unites all human beings is not just a given, but is invariably a task. A task that requires creativity and, therefore, freedom.

Actually, Pico seemed to be well aware that freedom is not in the least absolute. His idea of freedom was not abstract. The Latin text of his "Oratio" reads, "Medium te mundi posui, ut circumspiceres inde comodius quicquid est in mundo. Nec te celestem neque terrenum, neque mortalem neque immortalem fecimus, ut tui ipsius *quasi* arbitrarius honorariusque plastes et fictor, in quam malueris tute formam effingas." (our italics).

Pico is thus aware that humans are almost (*quasi*) entirely free to shape themselves: human freedom is not absolute, but always subject to conditions and restrictions – it is a situated freedom. Since its origin, Humanism has never been the defence of the reasons for an absolute freedom: it has been (and still needs to be) the affirmation that the task that awaits us to freely become human, to freely fulfil our humanity, is rooted in our concrete situation, whether it is biological, historical, cultural or other.

To grasp more clearly this task humanism that appeals to, it is interesting to further clarify the concept, which is highly polysemic. It is necessary to distinguish at least three ways in which the term can be used: the historiographic use, the cultural, and the axiological (Biancu, 2019a; Biancu, 2019b; Biancu, 2019c).

Humanism is above all a historical and historiographic term with a descriptive and interpretative function applied to certain points in European (and Western) intellectual history: Italian Humanism of the 15th Century, German New Humanism of the 18th (the *Goethezeit*), and the various humanisms of the 20th Century: pedagogical humanism (Jaeger, 1934-1947); Christian humanism (Maritain, 1936); Marxist humanism (Fromm, 1965; Merleau-Ponty, 1947); existential(ist) humanism (Sartre, 1946; Jaspers 1949); and the many humanisms of the Anglo-American humanist movements. Then there are the reactions to such humanisms: the anti-humanisms of the 20th Century (Althusser, 2018; Foucault, 2001; Lévi-Strauss, 1956, Lacan, 1978), and the various post- and trans-humanisms of the 21st century.

Still, humanism is also a broad term in culture. As such, humanism is not only a term representative of a given historical period, rather it

ISSN 0039-4130

serves as a catalyst for a worldview (*Weltanschauung*) and an *ethos* that implies social and political institutions of a certain form. Humanism is at this point nothing less than an eponym for European civilisation (Tognon, 2019). Humanism is here understood as that generative category which – for better or worse – gave rise to a particular civilisation, i.e. to a culture and its social and political institutions.

Apart from being a historical and cultural term, humanism also carries an axiological meaning; as such, it has performed the role of a regulative ideal. It is no coincidence that at each and every crisis that European civilisation has undergone, the term "humanism" has been evoked as a synthesising term standing for "civilisation" in a time of barbarism. This was the case with Italian Humanism in the aftermath of the crisis of medieval Europe, and afterwards with the various humanisms of the 20th century, in the wake of the two World Wars.

In its historical and historiographical meaning, the term "humanism" has a clear and definite referent, i.e. precise moments in the intellectual history of Europe and the West. This is no longer the case, however, when the term is applied in the broad cultural sense and with its axiological meaning. In these cases, the term "humanism" functions more as a mythical than a logical concept. "Humanism" is not understood here as a descriptive or informative term, but rather as a regulative ideal that aims to establish a space of reciprocal recognition and a just order of relationships (relationships with ourselves, with each other, with the world, and even with what is perhaps beyond the world or present at its foundation).

It is no coincidence that about seventy-five years ago, in the aftermath of the catastrophe of the war, the 1949 edition of the very famous *Rencontres Internationales de Genève* was entitled "Pour un nouvel humanisme" ("Towards a New Humanism"). The conference was attended by leading intellectuals of the time, such as the Swiss theologian Karl Barth and the German philosopher Karl Jaspers.

In his speech – which was given in German and translated into French by Jean Hersch – Jaspers said, "If we are seeking what our humanism might be, it is because we are concerned for ourselves, for the human being of today." (Jaspers 1949, p. 181). We are concerned about the task ahead of us of becoming human, of fulfilling our humanity.

The question of the test that AI and neuroscience imposes upon human freedom can therefore be reconsidered now as a reflection on the role new technologies and new knowledge can play with respect to this task of becoming human. As Jonas stated, the knowledge and tech-

ISSN 0039-4130

nology we produce can support us in our free and creative task of being and becoming human, but they can also hinder us. They are not neutral tools. In the same vein, what Jaspers already observed is still relevant today: "Today, the human being's destiny is played out in technology, and technology can serve either to save them or to destroy them – the die has not yet been cast. We need to seize this destiny and make it our task. To want a future humanism, then, is to agree to toil endlessly to assimilate and master technology – an unlimited field open to human endeavour." (Jaspers 1949, p. 190).

AI can support us in our free and creative task of becoming human, but it can also hinder us. These technologies support us in our calculating capacity, for instance, or in freeing us from burdensome or hazardous tasks. They hinder us, on the other hand, when we delegate our freedom and creativity to them, perhaps in the name of a utopian quest for objectivity or, even worse, in order to free ourselves from the burden of our freedom and the responsibility that goes with it.

In relation to scientific knowledge, Jaspers was also right: "We must first ask ourselves whether human beings are exhausted by the knowledge we have of them, or whether they are freedom, thereby escaping objective knowledge. [...] If the human being appears to me exclusively as a natural being who can be known by objective methods, then I am renouncing all humanism in favour of hominism (Windelband). [...] If, on the other hand, I see human beings in their freedom, their dignity imposes itself fully. Each individual, and I myself, are irreplaceable, and we are all asked to do a great deal". (Jaspers 1949, pp. 184-185).

Any knowledge that claims to be able to say anything that is at the same time objective and exhaustive about the human being runs the risk of distancing us from the task of becoming human.

We need to take the findings of neuroscience extremely seriously, but at the same time we cannot address them as if they were capable of saying the final word about our humanity. There is the bedrock of our humanity that makes each of us irreplaceable and on which our dignity depends. A bedrock that cannot be objectified and yet cannot be eliminated.

This is why new and powerful technologies such as AI and new fields of knowledge such as neuroscience require an ideal regulator, such as that which humanism has offered throughout history and can still offer, in the name of our common humanity and the task it represents for each of us and for us all together. This is why, in the times of AI and neuroscience, Julia Kristeva's invitation seems even more urgent: we must dare humanism (Kristeva 2011).

ISSN 0039-4130

These theoretical philosophical and anthropological insights could have very concrete consequences for political and societal issues. The questions raised on human freedom in our current era, and the answers we can (and must) develop, are closely bound up with our understanding of what it means to live as free individuals in societies whose institutions, and decision-making processes, are profoundly influenced by AI and the developments in neuroscience.

Our idea of political freedom as non-domination, which is one of the core elements of the Rule of Law, emphasises the protection of individuals from arbitrary or excessive control by others, including the State. It ensures that power is exercised in a manner that does not lead to the subjugation or domination of individuals or groups. Non-domination is closely related to the concept of individual autonomy and freedom, ensuring that people are not subject to the whims of those in power, but are governed by laws that are just, predictable, and fairly applied. At the same time, non-arbitrariness requires that the law and its enforcement should be predictable, transparent, and consistent. Non-arbitrariness ensures that decisions and actions by authorities are based on clear, objective criteria rather than on personal discretion, favouritism, or random choice. This principle is essential for maintaining fairness and justice within the legal system, as it guarantees that similar cases are treated similarly, and that the laws are applied impartially and consistently (Pettit 1997). Power is not arbitrary, in other words, to the extent that it is reliably controlled by effective rules, procedures, or goals that are common knowledge to all the persons or groups concerned (Lovett 2002).

As clearly stated by the EU AI Act, the questions of transparency and accountability are crucial when analysing the impact of AI on contemporary political systems and democratic societies. The massive use of AI in the decision-making mechanisms of public authorities and institutions can have a significant impact on power relations, the transparency of these mechanisms, and thus on citizens' rights. Therefore, we might ask what consequences AI technologies will have on decision-making processes and on the arbitrariness of these decisions, especially on transparency, accountability, and publicity. The rules behind AI decisions are obscured within proprietary 'black boxes', making them fundamentally opaque to their users and even to their developers. These rules are not accessible for reading, discussion, analysis, or reasoning, rendering the decision-making process of AI systems inscrutable. Consequently, AI-driven decisions lack transparency, undermining

ISSN 0039-4130

trust and accountability. Without the ability to explain and justify these decisions, it becomes difficult to ensure fairness, address biases, and provide recourse for those affected by AI's actions.

Additionally, replacing human decision-makers with AI systems can arguably impact the dignity of those who are subject to these decisions, as human decision-makers can exercise discretion, and compassion, and apply their moral perspective in nuanced ways. This is not to suggest that human decision-makers are inherently superior to algorithms. The same humans who offer compassion can also introduce errors or biases. Human decision-makers can exhibit racism, both explicitly and implicitly. However, the transition from human decision-making to AI or hybrid human-AI systems fundamentally changes the ability to detect these biases or discriminatory attitudes. The criteria for decisions are shaped by the algorithm designers and are obscured within the algorithm's black box. Consequently, the shift to AI decision-making complicates accountability and the ability to address and rectify biases, ultimately affecting the fairness and justice perceived by those subjected to these decisions. This evolution calls for robust mechanisms to ensure transparency and ethical integrity in AI-driven decision processes.

Finally, the questions of accountability and the very possibility of contesting decisions come to the fore. The Rule of Law not only requires that a decisional system should be fair, consistent, predictable, and rational but also that individuals can contest decisions affecting themselves. The right to challenge human decisions permits individuals to request the reconsideration of an adverse decision. This can involve seeking an internal review, where individuals interact directly with the decision-maker, or an external review, where the case is presented before an impartial third authority, such as tribunals or ombudsmen. This is why both the Council of Europe and the OECD have adopted recommendations setting out guidelines to address algorithmic decisions. The recommendation of the Council of Europe aims at providing "effective means to contest relevant determinations and decisions" by AI systems (Council of Europe 2020), and the OECD specifies that people adversely affected by these decisions should be allowed to challenge the outcome based on "plain and easy-to understand information on the factors and the logic that served as the basis for the prediction, recommendation or decision" (OECD 2019). Unfortunately, the black-box nature of AI considerably weakens the right to contest decisions. AI's decision-making processes are largely inscrutable and thus difficult to contest, and they are likely to become even less transparent the

more numerous the sectors and decision-making processes in which they will be used. AI's decision-makers risk being the source of a power which will remain largely obscure for humans, preventing them from understanding these decisions, and increasing uncertainty.

Again we see all the paradoxical character of the relationship between human freedom and sciences and technology. The texts presented in this issue of STUDIUM - Contemporary Humanism Open Access Annals 2024 all in their own way serve the overarching goal of contributing to the reflection upon human freedom in the age of AI and neuroscience. They are based on contributions to the international conference "Human Freedom at the Test of AI and Neurosciences" organised by LUMSA in Rome (2–5 September 2024) within the framework of the NHNAI project [1] and in collaboration with the international PhD programme network Contemporary Humanism, ATEM (Association de Théologiennes et de Théologiens pour l'étude de la Morale) and the University of Notre Dame Rome.

The exploration begins with a first series of contributions that highlight ethical and societal issues with AI and neuroscience in various domains such as democracy, education, and health, pinpointing the relevance and necessity of a background reflection upon humanism. Mario De Caro shows the extreme practical and political saliency of the philosophical debate on freedom and freewill, at a time when some people are invoking the results of cognitive science and neuroscience as evidence against the existence of free will, and therefore as arguments in favour of anti-retributivist legal systems. Fiorella Battaglia delineates AI's epistemic influence on the functional relationship between democracy and education. In the same vein, Angelo Tumminelli analyses the consequences of generative AI (as infodemics and post-truth enhancer) on democracy and political freedom and on the possibility to develop personal identity and freedom. In the light of the various ethical chal-

[1] NHNAI is a research action project that seeks to strengthen collective capacities for ethical orientation along two structuring and intertwined axes: 1) to contribute to the elaboration of an "ethical compass" by means of reflection on the theme of humanism, of a renewed exploration of what it means to "be human" in the age of AI and neuroscience. 2) to contribute to the development of this compass, not only through academic research but also with an action-research initiative aimed at (and with) concerned societal actors. The ethical and societal issues linked to AI and neuroscience are often described as "wicked problems" (Pohl et al. 2017), as they correspond to scientific problems (including human and social sciences) as well as to political matters. To respond to these kinds of wicked challenges, the contribution of technical and scientific expertise is essential, but not sufficient. A response will never be purely technical, but will also inevitably correspond to a political commitment (possibly implicit), to the adoption (voluntary or not) of a way of living together.

ISSN 0039-4130

lenges triggered by the disruptive impact of AI and neuroscience in health and healthcare, Laura Palazzani calls for an assessment of the theoretical consistency and practical applicability of the philosophical conception of freedom. In her turn, Martina Properzi warns against the danger of a neurocentric bias when discussing the ethical implications of augmentative technology, and argues in favour of a focus upon the embodied subject. Moreover, Helga Martins, Joana Romeiro, and Sílvia Caldeira argue that the use of AI and neuroscience in healthcare should serve a holistic approach that encompasses all the dimensions of the human person, especially the spiritual dimension.

The second series of contributions presents a critical perspective on humanism, discussing various limitations and difficulties with common anthropological and philosophical backgrounds. This critical perspective is paramount, given the ethical and societal relevance of humanism pointed out by the first series of texts. Justin Nnaemeka Onyeukaziri criticises the naturalistic background infusing western anthropologies and philosophies since modernity, a background inadequate to fully acknowledge the specificity of ethical and moral issues (which he claims are irreducible to computational questions). Similarly, Fernand Doridot points out, within philosophical backgrounds underlying contemporary developments in converging technologies, a tendency to drift away from the ambition of understanding humanity toward that of controlling, reproducing, and enhancing it. In addition, Sylvain Lavelle argues that the power humans possess to modify deeply what they are implies that philosophical anthropology must confront normative issues (by contrast with their traditional focus on factual-descriptive approaches).

These various critical discussions highlight the need to explore further the question of humanism and of human freedom. The last series of contributions seeks to contribute to meeting this need by offering various resources and avenues for exploration. Dominique Lambert recalls crucial distinctions when it comes to human intelligence, notably the difference between notional and real knowledge, warning against the danger of an AI-induced reduction of the important matters in our lives to the notional domain only. Sara Fernandes, Leonor Almeida, and Alexandre Castro Caldas draw on Kant, Ricœur, and Taylor to show that the findings of neuroscience concerning the unconscious brain processes that precede conscious intentions do not entail negating free will. Elad Magomedov proposes a Sartrean perspective that binds freedom to the ability to take responsibility for one's choices. Zsolt Almási argues that narrow AI challenges but does not discard human freedom

ISSN 0039-4130

and agency, to the extent that engagement with technology can and should come with a moral imperative to refine one's capabilities and achieve a unique, authentic voice. On the legal side, Yves Poullet proposes a thorough exploration of new facets of the fundamental right to dignity and of the associated new obligations for States. Marco Russo proposes to mobilise Aristotelian virtue ethics and the practice of *phronesis* to bridge the gap between general and abstract norms and values intended to structure AI ethics and concrete situations.

These philosophical, ethical and legal insights are complemented with contributions linked to religious traditions. Thierry Magnin develops the idea that "Christians can be vigilant technophiles", especially by mobilising the key principles of Christian social thought and a "tripartite anthropology of body-soul-spirit". He warns against the risk of sacrificing human uniqueness over the enhancement of some functionalities. Heup Young Kim harnesses insights from East Asian traditions (Daoism, Confucianism, and Theo-Dao) to pinpoint important limitations on transhumanism and posthumanism and to pave the way for reviving the legacy of Modernity and the Enlightenments through an inclusive humanism emphasising Confucian virtues. Buckling the loop, Michael D. Prendergast studies the frequency of the occurrence of various biases within ChatGPT's answers to faith-specific moral and ethical questions, as a function of faiths (Zen Buddhism, Sunni Islam, Catholicism, Orthodox Judaism and secular humanism) and GPT's versions.

References

Biancu, S. (2019a) 'The Human Measure and the (Impossible?) Legacy of Humanism', ETICA & POLITICA / ETHICS & POLITICS, 21, pp. 9-23.

Biancu, S. (2019b) 'L'humanisme : (im)pertinence d'une notion pour l'éthique', Revue d'éthique et de théologie morale, 303, pp. 11-26.

Biancu, S. (2019c) 'Competing Paradigms. A Century of Humanism and homo symbolicus', Munera. Rivista europea di cultura, special issue, pp. 111-127.

Council of Europe, Recommendation CM/Rec (2020) of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems 9,13 (Apr. 8, 2020), https://rm.coe.int/09000016809e1154 [https://perma.cc/2MMJ-WVVC]

Pico Della Mirandola, G. (1958) Oration on the Dignity of Man, Translated by A.R. Caponigri, Introduction by R. Kirk. Chicago: A Gateway Edition – Henry Regnery Company.

Foucault, M. (2001) 'Qui êtes-vous, professeur Foucault?', in: M. Foucault,

ISSN 0039-4130

Dits et écrits (1954-1988), tome I (1954-1975), pp. 601-620. Paris: Gallimard.
Fromm E. (1965) (ed.) Socialist humanism. An international symposium. Garden City: Doubleday.
Jaeger, W. (1934-1947) Paideia. Die Formung des griechischen Menschen. 3 vols. Berlin: Walter de Gruyter & Co.
Jaspers, K. (1949) 'Conditions et possibilités d'un nouvel humanisme', pp. 181-210 in: Pour un nouvel humanisme. Rencontres internationals de Genève (tome IV). Neuchatel: Éditions de la Baconnière.
Jonas, H. (1974) Philosophical Essays: From Ancient Creed to Technological Man. Chicago: University of Chicago Press.
Kristeva, J. (2011) 'Oser l'humanisme'. Revue des deux mondes. Septembre. pp. 79-102.
Lacan, J. (1978) Le Séminaire. II. Paris : Éd du Seuil.
Lévi-Strauss, C. (1996), 'Les trois humanismes' (1956), in Id., Anthropologie structurale II (1973), pp. 319-322. Paris : Plon.
Lovett, F. (2022) 'Domination and democratic legislation'. Politics, Philosophy & Economics. 21.2, pp. 97-121.
Maritain, J. (1984) 'Humanisme intégral. Problèmes temporels et spirituels d'une nouvelle chrétienté' (1936). In: Maritain J. and R., Œuvres complètes, vol VI (1935-1938), Fribourg-Paris: Éditions Universitaires Fribourg Suisse - Éditions Saint Paul.
Merleau-Ponty, M. (1947) Humanisme et terreur: essai sur le problème communiste. Paris : Gallimard.
OECD, Recommendation of the Council on Artificial Intelligence, § 1.3.iv, at 8, OECD Legal Instruments (May 22, 2019), https://egalinstruments.oecd.org/en/instrume nts/OECD-LEGAL-0449 [https://perma.cc/6K-CV-BL2R].
Pettit, P. (1997) Republicanism: A Theory of Freedom and Government. Clarendon Press: Oxford.
Pohl, C., B. et al. (2017) Addressing Wicked Problems through Transdisciplinary Research in: Frodeman R. (ed.), The Oxford Handbook of Interdisciplinarity, 2nd edn, Oxford Handbooks, 319–331.
Pope Francis (2024). Message for the 57th World Day of Peace - Artificial Intelligence and Peace (1 January 2024), https://www.vatican.va/content/francesco/en/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html
Sartre, J.-P. (1946) L'Existentialisme est un humanisme. Paris : Nagel.
Tognon, G. (2019) 'Humanism. Reflections on an Eponymous' Idea, in: Contemporary Humanism – Questioning an Idea: A Time of Fragility, a Time of Opportunity?. Munera. Rivista europea di cultura, special issue.

ISSN 0039-4130

Does Imputability Require Free Will? The Discussion in the Civil Law Tradition

Mario De Caro

Is there an essential conceptual link between fundamental legal notions – such as guilt, imputability, accountability, responsibility, and liability – and the metaphysically charged notion of free will? Philosophically and legally, this is a profoundly significant issue that has sparked extensive discussion in both civil law and common law traditions. While the discussions within these two traditions exhibit distinct nuances, their underlying substance is largely similar. Having addressed this topic in the context of common law elsewhere[1], I will here examine it through the lens of civil law jurisprudence.

In their *Criminal Law. General Part* – a text now regarded as a classic in Italian juridical thought, which is firmly established in the civil law tradition –, Giovanni Fiandaca and Enzo Musco address the assumptions underlying imputability as follows:

> «Everyday practical life rests on the implicitly accepted assumption that each individual possesses sufficient capacity to regulate their conduct in accordance with the expectations of others. Whether this is reality or illusion, each person experiences, as a psychological experience, the feeling of freedom to self-determine in alignment with their choices and desires. It is something similar that modern criminal law is satisfied with. That is, it assumes freedom of will not as an ontological given, but as a necessary presupposition of practical life; not as a scientifically demonstrable fact, but as the content of a legal-social expectation»[2].

[1] M. De Caro, "*Actus non facit reum nisi mens sit rea*": *The concept of guilt in the age of cognitive science*, in A. D'Aloia (ed.), *Neuroscience and Law*, Springer, Dordrecht 2020, pp. 69-79.
[2] G. Fiandaca - E. Musco, *Diritto penale. Parte generale*, Zanichelli, Bologna 2024[9], p. 331.

According to Fiandaca and Musco, the assumption inherent in traditional jurisprudence – that free will, implying the faculty of self-determination [3], serves as the foundation of imputability – must be rejected. They argue that the metaphysical speculations on this matter are too obscure and unusable in legal contexts. Instead, they observe, imputability can be sufficiently grounded in «the feeling of freedom to self-determine in a way that conforms to choices and desires». This pre-philosophical intuition, which suggests that we can, in fact, exercise free will, must be accepted regardless of whether it represents «reality or illusion», as metaphysics cannot conclusively prove its correctness or incorrectness. Thus, on a legal level, we can rely on this firm, common-sense intuition of free will, as Descartes argued in his *Principia Philosophiae*: «That there is freedom in our will [...] must be counted among the first and most common notions innate in us» [4].

Indeed, many traditional conceptions of free will lend legitimacy to Fiandaca and Musco's proposal. Throughout the history of philosophy, free will has been attributed to disembodied minds to which the laws of nature would not apply (Descartes) [5], to noumenal (i.e, outside space-time) agents who act as *causa sui* (Kant) [6], to entities with the prerogatives of a "first unmoved mover" (Chisholm) [7], and so on and so forth with such metaphysical fumistries. Actually, these conceptions are so abstractly metaphysical that they are both irrefutable and unverifiable. Consequently, if one relies on them, it may seem reasonable to conclude that we can never know whether free will is real or illusory. Viewed in this light, it Fiandaca and Musco's "as if" approach – being content with the "feeling of freedom to self-determine" embedded in common sense – may appear justifiable.

[3] According to the mainstream definition in current literature – which, however, has solid roots in traditional debates – free will is defined by two necessary and sufficient conditions: the freedom to do otherwise and the power of self-determination (see T. O'Connor - C. Franklin, *Free Will*, in E. N. Zalta - U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy,* 2022, https://plato.stanford.edu/archives/win2022/entries/freewill/>.

[4] R. Descartes, *Principia Philosophiae* (1644), Eng. transl. by J. Cottingham - R. Stoothoff - D. Murdoch, *The Philosophical Writings of Descartes*, I, Cambridge, University Press, Cambridge 1984-1985, p. 205.

[5] See especially the Fourth Meditation in R. Descartes, *Meditationes de prima philosophia* (1641), Eng. trans. by J. Cottingham - R. Stoothoff - D. Murdoch, *The Philosophical Writings of Descartes*, II, pp. 37-43.

[6] On the possibility of declining the Kantian perspective in perspectives more in keeping with contemporary philosophy, see *The Third Antinomy in the Age of Naturalism*, in L. Corti - J.-G. Schuelein (eds.), *Life, Organisms, and Human Nature*, series *Studies in German Idealism* 22, Springer Nature, Cham 2023, pp. 265-279.

[7] R.M. Chisholm, *Human Freedom and the Self: The Lindley Lecture*, Department of Philosophy, University of Kansas, Lawrence 1964.

ISSN 0039-4130

Put differently, the aforementioned metaphysical conceptions – along with other similarly abstruse ones – being neither demonstrable nor refutable, make it impossible to determine whether the feeling of freedom has a real foundation. In other words, we cannot definitively establish whether human beings (specifically, adults without psychological or cognitive impairments) are truly capable of regulating their «own conduct so as not to disregard the expectations of others». However, if we accept that the ontological foundation of free will will always remain elusive, it may seem reasonable to set aside the philosophical debate and base legal reasoning about accountability on our pre-philosophical intuition. This intuition, as Fiandaca and Musco argue, thus becomes the «necessary presupposition of practical life».

Things, however, are more complex than they may initially appear, for four distinct reasons. Let us consider them one by one. The first reason is historical: during the course of modernity, conceptions of free will have emerged that are significantly less metaphysical and much closer to ordinary common sense than those mentioned above (consider, for example, the Humean tradition, pragmatism, or logical empiricism) [8]. These conceptions, however, were primarily developed within the Anglo-Saxon philosophical context, which, until recently, has had relatively limited influence on the Italian legal sphere. This likely explains why Fiandaca and Musco do not take them into account.

Secondly, it is worth noting that contemporary philosophical debates on free will have taken a distinctly naturalistic turn [9]. The most prominent conceptions – whether they ground free will in a deterministic framework, an indeterministic one, or remain neutral on the issue – are now highly focused on reconciling philosophical perspectives with scientific insights [10]. In particular, these conceptions engage with widely

[8] A historical presentation of the debates on free will since antiquity to nowadays is in the Italian volume M. De Caro - M. Mori - E. Spinelli (a cura di), *Il libero arbitrio. Storia di una controversia filosofica*, Carocci, Roma 2014.

[9] M. De Caro, *How to deal with the free will issue: The roles of conceptual analysis and empirical science*, in M. Marraffa - M. De Caro - F. Ferretti (eds.), *Cartographies of the Mind. Philosophy and Psychology in Intersection* Springer, Dordrecht 2007, pp. 255-268. It is important to note that in the contemporary philosophical discussion, the concept of naturalism does not imply that normative, intentional, agential, moral and phenomenological concepts is not necessarily seen as reducible to the concepts of the natural sciences. See M. De Caro - D. Macarthur (eds.), *Routledge Handbook to Liberal Naturalism*, Routledge, Oxon - New York 2022 and M. De Caro, *The Indispensability of the Manifest Image*, in *Philosophy and Social Criticism*, 45, 2019, pp. 1-11.

[10] A programmatic example of the post-metaphysical trend in the current discussions on free will (and self-determination) is provided by A. Roskies, *Don't Panic: Self-Authorship without Obscure Metaphysics*, in *Philosophical Perspectives*, 26, 1, pp. 323-342.

ISSN 0039-4130

accepted theories from contemporary physics, biology, neurology, social sciences, and psychology. As a result, the more radical metaphysical abstractions have largely receded in contemporary philosophy, and legal discussions of imputability, especially in English-speaking contexts, are increasingly intertwined with philosophical considerations.

Third, another important development in the contemporary discussion of free will deserves consideration. While, as mentioned earlier, many scholars aim to frame this concept in scientifically credible terms, a significant number of others – both scientists and philosophers – strongly contend that free will is an illusion. Consequently, they argue, it cannot serve as the foundation for moral responsibility or criminal imputability.

The roster of contemporary critics of free will is, indeed, quite remarkable. Scientists include such luminaries as psychologists Steven Pinker (Harvard), Joshua Greene (Harvard), Chris Frith (University College, London), and the late Daniel Wegner (Harvard); neuroscientists Wolf Singer (Frankfurt), Michael Gazzaniga (University of California at Santa Barbara), Robert Sapolsky (Stanford), Patrick Haggard (University College, London), and John-Dylan Haynes (Humboldt-Universität, Berlin); biologists Richard Dawkins (Oxford) and Jerry Coyne (Chicago); physicists Lawrence Maxwell (Arizona State University) and the late Stephen Hawking (formerly at Cambridge). Among the philosophers there are Thomas Metzinger (Johannes Gutenberg-Universität Mainz), Derk Pereboom (Cornell), Galen Strawson (Texas at Austin), Saul Smilansky (Haifa), Gregg Caruso (Fairfield), as well as Paul and Patricia Churchland (San Diego) [11].

If these scholars are correct, Fiandaca and Musco's position – which assumes the indemonstrability of both free will and its denial and thus the practical irrelevance of philosophical debate – would no longer be tenable. If we were to be convinced that free will is illusory, the possibility of self-determination would vanish *ipso-facto*. And this would undermine the "as if" notion that we can unproblematically embrace the «feeling of freedom to self-determine in a way that conforms to our choices and desires».

If one were confronted with the impossibility of free will and yet persisted in adopting Fiandaca and Musco's "as if," which assumes an agnostic stance on the issue, one would fall into irrationality. This

[11] A list of publications that support skepticism regarding free will can be found here: https://philpapers.org/browse/free-will-skepticism.

ISSN 0039-4130

attitude would be comparable to arguing that it is both possible and necessary to maintain that the Earth is firmly at the center of the universe, despite overwhelming evidence to the contrary. It is, therefore, no coincidence that today – amid widespread skepticism about free will – radically revisionist legal theories have emerged, particularly in the realm of punishment [12]. These theories, marked by a strongly anti-retributivist stance, reject any reliance on concepts such as responsibility, merit, guilt, and *mens rea*, all of which, at least in their traditional interpretation, presuppose the existence of free will.

Let us summarize what has been said so far. Focusing exclusively on the hypertrophically metaphysical side of the traditional philosophical discussion of free will – which has long dominated cultural thought in Italy and more broadly in the Continental world – Fiandaca and Musco argue that legal discussions should accept our innate feeling of self-determination at face value, avoiding abstract disquisitions that cannot yield practical results. However, in doing so, they overlook another, far more concrete aspect of the philosophical debate on free will, one that is gaining traction as it moves from the English-speaking philosophical world to Continental Europe. In this context, models of free will have emerged that are not purely metaphysical but are instead informed by scientific insights. These models, therefore, lend themselves to practical applications, including in the legal sphere.

However, a further complication arises from the growing support among many scientists and some philosophers for the view that free will and self-determination are illusory. From this perspective, the very foundation of the traditional legal system is profoundly challenged. If this framework is to be properly defended, we can no longer rely on an "as if" approach – that is, the uncritical acceptance of our intuition about self-determination. Instead, it is necessary to confront these denialist views by challenging their underlying assumption: namely, the claim that free will, and the capacity for self-determination it entails, is merely an illusion. In short, the authority of contemporary deniers of free will has created an inescapable task for those who seek to preserve the traditional conceptual framework that upholds the notion of imputability.

[12] D. Pereboom, *Free Will Skepticism in Law and Society. Challenging Retributive Justice*, Cambridge University Press, New York 2019; E. Shaw - D. Pereboom - G. Caruso (eds.), *Free Will Skepticism in Law and Society: Challenging Retributive Justice*, New York 2019; G. Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*, Cambridge University Press, New York 2021.

ISSN 0039-4130

Although this is not the place to delve into a debate of such complexity, a few general points can be made. According to a frequently repeated argument, the illusory nature of free will is demonstrated by the fact that, in areas relevant to human action, science has proven the correctness of determinism (i.e., the thesis that all events, including human actions, are determined by sufficient causes according to the laws of nature) [13]. In this context, the deterministic thesis is articulated, as appropriate, with reference to physics, genetics, neuroscience, or even social psychology [14]. The truth of determinism, it is argued, would *ipso facto* prove that free will is nothing more than an illusion.

Let us see some examples of this attitude. Stephen Hawking, for example, wrote:

> The ultimate objective test of free will would seem to be: Can one predict the behavior of the organism? If one can, then it clearly doesn't have free will but is predetermined» [15].

Similarly, Lawrence Krauss writes that «As a physicist, I don't think there's free will [...]. At some level, the universe is deterministic» [16]. Sam Harris in turn states:

> «Free will *is* an illusion. Our wills are simply not of our own making. Thoughts and intentions emerge from background causes which we are unaware and over which we exert no conscious control. We do not have the freedom we think we have» [17].

[13] Various forms of determinism have been discussed throughout the history of philosophy and science, including *theological* determinism (God necessitates all events) and *logical* determinism (the notion that the truth value of statements about future events is determined *ab aeterno*). Here, however, I refer to *nomological causal* determinism, which is central to the current debate on free will: the idea that all events have a sufficient cause in line with the laws of nature.

[14] In this context, the indeterminism that most interpretations attribute to quantum mechanics is deemed irrelevant to the generation of human actions (see below). Other interpretations, such as the Bohm interpretation and Everett's "Many-Worlds Interpretation", however, conceive of quantum mechanics in deterministic terms: see L. Vaidman, *Quantum theory and determinism*, in *Quantum Studies: Mathematics and Foundations*, 1, 2014, pp. 5-38; see also S. Hossenfelder - T. Palmer, *Rethinking Superdeterminism*, in *Frontiers in Physics*, 8, 2020, art. 139, https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2020.00139/full.

[15] S. Hawking, *Black Holes and Baby Universes and Other Essays*, Random House, New York 1993, p. 133.

[16] Cited in S. Seckel, *Enlightenment and Reason Make the World a Better Place*, 10/20/2015, https://news.asu.edu/20151020-solutions-enlightenment-and-reason-make-world-better-place.

[17] S. Harris, *Free Will*, Free Press, New York 2012, p. 5.

ISSN 0039-4130

By a similar line of reasoning, Michael Gazzaniga goes so far as to argue that the idea of free will is as misguided as the flat Earth theory, even though it is much more widespread. The time will come, however, Gazzaniga adds, when we will finally be able to abandon this discredited notion and accept the fact that we are nothing more than a special kind of machine – a deterministic machine, in fact. Robert Sapolsky, as a biologist, frames the issue within the terms of his discipline. In his view, we are biologically determined in everything we do:

> «Studies show that if you're sitting in a room with a terrible smell, people become more socially conservative. Some of that has to do with genetics: What's the makeup of their olfactory receptors? With childhood: What conditioning did they have to particular smells?» [18].

And so, through scientific example after scientific example, Sapolsky concludes that we are always determined and, because of this, must abandon the idea of free will. However, far from regretting this conclusion, Sapolsky sees it as a very good thing because, in his opinion, «a large part of humanity's misery is due to the myth of free will». The evils of humanity, he argues, also include conceptions that, directly or indirectly, trace back to the conceptual framework of free will, including all forms of retributivism (a claim that directly contradicts Fiandaca and Musco's point of view).

Paul Bloom, finally, adds support to this kind of argument. In his view, probably we are completely determined in all our behaviors, and consequently free will is impossible. However, even if some relevant effect of quantum indeterminism were to manifest at the macroscopic level, this would not support the case for free will in any way:

> «Most scientists and philosophers agree that [free will] is an illusion. Our actions are in fact literally predestined, determined by the laws of physics, the state of the universe, long before we were born, and, perhaps, by random events at the quantum level. We chose none of this, and so free will does not exist» [19].

[18] An olfactory approach to free will, one would say: Robert Sapolsky, interviewed by Hope Reese, *New York Times*, 10/16/2023, https://www.nytimes.com/2023/10/16/science/free-will-sapolsky.html.

[19] P. Bloom, *Free Will Does Not Exist. So What?*, in *The Chronicle of Higher Education*, 18.3.2021, https://www.chronicle.com/article/free-will-does-not-exist-so-what/.

ISSN 0039-4130

This kind of argument, although often presented as a truism, is far less solid than commonly thought, and there are several reasons for this. Let us begin with the reference to indeterminism in the Bloom quote just cited, which also undermines theories that root free will in pure indeterminism. Patricia Churchland – who, like others, returns to the analogy with the flat Earth theory (evidently a favorite among critics of free will) – is also in this line of criticism.

> «A rigid philosophical tradition claims that no choice is free unless it is uncaused; that is, unless the "will" is exercised independently of all causal influences – in a causal vacuum. A philosophy dedicated to uncaused choice is as unrealistic as a philosophy dedicated to a flat Earth» [20].

The idea, in short, is that freedom cannot arise from mere indeterminism. This observation has been repeated for a long time: for example, developing an argument already hinted at by Hobbes in his polemic against Archbishop John Bramhall [21], Hume argued that the only thing that can arise from indeterminacy is chance – and chance is the opposite of freedom. The key point, however, is that today no serious defender of free will supports this thesis, much to the chagrin of Bloom and Churchland. It is true that some authors argue that indeterminism is a necessary condition for free will (the so-called "libertarians"), but no one claims it is a sufficient condition. Some (such as Robert Kane [22] and Mark Balaguer [23]) believe the crucial concept for free will is not pure indeterminism but probabilistic causation; others (such as Carl Ginet [24]), drawing on insight from Wittgenstein, insist that free will should be interpreted as the ability of rational agents to act rather than as a causal ability; still others (such as Tim O'Connor [25]) develop complex theories of agent

[20] P. Churchland, *The Big Questions: Do we have free will?*, in *New Scientist*, 11/15/2006, 2578, https://www.newscientist.com/article/mg19225780-070-the-big-questions-do-we-have-free-will/.

[21] *Hobbes and Bramhall on Liberty and Necessity*, ed. by V. Chappell, Cambridge University Press, Cambridge 1999.

[22] R. Kane, *The Significance of Free Will*, Oxford University Press, New York 1996.

[23] M. Balaguer, *Free Will as an Open Scientific Problem*, MIT Press, Cambridge (MA) 2010.

[24] C. Ginet, *On Action*, Cambridge University Press, Cambridge 1990.

[25] T. O'Connor, *Persons and Causes: The Metaphysics of Free Will*, Oxford University Press, New York 2000; Id., *Agent–causal power*, in T. Handfield (Ed.), *Dispositions and Causes*, New York 2009, pp. 189-214; E. J. Lowe, *Personal Agency: The Metaphysics of Mind and Action*, Oxford University Press, Oxford 2008.

ISSN 0039-4130

causation that, building on Thomas Reid's hypothesis, attribute to human beings a causal capacity beyond the normal causation between events – namely, the ability that one has to intentionally initiate new causal chains, being causally inclined by one's past without being determined by it. Finally, Roger Penrose has developed a detailed neurophysiological model to show how quantum indeterminism could potentially produce free will (and it would be bold to claim that Penrose does not understand science, given that he won the Nobel Prize in Physics in 2020) [26]. In short, Bloom and Churchland's argument attacks a strawman, since no serious thinker today defends the crude conception they so vehemently criticize [27].

Second point. Another assumption underlying scientific theories that criticize free will is that if determinism were proven true, the debate would be resolved, as free will is considered *obviously* incompatible with determinism [28]. A prominent proponent of this view is Robert Sapolsky, whose book *Determined: A Science of Life Without Free Will* has been a long-time bestseller on the *New York Times* list [29]. In this work, supported by a wealth of scientific examples, Sapolsky argues that the various sciences of the macroscopic world converge to demonstrate that human actions are deterministic in nature, consequently

[26] R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*, Oxford 1999.

[27] This, of course, does not mean that conceptions that base free will on quantum-mechanics indeterminism cannot be criticized: see, for example, M. De Caro and H. Putnam, *Free Will and Quantum Mechanics*, in *The Monist*, CIII, 4, 2020, pp. 415-426.

[28] In this context, the arguments of those who draw on experiments inspired by Benjamin Libet – such as the well-known study by C.S. Soon, M. Brass, H.J. Heinze, and J.D. Haynes, 'Unconscious Determinants of Free Decisions in the Human Brain' (Nature Neuroscience, 11, 2008, 543-545) – are often cited to assert the illusory nature of free will. Other authors, also skeptical of free will, draw on experiments primarily conducted in the field of social psychology, which purport to demonstrate the causal impotence of the conscious mind. (see D. Wegner, *The Illusion of Conscious Will*, NIT Press, Cambridge (MA) 2002; L. Hall, T. Strandberg, P. Pärnamets, A. Lind, B. Tärning, P. Johansson, *How the Polls Can Be Both Spot on and Dead Wrong: Using Choice Blindness to Shift Political Attitudes and Voter Intentions*, in *Plos One*, VIII, 4, 2013, https://doi.org/10.1371/journal.pone.0060554). Criticisms of these interpretations are offered in M. De Caro, *Is Emergentism Refuted by the Neurosciences?*, in A. Corradini, Tim O'Connor (eds.), *Emergence in Science and Philosophy*, Routledge, London 2010, pp. 190-211; A. Mele, *Free: Why Science Hasn't Disproved Free Will*, Oxford University Press, Oxford 2014; M. De Caro and S. Bonicalzi, *How the Libet Tradition Can Contribute to Understanding Human Action Rather Free Will* in C.J. Austin, A. Marmodoro, A. Roselli (eds.), *Powers, Time and Free Will*, Springer Nature, Cham 2022, pp. 199-22.

[29] R. Sapolsky, *Determined: A Science of Life without Free Will*, Penguin Books, New York 2023.

ISSN 0039-4130

leaving no room for free will[30]. This is not the place for an in-depth analysis of Sapolsky's book, as such an examination would require many pages. However, interested readers may find it useful to explore the critical review of the book by John Martin Fischer, a leading figure in the contemporary philosophical debate on free will. Fischer, in particular, argues that Sapolsky fails to provide a valid rebuttal to the majority philosophical view that determinism and free will are not necessarily incompatible[31].

However, a brief mention may be made here of another commercially successful pamphlet: *Free Will* by Sam Harris, which prominently features an image of a puppet on its cover. The idea behind this image is that, like puppets, human beings lack free will because they are determined by factors beyond their control – that is, the forces of nature, which act as the analogous of the puppet master. In this light, Harris writes that «we are all moved by chance and necessity, just as a puppet is made to dance on its strings[32]». This is a version of the thesis presented by Sapolsky and many others: we do not have free will because we are nothing more than cogs in the mechanism of nature. We might consider it the main thesis offered by the deniers of free will. However, the specific way in which Harris argues for this view is particularly relevant because it can be directly applied to Fiandaca and Musco's thesis.

For the past few decades, the classical view – defended over time by Hobbes, Locke, Leibniz, Hume, Mill, Ayer, P.F. Strawson, Davidson and Dennett – that free will can coexist with determinism has been known as "compatibilism". According to Harris, however, this view is «deliberately obtuse», as he believes it amounts to the (obviously absurd) idea that «a puppet is free insofar as it loves the strings that

[30] In the case of quantum mechanics, which deals with the subatomic world, things are notably different, as most interpretations suggest that the theory should be understood indeterministically: see W. Myrwold, *Philosophical Issues in Quantum Theory*, in. E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, 2022, https://plato.stanford.edu/entries/qt-issues/. But see above for some references to views that conceive of quantum mechanics in deterministic terms.

[31] J.M. Fischer, *Review of* Determined: A Science of Life Without Free Will, in *Notre Dame Philosophical Reviews* 2023, https://ndpr.nd.edu/reviews/determined-a-science-of-life-without-free-will/.

[32] S. Harris, *The Marionette's Lament. A Response to Daniel Dennett*, 12/2/2014, https://www.samharris.org/blog/the-marionettes-lament

ISSN 0039-4130

move it»[33]. This, however, is a rhetorical argument: it is psychological in nature and grounded in a rather unlikely form of armchair psychology. The issue at stake, instead, is not psychological; it concerns a conceptual analysis regarding two ontological notions, respectively about how we and the world are made.

In his book, however, Harris also offers another criticism of compatibilism, namely that this view conflicts with the common sense intuition about free will, which, according to Harris presupposes indeterminism[34]. To this, Dennett responded that even if one were to concede the correctness of this claim (which is far from obvious), it would not pose a problem for philosophical analysis[35]. After all, many of the most significant philosophical views are at least partially, if not entirely, revisionist. Consider, for instance, the Democritean-Galilean-Lockean perspective that secondary qualities do not exist independently in the world, Hume's ideas on personal identity, Kantian theses on the structure of reality, or, more recently, Putnam's and Kripke's semantic externalism.

In fact, to genuinely refute compatibilism, one would need to examine the core theses advanced by its defenders with far greater care and rigor. These theses include the following: for an action to be free, it is necessary that some conscious and relevant intentional states of the agent constitute part of the sufficient cause that generates the action; that causal chains of this nature are entirely compatible with causal determinism; and that the ability to act otherwise than one actually does must be understood in a conditional, rather than categorical, sense

[33] Ibid., p. 30. The view that Harris boldly attributes to compatibilism in general might be more convincingly ascribed to the Stoics – particularly Epictetus and Marcus Aurelius – as well as to Spinoza, and Nietzsche, all of whom, albeit with different nuances, adhered to the idea of "*amor fati*". This idea, however, was never conceived as a defense of free will (at least in the traditional sense of the term) since it suggests that to achieve self-control, we must relinquish it by accepting our destiny. See A. T. Kronman, *Amor Fati (The Love of Fate)*, in *The University of Toronto Law Journal*, XLV, 2, 1995, pp. 163-178; Han-Pile, *Nietzsche and Amor Fati*, in *European Journal of Philosophy*, XIX, 2, 2011, pp. 224-261.

[34] Unknown to Harris, studies in experimental philosophy have not reached a consensus view about whether the common sense view of free will tends to libertarianism (that is, links it to indeterminism) or to compatibilism (that is, sees free will as compatible with determinism): see S. Nichols - J. Knobe, *Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions*, in *Noûs*, XLI, 4, 2007, pp. 663-685; A. Roskies, *Neuroscientific challenges to free will and responsibility*, in *Trends in Cognitive Science*, X, 9, 2006, pp. 419-423; F. Cova, *A Defense of Natural Compatibilism*, in J. Campbell, K. M. Mickelson and V. A. White (eds.), *Blackwell Companion to Free Will*, Blackwell, Malden (MA), forthcoming (penultimate draft at: https://osf.io/preprints/psyarxiv/4bkwe).

[35] D. Dennett, *Reflections on* Free Will, 26.1.2014, https://www.samharris.org/blog/reflections-on-free-will.

ISSN 0039-4130

(e.g., «If I had wanted to order pasta instead of risotto, I could have done so»)[36]. In essence, compatibilism undertakes a rigorous conceptual analysis to clarify what is genuinely «worth wanting» in the notion of free will (to borrow Dennett's famous phrase)[37].

As is often the case with the staunch advocates of the anti-free will stance, Harris dismisses the legitimacy of such conceptual analyses. In this vein, addressing Dennett directly, he writes:

> «You believe that, along with the other compatibilists, you have purified the concept of free will by replacing the common notions by which it is conceived by common sense with rigorous, demystified concepts. In my opinion, however, in doing so you have simply changed the subject, ignoring the very phenomenon we should be talking about: the common and perceived feeling that I/we/you could have done otherwise (an idea generally known as "libertarian" or "countercausal" free will), with all its moral implications»[38].

Harris's point, in essence, is that the discussion should center on the common intuition of free will rather than the philosophical view of compatibilism, which makes that intuition rigorous. In making this argument, however, Harris merely reiterates the recurring theme we noted earlier – a hallmark of the imprecise critiques that several scientists, even prominent ones, have frequently directed at free will. Common sense intuitions are often far removed from the realities of the world; indeed, science progresses by correcting these intuitions when necessary. Why should philosophy be any different? The real question, then, is not whether the raw, intuitive conception of free will is accurate (it clearly is not), but whether a rigorous version of it can be[39].

And this is the fourth reason why Fiandaca and Musco's position is unsatisfactory. Today, one can no longer rely solely on the pre-philosophical intuition of self-determination (i.e., «the feeling of freedom to self-determine in a way that aligns with one's choices and desires»),

[36] For a discussion of the interpretation of the possibility of doing otherwise, see C. List, *Free Will, Determinism, and the Possibility of Doing Otherwi*se, in *Noûs* 2014, XLVIII, 1, 156-178.

[37] D. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting*, MIT Press, Cambridge (MA) 1984.

[38] S. Harris, *The Marionette's Lament*, cit.

[39] For a more extended criticism of Harris's view, in defense of Dennett's compatibilism, see M. De Caro, *In defense of avuncularity. Dennett and Harris on the relation between philosophy and science*, in *Rivista internazionale di filosofia e psicologia*, VIII, 3, 2017, pp. 266-273.

ISSN 0039-4130

as it has been demonstrated that such intuition is inadequate for the task assigned to it – namely, justifying the fundamental legal notion of imputability. Instead, we must turn to the most rigorous conceptions of free will currently available, which serve to refine and formalize that intuition.

Note: The thesis that free will is compatible with determinism may indeed be mistaken. However, much stronger arguments than those just presented are required to challenge it –

assuming we can even call them "arguments"[40]. Nor can the idea of free will be accepted as self-evident, as arguments have been raised against this thesis that is not plainly erroneous[41]. These arguments, however, are contentious and therefore not decisive, as is always the case with genuinely philosophical debates.

What is clear, in short, is that anyone seeking to defend the concept of imputability today – at least in a sense akin to how it has traditionally been understood in jurisprudence – can no longer do so by complacently ignoring debates about free will and relying solely on the vague feeling of self-determination.

[40] Much discussed in this context is the so-called "Consequence argument" (P. van Inwagen, *An Essay on Free Will*, Oxford University Press, New York 1983, 126-152), which asserts that, in order to freely perform a given action, an agent must control that action; however, for this to be the case, the agent would need to control at least one of the two factors that, if determinism is true, necessitate that action: namely, the laws of nature and the events of the remote past that are links of the causal chain that led to the performance of that action. But both factors are beyond the agent's control, as the past is unalterable, and the laws of nature are inescapable. Therefore, if determinism is true, no one can control the actions they take. As a result, free will is impossible and compatibilism is false. The discussion on this topic is vast and has not yielded definitive conclusions: see D. Speak, *The Consequence Argument Revisited* and T. Kapitan, *A Compatibilist Reply to the Consequence Argument*, in R. Kane (ed.), *The Oxford Handbook of Free Will*, cit., pp. 115-130 and 131-152.

[41] Philosophers who defend so-called "illusionism," the view that free will is conceptually impossible, develop sophisticated analyses to demonstrate the incompatibility of free will with both determinism and indeterminism (and note that, between determinism and indeterminism, *tertium non datur*). Among these philosophers are S. Smilansky, *Free Will and Illusion*, Oxford 2000; G. Strawson, *The Bounds of Freedom*, in R. Kane (ed.), *The Oxford Handbook of Philosophy*, cit., pp. 441-460; D. Pereboom, *Living without Free Will*, Cambridge 2017; G. Caruso, *Exploring the Illusion of Free Will and Moral Responsibility*, Lanham (MD) 2015.

ISSN 0039-4130

Democracy and Education at the Time of AI

Fiorella Battaglia

1. *Introduction*

The entanglement of democracy and education is central to Dewey's thinking. He identifies two elements that are constitutive of education and point to democracy. These two elements go beyond simply stating a principle. The first element, the criterion of pluralism, is not exhausted by the diversity of points of view, but rather, also implies a relational perspective in which it is important to recognise each other's interests. The second element has less to do with freedom than with the necessary evolution of conditions that require constant readjustment. With Dewey's pluralism and freedom in mind, I will frame the focus on democracy and education within a broader contemporary trend, according to which philosophy has shifted its focus towards investigating wrongdoings, particularly the occurrence of threats, rather than pursuing ideal theory. The reason for this shift in focus on specific harms is the greater ability of such theories to guide action in real-life circumstances. From a methodological point of view, I am not going to defend this new trend, because it would require quite a lengthy meta-analysis, and shifts the focus away from the initial concern of the paper - AI's effect on democracy and education - to methodological questions in the field in general. To this aim, the paper focuses on the question of the nature of knowledge: how does knowledge conveyed by AI differ from human knowledge? The epistemic dynamics activated by social media and predictive and generative AI lead to misinformation. According to Dewey's perspective, this has a negative impact on both democracy and education. Toward the end of the paper, I will also briefly discuss some more theoretical reasons for believing that there is an intimate relationship between democracy and education which is emphasized by the risks faced in embracing digitalization.

2. *Democratic knowledge*

Knowledge produced by universities is beneficial not only for science but also for innovation, citizens and society. This holds particularly true for the results in social sciences, ethics and political philosophy. Indeed, research in these fields has a wide-reaching societal impact, so assessing its value is something that should accompany research from the very beginning.

Futures that incorporate AI will involve a changed process of knowledge production. Already today we can observe a paradigm shift in the production of knowledge on social media. The shift goes this way: from top-down or one-to-many issuers to many-to-many; all placed on an equal footing potentially having equal dignity and credibility. The communication model has not always been this way (Shannon and Warren, 1949). From the advent of mass media in the 19th century through to the 20th century columnists voiced a certain view. Often this first expert was echoed by a different columnist. In the end, by being exposed to the various debates, it would be possible to form one's own opinion. In the terminology introduced in the theory of communication, constituents of communication occur either "as sender" or "as receiver". The message is a concept, information, communication or statement sent to the recipient in verbal, written, recorded or visual form (Biesta, 2006). The communication framework, designed by Shannon and Warren, has been subjected to one critical perspective which is explicitly oriented toward articulating and questioning messages which are judged to be untrue, dishonest, or unjust. Critical theory questions epistemic conditions of the production and communication of knowledge (Habermas, 2006). Truth matters for individual and social communication and, therefore, also for democracy. Nowadays, the environment in which we communicate has changed. We are all authors placed on the same level and potentially have equal dignity and credibility. This transformation, of which the online Encyclopedia, Wikipedia, is perhaps the most fundamental example, is considered the achievement of democracy. However, this view does not take into consideration that in this process certain desirable characteristics of the media landscape will die out. The character of the classic opinion leader, the journalist, and the expert, collapses, because we are all capable of becoming broadcasters. We can all produce and share content, and distribute our vision of reality "worldwide", without filters and control. Or at least that is how it seems. In all this, the importance of

ISSN 0039-4130

the quality of information, on which the vision of a given phenomenon is built, risks being severely compromised. This is particularly relevant to the truth requirement, which, in turn, is important for political communication and democracy. But for the new scenario dominated by the post-truth, truth seems to have a despotic character while "post-truth" (Oxford Dictionaries Word of the year 2016) relates to or denotes circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal beliefs.

3. *The truth in political theory*

Scholarship on truth, knowledge and democratic authority tends to regard epistemic values, which are independent of procedures, as being of a tyrannical character. Hannah Arendt makes this point very clear: since debate is the essence of political life, starting a debate with true claims does not facilitate debate, but rather makes the exchange of views impossible (1967; see also Yack 2012 and Urbinati 2012). On Hannah Arendt's account, truth «precludes debate, and debate constitutes the very essence of political life» (Arendt 1967, 114). As Michael Walzer puts it, in the democratic world «truth is indeed another opinion» (1981). In a completely changed scenario, dominated by AI, truth has acquired a new profile. Philosophers are ditching procedural standards for substantive ones.

Classic debates, which have recently resumed as alternative facts emerge in the context of digitalization, revolve around the question of whether our political outcomes should be either true or popular. Today, populism is transforming democracy and technology, especially predictive and generative AI, is facilitating the creation of misinformation and distortions in communication. In this different scenario the debate has evolved into a general debate about post-truth. This shows that the dichotomy between justice and truth, which has dominated the previous debate, is not sufficient to address the new challenges to democracy. As a result, we have several effects. First, the quality of information has been compromised. Misinformation comes from the attitude to accept someone's opinion even if they are not experts and do not show reasons for backing their idea. Their authority would need some basis. In the case of communication on social media and predictive and generative knowledge this authority comes from the consent of the recipients and not from the author's expertise or reasons.

ISSN 0039-4130

4. *AI in political theory*

What is more, AI is influencing political theorizing. Worries about claims that peremptorily demand to be acknowledged are giving way to users focused on contributing to knowledge and truth production, on having access to better data and their processing, and on sharing decisions and potential better benefits. The novel AI-induced process of knowledge production and communication will also have changed attitudes towards what was considered quite normal before the emergence of new practices and activities of citizens, workers and consumers - from individual to collective actors, including organisations - with generative AI technologies and applications in the social media scenario. Anxiety about truth has been replaced by the anxiety about substantive political arrangements, especially regarding the public sphere.

Some empirical research suggests that social media may encourage our problematic confirmation bias (Bessi, Coletto, Davidescu, Scala, Caldarelli, & Quattrociocchi 2015). This naturally comforts us, reassures us, and makes us safe from having to shake our certainties, as well as questions our points of view and categories of reference. It gives us the false perception of control, of knowing who we are and where we stand, what we know and to which community we belong. In such a new context, pluralism does not do anything. The first element in the democracy-education relationship i.e. the reliance upon the recognition of mutual interests as a factor in social control does not play a role. In general, this dynamic concerns the very processes of aggregation of human beings, and therefore we certainly cannot consider it as an exclusive characteristic of social networks. Predictive and generative AI may reinforce flawed ideas or biases by providing only the views or representations a user expects, reinforcing cultural biases, including prejudices and stereotypes. On social networks populated by chatbots, this mechanism is practically automatic. Echo chambers are created where we can do what we like best, meeting people who have the same interests and share the same narratives as us. It is precisely this mechanism that powerfully underlines the amplification and dissemination of even false information and misinformation online, which, once assumed to be credible, is rarely refuted or reassessed. In the same way, once misinformation has gone viral, it will hardly be possible to spread its correction as widely. The result is that everyone is talking to themselves. It is difficult at this point to get in touch with realities other than those we already know and feel close to, not to mention the possibility of adapt-

ISSN 0039-4130

ing reality to the new conditions and practices that are emerging in a democratic society. The second element characterising both education and democracy, which Dewey terms as freedom, will be lost. The need for the structures of communication and the narratives of society to evolve together in such a way as to make possible the re-adaptation of society to the new demands that are emerging cannot be satisfied. As such, it points to changes in social habits that may be seen as dangerous (Rawls, 1971). On the other hand, such evolution may be thought of as "imagination" (Carens, 1987). For instance, if I regret some past action, it is plausible to say that I wish I could change the past. It is a common type of attitude, opposed by the rigid structure of communication, with its closures that prevent the boldest imaginations from coming into contact, spreading misinformation and misunderstanding.

5. *Conclusion*

It would be superficial to underestimate the impact and importance of AI on the fundamental structures of knowledge that determine the relationship between education and democracy, especially with its initial effects on public communication through social media. If these trends continue, they have the capability to challenge democracy. In such a case, ethical and social oversight will need to be even more rigorous and meticulous than it has been to date. As highlighted in the analysis so far, intervention in the production and transmission of knowledge raises a number of sensitive issues, all of which need to be handled. For all these reasons, it is crucial that citizens, educational institutions and the scientific community remain vigilant in their criticism of a process that carries potential risks and opens debates which have not been adequately addressed such as that on democracy and truth. Dealing with misinformation turns out to be more effective in guiding action in real-life circumstances.

References

H. Arendt, *Truth and Politics. Philosophy, Politics and Society*, Blackwell, Oxford 1967.

A. Bessi – M. Coletto - G. A. Davidescu - A. Scala - G. Caldarelli - W. Quattrociocchi, *Science vs conspiracy: Collective narratives in the age of misinformation*. *PloS one*, X, 2, 2015, e0118093.

ISSN 0039-4130

J. J. Gert Biesta, *"Of all affairs, communication is the most wonderful." Education as communicative praxis*, in D.T. Hansen (ed.), *John Dewey and our educational prospect. A critical engagement with Dewey's Democracy and Education*, NY: SUNY Press, Albany 2006, pp. 23-37.

J. Carens, *Aliens and Citizens: The Case for Open Borders*, *The Review of Politics,* XLIX, 2*,* 1987, pp. 251-273.

J. Dewey, *Democracy and Education: An Introduction to the Philosophy of Education*, Macmillan, New York 1916, Retrieved 3 February 2025 via Internet Archive.

D. Estlund, *The Truth in Political Liberalism*, in J. Elkins - A. Norris (eds), *Truth and Democracy*, University of Pennsylvania Press, 2012, pp. 251-271.

J. Habermas, *Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research*, Communication Theory*,* XVI, 4, 2006, pp. 411-426. doi:10.1111/j.1468-2885.2006.00280.x

J. Rawls, *A Theory of Justice*, MA: Harvard University Press, Cambridge 1971.

C. E. Shannon - W. Weaver, *The Mathematical Theory of Communication*, IL: University of Illinois Press, Urbana 1949.

N. Urbinati, *Democratic politics and the lovers of truth.* In Elkins, J. and Norris, A. (eds.). *Truth and Democracy*, University of Pennsylvania Press: Philadelphia, Philadelphia 2012.

M. Walzer, *Philosophy and Democracy*, in *Political Theory*, IX, 3, pp. 379-399, , 1981. http://www.jstor.org/stable/191096

B. Yack, *Democracy and the Love of Truth*, in J. Elkins - A. Norris (eds.), *Truth and Democracy*, University of Pennsylvania Press: Philadelphia, Philadelphia 2012, pp. 165-80.

ISSN 0039-4130

AI and Democratic Citizenship

The Consequences of Infodemics for Human Freedom [1]

Angelo Tumminelli

Introduction

The process of identifying the risks of infodemics serves to define some strategic elements in understanding the relationship between human freedom and artificial intelligence: first, it is necessary to identify the political actors involved in the dissemination and circulation of GANs (Generative Adversarial Networks) by highlighting their interests in controlling public opinion and the ideological orientation of users. Furthermore, the paper aims to present the impact of infodemics on the democratic process and the expression of political freedoms [2].

Indeed, the dissemination of deepfakes is not only capable of influencing the exercise of democracy in individual states by conveying ideologised thinking and directing citizens' political consciences, but it can even influence international dynamics by producing conflicts and fuelling the polarisation of political viewpoints [3]. Thus, the paper also

[1] This research was funded by the European Union HORIZON-RIA SOLARIS project, grant number 101094665. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

[2] Cf. B. Fallery, *Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques*, in *HAL Open Science*, 2021 [Preprint]. Available at: https://hal.science/hal-03126059/document; L. Chiappini, *The geopolitical effect of Artificial Intelligence: The implications on Human Rights and Democracy*, Mondo internazionale APS - ETS, Pisa 2022; N. Miailhe, *Géopolitique de l'Intelligence artificielle : le retour des empires*, in *Politique étrangère*, in Automne (3), 2018, pp. 105–117. Available at: https://doi.org/10.3917/pe.183.0105.

[3] Cf. T. Weikmann - S. Lecheler, *Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network* in *Digital Journalism*, 2023, pp. 1–18. Available at: https://doi.org/10.1080/21670811.2023.2194665.

aims to highlight the need for a responsible use of GANs technologies in order to exercise democracy freely and free of ideological conditioning.

Through the use of an interdisciplinary methodology and transdisciplinary approaches, the paper is divided into three sections: the first deals with the relationship between artificial intelligence and the democratic process, the second investigates the issue of digital colonisation and, finally, the last paragraph deals with the issue of personal identity related to the use of algorithmic systems.

As is well known, the influence of AI on democracy is directly proportional to the protection/violation of certain human rights[4]. Freedom of thought is one of the main rights of a democracy: people must be able to think freely without being punished for it. This condition creates pluralism, which is a pillar of a democratic society. Artificial intelligence systems have the power to stimulate man's creative thoughts, presenting concepts that some may not have considered. However, they are also able to show only the content a person wants by recording their previous online behaviour, encouraging confirmation bias instead of facilitating critical thinking. Thinking critically about our surroundings is essential for pluralistic views and inclusive debates. Artificial intelligence can even create fake and realistic videos, audio and images that can challenge decision-making and be used as propaganda to influence public opinion and manipulate elections.

On the other hand, AI can be a useful nudge to direct human beings towards responsible choices if it is programmed to structure a good choice architecture that allows governments to protect the freedom of citizens by encouraging them to make wiser decisions. In this sense, the aim of the paper is also to define ethical strategies to curb negative consequences and promote, instead, the exercise of a free and responsible political democracy in which the contribution of individual citizens serves to the common good and the realisation of universal peace.

1. *Artificial intelligence and the democratic process*

One of the pillars of democracy is that governments should be chosen by those they will serve. But for a voter to make an informed decision, he or she must first be informed. The information space has

[4] Cf. E. Aizenberg - J., Van Den Hoven, *Designing for human rights in AI*, in *Big Data & Society*, 7(2), 2020, p. 205395172094956. Available at: https://doi.org/10.1177/2053951720949566.

ISSN 0039-4130

changed considerably in recent years due to the impact of artificial intelligence. These changes have influenced political communities [5] that have become more heterogeneous due to the development of social media and the use of AI in them.

This means that for every voter in a political community, there is a distinct opinion on the same issue. Especially in countries with a tradition of several political parties, this tendency can create a more fragmented political community, where no one party or group of parties is able to establish governing power and thus create a paralysed political community.

Numerous information sources, combined with the potential of AI to create indistinguishable false content and infodemics, can create the conditions for heterogeneity of opinions on political issues to prevail and hinder the creation of socio-political groups with similar ideas and opinions within a political community. In this way, the political community fragments into atomised individuals who do not have many relationships with each other in terms of political positions [6]. This phenomenon may be more pronounced in Western democracies, where religious and ethnic ties within the population have dissolved, whereas in countries where religious and ethnic ties are still very strong and decisive for collective identity, heterogeneity of information cannot dissolve cultural homogeneity.

Political communities in democracies have also become less informed. The fact that there is an abundance of information that everyone can access initially meant that voters would become more informed. However, with the growth of social media and the use of AI to generate videos, photos and texts, the opposite effect has occurred. AI, so far, has contributed to making information unreliable and not credible. Voters in a political community are no longer sure which information is true and which is not. Rather than prompting greater engagement in political processes, AI may contribute to making people less inclined to engage in them, because there is a lack of trust in the information they receive, and if the voter is unsure of the information they receive, then they are unsure of participating in a political community. Generative AI, with its power to generate deepfake images and videos, can put

[5] Cf. A. Alesina, - E. La Ferrara, *Participation in Heterogeneous Communities*, in *Quarterly Journal of Economics*, 115(3), 2000, pp. 847 – 904. Available at: https://doi.org/10.1162/003355300554935.

[6] Cf. B. Buchanan et al., *Truth, Lies, and Automation: How Language Models Could Change Disinformation* in *Center for Security and Emerging Technology*, 2021. Available at: https://doi.org/10.51593/2021CA003.

people in situations they have never been in, manipulate their perception of events, and potentially manipulate an entire section of the electorate into believing things that never happened. What may have an impact is not the fact that AI-generated content creates false information and tries to influence voters through it, but the fact that for a voter, the idea that false AI-generated content is now commonplace tends to create distrust of all content and therefore a less informed voter is less willing to participate in a political community. Surveys show that almost half of the respondents are unable to distinguish between real and manipulated videos, with a significantly higher percentage among the older generations[7], who in many countries are the most actively engaged in political processes.

At the same time, political communities have become more polarised between two usually opposing positions. In this case, artificial intelligence is used to generate fake content in order to strengthen the arguments of each side. People usually look for content that confirms their previous thoughts and ideas, but in the pre-social media and pre-internet world, the amount of content and its quality to reinforce previous ideas is relatively limited and slow to be distributed. The development of social media has created an explosion of this type of content and made its distribution almost instantaneous. The use of artificial intelligence to create this type of content has greatly improved its quality and power to convince those who see, hear or read it. This improved quality of false or distorted information created by AI can help increase political polarisation within a social community. People become more convinced of their beliefs due to the plausibility of AI-created videos, texts and photos and are less inclined to change their previous ideas and switch political sides within a community. The dangers of the use of AI, described above, cannot be primarily attributed to AI, but mainly to human nature and the way it views political engagement, democracy, elections, and the reception of information within society.

2. *Hi-Tech colonisation*

Another very relevant aspect related to the impact of generative artificial intelligence on political freedom is the so-called "Hi-Tech colonisation": this is a form of digital colonialism where the use of digital

[7] Cf. C. Helmus, *Artificial Intelligence, Deepfakes, and Disinformation: A Primer*. RAND Corporation, 2022. Available at: https://doi.org/10.7249/PEA1043-1.

ISSN 0039-4130

technologies is aimed at the political, economic and social domination of another nation/territory or ideological group. Whereas with classical colonialism, Western nations seized foreign lands and appropriated indigenous knowledge in order to incorporate it into industrial processes, with the advent of digital colonialism, the spread of digital technologies and artificial intelligence has become deeply integrated with the conventional tools of capitalism and authoritarian governance, such as labour exploitation, policy capture, economic planning, intelligence, ruling class hegemony and propaganda [8].

The privatisation of software has been accompanied by the rapid centralisation of the Internet in the hands of intermediary service providers such as Facebook or Google. Essentially, this shift to cloud services has nullified the freedoms that Free and Open-Source Software licences guaranteed to users because software is run from the computers of Big Tech multinationals. Corporate clouds expropriate people from the ability to control their own computers. Cloud services provide petabytes of information to corporations, which use the data to train their artificial intelligence systems. Artificial intelligence uses Big Data to "learn" - it needs millions of images to recognise, for example, the letter "A" in different fonts and formats. Applying this to humans, the sensitive data of people's private lives becomes a resource of incalculable value that technology giants relentlessly try to extract. In other words, the technology giants control the business relationships throughout the production chain, profiting from their knowledge, accumulated capital and hegemony of key functional components.

The ecological crisis created by capitalism is seriously threatening to destroy life on earth, and solutions for a digital economy must intersect with environmental justice and a broader battle for social equality. In order to eliminate the phenomenon of digital colonialism, we need an ethical paradigm capable of challenging the purely economic ends of hi-tech imperialism in order to put human beings in their universal value at the centre [9].

This means, as Paolo Benanti reminds us [10], to initiate an ethical transition within the use of digital technologies and, in particular, arti-

[8] Cf. M. C. Horowitz, *AI and the diffusion of Global Power,* November 16th, 2020. Available at: https://www.cigionline.org/articles/ai-and-diffusion-global-power/.

[9] Cf. Y. Katz, *Artificial whiteness: politics and ideology in artificial intelligence*, Columbia University Press, New York 2020.

[10] Cf. P. Benanti, *Human in the Loop. Decisioni umane e intelligenze artificiali*, Mondadori, Milano 2022; Id., *La condizione tecno-umana, Domande di senso nell'era della tecnologia*, EDB, Bologna 2022.

ISSN 0039-4130

ficial intelligence in such a way that the data collected do not serve to increase the economic gain of a few technocratic elites but are turned to the promotion of human flourishing that can involve everyone.

The consequences of digital colonialism in education should also be borne in mind here: it is spreading rapidly in the educational systems of many countries. Schools are good sites for Big Tech to expand control of the digital market. In those countries where governments provide students with a device at no cost, multinationals are able to acquire and capitalise on a significant amount of data. This is a way to retain the new generation in the use of specific software and platforms. In doing so, however, not only do students become true guinea pigs from which to obtain data, but they are also oriented towards the use of specific platforms that will most likely be preferred by them in the future. Faced with these scenarios, a reflection is required on the need to curb digital colonialism in order to foster a fair and transparent distribution of digital resources with the unequivocal aim of fostering the human in its fullness, beyond any social or economic discrimination. It is therefore necessary to rethink digital technologies no longer as a private service for the benefit of a few, but as a public good that requires to be disciplined and regulated in its use in order to ensure maximum transparency, social justice, and to avoid the exploitation of the weakest categories of the world's population.

3. *Algorithmic society and personal identity*

Having considered the risks of misuse of AI at the political level, it is now necessary to highlight its role in the process of personal identification, with particular reference to the expression of individual and democratic freedom. In fact, from our point of view, there is a close link between personal and political freedom, since only an inwardly free person can express his or her active and responsible participation in the political context. Faced with the increasingly complex scenario of technological societies inhabited by algorithmic systems, Simona Tiribelli, professor of Ethics of Artificial Intelligence and Global Justice and Technology at the University of Macerata, proposes a philosophical reflection on the relationship between personal identity and algorithms from an ethical perspective. In this last sectionof the paper I want to start from Tiribelli's reflection: in her recent volume, entitled *Identità personale e algoritmi. Una questione*

ISSN 0039-4130

di filosofia morale [11] the scholar asks what it means to be a 'person' in algorithmic societies and to what extent these systems affect the processes of personal identity formation and the exercise of freedom of choice. Rigorously analysing, and supported by the most recent theoretical research, the context of the "phygital" environment, the one characterised by an impressive hybridisation between physical and virtual space, Tiribelli on the one hand acknowledges the enormous potentialities connected to the infosphere [12], the possibility of establishing previously unseen connections and generating new relational modalities, and yet she does not hide the properly ethical concern related to the development of personal identities in the context of an ever-increasing datafication of reality. If, in fact, on the one hand algorithmic systems, those of a deterministic type, but even more so those of a predictive and probabilistic type such as machine learning or deep learning, increase the human capacity for agency by promoting the right of self-determination of subjects and the enhancement of their individual and collective possibilities of action, on the other hand they are not exempt from important ethical risks. Studying in particular the predictive algorithmic enhancement based on individual identity profiling techniques, Tiribelli highlights how such systems are able to guide human behaviour to the point of determining a true ideological manipulation of individuals [13].

Compared to the numerous research already active in the field of the right to privacy and the protection of personal data in the context of the cybersphere, the novelty of Tiribelli's theoretical proposal consists precisely in not considering personal identity in exclusively informational terms, but in assuming an ethical perspective that conceives identity as a process of maturation of the sense of existing in an integral anthropological perspective. Stressing on several occasions in her book the need to recover an ethical reflection on the theme of personal identity, Tiribelli proposes reformulating the right to privacy itself in terms of the right to freedom of choice and self-fulfilment. According to the scholar, in fact, when we speak of the constitution of personal identity

[11] S. Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Carocci, Roma 2023.

[12] Cf. L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina, Milano 2017; Id., *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide,* Raffaello Cortina, Milano 2022.

[13] Cf. A. Marwick - R. Lewis, *Data & Society Media Manipulation and Disinformation Online Case Studies*, Data & Society Research Institute, 2017. Available at: https://datasociety.net/output/media-manipulation-and-disinfo-online.

ISSN 0039-4130

we are referring to a dynamic that is deeper than the mere acquisition of data, but which concerns the constitution of subjectivity in its most radical ethical aspiration. In this sense, her contribution intends to offer an ethical analysis of personal identity in the context of algorithmic profiling starting from an anthropological vision that does not conceive of the human being as a mere informational aggregation but as a complex and historically given reality. As the scholar writes, «the philosophical framework in which this work moves assumes freedom of choice and autonomy as two fundamental elements in the process of construction and development of personal identity, in accordance with a tradition of Kantian inspiration, which sees in the choices and actions of individuals not only the place of expression but also the place of constitution and formation of personal identity» [14]. In continuity with the tradition of Kantian-inspired moral philosophy, Tiribelli bases her analysis on the constitutive link that exists between personal identity and freedom of choice, showing how a philosophical reflection of this nature is called upon to confront the way in which digital technologies influence the horizon of meaning and the constitution of personal subjectivities in their axiological openness.

Eschewing any metaphysical considerations on the subject of personal freedom [15], Tirabelli's reflections in the volume lean towards a practical dimension, attesting to the urgency of informing algorithmic design with an ethical perspective, without completely evading the scope of informational theories on identity but enriching them with a new perspective light, that of ethics.

The great gain of Tirabelli's proposal, which is useful for the argumentation of this paper, consists then in a reconsideration of the theme of personal identity in the context of algorithmic societies; a theme that cannot be the exclusive prerogative of the legal and normative sphere but must also be placed in an ethical perspective. Indeed, as the book concludes, it is necessary to establish ethical criteria and specific actions to mitigate algorithmic interference and to discern the possibilities and risks of such systems. Only an ethical design of them, an ethics by design, can guide technological societies to the promotion of full human identity, which cannot be reduced to an aggregation of data but

[14] S. Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, cit., pp. 14-15.

[15] Cf. A. Andrade Braga - M. Chaves, *A dimensão metafísica da Inteligência Artificial, The Metaphysical Dimension of Artificial Intelligence. La dimension métaphysique de l'Intelligence Artificielle*, in *Revista Crítica de Ciências Sociais*, 119 | 2019, Número semitemático.

ISSN 0039-4130

must be grasped in its bio-psycho-spiritual complexity and in its deep connection with the dimension of free ethical choice [16].

Conclusion

What is offered in this paper constitutes the outcome of a shared and transversal research path whose aim is to interrogate current phenomena in order to put them at the service of present and future humanity. For this reason, the theoretical gains of this research are aimed at shedding light on the risks of a manipulative use of such technologies that is directed towards the spread of infodemics and the assertion of authoritarian powers, in order to guarantee the exercise of free and responsible citizenship in contemporary societies and to foster an ethical use of digital technologies and artificial intelligence.

In conclusion, it can be said that artificial intelligence systems can undoubtedly be useful for human development. AI can be a gentle nudge to steer humans towards responsible choices, but only if it is programmed to structure a good choice architecture that allows governments to protect the freedom of citizens by encouraging them to make wiser decisions that respect individual and collective freedom.

Short bibliography

A pro-innovation approach to AI regulation, Dandy Booksellers Ltd, London 2023.

E. Aizenberg - J., Van Den Hoven, *Designing for human rights in AI*, in *Big Data & Society*, 7(2), 2020, ELocator: 205395172094956. https://doi.org/10.1177/2053951720949566.

A. Alesina, - E. La Ferrara, *Participation in Heterogeneous Communities*, in *Quarterly Journal of Economics*, CXV(3), 2000, pp. 847 – 904. https://doi.org/10.1162/003355300554935.

A. Andrade Braga - M. Chaves, *A dimensão metafísica da Inteligência Artificial, The Metaphysical Dimension of Artificial Intelligence. La dimension métaphysique de l'Intelligence Artificielle*, in *Revista Crítica de Ciências Sociais*, CXIX | 2019, Número semitemático.

[16] Cf. M. Coeckelbergh, *AI ethics*, The MIT Press, Cambridge 2020; Id., *Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence*, in *AI and Ethics*, 2022. Available at: https://doi.org/10.1007/s43681-022-00239-4.

P. Benanti, *Human in the Loop. Decisioni umane e intelligenze artificiali*, Mondadori, Milano 2002.
B. Buchanan et al., *Truth, Lies, and Automation: How Language Models Could Change Disinformation* in *Center for Security and Emerging Technology*, 2021. https://doi.org/10.51593/2021CA003.
L. Chiappini, *The geopolitical effect of Artificial Intelligence: The implications on Human Rights and Democracy*, Mondo internazionale APS - ETS, Pisa 2022.
M. Coeckelbergh, *AI ethics*, The MIT Press, Cambridge 2020.
M. Coeckelbergh, *Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence*, in *AI and Ethics*, 2022. https://doi.org/10.1007/s43681-022-00239-4.
B. Fallery, *Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques*, in *HAL Open Science*, 2021 [Preprint]. https://hal.science/hal-03126059/document.
L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina, Milano 2017.
L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide,* Raffaello Cortina, Milano 2022.
C. Helmus, *Artificial Intelligence, Deepfakes, and Disinformation: A Primer*. RAND Corporation, 2022. https://doi.org/10.7249/PEA1043-1.
M. C. Horowitz, *AI and the diffusion of Global Power,* November 16th, 2020. https://www.cigionline.org/articles/ai-and-diffusion-global-power/
Y. Katz, *Artificial whiteness: politics and ideology in artificial intelligence*, Columbia University Press, New York 2020.
A. Marwick - R. Lewis, *Data & Society Media Manipulation and Disinformation Online Case Studies*, Data & Society Research Institute, 2017. https://datasociety.net/output/media-manipulation-and-disinfo-online.
N. Miailhe, *Géopolitique de l'Intelligence artificielle: le retour des empires*, in *Politique étrangère*, in Automne (3), 2018, pp. 105–117. https://doi.org/10.3917/pe.183.0105.
C. Novelli et al., *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, in *SSRN Electronic Journal*, 2024 [Preprint]. https://doi.org/10.2139/ssrn.4694565.
C. Novelli et al., *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity* (January 14, 2024). Computer Law & Security Review, volume 55, 2024[10.1016/j.clsr.2024.106066], Available at SSRN: https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.1016/j.clsr.2024.106066.
S. Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Carocci, Roma 2023.
T. Weikmann - S. Lecheler, *Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network* in *Digital Journalism*, 2023, pp. 1–18. https://doi.org/10.1080/21670811.2023.2194665.

ISSN 0039-4130

Health in the Age of AI and Neuroscience: Ethical Challenges to Autonomy and Freedom

Laura Palazzani

1. *Emerging technologies, transformations in healthcare and implications for autonomy and freedom*

The use of technologies in medicine is certainly nothing new, nor is the ethical discussion of the application of technologies in medicine. For decades, bioethics, a discipline based on interdisciplinary and pluralist knowledge (characterised by the comparison between different moral views), has discussed and continues to discuss the moral limits of the use of technologies that interact with the human body/mind.

However, today we are faced with a "new biotechnological wave" (which includes neuroscience and AI, both as separate and converging fields), which presents specific characteristics, and which have a potentially "disruptive" impact on health also in ethical and legal terms [1]. New possibilities are emerging, driven by the speed of innovation, technological and scientific complexity (which can make interdisciplinary dialogue more difficult), forms of invasiveness/pervasiveness of technology (with uncertainties difficult to delineate), nuances of boundaries (between health and disease, therapy and enhancement), and breadth of applications (medicine and beyond).

The discussion of the topic is taking place among experts, within science and technology ethics committees at a global, European,

[1] J. F. de Paz Santana - D.H. de la Iglesia - A.J. López Rivero (eds.), *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*, Springer 2022; R. Strand, M. Kaiser, *Report on Ethical Issues Raised by Emerging Sciences and Technologies,* Norway: Centre for the Study of the Sciences and the Humanities, University of Bergen, XXIII, 2015.

and national level (UNESCO, OECD, WHO, European Commission, Council of Europe, National Bio/ethics Committees). The objective of the ethical discussion is to identify and discuss ethical challenges and analyze any regulatory gaps to provide governance and/or specific regulation solutions [2].

Here the focus of the discussion on the applications of Neuro AI in the field of medicine and the transformations in the concept of health will be defined with specific reference to the theme of human autonomy/freedom. Human freedom "at the test" means that new technologies force us to verify whether our conception of freedom on a philosophical level is theoretically consistent and applicable on a practical level; or if technology leads us to rethink it.

2. *Neurotechnology and AI applied to health: some preliminary distinctions*

First of all, we need to start with some definitions and distinctions.

Neurotechnology refers to devices and procedures used to access, assess, monitor, emulate or modulate the structure and function of the nervous systems of human beings. These are tools that allow us to directly visualize and record brain activity, and correlate this activity with human sensations, thoughts and actions. On the one hand, neural data can be correlated with sensations, thoughts and actions, with the aim of 'reading the mind' (knowing what one feels, thinks, and how one acts, with reference to oneself and the other) or predicting future outcomes using statistical projection (what one will feel, think, how one will act); and on the other hand, it can be used for influencing or modifying sensations, thoughts, behaviors ('writing' on the brain). Neurotechnology is wider than that: it also affects brain activity without collecting data,

[2] See: Commission nationale d'éthique dans le domaine de la médecine humaine (NEK-CNE), *L''amélioration' de l'humain par des substances pharmacologiques*, Berne, NEK-CNE 2011; M. Ienca, *Common Human Rights Challenges Raised by Different Applications of Neurotechnologies in the Biomedical Fields,* Report for the Committee of Bioethics of the Council of Europe, 2021; Italian National Committee for Bioethics, *Neuroscience and Human Experimentation: Bioethical Problems*, Rome, 2010; National French Consultative Ethics Committee for Health and Life Sciences, *Ethical Issues arising out of Functional Neuroimaging*, Opinion No. 116. Paris, 2012; Nuffield Council on Bioethics, *Novel Neurotechnologies: Intervening in the Brain*, London, 2013; OECD, *Recommendation of the Council on Responsible Innovation in Neurotechnology,* Paris, 2013; UNESCO, International Bioethics Committee, *Report on Ethical Issues of Neurotechnology*, Paris, 2021.

ISSN 0039-4130

like with deep brain stimulation or transcranial magnetic stimulation [3]. *AI systems* consist of algorithms trained to correlate large quantities of data (in this case brain data) using powerful electronic computers. AI systems trained on brain data can potentially -using neurotechnologies- know, predict and modify our brain activity. Advances in big data analytics and machine learning could enable a greater inferential capacity based on large-scale pattern recognition and statistical processing of information, and predict outcomes based on the combination of different data sources/complex data sets.

Neurotechnology and AI should be considered both as separate and convergent technologies. Their convergence (*Neuro AI*) is one of the fastest-growing fields in neuro-medicine research and innovation.

The ethical reflection on the application of technologies to health must be articulated at different levels (sometimes nuanced and not clearly distinguishable) in relation to the diversity of objectives. Some neurotechnologies are used to cure, i.e. prevent or repair damage and rehabilitate patients with compromised or damaged brain function (including mental disorders and other functional impairments); some to carry out research to acquire new knowledge and develop new technologies; and others to intervene and modify functions of healthy individuals (cognitive and emotional enhancement).

Additional technologies, in particular Neuro AI, are used to visualize brain functions or acquire "neural data" or "brain data" that can be used to predict a pathology in the field of "omics" medicine similar to precision medicine in the genetic field (identify predictive biomarkers of pathologies, with probability thresholds); acquire data to profile individuals and predict thoughts and actions (with some applications also in other fields such as teaching/education, gaming/entertainment, and marketing).

The ethics of technologies (neuro and AI) are used to analyze the context of proportionality (beneficence/non-maleficence), as a framework for evaluating the patient's autonomy, and also in the context of social justice. In this context, autonomy will be the center of the analysis.

[3] H. Chneiweiss, *Neurosciences et neuroéthique: des cerveaux libres et heureux*, Alvik Editions, Paris 2006.

ISSN 0039-4130

3. *Ethical challenges to autonomy/freedom in health in neurotechnologies, AI and neuro AI*

3.1. Neurotechnologies for treatment: autonomy as awareness (informed consent)

Neurotechnologies include different types of devices with different levels of invasiveness in the patient's body and mind[4]. Some technologies are "non-invasive"[5] and do not require opening the skull to directly access the brain; some are "invasive"[6]; some may be both invasive and non-invasive[7];

[4] UNESCO, International Bioethics Committee, *Ethics of Neurotechnology*, 2021.

[5] Electroencephalography (Scalp EEG); magnetic resonance imaging (MRI); positron emission tomography (PET); functional magnetic resonance imaging (fMRI) studying the functional anatomy of the human brain; transcranial direct current stimulation (tDCS) or transcranial electrical stimulation (tES) involve devices delivering continuous currents supposedly to enhance concentration or relaxation.

[6] Deep brain stimulation (DBS) involves implanting electrodes within certain areas of the brain. The amount of stimulation is controlled by a pacemaker-like device placed under the skin in the upper chest. Deep brain stimulation is approved to treat a number of conditions, such as Parkinson's disease, essential tremor, dystonia, epilepsy and obsessive-compulsive disorder, a potential treatment for major depression, traumatic brain injury, stroke recovery, addiction, chronic pain, cluster headache, dementia, Tourette syndrome, Huntington's disease and multiple sclerosis. The possible side effects: surgery complications, hardware (device and wires) complications, and stimulation-related complications.

[7] Brain Computer Interfaces (BCIs) are a type of neurotechnology that aims to translate brain processes that underlie thought and action into desired outcomes (e.g. enhancing mood in a depressed person or moving a prosthetic limb). This is made possible by collecting the data related to neural activity by sensors or electrodes placed in the brain, on the brain, or over the surface of the scalp, transforming them into a signal and then converting this signal into a mechanical or electrical action. BCIs are often directed at researching, mapping, assisting, augmenting, or repairing human cognitive or sensory-motor function (for example, in the case of a brain lesion resulting in hemiplegia, paraplegia or tetraplegia). They can be invasive, partially invasive or non-invasive. Invasive BCI requires surgery to implant electrodes within the grey matter of the brain, for directly relaying brain signals to device output, as in the case of treatments for non-congenital blindness. BCIs focusing on motor neuroprosthetics aim to either restore movement in individuals with paralysis or provide devices to assist them to communicate or physically interact with their environment, such as interfaces with computers or robot arms. Partially invasive BCI devices are implanted inside the skull but rest outside the brain, an example being Ecocorticography (EcoG) technology. Non-invasive EEG-based technologies and interfaces have been used for a much broader variety of applications. Neuroprosthetics is an area of neuroscience concerned with neural prostheses, using artificial devices to replace the function of impaired nervous systems and brain-related problems, or to replace the sensory organs themselves. The first neuroprosthetic device was the pacemaker. BCIs focusing on sensory prosthetics aim to restore the brain perception of sensory organs, such as eyes for sight or ears for hearing. Deep learning, partly modelled on biological processes occurring within the human brain, enables machines to recognize shapes and patterns. Brain Computer Interfaces use deep learning to decode brain activity, and some can help paralysed patients to regain speech or movement. Deep learning can also be applied for medical tasks, with prominent examples being convolutional neural networks (ConvNets) on EEG signals.

ISSN 0039-4130

some are pervasive, that is, not invasive in the body but they can be psychologically or socially.

Given the growing neurotechnological possibilities for intervening in the brain, and consequently the mind, invasive/non-invasive, it is necessary to consider the integrity of the brain and mind in the framework of the dignity of the human body as a condition of autonomy. Ienca and Andorno [8] recognize "mental integrity" as a value, faced with the neurotechnological possibility of provoking "direct harm" caused by "the alteration of a person's neural condition". In this perspective, the integrity of the body, and the brain/mind as part of the body, should be recognized, respected and protected by the physician, who is the only one who has the competencies to balance the proportionality of the intervention (benefit in relation to the therapeutic objective; harm considering the consequences on the body/mind).

Integrity of the body/mind is strictly connected to authenticity**,** with reference to the continuity of personal identity. Brain implants may alter the content of memory (memory editing); memory modification techniques can erase a memory, induce amnesia, reducing the emotional impact of a painful memory and the risk of post-traumatic stress disorder; deep brain stimulation can pose a threat to an individual's mind-to-body unity as their authentic self, because while the body regains appreciable autonomy in its movements, the mind can be disoriented by the active presence of the technical device. The individual may experience a feeling of alienation (subjugation to a technical device) that bodily improvement cannot eliminate. Added to this is the possibility that the device can be controlled remotely by a clinician, perhaps without the patient's knowledge. In Parkinson's disease, it aims to reduce motor symptoms. In obsessive-compulsive disorders, which appear earlier in life, it is the behaviour of the person, including their way of thinking and feeling, that is targeted. In addition, depending on the type of pathology and the duration of the patient's experience of this pathology, some people have integrated this pathology into their personal identity. They feel more authentically themselves with the pathology than without it. This indicates that there is no universally experienced correlation between the notions of health and authenticity, disease and alienation.

In this context, there is an ethical need to refine the definition of patient autonomy. In medicine, informed consent is one of the basic

[8] M. Ienca - R. Andorno, *Towards New Human Rights in the Age of Neuroscience and Neurotechnology*, in *Life Sciences, Society and Policy*, XIII, 5, 2017; R. Yuste et al., *Four Ethical Priorities for Neurotechnologies and AI*, "Nature News", Vol. DLI, 7679, 2017, pp. 159.

principles of bioethics closely linked to the principle of autonomy [9]. It also implies that the duty of the physician is to give all the relevant information in a complete, clear, comprehensible, and updated way in order to allow the patient to participate in the decision: participation means expressing the acceptance/refusal of treatment as an autonomous, aware and voluntary activity.

An individual should receive understandable and individually tailored information in a dialogue that makes it possible for that individual to decide on whether or not to accept medical intervention considering the benefit/risk ratio. It has also been established that one of the most important aspects to be provided is information about possible risks and benefits related to a proposed medical intervention; this is a key component in obtaining consent [10]. It should not have a long, incomprehensible, bureaucratic and legalistic form, to be signed only to defend the doctor.

Informed consent as an expression of the autonomy of the patient needs to be expressed at the beginning of a possible treatment, given all the information about benefits and risks. For neurotechnology, on many occasions the risks and benefits are still uncertain, therefore doubts about the validity of informed consent and autonomy may arise because of the lack of awareness. Currently, consent typically focuses only on the physical risks of (neuro)surgery (in the case of invasive technologies), rather than the possible effects of a device on mood, personality or sense of self, personal identity and authenticity; and on the direct benefits to patients, rather than to society as a whole (indirect benefit) [11].

The obligation to avoid harm requires an ongoing commitment to develop a robust body of evidence through research, attention to the needs and vulnerabilities of particular individuals, and a willingness to reflect upon and review clinical practices and the development trajectories of these technologies. In such circumstances, some recommend the application of the "principle of caution", referring to a «less restrictive standard of behaviour, one which is tempered by the recognition that

[9] T.L. Beauchamp - J.F. Childress, *Principles of biomedical ethics*, Oxford University Press, Oxford 2001.

[10] L. Palazzani (ed.), *Special Issue on iConsent - Improving the Guidelines for Informed Consent, Including Vulnerable Populations, Under a Gender Perspective*, in *BioLaw Journal-Rivista di BioDiritto*, Special Issue 1/2019, pp. 154.

[11] However, a recent article pointed out that there isn't enough evidence to talk about personality changes (in the case of DBS). See G. Frederic - J.N.M. Viaña - C. Ineichen, *Deflating the "DBS Causes Personality Changes" Bubble*, in *Neuroethics,* 14.Suppl 1, 2021, pp. 1-17.

ISSN 0039-4130

some risks, and some uncertainty about risks, may be tolerated where technologies could make a significant contribution both to individual patients and to the public good» [12]. In the context of neurotechnologies *the boundaries of research and treatment are blurred* [13].

3.2. Neurotechnologies for research: awareness of uncertainties

The invasiveness of neurotechnologies is justified only for therapeutic purposes. To the extent that neurotechnologies are invasive, thus exposing patients to particular risks both physically and psychologically, as well as discomfort (which can also cause pain in the physical, psychological and social sense), research is justifiable only if there is a proportion between the expected benefits for the treatment of the disease and the foreseeable risks. Such research, if invasive, is only justified in cases of severe disease and in the absence of less invasive non-experimental alternative therapies, applying the principle of the best interest and the minimization of risks and inconveniences.

Since the *Nuremberg code* (1997), through the *The Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine* (Council of Europe, Steering Committee for Bioethics, Oviedo, 1997) and the *Declaration of Helsinki* (World Medical Associations, last version 2013), consent is considered the main guarantee for the patient's dignity, autonomy and rights).

In research with neurotechnologies, one ethical challenge is the possibility of incidental findings or unexpected brain anomalies i.e. in functional magnetic resonance imaging (fMRI). The frequent discovery of unexpected anomalies as well as the difficulty of interpreting them, may pose ethical challenges such as determining appropriate strategies for communication (before and after), and how to handle clinically relevant incidental findings above all concerning minors [14].

The principle of respect for autonomy may be considered for adults: they can refuse to know incidental findings also of clinical rel-

[12] Nuffield Council of Bioethics, *Novel neurotechnologies: intervening in the brain*, London, 2013.

[13] E. Mullin, '*Neuralink's First Brain Implant Is Working. Elon Musk's Transparency Isn't*', in *"Wired",* 21 February 2024.

[14] C.N. Di Pietro - J. Illes, *Disclosing Incidental Findings in Brain Research: The Rights of Minors in Decision-Making*,
in *Journal of Magnetic Resonance Imaging*, XXXVIII, 2013, pp. 1009–1013..

ISSN 0039-4130

evance. But when they are found on minors, the parents' autonomy may be in conflict with the principle of beneficence of the children and their best interest, when persons subjected to brain imaging research renounce their right to know and claim the right not to know. In seeking to do good (beneficence), a researcher may deem it a responsibility both on the ethical and deontological level to disclose the potential consequences of incidental findings discovered during brain imaging research because they may be beneficial to health (at the level of prevention, treatment, care), but doing so may be a breach of respect for a person's right to autonomy.

Another challenge to autonomy is the possibility of predictive value of brain images and the possibility to diagnose certain dispositions in the brain (e.g. the likelihood of getting a certain disease): this issue is also of ethical concern in neuroscientific research. During research the subject should be aware of the possibility of ascertaining the probability or certainty of such a disposition, since the possibilities of false positives (diagnosing a pathology that is not there) or the possibilities of false negatives (the failing to identify or communicate a possibly life-threatening condition) may have major consequences on patients. The negative impact of false positives weighed against the potential consequences associated with failing to identify or communicate a possibly life-threatening condition needs to be considered in neuro-research.

This aspect is extremely delicate with minors, in a similar way to predictive genetic tests, because of the difficulty to manage an existentially problematic outcome which may cause anxiety in a scenario of predictive incurable illness, considering the uncertainty. Respect for the child's autonomy and future decision-making (the so-called right to an open future) sometimes call for a special management of the information that has no immediate relevance in the child's health or health management, above all when no treatment or preventive interventions are available. More robust protective measures might be required in order to preserve basic human rights, including the autonomy of vulnerable human beings, as children.

3.3. Neuro-enhancement

Neurocognitive enhancement means the use of biotechnologies in order to intervene in a healthy body/mind with the aim of improving

ISSN 0039-4130

mental and emotional performance (including psychotropic drugs, neuroimaging technologies, neurostimulation technologies, transcranial magnetic stimulation or transcranial direct current stimulation applied over the cortex, or brain implants and brain-computer interfaces). Neurocognitive enhancement encompasses diverse methods of intervention, more or less invasive with regard to the body/mind, with potential short- and long-term consequences. Despite their differences, they share common goals of intervention, which can be identified as enhancing human capabilities, "beyond therapy" [15].

The blurring between enhancement and therapy comes from the subjectivist view of health, considered a state of complete physical, mental and social well-being. In this perspective, enhancement is equated with therapy, insofar as a reduced capacity may be subjectively, socially and culturally perceived as a source of discomfort or illness. The subjective perception of "normality" plays a role in this blurring (in the case of cochlear implants for deaf children, which may be considered normal by deaf communities).

Arguments in favour of enhancement [16] start from the idea that enhancement and therapy are interchangeable, considering improvement as part of human development, whether natural or artificial. It is considered a "technological shortcut", or a stage of evolution to be replaced by the "deliberate choice" of the selection process, allowing the same result to be achieved rapidly and with much less effort. Although possible negative outcomes still remain unknown, to halt progress in this direction would imply hampering or preventing the possibility of accelerating human evolution. This theory of "self-evolution" and "enhancement evolution" would shorten the time required for millions of years of evolutionary progress, allowing human beings and humanity to attain and realize their full potential, in order to balance the effects of what, in physical and social terms, is a natural lottery. This approach

[15] See Reports and Opinions of Bioethics Committees: Health Council of the Netherlands, *Human Enhancement*, The Hague, 2003; Italian National Committee for Bioethics, *Human Rights, Medical Ethics and Enhancement Technologies in the Military*, Rome, 2013; Italian National Committee for Bioethics, *Neuroscience and Pharmacological Cognitive Enhancement: Bioethical Aspects*, Rome, 2014; National French Consultative Ethics Committee for Health and Life Sciences, *Recours aux techniques biomédicales en vue de 'neuro-amélioration' chez la personne non malade: enjeux éthiques*, Opinion No. 122, Paris, 2013; U.S. President's Council on Bioethics, *Beyond Therapy: Biotechnology and the Pursuit of Human Improvement*, Washington DC, U.S. 2003.

[16] J. Savulescu - T. Meulen - G. Kahane, *Enhancing Human Capacities*, Wiley-Blackwell, London, 2011; J. Harris, *Enhancing Evolution. The Ethical Case for Making Better People*, Princeton University Press, Princeton, 2007.

ISSN 0039-4130

justifies a "duty to enhance" as a "duty of beneficence", which is not only individual but also collective.

The critical approach[17] to neuroenhancement underlines the threats to the dignity of attempting to overcome (i.e., not accepting) the limits of nature. The use of technologies for enhancement purposes can cause serious harm, disproportionate compared to the expected benefits of the fulfilment of subjective desires. Excessively risky interventions in terms of the achievable benefits (deemed ineffective, costly and burdensome for patients), as well as irreversible and predictably inconclusive interventions, cannot be ethically or deontologically[18] justified, even if requested by patients. In this view enhancement is a "fraudulent misrepresentation" unfair towards the other. By contrast with enhancement, "achievement" encompasses the dimension of acquiring, in the sense of developing and realizing potentialities naturally, through an active effort and personal commitment that enable modification of one's own natural capacities.

In this sense, enhancement embodies the **hidden pressure** (or "social despotism", as Sandel calls it[18]) exerted by society on individuals to adapt to standards of mental/emotional efficiency in studying, working, and athletic performance in a competitive society. This practice could **restrict autonomy** and become **coercive**, directly or indirectly, for individuals and the population in general, or within specific categories (both in the public and private sector), in terms of possible discrimination, marginalisation and stigmatization of those refusing to use it. There is a high demand and a potentially broad market in our competitive society, where the aged population do not bear the loss of memory, parents want to stimulate their children to reach the best possible result, and professionals are strained by unsustainable work demands. The strong social pressure drives people to seek ways of raising the level of their performance in education and work using pharmaceuticals and technologies putting their health at risk, because of unknown possible adverse effects, that may be serious and irreversible[19].

[17] L. Kass, *Life, Liberty and the Defence of Dignity. The Challenge for Bioethics,* Encounter Books, San Francisco 2002; F. Fukuyama, *The End of History and the Last Man,* Free Press, New York 2006; M.J. Sandel, *The Case against Perfection*, Cambridge, Harvard University Press 2007.

[18] M.J. Sandel, *The case against perfection*, quoted.

[19] Implications for autonomy, agency and responsibility are associated with the malevolent misuse of neurotechnology by third parties, especially by external interventions that hijack control over a person's neurotechnological systems. It has been experimentally demonstrated that such neurotechnologies can be hacked by malicious actors in order to alter their control with deleterious consequences for the subject.

ISSN 0039-4130

The dual use of biotechnologies (the fact that they may have clinical applications in therapeutic settings and may also be applied for enhancement purposes) makes their ethical justification particularly sensitive and troublesome. A total ban on research and the use of technology may a priori hinder the development of a number of possible therapies; at the same time, the discovery of certain technologies may encourage their use for enhancement purposes. The dual-use argument, which was generally brought up by bio-conservatives to emphasize risks, is now being used by bio-progressives to justify some development methods.

There has, to date, been non-conclusive research study or proof of the safety and efficacy of the use of neuroenhancement [20]. A number of small studies, most of them occasional with non-systematic analysis and without an adequate statistical sample, using neurotechnology, report improvements in participants' performance in a laboratory (for example memory or language skills, or in their mood). Great care is needed in extrapolating from small studies conducted under laboratory 'artificial' conditions to lasting real-world effects; the potential use of neurostimulation for neural enhancement is still far from proven. Not only have no appropriate trials been carried out on this [21], but it would also be extremely problematic from an ethical point of view to experiment with such interventions on healthy subjects, given the absolute uncertainty and the possible high risks associated with non-therapeutic and moreover implausible objectives. Obtaining informed consent/autonomy in this context is another particularly delicate aspect, and is an indispensable requirement of all legitimate research.

3.4. Neuro AI: brain data, mental privacy and cognitive liberty

The convergence of neurotechnologies and AI opens new opportunities in the application to medicine, above all in research. It is possible to obtain data about the structure, activity and function of the human brain (human brain data) that can reveal information about a person's health status (e.g., neurological, or psychological health) and, to some

[20] C. Forlini, *Clearing the Bottleneck of Empirical Data in the Ethics of Cognitive Enhancement*, In F. Jotterand, M. Ienca (eds.), *The Routledge Handbook of the Ethics of Human Enhancement,* Routledge, London 2023.

[21] UNESCO, International Bioethics Committee, *Report on Ethical Issues of Neurotechnology*, Paris, 2021.

extent, support inferences about mental processes[22]. Current neurotechnologies, especially non-invasive techniques, are not yet able to decode thoughts, and are not able to provide a full and real-time account of the neural patterns of specific cognitive processes. However, they already allow researchers to infer the engagement of some mental (e.g. perceptual and cognitive) processes from patterns of brain activation: intelligent algorithms, that automatically learn the neural patterns associated with specific feelings, thoughts or actions, have shown the potential to detect distinctive patterns of brain activity that correlate with particular experiences, memories, hidden intentions, preferences and dreams.

By combining neurotechnology and AI, it is now possible to decode various components of the brain's extremely rich information content. Because of this information potential, these technologies have often been classified under the label of "brain reading". At the current stage of neurotechnology development, neither conceptual (semantic) nor non-conceptual mental content decoding is possible: the algorithm is not capable of revealing the content of a mental representation. A realistic assessment of the current limitations of neurotechnology-enabled mental decoding is necessary to avoid unrealistic public expectations.

Privacy is a primary ethical concern related to the collection, sharing and processing of brain data. While challenges to privacy arise from the processing of any human data, the processing of brain data raises new challenges to the notion of privacy for specific reasons: limited conscious control over one's own brain recordings (as unconscious and subconscious processing) and risk of neurodiscrimination, as brain signals make it possible to distinguish or trace the biometric identity of an individual and are potentially linkable to that individual with risk of discrimination based on a person's neural signatures (indicating, for example, a dementia predisposition), or mental health, personality traits, cognitive performance, intentions and emotional states.

In this context, mental privacy refers to the explicit protection of individuals against the unconsented intrusion by third parties into their cognitive information (be it inferred from their neural data or data indicative of neurological, cognitive, and/or affective information) as well as against the unauthorized collection of those data. The term privacy is used with a specific meaning, which is not confidentiality:

[22] P.R. Roelfsema - D. Damiaan - P.C. Klink, *Mind Reading and Writing: The Future of Neurotechnology*, in *Trends in Cognitive Sciences*, 2018.

ISSN 0039-4130

privacy as the right to reserve certain areas of an individual's private life for themselves.

Neurotechnology may transmit brain data and digital data related to the brain activity of its users. Implanted neurodevices such as those used in deep brain stimulation, and even non-implanted devices might also record patients' brain activity. Information collected and processed from neurodevices can be obtained and used to identify someone or reveal their brain activity, which is particularly problematic when this indicates a stigmatizing neurological or mental health condition or could otherwise be used for discriminatory purposes (health insurance or workplace) [23].

Brain recordings can be predictive of a neurological disease (for example early signatures of dementia that can be inferred from neuroimaging biomarkers). Some of those consequences have already been raised by genetic/genomic research in precision medicine [24], mainly those regarding the access to such information by third parties (employers, insurance companies), but can also include the problem of each individual's right to know or not to know (above all in cases of incurable diseases).

Perhaps the biggest ethical issue to emerge from all of this newfound knowledge and understanding of brain processes is the potential threat to cognitive liberty. Cognitive liberty includes the idea that we ought to be able to have freedom from interference with our mental processes, as well as the freedom to control our own mental processes. The right to freedom of thought is held to be fundamental not just to personhood, but also to democratic legitimacy. Given the abilities of emerging neurotechnologies, it is no longer clear that we can take such freedom for granted. The increased impact of digital technology and AI on the way we feel, think, and behave calls for a new perspective on regulation to protect our rights to freedom of thought and opinion in the "forum internum", the inner space of our mind [25].

Brain-derived data is not entirely voluntarily produced, as they may be out of their owner's control and choice regarding data processing. With the advances in Neuro AI, the ambition is to go beyond reading

[23] D. Susser - L.Y. Cabrera, *Brain Data in Context: Are New Rights the Way to Mental and Brain Privacy?,* in *AJOB Neuroscience*, XV (2), 2024, 122-133.

[24] L. Palazzani, *Innovation in Scientific Research and Emerging Technologies: A Challenge to Ethics and Governance*, Springer Nature Switzerland AG and G. Giappichelli Editore, Cham (Switzerland), 2019.

[25] S. Lighart et al., *Minding Rights: Mapping Ethical and Legal Foundations of 'Neurorights',* Cambridge University Press, Cambridge 2023.

ISSN 0039-4130

brain activity and predicting its course, but also to manipulate/modify it. In such cases, the predictive power of Neuro AI would be combined with a stimulatory augmentation in order to produce better (enhanced) brain and mind states.

But of course, the advent of such an ability would lead to clear questions about *who could exercise this control, in what contexts, and within what kinds of parameters.* For example, a state that could use technologies to record and intervene in cognitive processes would pose a severe threat to democracy. Private corporations could use such technology in order not just to predict consumer behaviour, but also nudge (without their awareness) it toward their own ends. It would be a serious attack on notions of self, personhood, freedom, and thought.

In the interaction of a human being with a neurotechnology application, the protection of freedom of thought of the person should cover both the *forum externum*, i.e., the expression or manifestation of thought through behaviour, and the *forum internum*, i.e. the underlying neurobiological, sensory, motor, mental state processing. This holistic interpretation of freedom ensures protection from undue interference with this aspect of cognitive liberty both at the level of internal neurobiological and mental processing, as well as at the level of externalizing those processes within society through verbal speech, activity or other behaviour.

People should have the right to self-determination to make free, informed, and voluntary decisions about whether they want to use a certain neurotechnology application or refuse to do so. These individual decisions, however, should be balanced against considerations relating to societal and collective well-being. Measures should be taken to ensure that neurotechnology is not advertently or inadvertently used to exert undue influence or manipulation on a person's neurobiological, sensory, motor, cognitive and affective processing in ways that interfere with their freedom of thought and self-determination.

The increasing use of machine learning and AI to optimise the functioning of Brain Computer Interfaces (BCI) also has implications for the ethical notions of autonomy, agency and responsibility.

For example, Haselager[26] hypothesised that when BCI control is partly dependent on intelligent algorithmic components, it may become difficult to discern whether the resulting behavioural output was

[26] P. Haselager, *Did I Do That? Brain–Computer Interfacing and the Sense of Agency*, in *Minds and Machines*, XXIII, 2013.

ISSN 0039-4130

actually performed by the user. This difficulty introduces a *principle of indeterminacy* within the cognitive process that starts from the conception of an action (or intention) to its execution, with consequent uncertainty in the attribution of responsibility to the author of this action. It could generate a sense of alienation in the user, the ethical relevance of which is all the greater in the case of a vulnerable individual such as a neurological patient. For example, a patient suffering from tetraplegia using a BCI: how will it be possible to determine which components of the patient's actions are attributable to the patient's volition and which to the AI? There is a possibility that BCIs may affect subjective experience, and thus personal identity. Therefore, it is difficult to determine in an absolute sense whether intelligent BCIs can increase the autonomy of the user. On the contrary, it is necessary to assess case by case and determine under which circumstances, in which time intervals, and in relation to which mental or physical domains a change (positive or negative) in the autonomy of the user is detectable. In carrying out such assessments, it is important to acquire not only quantitative and objective information (e.g., on mathematical measurements or behavioural observations) but also qualitative and subjective information. The latter category includes the user's introspective self-assessments.

3.5. Use of neurotechnologies and AI outside of the medical field

There are private companies that market non-invasive BCIs to an ever-increasing number of healthy users for purposes such as self-quantification, cognitive training, and neurogaming (the use of brain-controlled video games for recreational or competitive purposes). In the non-medical field (non-therapy derived) there are already games on the market using BCI technology that rely upon non-invasive brain imaging techniques such as electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). There is research activity to develop commercial games that are BCI-controlled. These neurotechnology-based approaches are used for recreational purposes. A large number of people use these applications, with the lack of any clear evidence on benefits/risks. Despite the risks associated with some of these unnecessary applications, they are used without medical monitoring in private settings.

In military settings, novel neurotechnology potentially has applications in treating physical and psychiatric injuries, enhancing fighters' physical, cognitive, and emotional capacities.

ISSN 0039-4130

Uses of non-invasive neurostimulation or BCIs either for recreational purposes or gaming do not generally pose serious health risks. However, the large number of people who use these applications and the lack of any clear associated benefits/risks mean that it is important to address several ethical concerns. In particular, to minimize the pursuit of unnecessary brain interventions, there is a need to ensure the originality and rigour of research investigating non-therapeutic uses in humans and also to disseminate existing evidence. There is a particular concern in children, in whom the effects of neurostimulation or BCIs on the developing brain are not well known. Observational research with children who are already using neurotechnologies are needed to address this, and advice should also be issued to teachers and parents about the current evidence on the efficacy of neurofeedback as an educational enhancement tool.

Consideration should also be given to the fact that cognitive function can be improved in a more lasting manner through instruction, education and continuous training, through a rich social life, relationships, study, learning, the continuous stimulation of hobbies and interests, and by leading a healthy lifestyle (in terms of nutrition, physical activity, etc.). It is a path that requires a lot of time but is (perhaps) more respectful of the opportunities for growth and development of personal and relational identity.

4. Ethical analysis – subjectivity neglect

These points all serve to fuel *a risk of generally undermining human dignity obscuring agency and autonomy*[27]. The emerging call of *neurorights*[28] aims to protect mental integrity, mental privacy, and freedom of thought, together with personal identity and authenticity from undue intervention through neurotechnological means. A useful way to collect these various risks and harms when it comes to data and the self is to use the umbrella term of "subjectivity neglect". This is a neglect of the subjective dimensions that serve to make human experience valuable.

There are dimensions of meaningful human action that cannot be understood except with reference to subjectivity. An **omics* approach

[27] Nuffield Council on Bioethics, *Novel Neurotechnologies: Intervening in the Brain*. London, 2013; Science and Technology Briefings, Parliamentary Office for Scientific and Technological Assessment (OPECST), *Neurotechnology: Scientific and Ethical Challenges*, 2022.

[28] M. Ienca, *On Neurorights*, in *Frontiers in Human Neuroscience*, XV, 2021.

ISSN 0039-4130

might predict well all sorts of descriptive facts about what a person may do or maybe what they might feel and think. A person might even learn things about themselves by scrutinising the predictions made. But the claim that Google could *know you through data* does not hold up beyond a metaphorical sense, amounting to *data hubris*. A crucial part of what is meaningful for human persons is their subjectivity, linked closely with their contexts of action and the concerns they have.

The orientation toward data behind Neuro AI development leads to a significant part of one's social experience being handed over to AI-processed brain data science for ultimate explanation. The collective resources for the identification of social experience are diminished if they are supposed to be captured by an **omics* approach as this would obscure collective understanding of experience by, for example, suggesting that the *real* explanation for this or that behaviour is to be found in the analysis of data and not in expressions of relevance, value, or meaning.

There is a clear *epistemic injustice*[29] in this, in that the knowledge claims of one party (the person) are overlooked in favour of another (the data processing system). The hermeneutic injustice can emerge when an **omics* approach competes with the sorts of self-expression. From any person's point of view, the predictions based on data would be inscrutable having been issued from impenetrable AI sources and opaque datasets.

Subjectivity neglect itself can be seen in terms of the risks it brings through being centred in data**.** Given this discussion of hermeneutic injustice, we can diagnose an important issue concerning Neuro AI as arising from data hubris. This label encompasses data practices that have an overconfidence in what they can represent. In this case, the remarkable capacity data practices have to represent behavioural and bio-signals taken at the same time to represent persons. By examining moral identification, and the role played in self-identity by reflection, deliberation, taking responsibility, cultural, and historical notions, it becomes clear that *data misses significant dimensions of identity* (we are not a mere sum of data). There are hazards that emerge from this too in that predictions made by Neuro AI might easily overstep some sensible boundaries concerning what matters to people about their subjectivity. Subjectivity neglect like this may damage the dignity of human beings

[29] J. Symons - R. Alvarado, *Epistemic Injustice and Data Science Technologies*,"Synthese", LXXXVII, 2022; M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, in *The Philosophical Quarterly*, LXIX (234), 2021, PP. 177-178.

ISSN 0039-4130

and their autonomy, since it could easily erase persons from their telling of their own stories in domains that nevertheless ought to matter to them, like in human research, clinical practices, and socio-political contexts.

5. *Governance: Potential ethical and legal gaps*

There are good reasons to think that loopholes may appear within existing regulations when it comes to Neuro AI[30]. The various convergences behind Neuro AI, and the potential ramifications from it, are part of a potential *regulatory disruption*. Reports commissioned by OECD, UNESCO, WHO, European Commission, Council of Europe, The French National Assembly, The Parliamentary Office For Scientific and Technological Assessment (OPECST), the Nuffield Council on Bioethics, and the IEEE[31], among others, discuss the complexity of regulatory challenges. Across these reports the various possibilities for responding to Neuro AI developments are discussed, ranging from the proposal of new human rights, to the re-evaluation of existing rights e.g. privacy, and freedom.

It was suggested above that classifying people through inferences from neural data could breach the provisions of the *General Data Protection Regulation* 2016 (data are not classified according to the purpose) and the *AI Act* (March 2024). If this were the case, Neuro AI would be placed in the "Unacceptable risk AI systems" category. It may not, however, be clear how to classify Neuro AI applications and devices. On the one hand, for example, consumer brain recording devices might be seen as low risk in being quite crude neural signal recording devices, without much genuine capacity for indicating or affecting users' cognitive, emotional, or other such states. However, given the discussion of subjectivity neglect, and the centrality of data for evolving unsupervised machine learning approaches to Neuro AI and related technology development, even low-grade brain signal recording could play a significant role in driving the Neuro AI sector. Taken as a whole, data from all sources might be expected to be aggregated and processed, towards goals not just unstated, but unformulated at

[30] S. O'Sullivan - h. Chneiweiss - A. Pierucci - k.S. Rommelfanger, *Neurotechnologies and Human Rights Framework: Do We Need New Rights?*, Council of Europe and OECD 2022.

[31] IEEE Brain Technical Community, Addressing the Ethical, Legal, Social, and Cultural Implications of Neurotechnology https://brain.ieee.org/publications/ieee-neuroethics-framework/.

ISSN 0039-4130

all. On the other hand, then, any brain recording device might fall under the category of unacceptable risk, summarised as follows: cognitive behavioural manipulation of people or specific vulnerable groups (for example voice-activated toys that encourage dangerous behaviour in children); social scoring or classifying people based on behaviour, socio-economic status or personal characteristics; biometric identification and categorisation of people; real-time and remote biometric identification systems, such as facial recognition.

OECD and IEEE have formulated recommendations and standards for responsible neurotechnology development. *EU Medical Devices Regulation* (2017)[32] may apply to medical BCIs, while consumer protection laws most likely govern others. As neurotechnologies like BCIs advance and especially as they converge with AI, new regulatory approaches may be necessary to foster responsible, ethical innovation and ensure device safety, autonomy and privacy.

[32] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.

ISSN 0039-4130

The Human Body and the Challenges of Augmentative Technology

Martina Properzi

1. *Introduction*

In recent years, the pervasive production and utilization of Artificial Intelligence (AI) systems have prompted academic and governmental institutions to promote a novel paradigm for research and innovation that centralizes the human user, its social and environmental contextuality. This paradigm shift has been driven by the growing recognition of the need to address the ethical implications of AI, as well as the limitations of current approaches to research and innovation. Notable initiatives from the academic community and the European Commission, respectively, include the Digital Humanism Manifesto [1] and the AI Act [2]. These have sought to elucidate and regulate the impact of AI on human life and society. The focus is frequently placed on generative AI tools that are progressively capable of emulating higher-level cognitive functions, such as text creation and verbal communication, which are typically regarded as hallmarks of the human condition. However, an increasing concern has emerged regarding another type of technology, namely, augmentative technology. In this case, the issue is not one of simulating human capabilities, but rather of exerting direct influence on them. In other words, the challenge is transformative, namely the transformation of the human user into a cyber-human [3].

The concept of the cyber-human is closely associated with the transhumanist movement, which has given rise to a popular debate sur-

[1] https://caiml.org/dighum/dighum-manifesto/.

[2] https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

[3] W. Barfield, *Cyber-Humans. Our Future with Machines*, Springer, Cham 2015.

rounding this notion. The transhumanist perspective postulates the advent of a novel evolutionary era for the human species, namely the era of the *homo sapiens technologicus*. This era is distinguished by the augmentation of the human condition in terms of cognitive and physical capabilities, longevity, and well-being through the deployment of augmentative technology[4]. The cyborg challenge extends beyond the boundaries of popular discourse, encompassing a scientifically pertinent dimension. This coincides with the monitoring and manipulation of human performance. This article aims to show that a significant portion of the ongoing scientific discourse regarding the implications of augmentative technology for individuals and societies is shaped by a "neurocentric" bias (§3). This impedes comprehension of the impact of this technology on the human condition. It is therefore proposed that the scientific debate must be realigned to prioritize not the mind/brain but the embodied subject in order to fully capture the implications of augmentative technology (§ 4). As an initial step, an overview of augmentative technology will be provided.

2. *Augmentative technology*

The field of augmentative technology encompasses a vast array of tools that are capable of directly tracking and/or modifying the human user's intelligence, decision-making, and behavior. These technologies can be classified according to multiple criteria, the most relevant of which are the technology's primary mode of action, its goals, and its relationship with the human body. In their review article on brain augmentation, Jangwan and colleagues differentiate between three categories of technology augmentation based on the manner in which technology functions[5]. From this perspective, there are three categories of augmentative devices: biochemical, behavioral, and physical. It is commonly accepted that traditional medicines and pharmaceuticals function as biochemical enhancers. Behavioral augmentation is contingent upon lifestyle modifications, such as those associated with meditation exercises. In contrast, non-invasive and invasive brain stimulation techniques are classified as physical augmentation technologies. As Jangwan et al. observe,

[4] F. P. Adorno, *The Transhumanist Movement*, Palgrave Macmillan, Cham 2021.

[5] N. S. Jangwan et al., *Brain Augmentation and Neuroscience Technologies: Current Applications, Challenges, Ethics and Future Prospects*, in *Frontiers in System Neuroscience*, 16, 2022, 1000495, doi: 10.3389/fnsys.2022.1000495.

ISSN 0039-4130

> «While invasive methods such as DBS [Deep Brain Stimulation – N.d.R.] have been shown to improve cognition in subjects with pathological conditions, several allegedly noninvasive stimulation strategies, including electrical stimulation methods [...] are increasingly used on healthy subjects» [6].

This observation, which differentiates between cyber-humans with and without health conditions, assumes another classification criterion: the objective for which the technology is designed and employed [7].

Augmentative technology may be considered restorative in nature, whereby it is designed and employed with the objective of restoring physical and cognitive functionality in individuals who have experienced disease or injury, or in cases of disability. Examples of restorative technology include prostheses designed to restore the functionality of limbs, senses, and cognition [8]. Nevertheless, the objective of augmentative technology may also be to enhance or augment sensorimotor and cognitive performance. This category of technology, which may be referred to as "enhancement technology", is exemplified by both invasive and non-invasive brain stimulation techniques [9].

Invasive techniques include brain-computer interfaces and deep brain stimulation (DBS). Transcranial electrical stimulation (tES) is a widely utilized technique that is typically regarded as non-invasive [10]. tES entails the attachment of electrodes to the scalp to administer a modest direct or alternating current (typically 1–2 mA in intensity) for up to 30 minutes. A brain-computer interface (BCI) is a nano-implant that receives signals and transmits them wirelessly to extracorporeal computers or smartphones. In contrast, deep brain stimulation (DBS) involves the implantation of neu-

[6] *Ibid*. pp. 4-5.

[7] C. Cinel, D. Valeriani and R. Poli, *Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects*, in *Frontiers of Human Neuroscience*, XIII, 13, 2019, doi: 10.3389/fnhum.2019.00013.

[8] In her work, De Preester puts forth a three-class partition of prosthetic technology, distinguishing between prosthetic technologies that are primarily geared towards restoring motor, sensory, and cognitive functionality. See H. De Preester, *Technology and the Body: The (Im)Possibilities of Re-embodiment*, in *Foundations of Science,* 6, 2011, pp.119-137.

[9] The concept(s) of (non-)invasiveness has (have) no fixed definition in the scientific literature. The concept of invasiveness, and indeed that of non-invasiveness, is not defined in a fixed manner within the scientific literature. Recent research underscores the normative dimension inherent to the concept(s), particularly in regard to the perceived dangers, intrusions, and disruptions associated with them. For a comprehensive conceptual analysis, refer to E. Klein, *What Does It Mean to Call a Medical Device Invasive?* in *Medicine, Health Care and Philosophy*, 2023, 26, pp. 325-334.

[10] Some have expressed reservations regarding the non-invasiveness of tES. See T. Reed and R. Cohen Kadosh, *Transcranial Electrical Stimulation (tES) Mechanisms and its Effects on Cortical Excitability and Connectivity*, in *Journal of Inherited Metabolic Disease*, 41, 2018, pp. 1123-1130.

ISSN 0039-4130

rostimulators in specific regions of the brain, which emit electrical pulses to disrupt neuronal activity at the targeted locations. Due to its invasive nature, financial costs, and the ethical concerns associated with its use, enhancement technology is currently employed exclusively in medical contexts to improve patient quality of life. Perhaps the most striking examples are brain-computer interfaces that compensate for language deficits by using methods like text editing and speech synthesis [11] and brain-spine interfaces that bypass a spinal cord injury to restore walking [12].

The third pivotal criterion for the classification of augmentative technology pertains to the human body and the ways in which technology interacts with it. Warwick differentiates between augmentative technology that is situated in close proximity to the human body but not integrated into it, augmentative technology that is implanted into the body but not the brain/nervous system, and augmentative technology that is linked directly to the brain/nervous system [13]. Building upon this classification, Barfield and Williams put forth a more nuanced distinction encompassing four categories: general external augmentative tools (e.g., limb prostheses), implanted technology (e.g., biometric sensors), brain augmentative tools (e.g., DBS and brain-computer interfaces), and exoskeletons and mobility aids [14].

As will be demonstrated in the subsequent section, the scientific discourse on the ethical and social implications of augmentative technology is often conflated with a debate concerning the potential risks associated with brain augmentation. The holistic conceptualization of the cyber-human as an embodied subject is often overlooked in favor of a reductionist view that identifies the human person with its mind/brain. This neurocentric bias hinders the comprehensive endeavor of harmonizing technological advancement with humanistic values.

3. *The neurocentric bias*

The most recent World Congress for NeuroRehabilitation (WCNR 2024), held in Vancouver, revealed deficiencies in the reflection on the

[11] N. Birbaumer et al., *A Spelling Device for The Paralysed*, in *Nature*, 398, 1999, pp. 297-298.

[12] H. Lorach et al., *Walking Naturally After Spinal Cord Injury Using a Brain-Spine Interface*, in *Nature*, 618, 2023, pp. 126-133.

[13] K. Warwick, *Homo Technologicus: Threat or Opportunity?* In *Philosophies*, 1, 2016, pp. 199-208.

[14] W. Barfield and A. Williams, *Cyborgs and Enhancement Technology*, in *Philosophies*, IV, 2, 2017, doi:10.3390/philosophies2010004.

ISSN 0039-4130

impact of the current generation of augmentation technology [15]. In this comprehensive research and innovation area, the ethical and social concerns that have been identified are almost exclusively oriented towards the field of neurotechnology and brain augmentation. The research is focused on the mind/brain, which represents a single dimension of the cyber-human. It would appear that the broader context is being overlooked. The scholars have called for a reinvigorated approach to technology design and use in clinical settings, one that aligns more closely with the fundamental understanding of humanity as a complex existential condition, rendering each person a singular and irreproducible value.

This appeal for humanism aims to address the prevailing approach to research and innovation in the field of augmentation technology, as well as its practical applications. This approach is characterized by what we term a "neurocentric bias". In their 2021 recommendations, the Neurotechnology Ethics Taskforce (NET) explicitly identifies this pervasive assumption that risks reducing the cyber-human to a mere brain or the instantiation of cognitive functions. In considering the challenges posed to personal identity by BCI and DBS, the NET notes that

> «While the body and its functions are significant for the understanding of the individual's narrative identity, their psychological states more directly provide the interpretive frames through which their experiences are comprehended, and their narratives are shaped. Changing these psychological states, then, potentially more fully transforms the narrative identity of the person. Features sometimes ascribed to personality [...] may be altered through neural interventions» [16].

The argument seems clear and logical, presented in a way that suggests it is self-evident. The mind/brain is the entity that endows experiences with meaning for the individual undergoing them. Consequently, monitoring and manipulating the mind/brain may be perceived as a threat to a stable and enduring sense of self or personal identity. In essence, mind/brain alteration affects the constructive process of identity formation at both the individual and relational levels.

The argument for neurocentrism may be expanded to encompass additional issues, including those pertaining to agency, the privacy of brain data, and cognitive and social biases. As evidenced by the NET's

[15] Cf. https://wfnr-congress.org/.

[16] S. Goering et al., *Recommendations for Responsible Development and Application of Neurotechnologies*, in *Neuroethics*, 14, 2021, pp. 365-386, p. 369.

ISSN 0039-4130

2021 recommendation paper, these issues are interrelated and require collective attention. In light of the centrality of the mind/brain to subjective sensemaking, it is imperative that the purported risks associated with augmentative technology and the strategies to overcome them be conceived with almost exclusive reference to this dimension of the cyber-human. From an ethical and social perspective, the cyber-human must be reduced to her mind/brain. It is not a coincidence that, for the NET, augmentation overlaps with a subcategory within the overarching topic of neurotechnology and brain augmentation [17]. However, it is worth questioning whether this partial picture is an inevitability for the cyber-human. In a recent article, Ardaillon and colleagues maintain that this is not (and should not be) the case [18].

4. *Reevaluating corporality*

In their viewpoint paper, Ardaillon et al. recall two strategies that human-centered studies have recently pursued to identify, characterize, and displace the dominant view of the cyber-human as a technology-enhanced mind/brain. The initial strategy is neuroskepticism. This is an epistemological stance that questions the validity, utility, or safety of neuroscience. In this context, skepticism can be interpreted in a number of ways. It may be regarded as a repudiation of the naturalistic paradigm that equates humans with minds and minds with computational machines or metabolic organs. An alternative interpretation is that it facilitates interdisciplinary communication. This would entail situating the neuroscientific discourse within a broader and multifaceted theoretical framework that aims to comprehend not only brain mechanisms and mind functions, but also the profound «mystery of human subjectivity» [19]. In simple terms, according to this interpretation, neuroskepticism facilitates interdisciplinary collaboration, which is how neuroscience can grasp the human person as a whole, beyond the neural structures and functions of the mind/brain.

The second strategy for contrasting biased research in the field of augmentation technology is closely associated with the construc-

[17] *Ibid*. pp. 375-377.

[18] H. Ardaillon et al., *Striking the Balance: Embracing Technology While Upholding Humanistic Principles in Neurorehabilitation*, in *Neurorehabilitation and Neural Repair*, 38, 2024, pp. 705-710.

[19] E. Husserl, *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*, translated by D. Carr, Northwestern University Press, Evanston 1989, p. 3.

ISSN 0039-4130

tive path of neuroskepticism. It is recommended that interdisciplinary communication be pursued in order to integrate disciplines such as philosophy, ethics, psychology, and sociology into technology research programs. This integration would facilitate the gathering and analysis of qualitative datasets that can reveal phenomena such as affect, emotion, value, and norms that cannot be directly observed or reduced to experimentally reproducible setups. The integration of qualitative research in augmentation technology necessitates the establishment of a balance between the scientific and humanistic perspectives. The challenge is to determine the most effective means of implementing this balance within the context of ongoing technological advancement.

Ardaillon and colleagues put forth the proposition that the nexus between augmentation technology and humanism may be identified in the formulation of tangible actions that integrate humanistic values into the design and utilization of technology. Furthermore, they highlight the advantages of participatory design, which entails the active involvement of patients and caregivers in the creation and assessment of novel technologies. Furthermore, they highlight the significance of education, particularly the necessity of comprehensive training for therapists and clinicians that encompasses both technological and humanistic approaches. This is a comprehensive action program that necessitates significant financial and temporal resources, as well as scientific and educational expertise and capabilities.

In order to advance in the direction proposed by Ardaillon and colleagues, the article recommends cultivating a novel perception and practice of the body. This is done in order to provide contrast to the particular form of alienation from the somatic substrate that is distinctive of the cyber-human. It is imperative to prioritize bodily experience and the multifaceted ways in which the body is manifested, expressed, and performed in scenarios where humans and technologies are profoundly intertwined[20]. However, the question of which content and format are most conducive to achieving these objectives remains open, as does the question of which institutions, scientific and educational programs are best positioned to facilitate a new centrality of the body in the cyber age.

[20] Cf. M. Properzi, *The Technology-Supplemented Bodily Subject. A Case Study in Biomimetic Prosthetics*, in *Rivista di Estetica*, 87, 2024, in press.

ISSN 0039-4130

Unlocking the Soul: AI and Neuroscience Insights into Spirituality

Helga Martins[1], *Joana Romeiro*[2], *Sílvia Caldeira*[3]

This paper explores the intersection between artificial intelligence, neuroscience, and spirituality. To begin, let's explore the roots of the word "spirituality". Etymologically, it derives from the Latin word *spiritus*, which means "soul, courage, vigor, breath, or life force" (Lepherd, 2015). This origin reflects the profound and intrinsic connection between spirituality and our very essence of being.

Spirituality is not confined to a single definition (Koenig, 2012). Instead, it represents an individual experience that is both complex and universal (Weathers et al., 2016) Spirituality is wrapped with terms associated with meaning in life, connection and transcendence experience (Murgia et al., 2020; Weathers et al., 2016).

One of the most widely accepted definitions of spirituality comes from Best et al. (2020), who describes spirituality as a dynamic dimension of human life that relates to the way persons (individual and community) live and experience life, express and/or seek meaning, purpose

[1] Helga Martins, Pos-doctoral Fellow at Integral Human Development Program, PhD, RN, Universidade Católica Portuguesa, Doctoral School (CADOS). Faculty of Health Sciences and Nursing, Centre for Interdisciplinary Research in Health, Lisbon, Portugal. Instituto Politécnico de Beja, Escola Superior de Saúde. R. Dr. José Correia Maltez, 7800-111 Beja, Portugal. ORCID: https://orcid.org/0000-0001-5804-7934 E-mail: hemartins@ucp.pt.

[2] Joana Romeiro, Pos-doctoral Fellow at Integral Human Development Program, PhD, MSc, RN, Universidade Católica Portuguesa, Doctoral School (CADOS). Faculty of Health Sciences and Nursing, Centre for Interdisciplinary Research in Health, Lisbon, Portugal. ORCID: https://orcid.org/0000-0001-8867-2183. E-mail: jromeiro@ucp.pt.

[3] Sílvia Caldeira, PhD, MSc, RN, Associate Professor, Universidade Católica Portuguesa, Faculty of Health Sciences and Nursing, Centre for Interdisciplinary Research in Health, Lisbon, Portugal. ORCID: https://orcid.org/0000-0002-9804-2297. E-mail: scaldeira@ucp.pt.

and transcendence, and the way they connect to the moment, to self, to others, to nature, to what is significant and/or the sacred. In addition, to further understand spirituality, it is helpful to explore its various dimensions.

These dimensions of spirituality are categorized into intrapersonal, interpersonal, and transpersonal aspects, as described by Fombuena et al. (2016). The intrapersonal spirituality focuses on the internal aspects of the Being. It involves self-reflection, personal growth, and the inner journey of discovering one's values, beliefs, and purpose. Intrapersonal spirituality is about the connection we have with our own inner self, leading to greater self-awareness and inner peace (Fombuena et al., 2016). The interpersonal spirituality emphasizes the connections we foster with those around us—family, friends, and the broader community (Fombuena et al., 2016). At last, the transpersonal spirituality involves a connection to something greater than oneself, whether it is a higher power such as a Deity, Nature, or Cosmos (Fombuena et al., 2016). It addresses the quest for ultimate meaning and the desire to connect with the transcendent or the divine (Fombuena et al., 2016).

Over the past two decades, there has been a concerted effort within healthcare to include the spiritual dimension in clinical practice to foster a holistic approach (Puchalski, et al., 2014). Currently, there are studies that confirm that spirituality plays a relevant role in health outcomes, particularly in coping with adversity, having a impact on fostering positive emotions, addressing depression, suicide, anxiety, psychotic disorder/schizophrenia, bipolar disorder, substance abuse, personality traits and social problems (Koenig, 2012).

Despite notable advancements in healthcare, exploring spirituality as a fundamental human dimension still requires further investigation and understanding in some particular aspects (Martins et al., 2017; Romeiro et al., 2018).

Regarding spirituality and neuroscience, we can identify several major milestones that have shaped our understanding. It all started in the 1960s-1970s with the electroencephalography studies to examine brainwave patterns associated with meditation and other altered states of consciousness. Researchers observed distinct brainwave patterns, such as increased alpha and theta waves, when individuals meditate. Next, in the 1980s Neurotheology gained greater significance due to the amazing work conducted by the neuroscientist Andrew Newberg who explored the relationship between neural processes and spiritual experiences. The pioneering work in this area aimed to understand

ISSN 0039-4130

how spiritual experiences might correlate with specific brain activity and structures. From 1990s to nowadays, improvements in neuroimaging technologies, including Functional magnetic resonance imaging and Positron emission tomography, have been delved and refined. These techniques enable researchers to track real-time brain activity and pinpoint brain regions associated with spiritual experiences.

As we continue to explore spirituality and neuroscience, we come across a number of well-known scholars who have significantly advanced the discipline. For instance, David Lewis-Williams is indeed famous for his work on the neuropsychological aspects of prehistoric art, particularly in his influential book "The Mind in the Cave: Consciousness and the Origins of Art". David Lewis-Williams was the first researcher to unveil a neuropsychological explanation. Also, António Damásio, a contemporary neuroscientist, has made significant contributions to our understanding of consciousness and the neural mechanisms underlying subjective experience. His work, includes books like "Descartes' Error" and "The Feeling of What Happens", also the book "Self Comes to Mind: Constructing the Conscious Brain"(Damásio, 2010) which investigates the role of emotions and feelings as integral to our cognitive processes and sense of self. For instance, António Damásio's work provides a highlight of the emergence of a conscious mind where the assimilation and reaction to "environmental images" occurs through a set of "internal images", in the process of regulating the "self", its needs, and its affections. Simultaneously, the evolution of consciousness enables the improvement and complexity of memory, language, and the ability to communicate (Damásio, 2010). Damásio's contemporary ideas, not only describes the conscious and non-conscious aspects of mental processes that ultimately lie at the foundations of the regulation of the "self", but, in a broader context, his ideas also relate to "sociocultural homeostasis" (Damásio, 2010, p.330). The "self" finds its balance from social interactions and in the cultural mechanisms developed towards creating a sense of groups and individual wellbeing.

While Lewis-Williams focuses on the historical and psychological aspects of early art, Damásio provides a broader framework for understanding consciousness and emotional experience from a modern neuroscientific perspective. Both researchers contribute to our understanding of the mind, though from different angles: one through the lens of ancient art and altered states, and the other through contemporary neuroscience and cognitive theory.

ISSN 0039-4130

At first glance, spirituality and AI might not appear related, but this emerging field holds great potential for expansion nonetheless. This raises many profound questions. For example, is AI leading us towards reducing human existence to a series of algorithms? Is AI capable of predicting the complexities of our human needs with precision? Can AI offer meaningful responses that genuinely alleviate human suffering?

Above all, we must mention the advantages and potential. For instance, AI and spirituality offer a unique convergence where science and inner experience meet. In addition, AI's ability to analyze vast data and recognize patterns could help identify states associated with spiritual experiences, offering insights into meditation, mindfulness, and human consciousness. In addition, AI makes it possible to develop algorithms that capture the essence of spiritual experiences. Furthermore, AI should complement, not replace, human decision-making, enhancing diagnostics, treatment options, and holistic care.

Delving deeper into the topic of suffering, Steeves and Khan (1986) argued that meaning plays a crucial role in shaping how individuals perceive suffering and cope with it, reinforcing the significance of "The Meaning Theory", which is connected to Logotherapy, also known as the "Psychotherapy of the Meaning of Life" (Moreira, 2011). In this context, a key question arises: how can AI grasp and integrate the evolving complexity of human thought, particularly the abstract, metaphorical, spiritual, and transcendent dimensions of human existence?

Spirituality is deeply subjective and personal, and these two traits highlight the challenge in research and in merging AI specificities. AI and Neuroscience may unlock the possibility of understanding spirituality in patients in the healthcare setting. These new technologies open up a new set of opportunities to give a more concrete understanding of the spiritual realm. There is a deep need to unlock the soul and gain a more comprehensive understanding of spirituality, with the help of AI and neuroscience. However, AI and Neuroscience brings us challenges in particular regarding misrepresentation in the decoding of spirituality since it is a subjective and personal piece of the individual. Furthermore, ethical dilemmas emerge when addressing matters of spiritual belief, which is a private matter of the individuals.

As technological and scientific advancements progress, particularly in the realm of artificial intelligence, there is a concerning tendency to go back to a biomedical reductionist view of the individual. This perspective contrasts sharply with the demands of the holistic movement,

ISSN 0039-4130

as noted by Rogers (1992), which emphasizes the importance of considering the whole person. In addition, this reinforces the importance of acknowledging the positive impact that comes from a unitary and integral human development approach to one's health and wellbeing (Barrett, 2010; Papathanasiou et al., 2014). To conclude, Lewis-Williams brings an interesting and wise perspective since he once said, «We do not have to explain everything in order to explain something» (2002, p. 7).

References

S. Barrett, *The integral human development approach: Implications for health and well-being*, 2010.

M. Best et al., *An EAPC white paper on multi-disciplinary education for spiritual care in palliative care*, in *BMC Palliative Care*, XIX, 9, 2020. https://doi.org/10.1186/s12904-019-0508-4

A. Damásio, *Self comes to mind: Constructing the conscious brain*, Pantheon Books, 2010.

M. Fombuena - L. Galiana - P. Barreto - A. Oliver - A. Pascual - A. Soto-Rubio, *Spirituality in patients with advanced illness: The role of symptom control, resilience, and social network*, in *Journal of Health Psychology*, XXI (12), 2016, 2765–2774. https://doi.org/10.1177/1359105315586213

D. L. Kahn - R. H. Steeves, *The experience of suffering: Conceptual clarification and theoretical definition*m in *Journal of Advanced Nursing*, XI (6), 1986, 623–631. https://doi.org/10.1111/j.1365-2648.1986.tb03379.x

H. G. Koenig, *Religion, spirituality, and health: The research and clinical implications*, in *ISRN Psychiatry*, 2012, Article 278730. https://doi.org/10.5402/2012/278730

L. Lepherd, *Spirituality: Everyone has it, but what is it?*, in *International Journal of Nursing Practice*, XXI (5), 2015 566–574. https://doi.org/10.1111/ijn.12285

D. Lewis-Williams, *The mind in the cave: Consciousness and the origins of art*. Thames & Hudson, 2002.

H. Martins - J. Romeiro - S. Caldeira, (2017). *Spirituality in nursing: An overview of research methods*, in *Religions*, VIII (10), 2017, 226. https://doi.org/10.3390/rel8100226

A. Moreira, *The meaning theory: A study of Viktor Frankl's logotherapy and its applications*, 2011.

C. Murgia - I. Notarnicola - G. Rocco - A. Stievano, *Spirituality in nursing: A concept analysis*, in *Nursing Ethics*, XXVI (5), 2020, 1327-1343. https://doi.org/10.1177/0969733020909534

A. Newberg - E. D'Aquili - V. Rause, *Why God won't go away: Brain science*

and the biology of belief, Ballantine Books, 2001.
I. V. Papathanasiou et al., *A unitary and integral approach to human health and well-being: A review of the literature*, 2014.
C. M. Puchalski - R. Vitillo - S. K. Hull - N. Reller, *Improving the spiritual dimension of whole person care: Reaching national and international consensus*, in *Journal of Palliative Medicine*, XVII (6), 2014, 642-656. https://doi.org/10.1089/jpm.2014.9427
M. E. Rogers, *Notes on the future of nursing: A unitary perspective*, in *Nursing Science Quarterly*, V (1), 1992, 8-14.
J. Romeiro - H. Martins - S. Pinto - S. Caldeira, *Review and characterization of Portuguese theses, dissertations, and papers about spirituality in health*, in *Religions*, IX (9), 2018, 271. https://doi.org/10.3390/rel9090271
E. Weathers - G. McCarthy - A. Coffey, *Concept analysis of spirituality: An evolutionary approach*, in *Nursing Forum*, LI (2), 2016, 79-96. https://doi.org/10.1111/nuf.12128.

ISSN 0039-4130

Artificial Intelligence and the Question on Ethico-Moral Algorithmic Representation

Justin Nnaemeka Onyeukaziri

1. *Introduction*

As the science and design of artificial intelligence (AI) systems advance, the philosophy of AI and cognition in general becomes more cogent. This paper is an interrogation within the scope of the philosophy of AI and the science of cognition in general. It considers the question of moral and ethical knowledge, its representation, and processing or manipulation in cognitive systems, natural or artificial. Hence, at the heart of the problematic in this discourse are these questions: How is moral knowledge represented in humans? Can moral knowledge be represented in AI systems? In other words, the question is: Is moral and ethical knowledge automation possible?

J. N. Onyeukaziri, in a paper entitled, *Action and Agency in Artificial Intelligence: A Philosophical Critique*, interrogates the question of the notions of action and agency in AI systems and contends that: «AI systems do not and cannot possess free agency and autonomy, thus, [they] cannot be morally and ethically responsible» [1]. This paper has an epistemic and cognitive presupposition that there is a clear and certain knowledge of how moral and ethical knowledge is known, represented, and processed in human systems. Based on this presupposition, the aforementioned paper focused on the phenomena of free agency and autonomy in humans in relation to the question of moral and ethical responsibility in AI systems.

[1] J. N. Onyeukaziri, *Action and Agency in Artificial Intelligence: A Philosophical Critique*, in *Philosophia: International Journal of Philosophy*, XXIV, 1, 2023, pp. 73-90.

Moral knowledge is one of the implications of the rational capacity—which implies intelligence. One of the consequences of human intelligence is the ability to know moral good and bad, which is complemented by the ability to execute moral and ethical actions. As cognitive research on non-human intelligence progresses, one of the evolutionary distinctions of humans is the intelligence for moral and ethical knowledge, formulations, and judgment. Only the human race has been able to establish moral institutions and enact ethical codes (though some contemporary scholars such as Patricia S. Churchland[2] are advancing a claim for moral intuitions and capability in animals, especially primates, but only humans have been able to establish moral, ethical and legal institutions). This paper deals with what precedes moral volition, which is the question of moral intelligence (the representation and formulation of moral knowledge).

Thus, this paper is a discourse on the question of ethical and moral algorithmic representation in artificial intelligence (AI) systems. Hence, it raises questions that border on moral metaphysics and ethical epistemology, such as free agency and ethical determinism on one hand and moral apprehension and ethical cognition on the other hand. This paper argues that considering the metaphysical nature of free agency in the intrinsic relations between reason and desire in moral cognitive operations, at the root of ethical and moral actions, the question of the algorithmic representation of the human capacity for ethical and moral operations in AI systems is a possibility that cannot be automatized in AI systems.

2. *Investigation into Ethico-Moral Representations in Humans*

The ethico-moral question has been a central aspect of philosophy. It does not only deal with the question of *the good* one ought to do or *the bad* one ought not to do; more importantly, central to this question is the nature of the good *eo ipso*. The investigation of the nature of the good *per se*, from a philosophical angle, can be both metaphysical and epistemological. The nature of the moral action can be both neurological and psychological from a cognitive science perspective. But can it be both algorithmic and computational? This is the central problematic of this paper.

[2] See P. S. Churchland,*Conscience: The Origins of Moral Intuition*, W. W. Norton & Company, New York 2019.

ISSN 0039-4130

On the metaphysical nature is the question: What is the *quiddity* or essence of the good (action)? That is to say, what makes an action morally good? This metaphysical question, because it is necessarily anthropological, has both a socio-political implication and a legal *cum* jurisprudential implication because it raises the question: Why ought a person act morally or ethically? The metaphysical nature of the question was taken seriously by metaphysical conscious philosophers, such as the Classical and Scholastic philosophers. This is unlike in contemporary philosophical discourse, in which most philosophies attempt to be consciously ripped of metaphysis as the investigation of the essence of things. Hence, the attempt to "scientificize" moral questions by seeking a naturalistic ground for ethico-moral explications, an approach in philosophy that began during the modern period in Western philosophy is now assumed to be the epistemic norm for "philosophy" and ethics in particular. But it will be appropriate to maintain, therefore, that this is a change in the question of the nature of the good, from metaphysics to epistemology. For the epistemological nature of the good, the question is: How do we know the good (action) and choose to do or not to do the good action? Attempts to proffer answers to this question, have developed the different kinds of ethics or ethical approaches studied today. For instance, there are virtue ethics, divine law ethics, natural law ethics, deontological ethics, consequentialist ethics, utilitarian ethics, care ethics, and others.

On the metaphysics of the good action or morality, for Plato, it is important to enquire into the essence of a thing-including concepts and notions. In several of the dialogues of Plato, Socrates engages different persons in seeking the essence of beliefs they hold and concepts they employ in ordinary conversations, without a critical examination of what they actually are. A good example is the question of *piety* in the dialogue *Euthyphro*. Euthyphro was accused by his relatives of being *impious* for deciding to prosecute his own father in a court of law for killing their household slave. Euthyphro, on the contrary, maintains that his relative's belief in *piety* is wrong. This raises the philosophical questions in the dialogue: What is piety? How does one come to the knowledge of the belief of *piety*?

First of all, Euthyphro believes that piety or to be pious is to prosecute the wrongdoer, irrespective of who the person is and what the wrong action is, and to do the contrary is impious and not piety[3]. Euthyphro

[3] Plato, *Euthyphro*, in *Plato: Complete Work*s, edited by Cooper M. John, Hackett Publishing Company, Inc., Indianapolis 1997, 5d-e.

ISSN 0039-4130

based his belief on what piety is, on both the law and religious story on the god Zeus who punished his own father for unjustly swallowing his son[4]. In response, Socrates questions the truth-value or authenticity of the religious story on which he based his knowledge of piety, on which he is strongly convinced that the act of murder by his own father is an impious act[5]. More importantly, in questioning the truth-value or authenticity of religious stories as bases for the knowledge of piety, Socrates is interested in the *quiddity* of piety or a pious action-what makes a pious action pious. In other words, Socrates seeks to engage in the metaphysics of piety, as in the ultimate principle or cognitive model of piety, by which every pious action can be understood.

For want of providing adequate proof as in mathematics by Euthyphro for basing the knowledge of the essence of piety on the gods, appeal to the gods or religious account was dismissed as lacking epistemic truth-value for the knowing and judging of moral and ethical actions. In seeking a metaphysical essence or a cognitive model for each moral action, Socrates-Plato attempts to formalize or mathematicize moral and ethical knowledge. Hence, as in several dialogues, such as in *Apology*, Socrates maintains that human reason or rationality should be the ground and final arbiter for moral and ethical knowledge and judgment.

For Aristotle, morality or ethics is and should be based on reason; thus, it is a rational activity. However, his focus is not on the metaphysics of the good or the moral-ethical act but on being actually a moral or an ethical person[6]. Aristotle maintains a necessary connection between the good/end of a thing and its *quiddity*. Hence, since rationality or the possession of a rational soul is the *quiddity* of the human person, he contends that the ultimate good of the human person necessarily is a rational activity of the soul. The human person has many goals, but he maintains that the ultimate good/goal is eudaimonia—happiness, well-being, or flourishing, which has to do with the rational activity of the soul in accordance with virtue or excellence. Important to the discourse of this paper is his assertion that virtue or excellence, both as intellectual and as moral excellence, is not innate but is both learned by teaching or by habit[7].

[4] *Euthyphro*, cit., 6.

[5] *Euthyphro*, cit., 6c.

[6] See Aristotle, *Nicomachean Ethics*, in *The Complete Works of Aristotle*, vol. 2, edited by Jonathan Barnes, Princeton University Press, New Jersey 1984, 1103b, 25-30.

[7] See *Nicomachean Ethics*, cit., 1103a, 15-25.

ISSN 0039-4130

Even more important is his claim that humans become moral, ethical, or virtuous by doing moral, ethical, or virtuous acts. Hence, for Aristotle, the moral good, or the excellent or virtuous act, is to act in accordance with the right reason—which is the mean between two extremes, excess and defect [8]. Fundamental to the existence of the extremes, excess, and defect is the existence of the extremes of pleasure and pain [9]. This follows that the natural desires for pleasure and pain are central to moral calculus—moral representation and computation. [10] Hence, pleasure and pain determine human action toward the good or the bad, but they do not define or proscribe the moral good or bad.

3. *The Question of Ethico-Moral Representations in Artificial Intelligence*

According to Allen Newell and Herbert A. Simon: «Since ability to solve problems is generally taken as a prime indicator that a system has intelligence, it is natural that much of the history of artificial intelligence is taken up with attempts to build and understand problem-solving systems» [11]. Critical in this statement are the notions of *intelligence* and *problem-solving systems*. They fundamentally connect intelligence to problem-solving. The human race has persisted or survived by its ability to solve diverse existential and cognitive problems. A reduction of the notion of intelligence to problem-solving will no doubt open the door of intelligence to other living systems, plants, and other animals. The science and technology of AI attempt to open the door of intelligence even wider to non-biological systems. Today, there are several non-biological systems designed and developed by human beings that solve problems. They are called artificial intelligence systems (AI systems).

One kind of problem-solving demonstration that hitherto has been exclusively attributed to the human person is the knowing, processing, and solving of moral problems. The capability for moral apprehension, discernment, and judgment, thus, has been marked as one of the main

[8] See *Nicomachean Ethics*, cit., 1104a, 1-30.
[9] See *Nicomachean Ethics*, cit., 1104b, 5-10.
[10] See *Nicomachean Ethics*, cit.,1105a, 10-15.
[11] A. Newell and H. A. Simon, *Computer Science as Empirical Enquiry: Symbols and Search*, in *The Philosophy of Artificial Intelligence*, Margaret A. Boden (edited by), Oxford University Press, Oxford 1990[1976], p.120.

ISSN 0039-4130

specific differences in human intelligence. This is notwithstanding hypothetical claims in the contemporary field of neuroscience of ethics and neurophilosophy that attributes morality to other primates [12]. As the design and development of AI systems advance, the question that is central to this paper is: Whether moral apprehension, discernment, and judgment can be automated? Simply put, this paper examines the questions: What constitutes moral intelligence, and how does moral intelligence operate? And whether or not the operation of moral intelligence can co-exist both in the human intelligence systems and AI systems. These questions challenge the claim for the specificity of ethico-moral representation and knowledge in the human person. Or, is it the case that the representation of ethico-moral actions is of the kind that cannot be automated or computationalized? Thus, it is important to examine general knowledge representation and the automatization and computationalization of knowledge of moral actions.

However, the question of the representation of desires and beliefs that determine the outcome(s) produced by moral and ethical action(s) is of the greatest importance in this paper. For, all moral and ethical theorists seem to be in agreement that moral and ethical actions are determined by agents' desires and beliefs. Between the two determinants: desires and beliefs, it seems desires could be reduced to beliefs, since desires in their final analysis produce information, while beliefs in themselves are information coded in the mind. Both biological and psychological desires produce neurological information in the brain/mind. Hence, information theorists could argue that the determinant of the representation of moral knowledge deals with beliefs. If this is the case, the question is: Can all moral beliefs be computationalized? Theoretically, any phenomenon that can be formalized can be computationalized, and any phenomenon that can be computationalized can be automated by AI systems. Whatever phenomenon can be explained or thought of in a coherent and systematic manner can be logicalized and thus can be formalized. If moral knowledge is represented in the human mind/brain not as that inherently in the nature of the mind but as the accumulation of moral instructions through socio-cultural education, then a case can be made for the computationalization of moral knowledge or belief.

[12] See P. S. Churchland, *Conscience: The Origins of Moral Intuition*, New York: W. W. Norton & Company, 2019; Patricia S. Churchland, *Braintrust: What Neuroscience tells us about Morality*, Princeton University Press, New Jersey 2011.

ISSN 0039-4130

4. *Critique of Ethico-Moral Automation*

Moral and ethical knowledge is obtained by teaching. Though the mind is designed to store and process data of moral and ethical knowledge, they are not, by nature, contained in the mind. Hence, humans learn to be moral and ethical; we learn to do good and avoid bad based on the moral and ethical knowledge at our disposal. Hence, the questions that need to be interrogated are: Is it only human systems that can be morally and ethically coded by external agents? Can non-human biological systems be morally and ethically coded? Can non-biological, that is, artificial systems, be morally and ethically coded? In other words: Can there be moral and ethical automation? The answer to these questions following the discourse above depends on two factors: The first concerns the nature of the human person, and the second concerns the nature of symbolic representation or the question of representationalism in AI and in cognitive science in general.

Morality or ethics, as properly investigated by Aristotle, is a practical science. This is because it involves the manifestations of humans as social and political animals. Hence, it is not an abstract science that deals with demonstration from or to first principles. This is because at the heart of morality and ethics, human rationality and human desires, will, and passions are equally present. Thus, desires, pleasure, and pain are two fundamental principles that humans share with other animals. Morality and ethics are actually special operations of humans whereby humans demonstrate their difference from other animals: The assertion of reason as master over desires for pleasure and against pain. Hence, morality and ethics are actually the rational management of the animalistic nature of pleasure and pain in humans. Hence, all moral principles or ethical theories, one way or the other, centrally deal with desire—biological or psychological.

In the human person, it is not only (moral) beliefs that are represented in the mind/brain; desires are also represented. In fact, it could be argued that one or more desires accompany every belief in the human person. Hence, it could be argued that there are no pure moral beliefs. So, if the human mind in a general sense is a computational system, as Newell, Simon, Marvin Minsky, and other AI enthusiasts and computationalists in cognitive science argue, the computation of moral knowledge in the human system is *sui generis*. For example, in the *mind* of Euthyphro exposed above, the computation of the moral knowledge of *piety* does not only deal with the representations and manipulation

ISSN 0039-4130

of moral beliefs in the Greek gods and religious mythologies thought to him, but it also deals with his desire of filial love to his father and his entire family and relatives aimed at constraining him not to prosecute his own father. The computation of these struggles in the mind of Euthyphro is only unique to the human person; even other animals that are also equipped with the possession of the passion of pleasure and pain cannot have this kind of computation. AI systems cannot have this kind of moral computation of the human person because of the unique nature of the human person: Situated, embodied, and dynamic.

As Karol Wojtyła argues, to be a human person is to be a moral or an ethical being [13]. Morality and ethics necessarily need reason or rationality, but they do not exclusively need reason or rationality. They also need the entire self-conscious experience of every individual human person. The moral experience of every individual human person is different. And this is why morality and ethics demand individual autonomy and responsibility. This explains why morality and ethics are always connected to the questions of free will, free agency, and autonomy, without which no one should and can be held responsible for any action. This is why AI systems cannot be held responsible for any action but especially for any moral or ethical action.

5. *Conclusion*

This paper has attempted to investigate the question of ethical and moral algorithmic representation in artificial intelligence (AI) systems. It maintains that there is a uniqueness in the representation of moral knowledge that distinguishes it from other forms of knowledge representation. This uniqueness is that moral knowledge represented in the human mind is not pure rational belief; it inherently contains the representation of certain desires of passion and pain.

Based on two fundamental factors: first, on the nature of human mental representation and operation of ethico-moral knowledge, and second, on the critique of representationalism in AI research and cognitive science in general, this paper argues that the question of the algorithmic representation of the human capacity for ethical and moral operations in AI systems is a possibility that cannot be automatized in AI systems.

[13] See, Karol Wojtyła, *The Acting Person*, trans., by Andrzej Potocki, D. Reidel Publishing Company, Dordrecht 1979.

ISSN 0039-4130

About some Characteristics of Contemporary Discourses on Converging Technologies

Fernand Doridot

1. *Introduction*

Inquiring into the future of human freedom and humanism within the context of advancements in neuroscience and artificial intelligence naturally leads to a questioning of the notion of "technological convergence" as described and advocated by certain observers. Is there, within the technologies developed in our era, and more broadly in emerging technologies, a common trend and an underlying joint program that harbors the seed of irreparable harm to what has thus far defined our shared humanity? Despite their differences, do contemporary sciences and technologies (such as neuroscience and artificial intelligence) share a set of methods, objectives and metaphysical presuppositions, which ultimately bring them together, and make them converge towards a possibly anti-humanist goal? This concern has already been raised by various commentators, particularly in Europe, in response to the publication over twenty years ago of the seminal report *Converging Technologies for Improving Human Performance* by Rocco and Bainbridge [1], purportedly heralding the age of "NBIC" (for Nano-Bio-Info-Cogno) convergence. Now, twenty years later, as various works offer a current assessment and critical review of the forecasted technological convergence, it is worthwhile to examine how the "grand visions" that had infused this report and the critical discourses it had provoked have evolved. In this article, we will attempt to provide various evaluative

[1] M. C. Roco - W. S. Bainbridge (edited by), *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, Springer, Dordrecht 2003.

elements of this broad question, particularly in light of how the agenda of technological convergence, while having become more subdued in the West, encounters distinct interest in other regions of the world, where it is viewed through the lens of different scientific and philosophical traditions. Do emerging technologies lead us somewhere, and is this destination conducive to humanity's well-being? Ultimately, this is what we will attempt to provide some answers to.

2. *About some of the main features of the NBIC report*

The report *Converging Technologies for Improving Human Performance*[2], edited by Mihail C. Roco and William Sims Bainbridge, was first published in 2002 under the aegis of the National Science Foundation (NSF) and the US Department of Commerce. It presents a forward-looking vision of the convergence of four major scientific fields: nanotechnology, biotechnology, information technology and cognitive science (collectively known as "NBIC"). It anticipates a revolutionary fusion of these disciplines, leading to significant advances in science and technology. The report proposes the establishment of a unified framework for systems relating to matter, energy and biology, while stressing the importance of governance that anticipates the societal implications. Its main objective is to improve human performance, with a particular focus on health, energy efficiency and cognitive abilities.

Among other things, the report envisages the advent of a society of "total communication", marked by direct interaction between human brains and machines. The general conception of the human being expressed in this report is extremely reductionist; man is reduced to his genes and neurons, and envisaged as an informational machine. The report also advocates improving individual performance through technology as part of a highly individualistic vision of society. The transhumanist influence of this report is palpable, even if concepts such as the technological "singularity", popularised in particular by Ray Kurzweil, are not mentioned explicitly.

In a passage that has gained some notoriety, the report prophesies that, given the right decisions and investment, radical transformations could materialise in the space of twenty years or so, leading to a golden age of world peace, universal prosperity and evolution towards a higher

[2] *Ibid.*

ISSN 0039-4130

level of compassion and fulfilment. He even envisages the future of humanity as that of a "distributed and interconnected brain", enhancing both the productivity and independence of individuals, while offering them new opportunities to achieve their personal goals.

3. *About initial criticisms of the NBIC report*

The ambitions expressed in the NBIC report quickly came in for a great deal of criticism, particularly in Europe. These criticisms focused mainly on the links between the report and the transhumanist project, the expected impact on human identity and respect for nature, the accessibility and fair distribution of the expected benefits, and the risks inherent in the technologies mentioned in terms of the potential for dual use and widespread surveillance. The French philosopher Jean-Pierre Dupuy has expressed concern[3] that the project thus attributed to nanotechnologies is the expression of a radically new "metaphysical research programme", guided by unverifiable hypotheses about the structure of the world. Dupuy traces the origins of this programme precisely to a lecture given in 1948 by John von Neumann, who advocated a "bottom-up" approach to complexity, focusing on what structures can achieve rather than on how to control them. Dupuy points out some of the major philosophical risks associated with this programme, including the redefinition of nature and life, and the potential challenge to key concepts such as transgression, external reality and human consciousness. According to Dupuy, the scientists of the future could create self-replicating systems that would blur the boundaries between the natural and the artificial, heralding a reshaping of life itself.

Europeans as a whole finally felt the need to distinguish themselves from the prospects outlined in the NBIC report. This was expressed in a kind of "counter-report" written in 2004 by the German philosopher Alfred Nordmann. This report, entitled *Converging Technologies: Shaping the Future of European Societies*[4], was commissioned by the European Commission to assess NBIC convergence. It warns against the speculative approach of the American report, and argues in favour of responsibility and sustainability. It differs from the American ap-

[3] J.-P. Dupuy, *Nanotechnologies*, in M. Canto-Sperber (edited by), *Dictionnaire d'éthique et de philosophie morale*, Presses Universitaires de France, Paris 2004, pp. 1319-1322.

[4] A. Nordmann, *Converging Technologies: Shaping the Future of European Societies*, Office for Official Publications of the European Communities, Luxembourg 2004.

ISSN 0039-4130

proach by giving pride of place to questioning the social and ethical implications of the innovations envisaged, by calling for caution with regard to the promises of human modification and the potential risks, and by recommending inclusive and fully transparent governance of the associated initiatives. Any idea of "enhancement" is carefully banished. The notion of "human performance" is replaced by a reference to the "socio-cultural reality of European societies", and the expression "engineering of the mind" is replaced by "engineering for the mind".

4. *About some relevant contemporary literature*

We propose here to assess the way in which, some twenty years after these initial developments, the concept of technological convergence, and the debates that accompany it, have spread and found an echo in different parts of the world. We approach this objective with a study of the academic literature that explicitly mentions this concept or claims to be based on it.

To appreciate the geographical extension of this concept, and the diversity of sensibilities with which it is approached today in different parts of the world, we have attempted to select academic articles representative of different geographical areas, including the USA, Europe, and the rest of the world. An abundance of literature is *de facto* available. Nevertheless, it is possible to quickly identify recurring themes and major approaches characteristic of the different cultural areas concerned. Starting with an initial selection of forty academic references, we were thus able to select a reduced panel of eight academic articles, sufficiently representative of different geographical and cultural approaches to enable us to draw some general lessons. It's impossible to ignore, or claim to have avoided, the obvious biases that can be associated with such an approach: a form of subjectivity in the choice of references, the risk of a reductive or caricatural approach, an illusion of exhaustiveness. Nevertheless, we believe that, in the necessarily restricted format of a work such as this, our selection is relevant and can be profitably analyzed.

Our panel includes:

a) A North American article[5] by M.C. Roco himself, evaluating the National Nanotechnology Initiative twenty years after its launch.

[5] M. C. Roco, *National Nanotechnology Initiative at 20 years: enabling new horizons*, in *Journal of Nanoparticle Research*, XXV, 2023, Art. 197, https://doi.org/10.1007/s11051-023-05829-9.

ISSN 0039-4130

b) An article [6] by an Italian researcher who defends the importance of the concepts of vulnerability and embodiment in the face of the development of emerging technologies.
c) An article [7] by a Turkish researcher dedicated to the paradoxes of the notion of freedom in the context of transhumanism, artificial intelligence, digitalisation and robotics.
d) An article [8] by several Iranian researchers on the definition of a national technology assessment framework for converging technologies.
e) An article [9] by two Burkinabe researchers questioning the potential of NBIC convergence to serve as a springboard for transhumanism and posthumanism.
f) An article [10] by a South African researcher questioning the contemporary advent of a new technological revolution or a fourth industrial revolution.
g) An article [11] by a Chinese researcher analysing the "ethical turn" of design practices in the era of the development of technosciences.
h) An article [12] by several Indian researchers exploring the possibilities for innovation in the coupling of AI and robotics.

[6] A. Fasoli, *Vulnerability, Embodiment and Emerging Technologies: A Still Open Issue*, in *Philosophies*, VIII, 6, 2023, Art. 115, https://doi.org/10.3390/philosophies8060115.

[7] A. DaĺF, *Freedom as an Issue in the Context of Transhumanism and Artificial Intelligence, Digitalization, and Robotics (AIDR)*, in *Ilahiyat Studies*, XIV, 1, 2023, pp. 51-84, https://doi.org/10.12730/is.1261876.

[8] S. Ghazinoory-M. Fatemi-F. Saghafi-A. A. Ahmadian-S. Tatina, *A Framework for Future-Oriented Assessment of Converging Technologies at National Level*, in *NanoEthics*, XVII, 2, 2023, pp. 1-28, https://doi.org/10.1007/s11569-023-00435-4.

[9] J. Sawadogo-J. Simpore, *Would the Convergence of Nanotechnology, Biotechnology, Information Technology and Cognitive Science Be a Springboard for Transhumanism and Posthumanism?*, in *Open Journal of Philosophy*, XIII, 2023, pp. 681-695, https://doi.org/10.4236/ojpp.2023.134043.

[10] I. Moll, *Why there is no technological revolution, let alone a 'Fourth Industrial Revolution'*, in *South African Journal of Science*, CXIX, 1/2, 2023, Art. #12916, https://doi.org/10.17159/sajs.2023/12916.

[11] L. Zhang, *The Ethical Turn of Emerging Design Practices*, in *She Ji: The Journal of Design, Economics, and Innovation*, IX, 3, 2023, pp. 311-329, https://doi.org/10.1016/j.sheji.2023.09.002.

[12] N. Prakash-A. Atiq-M. Shahid-J. Rani-S. Dikshit, *Merging Minds and Machines: The Role of Advancing AI in Robotics*, in *EAI Endorsed Transactions on Internet of Things*, X, 2023, https://doi.org/10.4108/eetiot.4658.

ISSN 0039-4130

5. *About the contemporary North American vision*

The North American article is very self-congratulatory about the results achieved by the launch of the "NNI" (National Nanotechnology Initiative) in the United States some twenty years ago. It notes that the impact of this initiative (of which the NBIC report was a kind of manifesto two years later) has been profound on research and innovation in the USA, and fruitful at many levels. In particular, the implementation of NBIC convergence has led to the advent of emerging technologies (such as platforms for quantum information systems, AI systems, advanced semiconductors), as well as new concepts (such as wireless communication, modern bioeconomy, or advanced manufacturing). Every day, these innovations provide new tools to tackle challenges such as the sustainable society, nanomedicine, personalised learning, increasing human capacity and the fight against age-related dependency. With regard to the objectives set out by the NNI twenty years ago, Roco proposes to distinguish between a) objectives that have not been fully achieved over the last twenty years (such as widespread public awareness of nanotechnology); b) objectives that now seem within our grasp, whereas they seemed unattainable twenty years ago (such as research into the health and environmental risks associated with nanotechnology, approached across the board by the various government agencies concerned); c) objectives that, after twenty years, are better than initially expected (such as the formation of a flourishing "nano-community" that includes the issue of risks and ethical and environmental aspects). Rocco also endeavours to draw lessons from past experience (for example, on how nanotechnology research can also find a place within traditional industries and economic sectors), to attempt to characterise the contours of governance adapted to the development of nanotechnologies (which should be "visionary, anticipatory, transformative, responsible, inclusive and convergent"), and to draw up bright new prospects for the years to come.

6. *About viewpoints from the rest of the world*

Faced with these optimistic visions, the viewpoints of other authors can broadly be summarised as follows.

The South African article denounces the "hype" surrounding convergence and questions the reality of the so-called "revolution" of con-

ISSN 0039-4130

vergent technologies. For the author, convergence is an illusion, and our era is not one of technological and industrial revolution but one of continuity since the beginnings of digitalisation in the 1960s.

The Turkish and Italian articles take convergence for granted but offer a deep critique of its underlying "philosophical programme," which presents itself as liberating but in fact contains the seeds of new forms of alienation. Techno-prophetic trends and transhumanism promote autonomy, morphological liberation, and greater freedom. However, the development of the associated technologies also augurs increased dependence on technology, exacerbated power relations, and regressions in the realm of social freedoms. For the Italian author, the human being will remain an embodied entity, marked by great vulnerability, which can only be alleviated by the historical emphasis on practices of "care".

The Iranian and Chinese articles focus more on the mode of governance adapted to this technological convergence and aim to propose universal models of governance (less context-specific than the current US models), within a general framework not very different from the European concept of "Responsible Research and Innovation" (RRI). The Iranian article focuses on Technology Assessment, while the Chinese article seeks to define and promote a kind of "ethical design" inspired by and based on a Habermasian approach.

The Indian and Burkinabe articles, for their part, are somewhat "on the edge". They demonstrate and highlight a global development of this concept of technological convergence (and the associated innovations) while questioning it or attempting to position themselves in relation to it. The Indian article is somewhat naive and clumsy in its purely technical promotion of technological convergence, aimed at attracting investors and funding. The Burkinabe article expresses sincere concerns about the various ethical issues raised by converging technologies and attempts to identify them, drawing heavily on French and European literature.

7. *About the main lessons to be learned from these comparisons*

From this rapid overview, we can draw the following general lessons:

It appears, first of all, that the scientific concept of "technological convergence" has proven fertile and effective from a scientific perspective, and this concept has been exported worldwide. Its achievements

ISSN 0039-4130

and goals are now being discussed in all corners of the globe, and it is striking to note that the debates and attitudes it provokes today in some developing countries sometimes replicate almost word-for-word those that marked its inception.

It also appears that this development has been accompanied, to some extent, by the export of the reductionist, individualistic, and market-oriented philosophical tendencies that characterized its origins. The need for a demarcation, or substitution by another "metaphysical program", is still felt in many countries today. Even in Europe, this task has not yet been fully accomplished, and several voices are still calling for its completion. There are recurrent tendencies to challenge the fundamental principles of the NBIC framework. To the optimism of technological transcendence, which aimed to solve humanity's fundamental problems, is opposed a more cautious and humanistic approach, mindful of human realities. To the promotion of the scientific and economic potential of new technologies are opposed their complex ethical and social implications, as well as their potential deleterious effects on freedom, social justice, and the human experience. To the transhumanist ideal of surpassing human biological limitations is opposed the preservation of fundamental human characteristics such as vulnerability, embodiment, and ethical freedom.

Finally, another clearly expressed need throughout the literature reviewed is the implementation of governance procedures suited to this technological convergence. It seems widely acknowledged that the governance models currently available, while laying solid foundations, suffer from biases related to their specific cultural, economic, and political frameworks. (For instance, they either minimize regulation, as in the United States, at the risk of exacerbating social and environmental inequalities, or, as in Europe, they favor an upstream ethical inquiry that remains deeply influenced by Western culture.) The contributions considered aim, in particular, to adapt governance tools to the local dynamics of emerging countries and non-Western economies, to promote a prospective vision useful for policymakers through the combination of different evaluation criteria, and to universalize and pluralize models by integrating ethics in a way that is particularly open to contexts where social and environmental justice issues are most pressing. They also emphasize the importance of a unified approach to social, environmental, and ethical concerns related to the emergence of convergent technologies, despite the specialization and dispersion of the various communities involved.

ISSN 0039-4130

8. *By way of conclusion*

It seems beyond doubt that, on a deep philosophical level, contemporary developments in converging technologies (including neuroscience and AI) remain, at least partially, driven by ambitions such as understanding humanity (which remains an object of fascination), reducing it through analytical methods, controlling it, reproducing it, and why not extending and enhancing it. So the combined progress of neuroscience and AI and their gradual coupling could, for example, give rise to both virtuous applications (such as the treatment of cerebral diseases) and devastating attempts (such as influencing or taking control of subjects as a whole). Although these trends are centuries old, they still carry their share of challenges. If the philosopher wishes to make a useful contribution, it is undoubtedly their role to help make these trends more explicit, bring them to public attention, and discuss them collectively.

It is also beyond doubt that mercantilism has long been one of the drivers of technological development. In the absence of a world and a political system capable of countering its most pernicious contemporary developments, can we at least hope that these developments remain framed by robust public regulations? The author of these lines believes that, despite these concerns, it remains within our reach to develop sciences and technologies that would be democratic, conducive to progress, and in the service of publicly chosen and selected objectives and futures. He believes in our humanity and in our technologies, and he wishes to remain optimistic about our technological future.

ISSN 0039-4130

What a Human is, Could be and Should be

The Anthropology of the Human and the Philosophy of Humanism

Sylvain Lavelle

> «Our nature is a fact of the real world just like the rest, and there is no evidence that it will remain the same.»
> Bertrand Russell, *Principles of Philosophy*

The Image of the Human in Philosophy and in Anthropology

The question of defining the humanity of Man is an old issue since the dawn of philosophy and the time of the founding fathers, Socrates, Plato and Aristotle, who conducted several inquiries on this topic. This question was taken up by Kant who defined the task of philosophy with his three famous questions: *What can I know?* (theoretical philosophy); *What should I do?* (practical philosophy); *What may I hope for*? (teleological philosophy); the fourth question was supposed to sum up the three others: *What is a human being?* (anthropological philosophy). It seems that the development of some techniques and 'anthropo-techniques' (eg: neural and genetic science and engineering, robotics, automata and algorithms, artificial intelligence and sentience…) comes to challenge the common definition of the human [1]. It appears to some historians and philosophers that these powerful and sometimes revolutionary techniques could make of *Homo Sapiens* a kind of *Homo*

[1] On this later point, see S. Lavelle (2020), *The Machine with a Human Face: From Artificial Intelligence to Artificial Sentience*, in S. Dupuy-Chessa - H. Proper (eds), *Advanced Information Systems Engineering Workshops*, CAiSE 2020. *Lecture Notes in Business Information Processing*, CCCLXXXII. Springer

Deus[2]. It is assumed that these artificial productions are likely to bring about a radical change of the mankind and then lead to modify the very "Essence of Man" or, at least, a certain Image of the Human. The stream of "trans-humanism" and its correlative "post-humanism" plays an ambiguous role in this game insofar as they can be viewed either as a continuity or as a break with a more classical humanism.

It seems that whoever wishes to fix some limits to the development of the anthropo-techniques could benefit from the contributions of a *Philosophical Anthropology*. The synthetic view on the Human that this discipline has been providing could then be useful as a foundation for a *New Humanism*. Thus, an overall picture of what a Human is, of its common natural and cultural features, would allow to determine what those limits of the humanity of the human are or should be. However, the use and the value of this contribution also depends on how philosophical anthropology is conceived, for a synthetic view derived from *Scientific Anthropology* only provides a descriptive account (*what a human is*). And if the philosophical anthropology derives its synthetic view from a *Moral Anthropology*, it must be shown whether the latter differs from the former so that it can ground a prescriptive account (*what a Human should be*).

The questions of this philosophical inquiry about the relationship between facts and norms, between descriptions and prescriptions in anthropology and philosophy are the following: (1) Is it still relevant to inquire about the "Essence of Man" after the historical shift of Modern Humanism and of Pragmatic Anthropology (Kant)? (2) Is the Philosophical Anthropology based on science and concerned with morals relevant to fix a reference "Image of the Human"? (3) Is the classical scientific and moral Image of the Human really challenged by the development of new techniques and anthropo-techniques? (4) Is the development of new anthropo-techniques as suggested by the stream of Trans/Post-humanism coherent with the classical doctrine of Humanism? (5) Can Humanism be grounded on a Philosophical Anthropology that would help fixing some limits to the technical developments?

My hypothesis is that the classical Is-Ought distinction which prohibits any derivation from description (factual properties) to prescription (normative properties) is not the whole picture. It must be completed with the notion of *inscription* (factual-normative properties)

[2] See for instance: Francis Fukuyama (*The End of History and the Last Man*) and Yual Noah Harari (*Sapiens* and *Homo Deus*).

ISSN 0039-4130

that is specific to the design and production of human artefacts and provides an *inscriptive* account of *what a human could be*. I also suggest that the Philosophical Anthropology, be it scientific or moral, needs something else (or more) than a mere account of what Humanity is about. It needs a philosophical interpretation of the insights provided by philosophical anthropology, that is, an interpretation that can be expressed in terms of an *Anthropological Philosophy* and conceived as a foundation for a *Critical Humanism*.

I. *The philosophical question of the "Essence of Man"*

The question of the "Essence of Man" is an old issue in philosophy, and it has been taken over further by the discipline of Anthropology. This question is still present at the background of the Philosophical Anthropology which endeavours to give an overall view on what Man *is*. But the emergence of the discipline called Anthropology during the Enlightenment drove some philosophers like Kant to modify the object of it. Then the Anthropology is not aimed at revealing the "Essence of Man", but it rather seeks to study what Man actually *does* – although the question about Man is still there.

Some ancient and modern disputes on the Essence of Man

For Plato, there is no clearcut definition of what an animal is, and therefore, of what a human is, but it is assumed that the Man combines three parts : the soul, the body and the two together. According to the legend, when Plato gave the definition of man as "featherless bipeds", Diogenes the Cynic plucked a chicken and brought it into Plato's Academy, saying "Here is Plato's man". Then the Academy added "with broad flat nails" to the definition…Among the heretic philosophers, Diogenes famously challenged the definition of Man as elaborated by Plato, thus raising the problem of the necessary and sufficient conditions of humanity.

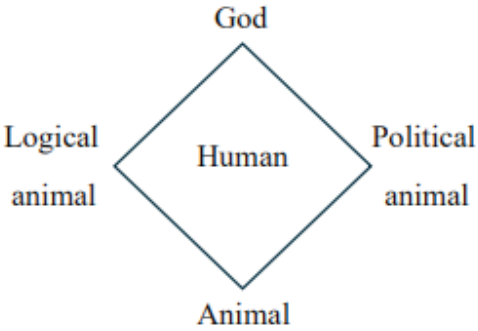
For Aristotle, if man is a being who is situated between the animal and the god, what distinguishes him from other animals is the moral sense[3]. This judgmental capacity is derived from the human language

[3] Aristotle, *Politics*, 1253a.

ISSN 0039-4130

which allows man, in addition to the true and the false, to distinguish the good and the evil, the just and the unjust[4]. The philosopher therefore attempts, starting from the natural and social situation of man, to identify two criteria, one political, the other logical, which distinguish man from the animal and the god. Moreover, these two criteria, political and logical, are not unrelated, because life within the human city presupposes being able to make use of rational language. But it seems that it is an ethical criterion, derived from the logical criterion, which constitutes in the clearest way the characteristic of man and distinguishes him from other animals. In this sense, human society and human language may be necessary conditions of humanity, but they are not sufficient conditions, because they must be supplemented by a moral sense derived from them.

The synthetic view on the Human that result from this ancient philosophical doctrine can be summarized in the following scheme:



The ancient disputes about the Essence of Man found an echo in the emerging stream of modern humanism whose members elaborated a new vision of the Human as a free creature of God. This vision would take up the ancient distinction between the bodily and the spiritual part of the human being, as well as the cleavage between his animal tendency and his divine tendency. But there was undeniably something

[4] Keitzmann would not agree with this view that Aristotle gave a definition of the human in his *Politics*, for he only gave some characteristic properties of it, namely the *logos* as a rational speech, as opposed to the animal voice. He did not provide a unique or unified definition of humanity in the rest of his work, for the human being is divided in two parts, the body and the mind. The material-bodily part can be studied by zoology, as a branch of physics, while the spiritual-divine part must be studied by theology, usually sorted as a branch of metaphysics. See C. Kietzmann (2019), *Aristotle on the Definition of what it is to be a Human*; G. Keil, N. Kreft, *Aristotle's Anthropology*, Cambridge, Cambridge University Press.

ISSN 0039-4130

new under the sun, for the human being, as a creature of God, has the freedom to create his own nature, which, consequently, is not determined in a fixed and eternal way. However, this time does not offer a unique view, and if we consider those of the icons of the Renaissance, such as Mirandola and Vinci, they conflict on their very image of the Human.

Pico de la Mirandola is like the symbol of humanism at the time of the Renaissance, and in his *Oration on the dignity of Man*, he propagates the idea that Man is a creature of God whose dignity lies in the exercise of the mind and in the ability to change oneself [5]:

> «All other things have a limited and fixed nature prescribed and bounded by our laws. *You, with no limit or no bound, may choose for yourself the limits and bounds of your nature*. We have placed you at the world's centre so that you may survey everything else in the world. We have made you neither of heavenly nor of earthly stuff, neither mortal nor immortal, so that with free choice and dignity, you may fashion yourself into whatever form you choose. To you is granted the power of degrading yourself into the lower forms of life, the beasts, and to you is granted the power, contained in your intellect and judgment, to be reborn into the higher forms, the divine».

The only constant in Man's history is change, as shown by the evolution of the human doctrines and institutions, and this secular trend reveals the "auto-poietic" nature of the Human. The other changes in nature are the result of an outer force, but for the human beings, it is the result of their inner free will, that then appears as a proper force, for the better or the worse.

However, if one refers to another icon of the Renaissance, Leonardo da Vinci, it appears that the capacity of creation of the Human is interpreted by him in a much less optimistic way. Vinci used to be very critical with the Humanists who fancied adding «tons of superficial comments on the works of the Ancients», instead of learning directly from the experience of nature. He also rejected severely their common view on the singularity or the superiority of men over animals, and he regarded most of the humans as some useless poor creatures. Moreover, Leonardo often blamed the evil nature of men who would certainly make bad use of his discoveries and his inventions and more broadly, of the progress of science, industry or medicine. Yet, quite paradoxically, his relationship with

[5] Pico della Mirandola (1486), *Oration on the Dignity of Man*, Gateway Editions, 1996.

ISSN 0039-4130

some potentates such as Sforza or Borgia was far from clear, for he presented and acted himself as a talented designer of powerful destructive weapons. It remains that his vision of Man is much darker than the one that the *Vitruvian Man* suggests, and as to Vinci himself, as surprising as it may be, he can hardly be labelled as a humanist.

Therefore, the Image of the Human at the time of emerging modern Humanism during the Renaissance is more complex than the cliché of an early Enlightenment. Behind the idea of the "Self-made Man" who exercises his freedom and who creates his own nature, the Image of the Human is oriented in two conflicting directions: a positive tendency (scientific and moral progress, improvement of life) and a negative tendency (power and domination over humans and non-humans).

The ambivalences of Kant's Anthropology

The question of the Essence of Man takes a new step with the development of an Anthropology as a rational discourse on Man, possibly as a "science of man" to be compared with a "science of nature". Kant in his *Anthropology from a Pragmatic Point of View* redirects the question "what is a human being?" from defining a human in terms of what he or she *is* to defining him or her in terms of what he or she *does*[6]. The philosopher makes a difference between *physiological anthropology* concerned with the *nature* of the Human and *pragmatic anthropology* concerned with the activities and the *actions* of the Human. For that, he makes a distinction between three levels of "man's praxis", of his acting in the world: technical (power), pragmatic (prudence) and moral (duty). Insofar as man is defined in terms of his praxis, he becomes the product of his own making rather than an immediate given: he «has a character that he himself creates», he is the on-going and never-ending result of his own making. In addition, Kant is supposedly respectful with the Is-Ought distinction, but far from being a mere descriptive account, several parts of his work contain some prescriptive advices[7].

[6] D. Sardinha, *Différence entre l'anthropologie pragmatique et l'anthropologie métaphysique*, in *Rue Descartes*, III, 75, 2012

[7] As Cohen suggests, «Kant's anthropology is pragmatic in three fundamental senses: its object is pragmatic insofar as it studies man in terms of his actions in the world, and thus as a freely acting being; second, its method is pragmatic in that it involves interaction rather than observation; and third, *its aim is pragmatic inasmuch as it is not only descriptive but prescriptive*». See A. Cohen (2008) *Kant's answer to the question 'What is man' and its implications for anthropology*, in *Studies in the history and philosophy of science*, XXXIX, 4.

ISSN 0039-4130

The descriptive approach in Kant's anthropology: the characteristics of the species

The use of the descriptive approach that is supposed to be a common feature of the scientific stance in anthropology is well illustrated by the study of the human species. Kant states that the human is a rational conscious animal who possesses as a human being the ideal of humanity, produces its own character as a person and for society, and can reach its vocation by developing his natural trends. The rationality of Man is not enough, one also need the reason, as a proper capacity that differs from the animal's instinct, so that a human being can distinguish the good and the evil [8]:

> «In order to appreciate this character of his species, the comparison with a standard that can be found anywhere else but in perfect humanity is necessary. One can therefore say that *the first character of the human being is the capacity as a rational being to obtain a character as such for his own person as well as for the society in which nature has placed him*. This capacity, however, presupposes an already favorable natural predisposition and a tendency to the good in him; for evil is really without character (since it carries within itself conflict with itself and permits no lasting principle in itself). The character of a living being is that which allows its vocation to be cognized in advance. – However, for the ends of nature one can assume as a principle that nature wants every creature to reach its vocation through the appropriate development of all predispositions of its nature, so that at least the species, if not every individual, fulfills nature's purpose.»

It can be debated whether the philosopher gives a genuine descriptive account of the human characteristics or provides a speculative general interpretation of the innate properties of the human mind or of

[8] F. Van de Pitte (1971) *Kant as a Philosophical Anthropologist*, The Hague, Martinus Nijhoff : «Kant realized that man's rational capacity alone is not sufficient to constitute his dignity and elevate him above the brutes. If reason only enables him to do for himself what instinct does for the animal, then it would indicate for man no higher aim or destiny than that of the brute but only a different way of attaining the same end. However, reason is man's most essential attribute because it is the means by which a truly distinctive dimension is made possible for him. Reason, that is, reflective awareness, makes it possible to distinguish between good and bad, and thus morality can be made the ruling purpose of life. *Because man can consider an array of possibilities, and which among them is the most desirable, he can strive to make himself and his world into a realization of his ideals*».

ISSN 0039-4130

the ends of human nature. Nevertheless, it is quite clear that when he presents the ideal of humanity, he does not word any prescriptive statement that would specify what the human should be or what the human should do. In this respect, this part of his anthropology remains "scientific", in the broad sense of the word, but obviously, it ceases to be when he deals in the other part, the 'moral' one, with the characteristics of the sexes.

The prescriptive approach in Kant's anthropology: the characteristics of the sexes

The use of the prescriptive approach that is supposedly no allowed in anthropology as a science is clearly present in Kant as for what concerns the difference of sexes, the male and the female. Kant states the following :

> «Who, then, should have supreme command in the household? – for there certainly can be only one who coordinates all transactions in accordance with one end, which is his. – I would say, in the language of gallantry (though not without truth): *the woman should dominate and the man should govern*; for inclination dominates, and understanding governs. – The husband's behavior must show that to him the welfare of his wife is closest to his heart. But since the man must know best how he stands and how far he can go, he will be like a minister to his monarch who is mindful only of enjoyment.»

What the philosopher does in his *Anthropology* is a typical Is-Ought speech drift, insofar as he moves from a descriptive to a prescriptive speech. The derivation is limited by its division between the *Didactics* (Part I) and the *Characteristics* (Part II), but his anthropology can be said to be twofold. It contains indeed both a descriptive approach and a prescriptive approach, though he assumes that the philosopher can hardly ground the morality on science[9]. In this respect, Kant's anthropology gives a synthetic image of what a Human is: this image is descriptive, and to some extent, scientific, while it is also moral in a way that is both descriptive *and* prescriptive.

The prohibition of any inference from the fact to the norm is commonly presented as the Law of Hume, and it entails that it is not possible from one description to infer one prescription. This problem of

[9] Warden states that «Kant believes the traditional male-female distinction is unlikely to disappear, but he never proposes the traditional gender ideal as the moral ideal. *He rejects the idea that such considerations of philosophical anthropology can set the framework for morality*». See H. Warden (2015) *Kant and Women*, in *Pacific Philosophical Quarterly*, XCVIII, 4.

ISSN 0039-4130

the Is-Ought gap remains for any development of anthropology that endeavors to be scientific while dealing with moral matters and which is asked to provide some normative statements or assessments.

II. *What a Human is and should be: the scientific and moral Image*

Philosophical Anthropology is supposed to identify what a Human is, but it sometimes overflows the limit of a descriptive-scientific approach and suggests a more prescriptive-moral approach regarding what a Human should be – or, at least, should do. Rather than an "Essence of Man", it is more appropriate to search for a generic "Image of Humanity", but it needs to be clarified whether this anthropological image is scientific or can also be moral. One can suggest that the Philosophical Anthropology has kept its ambition to provide with a unified image of the Human. But, first, the development of a scientific anthropology has challenged the ambition of unification of Philosophical Anthropology, and second, the same occurred in the development of moral anthropology – that furthermore appears to be a mere branch of the former.

Philosophical and scientific anthropology

The research programme of the Philosophical Anthropology was taken over by some philosophers in Germany, such as Max Scheler, Helmut Plessner and Arnold Gelhen [10]. Despite their multiple differences, they share a common theoretical core that gives a kind of unity to this field located at the intersection of philosophy and anthropology as a science. They proposed several different approaches to the Subject-Object relation, nevertheless the method of comparison of the philosophical anthropology moves from the non-human (vegetal, animal) to the human. This "bottom-up" approach allows to grasp the "functional (or vital) circle" by and through which an organism is correlated with its environment. Furthermore, the move from the Non-Human to the Human shows a break in the "functional circle" as far as Man is concerned as a spiritual being. According to those philosophers, the Human is twofold in his relation to the environment:

[10] M. Scheler (1928) *The Human Place in the Cosmos*; H. Plessner (1928) *Levels of Organic Life and the Human*; A.Gelhen (1940) *Man: His Nature and Place in the World*.

ISSN 0039-4130

an "internal" relation to the self, and an "external" relation as a body living among other bodies [11].

As to the development of the scientific anthropology, it gives the priority to some specific field studies, such as, for example, one tribe with its own language, customs and ritual. This is typically the kind of studies that are achieved by some scientific anthropologists, among them Malinovski with his study of New Guinea tribes, or Evans-Pritchard with his study of the Nuer in East Africa [12]. The scientific anthropology makes a common difference in cultural and social studies between the (global) level of anthropology (the "Theory of Man") and the (local) level of ethnography (the "Practice of Men"). In this respect, there is no unified scientific image of Man but a plurality of scientific images of Men ("human perspectivism") that grounds a principle of radical ethno-diversity. But some anthropologists keep the ambition of producing some more general reflections or conclusions, such as Levi Strauss, for instance, with his thesis on the prohibition of incest as a universal law of human societies [13].

Philosophical and moral anthropology

In contrast with the culture of virtues (*arete*) and duties (*officii*), the modern Image of the Human praises the ideal of human freedom and of a moral improvement of humanity through progress (eg: Declaration of Human rights). However, one can state that there is a historical failure of evolutionary and revolutionary moral (normative) anthropology, whether in the "hierarchy of races" (Gobineau) or in the ideology of the "New Man" in totalitarian regimes. The examination of the moral features of humanity is more easily accepted if one moves forward to an approach of philosophical anthropology, although a gap remains between the description and the prescription. However, the field study of moral anthropology is also debated regarding the moral Image of the Human, especially if the morals of a population are scrutinized through the lens of the social sciences.

[11] J. Fischer (2017) *Le noyau théorique propre à l'Anthropologie philosophique (Scheler, Plessner, Gehlen)*, *Trivium*, XXV.

[12] B. Malinovski (1912) *Argonauts of the Western Pacific: An account of native enterprise and adventure in the Archipelagoes of Melanesian New Guinea;* E. Evans-Pritchard (1940) *The Nuer: A Description of the Modes of Livelihood and Political Institutions of a Nilotic People*.

[13] Claude Lévi-Strauss and his heir Philippe Descola both claim to support a method of structural anthropology that seeks to identify some invariant structures of human societies.

ISSN 0039-4130

Some anthropologists pay tribute to the "ethical turn", some others suggest a "critical moral anthropology" [14], while some recent works suggest that moral anthropology is gaining some legitimacy as an academic field [15]. However, they face the problem of the epistemic reduction due to the scientific tropism of the "science of morals", as Moritz Schlick once suggested: "ethics is a branch of anthropology" - which means a scientific-descriptive study of the human. Alike scientific anthropology, that moral anthropology belongs to as one of its branches, this leads to a principle of radical "etho-diversity" in which morality frames are taken to be culture-specific. But it seems that some moral anthropologists, based on some comparative empirical works, would emphasize the common moral patterns throughout the variety of cultures and societies [16]. As Klenke suggests, 'the ethical turn uncovers a richer picture of moral phenomena on the intersubjective level, one akin to a virtue theoretic focus on moral character, with striking similarities of moral phenomena across cultures' [17].

What a Human is and should be: an experiment on the Image of the Human

The approaches of the scientific anthropology or those of the moral anthropology can be confused as regards their common relation to the facts and norms of the human life. But it is legitimate if one keeps in mind the stake of the philosophical anthropology (providing an Image of the Human) as well as that of the Law of Hume (avoiding the Is-Ought confusion) to reflect on some criteria that could operate in both side for defining the Human. The idea is not to make a plea for a deductive inference between the two discourse regimes, but to show (a) the legitimacy of each approach in its own sector or field of relevance : the epistemic one (scientific, factual, descriptive) and the ethical one (moral, normative, prescriptive); and (b) the necessity of not confusing

[14] M. Klenke (2019), *Moral philosophy and the Ethical Turn in Anthropology*, in *ZEMO*, II; D. Fassin (2012) *Towards a Critical Moral Anthropology*, D. Fassin (ed.) *Moral Anthropology*, Wiley-Blackwell, Malden.

[15] M. Lambek (2010) *Ordinary Ethics. Anthropology, Language and Action*, Fordham University Press; J. Faubion (2011) *An Anthropology of Ethics*., Cambridge University Press; D. Fassin (2012) *Moral Anthropology. An anthology*, Wiley-Blackwell, Malden.

[16] They then open a set of options that makes some of the supposedly essential features of the human some mere *accidental* (optional or potential) features that can be framed as a series of alternatives (eg: a human being can be X *or non-X*):

[17] M. Klenk (2019) Moral Philosophy and the Ethical Turn in Anthropology, *ZEMO*, 2.

ISSN 0039-4130

the two sectors or fields by an operation of reduction of the one to the other (like in a "science of morals", for instance).

This is the purpose of the following thought experiment that tackles the issue of what a human is and use the criteria to define what a human should be:

What a Human is	What a Human should be
A being is human if he or she is a being who: (i) is born, lives and dies (ii) is descended from a man and a woman (iii) needs air, water and food to live (iv) is provided with a body, a mind and a language (v) possesses functions linked to this body, this mind and this language (motion, consciousness, dream, thought, rite...) (vi) leads an existence within a human group (vii) develops social relations with other human beings (family, work, love, friendship...) (viii) reproduces by having children (ix) receives and gives an education which allows to live in the world.	A being is human if he or she is a being who should: (i) be born, live and die (ii) be descended from a man and a woman (iii) need air, water and food to live (iv) be provided with a body, a mind and a language (v) possess functions linked to this body, this mind and this language (motion, consciousness, dream, thought, rite...) (vi) lead an existence within a human group (vii) develop social relations with other human beings (family, work, love, friendship...) (viii) reproduce by having children (ix) receive and give an education that allows to live in the world, or not

In the perspective of the ordinary life, or in that of a religious doctrine, the two speech regimes are submitted to a rule of deduction: "what a human is" implies "what a human should be". But if we accept the distinction between the two ranges of discourse, then this rule of reasoning cannot operate in a direct and simple way. For instance, the rule of reasoning could function without much restriction for the criterion (iii): if the human needs air, water and food to live, then the human should be considered as a being who needs air, water and food to live. But it would be probably more difficult to use the rule of reasoning for the criterion (viii): if the human reproduces by having children, then the human should reproduce by having children – for not everybody agrees that as humans, they *should* have children.

This issue of the scientific and moral Image of the Human is made even more complex with the new possibilities opened by the development of anthropo-techniques and by the question relating to what a human *could* be.

ISSN 0039-4130

III. *What a Human could be: the Artificial Challenge*

The paradox of the "Essence of Man", if it equals the problem of defining the bonds of humanity, is made obvious by the change-inducing technical artefacts. The use of techniques produces an *inscription* of some factual-normative properties in a thing or a being, like what a potter does with a statue that is supposed to conform to a certain model. In this respect, a human is shown to be a "potential" creature whose nature since the outset is artificial, so that a human can be said to be an "artefact of nature" - or of a "supernature". The use of artefacts including for the modification of the body or the mind of one human is just a development of a fundamental feature of humanity. But it seems that some techniques, and especially some new anthropo-techniques, push for some deeper modifications of the human that really question the limits of their use. In some futuristic (but perhaps, realistic…) visions, this evolution becomes disruptive, so much so that *Homo Sapiens* becomes a kind of *Homo Deus*.

From Homo Sapiens to Homo Deus

Human technology takes a further step when, from changing the environment which indirectly produces in return a change in human life, it goes directly to changing the human being itself. It is then a deliberate and assumed plan of modification of the human being, and the work of the anthropo-techniques can then cover several aspects or dimensions. It can extend from bodily change to mental or spiritual change, from partial change in the repair of the body or possibly of the mind to the change of humanity itself, in the name of its "enhancement". It seems that there is a difference in the judgment about the scope of possibilities that technical change allows depending on whether it is a question of repair or improvement of the body or the mind. Thus, the judgment on the status of artefact seems to vary according to the use that is made of it, as if there were an essential difference, of an ontological nature, between a legitimate artefact and an illegitimate one.

There is no doubt a wide range of technical options as regard the *human enhancement*: modification of body or mind states, stimulation and increase of body and mind capacities, coupling of a human body or mind with the machine power (hybrid), modification of the human genome possibly for the purpose of human reproduction, etc. On the other side, one

ISSN 0039-4130

finds the interactions of humans with human-like machines, which require from the machines that they think, act and possibly feel like humans (artificial intelligence and sentience)[18]. This perspective of a radical human change grounded in some anthropo-techniques is made very explicit by Harari who envisages a shift from *Homo Sapiens* to *Homo Deus*[19]:

> «Humanity will set itself the third great project of acquiring divine powers of creation and destruction, and of elevating *Homo sapiens* to the rank of *Homo deus*… Above all, we want to be able to rearrange our bodies and minds to escape old age, death and misery, but once this goal is achieved, who knows what we will then use this capacity for? It is therefore not extravagant to think that the new human agenda consists basically of a single project, with multiple branches: achieving divinity…When we talk about elevating humans to the rank of gods, we must think more of the Greek gods or the Hindu devas than of the almighty father of the biblical sky. Our descendants would still have their weaknesses, their oddities and their limits. But they could love, hate, create and destroy on a much larger scale than we can. Throughout history, most gods have been credited not with omnipotence, but with rather specific super-abilities, such as the ability to design and create living beings, to transform their bodies, to control their environment and time; to read minds and communicate over distances; to travel at super-high speeds; and of course, to escape death and live forever. Humans are working to acquire all of these abilities, and more... In the quest for health, happiness, and power, humans will gradually change one of their traits, then another, and another, until they are no longer human».

The vision of the human's future, who would evolve from the realm of *Homo Sapiens* to that of *Homo Deus*, can be accepted as a mere hypothesis, or rejected as a speculative prophecy. But if it can be shown that it significantly challenges the classical Image of Human, it has the merit of posing the problem of the anthropo-techniques and of the limits to their development.

What a Human could be

It is not enough to identify through several criteria what a human is or what a human should be, it is also requested that one can identify *what a human could be*. The reason for that is the possibilities brought

[18] See Lavelle (2020).

[19] Y. N. Harari (2015) *Sapiens. A brief History of Humankind*, Harper; Y. N. Harari (2016) *Homo Deus. A brief History of Tomorrow*, Harper, pp. 56-58.

ISSN 0039-4130

in by the development of techniques modify the scope of the options and therefore the traditional features and limits that are inherent to the Image of the Human. The humanist view assumes that the human as a potential creature can modify the world, the life and the self, from a body as well as from a mind point of view. But the power and the impact of the anthropo-techniques is such that they can entail a degree of change that concerns some essential aspects of the human physical and psychic constitution. The scope of potential changes opens up a set of options that makes some of the supposedly essential features of the human some mere *accidental* (or optional) features that can be framed as a series of alternatives (eg: a human being can be X *or non-X*).

What a Human could be	
A being is human if it is a being who can: (i) be born, live and die, *or not* (ii) be descended from a man and a woman, *or not* (iii) need air, water and food to live, *or not* (iv) be provided with a body, a mind and a language, *or not* (v) possess functions linked to this body, this mind and this language (movement, consciousness, dream, thought, rite...), *or not* (vi) lead an existence within a human group, *or not* (vii) develop social relations with other human beings (family, work, love, friendship...), *or not* (viii) reproduce by having children, *or not* (ix) receive and give an education that allows it to live in the world, *or not*	The artificial image of the human is located so to say 'in between' the scientific image and the moral image of the Human: Scientific Image of the Human ↕ Artificial Image of the Human ↕ Moral Image of the Human

It can be that technology will modify the scale of the perspective, with a radical mutation of the Human that requires to inquire about the "Ought" and not only about the "Is". This anthropological concern could be one of the most crucial tasks of humanism if the human beings become some essentially artificial beings. It can be questioned whether the potentially radical change of the Human is a mere achievement of Humanism or a new development of the human evolution that fits the plans of Trans-Humanism and of Post-Humanism.

ISSN 0039-4130

IV. *The Philosophies of Humanism and of Trans / Post-Humanism*

Humanism is a complex philosophical and cultural (scientific, artistic, moral…) stream that has developed in Europe and more broadly in the West over several centuries since the time of the Renaissance[20]. Humanism as a historical movement appropriated the heritage of the Greek and Roman civilisations, developed critical mind and knowledge, and endeavors the flourishing of humanity that is made more humane through culture. Humanism as a philosophical attitude gives the Human a supreme value and claims for every human the possibility to flourish freely his or her humanity and his or her proper human capacities. One could summarize the spirit of Humanism in one sentence: *Humans are free with culture to develop, create or change their nature*. Nowadays, humanism is challenged by the current streams of trans / post-humanism, and one can wonder to what extent these streams are a development or a disruption of it.

On the principles of Humanism

Humanism is a polysemic notion, and as Cheyney suggested, it may be «the reasonable balance of life that the early Humanists discovered in the Greeks; it may be merely the study of the humanities or polite letters; it may be the freedom from religiosity and the vivid interests in all sides of life of a Queen Elizabeth or a Benjamin Franklin; it may be the responsiveness to all human passions of a Shakespeare or a Goethe; or it may be a philosophy of which man is the center and sanction. It is in the last sense, elusive as it is, that Humanism has had perhaps its greatest significance since the sixteenth century»[21].

Some historians and philosophers attempted to circumscribe the sources of humanist philosophy, and Larmont proposed ten principles that enable to characterize it[22]. It is tempting to contrast these classical principles of humanism with those of contemporary trans-humanism and then show the clear differences between the two of them. But looking at the historical philosophy of Condorcet, one of the most

[20] But there is of course a non-Western humanism: Asian, African...

[21] E. P. Cheney, *Encyclopaedia of the Social Sciences* Macmillan, New York 1937, IV, p. 541.

[22] C. Larmont, *The Philosophy of Humanism*,, The Humanist Press, 8th Edition, Amherst 1997, p. 38.

ISSN 0039-4130

representative doctrines of humanism, rooted in the idea of human progress, it is not so obvious that trans-humanism marks a break with humanism. Condorcet's view on human progress combines scientific, industrial and medical progress together with moral progress, but he also suggests that «the human species must improve itself…the real improvement of intellectual, moral and physical faculties, which can also be the result, or of that of the instruments which increase the intensity and the direct use of these faculties, or even of that of the natural organization»[23].

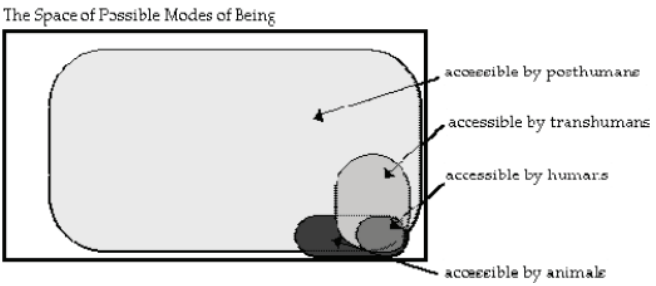
On the principles of Trans-Humanism

It should be noted that, unlike Condorcet's text, the transhumanist declaration considers the improvement of humanity from the angle of progress derived from advances in the sciences and the arts, but not from advances in the morals[24]. In any case, the manifesto of transhumanism highlights a few salient points that deserve the philosopher's attention. On the one hand, it carries a plan of modification of the human being that proposes an improvement or an increase in its capacities, both physical and psychic (health, youth, reproduction, intelligence, memory, etc.). On the other hand, the meaning given to this project is the blossoming of the human being by overcoming its limits, essentially bodily, but also mental ones, which must be transcended in order to acquire greater control over ones' life. Finally, it is important in this project not to immediately set limits, particularly moral ones, to its accomplishment, but to make a rational examination, on a case-by-case basis, of the advantages and risks that each of its options entails.

[23] N. de Condorcet (1795) *Sketch of a historical table of the progress of the human mind*. For the full quote : 'Our hopes for the future state of the human species can be reduced to these three important points: (1) the destruction of inequality between nations; (2) the progress of equality within the same people; finally, (3) the real improvement of man…the human species must (…) improve itself, either by new discoveries in the sciences and in the arts, and, by a necessary consequence, in the means of particular well-being and common prosperity; or by progress in the principles of conduct and in practical morality; or finally by the real improvement of intellectual, moral and physical faculties, which can also be the result, or of that of the instruments which increase the intensity and the direct use of these faculties, or even of that of the natural organization'.

[24] https://www.humanityplus.org

ISSN 0039-4130



Possible modes of being (Nick Bostrom)

A balanced judgment is necessary, especially if transhumanism can be considered, according to the authors, as an extension of humanism. It aims, alike for the well-being of humanity and of any being that possesses "feelings", whether it is a "human, animal, artificial or post-human brain" (*sic*). What is striking in this declaration, if we make a short exegesis of it, is as much the explicit as the implicit of the text, the frankness of the project and the vagueness that surrounds most of the terms it uses. Beyond its guiding principle, this program of directed human evolution leaves large areas of shadow, but they are perhaps even more interesting than the areas of light. Thus, it seems that the transhumanist has a problem with the human body, of which he points out the weaknesses, rather than the strengths, but he leaves to the imagination or to debate what the limit of the enterprise of surpassing its limits could be. It cannot be excluded from the outset, if technology allows it one day, to envisage for the human the exit from the human condition characterized by its mortality. It remains significant that, in all the technical projects carried by the streams of transhumanism, the improvement of the human being passes through an increase in the power of the human being.

There is hardly any evidence among these streams of an overall plan that would propose as a priority, as a means of improving humanity, a growth in reason or in wisdom. And even if this were the case, it is likely that this improvement of humanity would be conceived as a subsidiary or derivative aspect of an increase in human capacities. Basically, the spirit of the transhumanist program, however vague it may be, if we base ourselves on the ancient image of the human situated between the animal and the god, is the dream, not to say the fantasy, of an exit from the animal condition of the humans. It is a question, in view of the range of possibilities for modifying the humans that are opened thanks

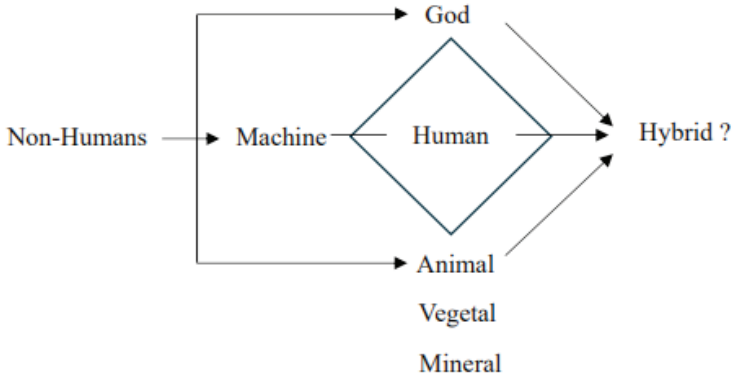
ISSN 0039-4130

to technology, of taking a significant step towards the divine condition, which could be summarized as follows: *failing to be gods, humans are allowed to seek to get as close to them as possible*.

Critical Humanism as an alternative

One could suggest that, if the achievements of Trans / Post – Humanism are not considered as a mere fatality of the human evolution, then there is room for an alternative view. *Critical humanism* is a version of humanism that is responsible for taking charge of the crisis of humanity, which is none other than the crisis of humanism itself. Thus, critical humanism is the philosophical position according to which the definition of anthropological identity and difference, of the place of man in nature and culture (a) is an open question that calls for a critique of humanity (b) is a question that leads humans to define what humanity is, could be and should be. Thus, critical humanism suggests that the essence of humanity is open to critical discussion and judgment and can therefore take refuge in neither a dogmatic nor a skeptical position. In this version of humanism, human beings are called upon to say what humanity is, what it could be and what it should be, but they cannot say this by relying on an Image of the Human, an *imago hominis*, which would be frozen in and by a tradition. For, if so, it would condemn in a dogmatic way the benefits as well as the losses, the advantages together with the disadvantages of the use of artefacts in the future evolution of humans.

The synthetic view on the Human in Critical Humanism that is derived from the modern ontology, in contrast with that of the ancient philosophy, can be summarized in the following scheme:



ISSN 0039-4130

The Image of the Human in the light of Critical Humanism bears on the distinction between Humans and Non-Humans (the God, the Animal and the Vegetal, the Mineral - and the Machine). And it assumes that whatever the techniques that he or she uses, from the less sophisticated (ex: food, clothes, tools…) to the most sophisticated ones (ex: prosthesis, drugs, chips...), a Human as a user of artefacts is a kind of hybrid anyway. The question then is to determine to what extent a Human can be modified by the development of some anthropo-techniques that incorporate some non-human elements. The use of these anthropo-techniques can lead to increasing the distance with a common natural and cultural given that is nevertheless taken to be a standard for a normal or genuine human life.

From Philosophical Anthropology to Critical Humanism

The Philosophical Anthropology can provide some interesting insights to give credit and relevance to a generic Image of the Human – rather than a universal 'Essence of Man'. However, if (1) the philosophical anthropology is developed on the basis of a scientific ground (*scientific anthropology*), and if (2) it can hardly goes beyond that limit, including when it deals with moral issues (*moral anthropology*), then it implies that philosophical anthropology doesn't provide as such the appropriate grounding for Critical Humanism. Philosophical anthropology can certainly be helpful in providing some interesting elements for a generic Image of the Human, but if and only if they are used to sketch it out in the perspective of an *Anthropological Philosophy*. Nevertheless, this Anthropological Philosophy, as both factual and normative grounding for Critical Humanism, requires an *Onto-deontology of Limits* and therefore a composition of epistemic, technical and ethical possibilities (*multi-modal compossibilities*).

Epistemic possibilities	Technical possibilities	Ethical possibilities
What a human is	What a human could be	What a human should be

We could then imagine, as a consequence of the evolving nature of the Human, that Kant's question on anthropology *What is a human being?* would split back into three modified questions:

ISSN 0039-4130

(1) Theoretical question: *What can we know about the (new) human being?*
(2) Practical question: *What must we do with the (new) human being?*
(3) Teleological question: *What may we hope from the (new) human being?*

The question of anthropology as an anthropological philosophy (and not as a philosophical anthropology…) could insist on the multi-modal compossibility and would then be the following: *What is, could be and should be a human being?* In a somewhat radical interpretation of the anthropological philosophy, one assumes that we will have to choose and decide on what kind of humans we will produce in the future. If we follow the path of a negative onto-deontology, then it entails that we are not able to determine what a Human is in positive ontological terms. We are only able to express in negative onto-deontological terms *what a Human should not be if he or she wants to remain humane*.

Appendices

Principles of Humanism (Larmont)

First, Humanism believes in a naturalistic metaphysics or attitude toward the universe that considers all forms of the supernatural as myth; and that regards Nature as the totality of being and as a constantly changing system of matter and energy which exists independently of any mind or consciousness.

Second, Humanism, drawing especially upon the laws and facts of science, believes that we human beings are an evolutionary product of the Nature of which we are a part; that the mind is indivisibly conjoined with the functioning of the brain; and that as an inseparable unity of body and personality we can have no conscious survival after death.

Third, Humanism, having its ultimate faith in human kind, believes that human beings possess the power or potentiality of solving their own problems, through reliance primarily upon reason and scientific method applied with courage and vision.

Fourth, Humanism, in opposition to all theories of universal determinism, fatalism, or predestination, believes that human beings, while conditioned by the past, possess genuine freedom of creative choice and action, and are, within certain objective limits, the shapers of their own destiny.

Fifth, Humanism believes in an ethics or morality that grounds all human values in this-earthly experiences and relationships and that holds as its highest goal the this-worldly happiness, freedom, and progress-economic, cultural, and ethical-of all humankind, irrespective of nation, race, or religion.

ISSN 0039-4130

Sixth, Humanism believes that the individual attains the good life by harmoniously combining personal satisfactions and continuous self-development with significant work and other activities that contribute to the welfare of the community.

Seventh, Humanism believes in the widest possible development of art and the awareness of beauty, including the appreciation of Nature's loveliness and splendor, so that the aesthetic experience may become a pervasive reality in the lives of all people.

Eighth, Humanism believes in a far-reaching social program that stands for the establishment throughout the world of democracy, peace, and a high standard of living on the foundations of a flourishing economic order, both national and international.

Ninth, Humanism believes in the complete social implementation of reason and scientific method; and thereby in democratic procedures, and parliamentary government, with full freedom of expression and civil liberties, throughout all areas of economic, political, and cultural life.

Tenth, Humanism, in accordance with scientific method, believes in the unending questioning of basic assumptions and convictions, including its own. Humanism is not a new dogma, but is a developing philosophy ever open to experimental testing, newly discovered facts, and more rigorous reasoning.

Transhumanist Declaration

Humanity will be radically changed by technology in the future. We foresee the feasibility of redesigning the human condition, including such parameters as the inevitability of aging, limitations on human and artificial intellects, unchosen psychology, suffering, and our confinement to the planet earth.

Systematic research should be put into understanding these coming developments and their long-term consequences.

Transhumanists think that by being generally open and embracing of new technology we have a better chance of turning it to our advantage than if we try to ban or prohibit it.

Transhumanists advocate the moral right for those who so wish to use technology to extend their mental and physical (including reproductive) capacities and to improve their control over their own lives. We seek personal growth beyond our current biological limitations.

In planning for the future, it is mandatory to take into account the prospect of dramatic progress in technological capabilities. It would be tragic if the potential benefits failed to materialize because of technophobia and unnecessary prohibitions. On the other hand, it would also be tragic if intelligent life went extinct because of some disaster or war involving advanced technologies.

ISSN 0039-4130

We need to create forums where people can rationally debate what needs to be done, and a social order where responsible decisions can be implemented.

Transhumanism advocates the well- being of all sentience (whether in artificial intellects, humans, posthumans, or non- human animals) and encompasses many principles of modern humanism. Transhumanism does not support any particular party, politician or political platform.

ISSN 0039-4130

AI and Freedom: some Ideas for a Debate

Dominique Lambert

1. *A Brief Definition of AI and Freedom*

We will not enter here into details about the various and possible definitions of AIs. The ones that interest us here are not expert systems (whose results are the products of perfectly transparent classical programming) but deep learning systems based on formal neural network technology [1]. We will therefore only consider here the problems raised by AI system considered as:

> «a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.» (OECD definition).

Today, in the civil or military field, AIs are likely to control different types of robots (meaning physical or electronic machines equipped with sensors, processors – this is where AI comes in – and effectors allowing action on physical or electronic environments), endowed with autonomy.

The term AI, as we know, is an abuse of language, very often underlined by AI specialists themselves. Indeed, AI is a simulation, a representation of the productions of human intelligence and not intelligence as such. The philosophy of science must recall this well-known fact since the Greek epicycles: it is not because

[1] Y. Le Cun, *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*, Odile Jacob, Paris 2023.

a model "saves phenomena" (as Pierre Duhem would have said), it means reproduces exactly the observable data of a phenomenon that it understands it and exhausts its reality. To claim this is a lie, and is perhaps the first moral problem posed by an ideological use of the term AI. Human intelligence is not exhausted in the computational and logical dimension implemented by neural networks. Not only because reasoning (including mathematical) reasoning is not exhausted in algorithmic procedures, as the results on the internal limitations of formalisms have clearly shown, but because a reflection on human intelligence cannot be reduced to a single facet, namely, analytical and notional.

2. *The two facets of intelligence and freedom*

For our purposes, it is interesting to return to a fundamental analysis of what intelligence is, such as the one given to us by Maurice Blondel in his book *La Pensée* or in *Le Procès de l'Intelligence*. Blondel distinguishes two essential facets of intelligence that lead to two types of knowledge. Following John Henry Newman, he calls the first "notional" knowledge and the second "real" knowledge.

The first is the one based on an approach rooted in concepts, abstract representations (extracted) from reality, made up of systems of signs (for example formal languages) seeking to reconstruct a duplicate of real objects and to unify the empirical data that come from them. It aims at linking the latter in order to reveal a unity (this is intelligence in the Latin sense of *inter legit*). This knowledge proceeds by analysis and deduction. It aims at problem solving and prediction based on laws expressing necessities, regularities, invariants, and universal properties. This type of knowledge is based on a sum of past data that constitutes a reified (formalized) and autonomous "double" of reality (the models that produce a simulation of reality appear as realities functioning by themselves, with relative autonomy: the program runs by itself!). It is typically this kind of knowledge that constitutes the scientific intelligence of the world, in the modeling of engineers and physicists, and it is also that we found in the approaches of artificial intelligence that seek to abstract invariants, correlations and regularities from Big Data to provide a basis for future prediction, with a view to a particular utility. But notional knowledge alone does not exhaust all that we understand by knowledge.

ISSN 0039-4130

Indeed, there are domains where notional knowledge is inoperative: the domain of contingent phenomena, where pure singular events, unpredictable and "unpreconceivable" as Stuart Kaufman[2] puts it, determine the real. The understanding of history (social, political or economic) is part of this case. The knowledge of human persons and their intentions escapes also, for the most part, abstract and conceptual analysis. One can even say that such an analysis could, through its objectification, destroy the human relationship by which one aims to know a person (conceptualizing love and objectifying it destroys it). Here we could say that in order to know we must no longer distance ourselves from the world, but adhere to it. Knowledge proceeds here without concept and without necessarily direct utility. It is based on affect, on emotion, on intuition more than on reflection. Knowledge through empathy, through the fact of feeling in one's body something of the misery, of the suffering of the other, belongs to this type of intelligence now understood in the Latin sense of *intus legit* (the intelligence that sees and reads within, in the heart). The relationship to the body is decisive here, as well as an intuitive (global) understanding of situations based on experiences made in contact with the world, what could be called a "common sense" of reality. We can therefore realize the nonsense of these AI systems that offer supposedly dialogues with avatars of deceased people. It is a dialogue that completely erases life and the real body in order to propose simulacrums.

The knowledge of the world and of the human being that great literary and artistic works give us, is part of this type of intelligence that is not entirely deductive and notional. Creative intelligence is also of this order, there is no formal and predictive model of mathematical or musical creation. We can make algorithms to prove theorems from systems of axioms, we can build algorithms to produce paintings or music in the manner of the great composers. But it is impossible to predict what will be, in the next century, the great and authentic theoretical creation in physics or the great musical work!

In the history of philosophy, we find these two irreducible facets of intelligence, the biological conditions of possibility of which could be

[2] Cfr S. A. Kauffman, *A World Beyond Physics: The Emergence and Evolution of Life*, Oxford University Press 2019. An event is said to be "unpreconceivable" when its occurrence cannot be the subject of any prediction, even a probabilistic one, because the set of possibilities from which it stems cannot be conceived or characterized (because it is too "large" to be apprehended and no measure of probability can be attributed to it). Some crucial events in human history or biological evolution are of this nature.

ISSN 0039-4130

highlighted by the most recent research in neurophysiology [3]. The distinction between *noesis* and *phronêsis* in Aristotle, between *ratio* and *intellectus* in the Scholastics, between the *knowledge of the 2nd* and *the 3rd kind* in Spinoza, between the *spirit of geometry* and the *spirit of finesse* in Pascal and perhaps also, as Fr. Pascal Marin [4] has clearly shown, between *language* (which can be coded) and *speech* (which, always personal and unpredictable, is improgrammable) [5].

This detour through a fundamental reflection on what intelligence is, happens to be very important for evaluating artificial intelligence from an ethical point of view. Indeed, the latter unilaterally overvalues notional intelligence by obscuring the other. With large language models, with ChatGPT, we are indeed in the realm of prediction based on the acquisition of an enormous amount of past information that is represented in systems of manipulable signs, forming like formal doubles of reality, but we have not at all entered, in the slightest, into the crea-

[3] *Cfr* Ch. C. Ruff, E. Fehr, *The Neurobiology of Rewards and Values in Social Decision-Making*, in *Nature Reviews Neuroscience*, XV, 8, 2014, p. 549-562 ; D. Kahneman, *Système 1, système 2. Les deux vitesses de la pensée*, Flammarion, Paris 2012. An illustration: Neurophysiology has highlighted two pathways leading to decision-making, involving different but connected regions of the brain. One intuitive and fast, based on emotions and the other, analytical and slower, based on conscious knowledge and reflexive (logical) processes.

[4] P. Marin, *Le robot et la pensée. Contre-philosophie de l'homme-machine*, Cerf, Paris 2019.

[5] Bergson sums up this difference well by calling «intuition the sympathy by which one transports oneself into the interior of an object in order to coincide with what is unique and consequently inexpressible in it. On the contrary, analysis is the operation that reduces the object to elements already known, that is to say, common to this object and others. To analyze is to express something in terms of what it is not» («intuition la sympathie par laquelle on se transporte à l'intérieur d'un objet pour coïncider avec ce qu'il a d'unique et par conséquent d'inexprimable. Au contraire, l'analyse est l'opération qui ramène l'objet à des éléments déjà connus, c'est-à-dire communs à cet objet et à d'autres. Analyser consiste à exprimer une chose en fonction de ce qui n'est pas elle». In *La pensée et le mouvant*, Alcan, Paris 1934, pp. 178-179). Both types of intelligence, and the knowledge they provide us with, are both important. As this student of the Master of Aix, Paul Archambault, says: «Between these two forms of knowledge, there is no choice for us: they are both necessary, one to the other [...] thanks to their union alone, can we reconcile the universality of a horizon coextensive with the whole being and the singular point of view of a personal conscience...» (« entre ces deux formes de connaissance, *il n'y pas pour nous à opter* : elles sont nécessaires l'une et l'autre, l'une à l'autre [...], grâce à leur union seule, peuvent se concilier en nous l'universalité d'un horizon coextensif à tout l'être et le point de vue singulier d'une conscience personnelle…». In *Initiation à la philosophie blondélienne en forme de court traité de métaphysique*, Bloud & Gay, Paris 1941). For Blondel, these two facets of intelligence, although they are inseparable, never manage to come together completely. The "*noetic*" (it means: conceptual, analytical, calculative) aspect of notional intelligence never succeeds in absorbing the "*pneumatic*" aspect of intelligence that John Henry Newman calls *real* but which could be called to be understood today *intuitive* or *lived*. And it is in the fissure between these two intelligences, which, we feel, should be unified and equal in order to meet the very requirements of thought, that what we might call a true transcendence of the act of thought appears as if implicitly.

ISSN 0039-4130

tive act, we have only imitated it. We can dream of transforming justice into the application of deontic algorithms, but we know that notional and formal knowledge of legal systems never replaces the knowledge acquired, over the course of experience, by a magistrate and which allows him, through a kind of intuition, wisdom and common sense, to qualify, even in uncertainty. The same is true for any decision-maker who has to act in a context where information about the world is partial or blurred and where an intuitive knowledge of the world makes it possible to risk actions whose possible decision was not deducible from a general law. By obscuring an essential facet of intelligence, the necessity of which is felt to be essential in the knowledge of everything that has to do with the life and death of the human being, and by aiming at a reduction of thought to this unilaterally *noetic* facet, the designers and users of artificial intelligence undermine the richness of intelligence and risk distorting what we know about reality. As such, the voluntary and ideological concealment of a component of thought is, for us, one of the roots of the ethical questions raised by the use of artificial intelligence.

The question of freedom in the context of AI can as well be approached from two angles. The first is that of individual freedom, which is susceptible to manipulation. The second angle is that of delegating our decision-making or action powers to AI systems or autonomous robots piloted by them and which would be considered as the equivalent of autonomous agents, of a system endowed with a kind of freedom.

Freedom, like intelligence, has two inseparable facets[6]: freedom and free will. We are moved by a powerful impulse, by a (willing) will that seeks happiness, that is to say, the perfect adequacy with what we are deeply. A will that is fully free from the constraints that would prevent it from attaining this end is precisely what we call *freedom*[7]. But

[6] We refer here to the fine analyses of the will of François-Xavier Puttalaz in the edition of the *Somme Théologique. L'âme humaine (traduction et notes de François-Xavier Puttalaz)*, Cerf, Paris 2018, pp. 713-728.

[7] When *freedom* is true, the will cannot hesitate for a single second in the face of the ultimate good it aims at, like that of the rescuer who gives his life freely and instantaneously. As the philosopher Jacques Beaufay says, evaluation and free will often intervene : «lorsque je ne veux pas assez une chose pour que la décision soit emportée sur le champ». «La liberté comme "petit mouvement de l'âme". Responsabilité morale et métamorale », Cahiers de l'Espace Philosophique (Université de Namur), n°24, octobre 1997; «La liberté culmine quand elle n'a plus le sentiment d'être "libre" (de choisir), elle produit alors un acte qui vient vraiment de moi, m'est naturel comme une nécessité intérieure» (p. 8); «… la liberté intervient à *tous les stades de la volition*, elle s'exerce non seulement dans le choix mais dès l'admission des possibles» (p. 7).

ISSN 0039-4130

in order to arrive at the end we must choose the appropriate means that our reason proposes to us, that we judge the one that is the best. *Free will* is, on the other hand, this ability (this particular mode of the human will) to choose one of these means in accordance with the deliberation of our (practical) reason. *Free will* is the way in which our will gradually chooses partial goods, which form what we might call the *desired will* ("volonté voulue"), gradually allowing us to reach the end that its deep impulse, *the aiming will* ("volonté voulante") has given it to apprehend. These choices of goods or partial means gradually help the will to become free, that is, to coincide fully with what we must be in depth (by nature could one say, using classical terms).

We can then see a problem related to two reductions. On the one hand, it is a question of reducing the question of *freedom* to that of *free will*, to the choice between partial goods. On the other hand, it is a question of the reduction consisting in identifying the judgment that is associated with *free will* with a purely *noetic*, formal, and even mathematical reasoning (of probability, optimization, deduction in modal logic, etc.).

The problem here is that judgment (of prudence, phronesis, that of the magistrate, of the decision-maker in general) cannot be identified with a purely logical-deductive, algorithmic procedure (2nd reduction), and that the focus on particular goods (1st reduction) makes us completely forget the deeper purpose that carries the will and risks absolutizing particular goods (aiming for example the profit of a small group while deep happiness would aim at the common good, ...).

In fact, the central question that AI poses to ethics is the confinement of intelligence and freedom in the coherent, formally closed, but partial world of notional and computational intelligence, digital procedures, etc. To raise the question of freedom in the face of AI, we will have to ask ourselves at all times both whether we have not restricted *freedom* to *free will* (forgetting the relationship to the deeper ends of the human, what is his individual or common good [8]) and whether, on the other hand, we have not associated this *free will* with a procedure of choice considered only from the formal (*noetic*) point of view: logical, mathematical, algorithmic.

[8] It would be interesting here to return to the dialectic of desire in Bruaire, who sees in freedom the mediation between desire and language, avoiding desire (Cl. Bruaire, *L'affirmation de Dieu. Essai sur la logique de l'existence*, Seuil, Paris 1964, pp. 154-168).

ISSN 0039-4130

3. *AIs at the service of humans, their intelligence and their willpower*

But let us clarify our thought. Artificial intelligence, like notional intelligence, is linked to the nature of the human. There is no human intelligence that would be conceivable without the "*noetics*" and all that stems from it: as theoretical representations, formal systems, robots, algorithms, ... etc. The applications of AI are therefore essential, and to refuse AI would be to deny what makes us a human being! All what can be linked to intelligence inside AI comes from the human being, there is no intelligence of the machine as such, but only a trace, a facet of the human intelligence inscribed in the machine, just as there is a trace of the intelligence of the scribe in the papyrus, itself not being intelligent. The question is therefore not "how to escape AI?", but how to prevent it from depriving us of the freedom to approach the world in ways that are not only notional, and of the freedom to preserve other ways of accessing reality and apprehending the human person.

We must therefore avoid technophobia and emphasize that human freedom can be served by AI: freedom is ensured by security and today AI systems make it possible to ensure the security of communication systems, sensitive installations, etc., in real time, but also the safety of surgical gestures in medicine. But it also gives humans more freedom by relieving them of very tedious tasks of classification, research, etc. Not to mention the services that AI can provide in the field of disability: restore mobility when AI controls the interface between the brain and an exoskeleton or when AI manages character recognition in a text that a blind person would not have been able to read without it.

4. *Freedom hindered by AI*

But of course, it's important now to highlight AI's obstacles to freedom. We will limit ourselves to briefly recalling some of these obstacles which are well known.

AIs, with the data that we freely offer them (or in the insouciance or unconsciousness of our quick acquiescence made by the least effort) plunge us into a society of permanent surveillance (this is *The Age of Surveillance Capitalism* described by Shoshana Zuboff[9]). And this tacitly consented surveillance leads to a channeling of our choices by incen-

[9] New York, Profile Books, 2018.

ISSN 0039-4130

tive techniques (nudging: our will is pushed imperceptibly in certain directions dictated by commercial imperatives). Our freedom of choice is restricted without us noticing. We are also trapped by well-studied addiction techniques (rewards, ...) that make us pathologically dependent on applications that make us deliver more and more personal data, which, in turn, will restrict our freedoms and the limits of our privacy.

We are also trapped by our own conceptions and desires, in digital bubbles that induce a phenomenon of social fragmentation at the heart of a network that is supposed to unite us. AIs monitor you and detect some of your preferences (example: political, ideological, artistic, etc.) and will automatically offer you products or contents that correspond to them and that reinforce your desires and ideas. Little by little, you are only confronted with "your" world. You imprison yourself freely in yourselves, in your fantasies, in your dreams. But it is here that an in-depth reflection on freedom must come into play: freedom is the fact of acting in coincidence with what we are in depth and not doing everything that is possible for us. However, what we think, we want and what we dream, and which we express by researching a particular activity, a particular hobby, etc., does not necessarily correspond to what would make us happy (precisely because it does not correspond to who we are, in our deep nature).

The constraints on our freedom can be seen clearly and more simply in the "assistances" to spell correction or writing in various applications of our computers and smartphones. The AI automatically corrects you and suggests words or phrases that you accept to go faster but that you might not have chosen if you thought about it. Our creative freedom is diminished... supposedly for our good (defined of course externally to our will).

Let's note in passing that generative AIs such as ChatGPT can, on the one hand, increase our freedom (to do something else!) by relieving us of tedious tasks such as searching for references or writing administrative texts. But, on the other hand, they lock us in the past. Why? Because ChatGPT only knows what it has been taught, namely all the literature available now, i.e. the literature that belongs to the past! There is no creativity in this kind of AI (but of course some tasks don't require it!) but by dint of using it, we become nothing more than parrots repeating the past, condemned to read what has been written but gradually becoming unable of thinking anything else. It is interesting here to quote this beautiful sentence of Blondel in his first book *L'Ac-*

ISSN 0039-4130

tion (1893) in a chapter that touches on the question of freedom [10] :

> «True knowledge is that reflection which carries forward the inner gaze towards the ends which solicit the will, because there alone is the sufficient reason for free determinations. Whoever is born for action looks ahead; or if he seeks where he comes from, it is only to know better where he is going, without ever locking himself up in the tomb of a dead past. Forward and upwards, the action is only through this […] the voluntary action appears as a creation within the creation »

It is interesting to see the link between creation and truly free will. We are no longer totally ourselves if we only repeat what the past has given us, we are no longer truly free if we are enslaved to a past knowledge without having the possibility or the courage to produce something truly different!

One of the problems of AIs is that they paint a digital portrait of you (with information obtained thanks to a very large number of "small" agreements given for ease when you access a site, when you use this or that service: browsing, music, ,...), a portrait that traps and crystallizes, in the eyes of the world, your personality for a more or less long time... time that will be enough for you not to be given a job, insurance or constantly offered things that do not interest you. This reduction to a numerical profile also corresponds in the academic or business world to rankings that will serve as a judgment to exclude you from a nomination or to give arguments to the jury to oust you when there would be no deep arguments to do so. It should be noted here that we are not free to escape these numerical evaluations since at some job interviews, it will be asked to provide them under penalty of being disqualified.

One of the important aspects of how AIs restrict freedom is the ability to disseminate information widely at very high speed. Today, all academic talks tend to be filmed and placed online (more or less with your consent, extorted on the sly after the conference). In a sense, this is important, because a wider audience can hear you and students who

[10] M. Blondel, *L'Action (1893)*, PUF, Paris 1950 (1973), p. 123. The original French text says : «la véritable connaissance c'est cette réflexion qui porte en avant le regard intérieur vers les fins qui sollicitent la volonté parce que là seulement est la suffisante raison des déterminations libres. Quiconque est né pour l'action regarde devant soi ; ou s'il cherche d'où il vient, c'est seulement pour mieux savoir où il va, sans jamais s'enfermer dans le tombeau d'un passé mort. En avant et en haut, l'action n'est que par là [...] l'action volontaire apparaît comme une création dans la création».

ISSN 0039-4130

would not be able to afford to travel can benefit from your teaching (which they can also watch at their own pace, which is valuable when it comes to presentations in languages they are not familiar with). But the disadvantage is that the larger the physical or virtual audience (and is no longer controlled), the more careful you have to be. Sometimes, in teaching, you have to be able to shake up audiences, to risk thoughtful deviations. And as a result of the transfer, the high-speed percolation of teaching to all kinds of platforms managed by AIs leads to inhibited thought and creativity. Freedom of speech can take a serious hit!

Another aspect concerns the restriction of freedom linked to the exclusion of people who do not have the cognitive or financial means to access AI services. What if to access my pension file or my medical analysis results, I have to go through an authentication and security application managed by AI accessible only from a smartphone? Are we free in the face of AI if we are in a country where access to electricity is complicated? Digital divides diminish the freedom of a number of vulnerable citizens.

Let's also note a phenomenon of "digital colonization" that consists of being cognitively and financially enslaved by great Powers or Nations that, from afar, capture your data and guide you remotely with your consent. And of course, this does not affect only poor countries. On the contrary, it is those who have access to AI who can become the most vulnerable.

One of the most crucial points where we perceive that AIs could be an obstacle to freedom (which is a major principle of democracy) are those applications that will strengthen your political inclinations or those that will automatically and with malicious intentions produce deepfakes, rumors, contents or images that discredit candidates in elections. The disruption of the democratic process by this kind of technique is all the more powerful because most people do not take the time to reflect or analyze sources. It is worth noting that AI tools that help you make your choices as voters can also infringe on your freedom. The algorithms based on your options (for or against certain things) can recommend this or that party to you... But this recommendation is not neutral and one may have had the experience (as a believer) of refusing or accepting this or that ethical option (concerning the dignity of persons) without wishing to find oneself with the members of a particular party advocating the same options in passing, but being very far from your political preference!

This allows me to raise a very important ethical problem which is

ISSN 0039-4130

often questioned by those who, with good intentions, want to save AI from technophobes by pointing out that the AI tool is neutral in itself, and that only their applications would have a determined moral charge. We disagree with that. Not because we are technophobic, but because the only way to defend AI is to measure adequately its potential benefits and also its real potential risks. However, there are risks that do not come from AI applications but that exist upstream, insofar as the writing of algorithms and the Big Data used to train AIs carry conscious or unconscious biases. The tool is not neutral because it already carries within it an inclination towards this or that valorization or discrimination. In the 1970s, there was a lot of reflection in the ethics of science around the non-neutrality of certain technologies. The idea was that certain technological tools already reflect into their very constitution the ideological options of a type of society (and therefore of political choices) and a type of relationship between people that are not neutral. In fact, AI is not a neutral tool: it really builds and frames a type of society and relationship between people, it also implies a relationship with the environment and therefore an influence on climate and on Earth resources (choosing AI means choosing to store large amounts of data in energy-intensive clouds and dependence on rare chemical elements concentrated in certain regions); choosing AI is not choosing a neutral technique whose applications only would raise an ethical question.

Thus, before any use, in the development process, the environment in which AIs are trained can a priori induce a restriction of freedom for certain classes of users (for example intelligent taps that recognize the presence of hands to deliver water in an economical way and whose AIs have only been trained on the hands of white people) or prevent conditional release for certain classes of prisoners.

Let us conclude our overview of the problems that AIs address to the respect of individual or social freedoms with an observation.

By using AI Applications, we believe we are free, but deep down, through less effort and recklessness, we are unknowingly subservient to the recommendations of economic or financial groups. Our behavior is becoming more and more robotic. But, at the same time, we renounce, knowing it, our freedom, by delegating important powers of action and decision power to autonomous machines that, in fact, have neither intelligence nor will. It is therefore a double renunciation of our fundamental freedoms and a double risk of submission of humans to machines, with the loss of responsibility that this entails. We are enslaving ourselves to the AIs and we are abandoning our powers (including

ISSN 0039-4130

sovereign powers) to the AIs. It is therefore important to protect our freedoms and to put people back at the heart of the processes that manage the use of AI. Regulation is necessary, but how should it be conceived? This is the subject of the last part of our reflection.

5. *Ethics, freedom and AI regulatory models*

How should we think about the relationships between AI and freedom? Several models are possible and well known.

A first type of model is collectivist.

Here the idea is that individual freedom does not have to be protected because what counts is the usefulness (global well-being, security, stability, health...) of a group, whether it is a nation (as in China) or of a given community (as in the philosophy of Silicon Valley in the USA). In this context, there is no longer a distinction between private and collective data, because it is assumed that everything that can be used for the usefulness of the group, takes precedence over what an individual might think or possess. This kind of perspective goes hand in hand with a monistic (in Asia) or utilitarian philosophy ("à la Bentham", in the West). The choice not to take into account individual freedom (to share personal data or not, for example; to refuse permanent geographical or health surveillance, etc.) is a choice that leads to a normalization of thought and action with an overvaluation of individuals' compliance with collective norms.

From a positive point of view, there is a resolutely open will to technologies and innovations that can serve humanity, with a concern for the collective good. But this is done with a total ignorance of the value of the person and his/her dignity and a contestation of the values of democratic debate. This kind of perspective can only be totalitarian. Individuals lose all degree of thought in the face of the "governing machine" that Fr. Marie-Dominique Dubarle already denounced in 1948, in the context of the nascent cybernetics[11]. It can also lead to overconfidence in progress leading to a lack of criticism of the limits

[11] *Le Monde*, 28 décembre 1948, pp. 47-49. For an analysis of Wiener's reaction to this article, see R. Le Roux from the translation of N. Wiener's book, *Cybernétique et société. L'usage humain des êtres humains* (trad. par P.-Y. Mistoulon), Seuil, Paris 2014, pp. 29-30.

ISSN 0039-4130

and risks of the technology itself (including in the long term, because of compliance with respect to the collective norms, for the development of creative thinking for AI). One could also say that this collectivism remains contradictorily a prisoner of the world thought by a few (and not by the collective as a whole) who believe they know what the collective good is.

A second type of model is individualistic.

On the other hand, we can defend the rights of the individual. This point of view is very well described and defended in France by Gérald Koenig (in his book *La fin de l'individu*[12]: "The end of the individual"). The starting point here is that the protection of the individual has always been a factor of emancipation. It is therefore necessary to defend the freedom of the individual (and his/her capacity to deliberate) by allowing him/her to diverge from the norm and from the collective well-being (Koenig insists on "the right to make mistakes, to diverge") inspired by Stuart Mill's philosophy. Here we have another argument for the defense of the position: wandering, mutation has always been a driving force of evolution and life, so we must preserve this aspect on pain of dying or seeing human history stagnate. This author argues that the protection of the individual can only be conceived by imposing the choices of the individual on AIs (a choice framed by an appropriate legal mechanism). Thus, personal data remain the exclusive private property of the individual who can monetize or retain it, even if it is to the detriment of the group or if it is not optimal for him/her or others. Platforms must also remunerate the individuals who provide them with the data.

This approach is very interesting because it gives back his place to the individual who resists the enslavement of his/her life to the interests of political or financial groups. It is also interesting because it restores, to thought and will, their true creative dimensions: it is the freedom to be "decoincident" ("not coincident" to use François Jullien's terms[13]), to transgress the canons, which makes humanity move forward (art but also science: Cantor said «The essence of mathematics is freedom!») and which prevents it from being a prisoner of the "closed societies"

[12] Paris, Editions de l'Observatoire, 2019.

[13] F. Jullien, *Dieu est dé-coïncidence*, Labor et Fides, Paris 2024. I thank Thomas Antoine for introducing me to this book.

ISSN 0039-4130

denounced by Bergson in *Les deux sources de la morale et de la religion* (precisely because their closure stops evolution). But, the limits of such a coherent, realistic and stimulating approach is that it refuses (by definition could one say) any relationship to the common good. One might think that an individual's refusal to provide (for free?) personal information that could help protect the group could be seen as a mistake and at the limit as a self-contradiction (because being part of this group he could eventually benefit from it). The overvaluation of a freedom considered as a free-will, measured only from the individual point of view, could lead to the destruction of the foundations of solidarity. Of course, the defenders of this conception are coherent, but they must then accept a society without solidarity, where the interests of the most fragile (their right to health, to access to knowledge) could be undermined by personal decisions to withhold knowledge, data, etc.

A third type of model is personalistic.

Here, I start from the principle that the human person (who cannot be defined only by characteristics, quantified and calculated performances and standards, but by an intrinsic dignity) is also a being of relationships, of relationships with his/her natural environment, but also with others in a society. Since the beginning of its evolution, humans have been characterized by their very high tendency to sociability. What contributes to the good of the person lies both in what he or she owns in his or her own right, but also in the goods that he or she shares with others, which, as such, cannot be appropriated for partisan purposes. The natural environment (the seas, the Antarctica, the Earth atmosphere, etc.) and even the Outer Space environment are constantly coveted commodities, but whose defense is imposed as something not appropriable. Some data that are essential for the security of health, for the survival of our planet could be considered (and defined by a legal instrument) as common goods, which must be shared (of course under appropriate legal regulations).

In this conception, innovation is not renounced (on the contrary, since the capital of important data available to train AIs is increased), but the rights of the individual are perfectly protected by tolerating this sharing of data, only for the common good (precisely the one that allows the person to achieve his or her full potential, in relation to others and to his or her environment). On the other hand, the integral protection of

ISSN 0039-4130

personal freedoms consists in ensuring that the concern for the common good does not turn (through increased surveillance, through massive intrusion into the private sector) into a kind of primacy of the collective (in fact the primacy of utility for a partisan group). It should also be noted that a lack of regulation (by overvaluing the individual dimension) would amount to admitting any content on the network (including hateful, racist, etc.), but an excess of regulation would annihilate any different (original, marginal) thought, any personal creativity [14]. This is why I think that the relationship to the common good and to the deep goals (the willing!) of the individual constitutes a path towards a regulation of AI that respects the person in all his/her dimensions.

In a certain way, this way of conceiving the protection of freedoms in the face of AI takes up what is positive in the concern for the collective: humans can only survive by respecting the good of all. But it puts barriers to this requirement, by refusing anything that could undermine what makes people unique.

6. *Conclusion: respecting freedom implies a regulation in tension between the collective and the personal. A common good personalism proposition!*

The position that I would like to value and submit to the debate is in line with that of Blondel, defending the person and the dynamic link between his/her "*volonté voulante*" and his/her "*volonté voulue",* as well as the articulation between the *noetic* and the *pneumatic* facets of the human thought, which constitutes the life of his/her intelligence (refusing to reduce the person to a homogeneous numerical individual, to an intelligence without body and without singularity or originality). But my position could also be compared to that of Fr. Teilhard de Chardin, who defends a personalization that goes hand in hand with an intensification of relations in the collective network of the "*noosphere*", in the context of a resolute faith in the future. Deep attention to the person cannot be conceived without being conscious to belong to a whole, without the bond with others in a deep fraternity [15]: "The human sense, on pain of being inhuman, must be of the order of love".

[14] *Cfr* A. Grinbaum, *Parole de machines. Dialoguer avec une IA*, humenSciences, Paris 2023.

[15] P. Teilhard de Chardin, *Esquisse d'un univers personnel*, in *L'Énergie Humaine. Œuvres de Teilhard de Chardin 6*, Seuil, Paris 1962. , p. 101.

ISSN 0039-4130

The question of freedom in the age of AI must be thought of in the balance between a conception of the collective (without being collectivist or totalitarian) that does not exclude any personal singularity, however fragile and vulnerable it may be, and symmetrically, a vision of the individual (without being individualistic), of the person, who is constantly opening up and sharing, to those with whom he wants to be close, all that is common to them, because they are brothers and sisters, belonging to a world where everything is held together and where nothing can be radically atomized without disappearing [16].

[16] The operational or practical point that is necessary at this stage would be to know how to think of a legal tool that preserves this conception of the human, which is both personalist (without being individualistic) and collectivist (without being totalitarian). It seems that recent European legislation (AI Act, 2024) is partly moving in this direction. The approach in terms of risk levels makes it possible to think of a defense of the person (including the most vulnerable), but it does not leave all the power of data retention in the hands of the isolated individual. Something of the common good is implicit, even if it should be even more marked.

ISSN 0039-4130

Free Will, Neurosciences & Robotics

Sara Fernandes, *Leonor Almeida* and *Alexandre Castro Caldas*

1. *The Neuroethical Problem of Human Freedom*

The neuroscience of ethics is the domain of neuroethics that investigates the problem of human freedom by studying the brain of the brain. Both philosophy and neurosciences strive to understand the biological and mental distinctions between behavior driven by freedom of will and behavior that lacks it. However, it is essential to reflect on the value attributed to neuroscientific discoveries in understanding human beings, particularly concerning the phenomenon of agency in the world. Are the findings of neuroscientific research on human behavior sufficient to offer a philosophical answer to the problem of human freedom and determinism? [1]

Recent advances in neurosciences have enabled a more rigorous systematization of the brain areas involved in reasoning and ethical decision-making. For example, the amygdala detects, evaluates, and assigns emotional significance to an individual's options, while the hippocampus complements this emotional evaluation with autobiographical memory. The anterior cingulate cortex helps us anticipate and solve practical problems, recognize mistakes, and manage situations of uncertainty. Meanwhile, the hypothalamus regulates the body's internal stability and links it to survival behaviors. Finally, mirror neurons allow us to internally simulate actions performed by others or ourselves without carrying them out. As a result, mirror neurons activate whenever an intention is present in an individual's mental process or when they per-

[1] A. Lavazza, *Free Will and Neuroscience: From Explaining Freedom Away to New Ways of Operationalizing and Measuring It*, In *Frontiers in Human Neuroscience*, X, 262, 2016, pp. 1-17.

form intentional behavior [2]. Thus, we understand that the orbitofrontal region is mainly responsible for the social brain. When it is damaged, such as by trauma, the individual may no longer exhibit appropriate behavior (behavior that is socially expected of oneself and others).

We need to revisit an important question: How should we interpret these neuroscientific findings from a philosophical standpoint? Do advances in our understanding of neurobiology undermine or even eliminate the concepts of personal freedom and responsibility? Do these findings imply that humans—whether they are healthy or experiencing a central nervous system disorder—are essentially prisoners of their brains, with their decisions completely determined by this organ, thus invalidating the idea of free will? In essence, to what degree do the findings from neuroscience about human behavior and decision-making challenge the widely held belief in freedom? Or, do they simply help us understand the neurobiological processes that underpin human actions and the concept of freedom? [3]

Rizzolatti's team's unexpected discovery of *mirror neurons* represent one of the most significant breakthroughs in neurosciences in recent decades. This finding has provided significant insights into how the social brain functions, including social skills, learning processes, and the emotional and cognitive aspects of empathy. While studying neurons in a region of the Rhesus monkey brain that controls the hand muscles, the team expected to find neurons that activated when the monkey engaged in specific actions, such as catching a ball or reaching for a banana. They discovered that certain neurons did activate during these actions. However, a surprising observation occurred when the researchers had lunch in the same room as the monkeys. They noticed that some of these neurons also fired when the monkeys watched an experimenter perform the same action, like bringing food to their mouth.

In summary, mirror neurons activate when a monkey either performs an action or observes another—be it a person or another monkey—doing the same action.

This discovery scientifically justifies the importance of intersubjectivity for the constitution of the person and human sociability in general, as Greco-Roman civilization had emphasized on a philosophical

[2] L. Tancredi, *Hardwired Behavior. What Neuroscience reveals about morality*, Cambridge, Cambridge, Cambridge University Press, 2005.

[3] A. Lavazza - S. Inglese, *Operationalizing and Measuring (a kind of) Free Will (and Responsibility). Towards a New Framework for Psychology, Ethics, and Law*, In *Rivista Internazionale di Filosofia e* Psicologia, VI, 1, 2015, 37-39.

ISSN 0039-4130

level for considerable centuries in the West. Mirror neurons, activated whenever someone acts, also enable the internal simulation of behavior practiced by others or oneself without carrying it out. Simply remembering an action or imagining that you or someone else will act can activate these neurons, underscoring their crucial role in understanding human behavior. Mirror neurons are activated whenever there is a mental intention, whether intentional behavior is performed or observed in someone else. Their activation in real and purely imagined scenarios (representing an "as if" experience in others or oneself) suggests they play a significant role in fundamental human traits. These include feeling, sharing, and recognizing emotions in others—traits closely tied to empathy. The ability to empathize, which involves identifying and sharing in others' sadness and joy, is essential for forming distinctly human relationships like friendship and love. It also underpins various forms of learning, both cognitive and social, as imitation serves as a core mechanism in the learning process. [4] As A.Lavazza argues:

> «Until a few years ago, empathy was mainly an object of philosophical and psychological research; then the discovery of mirror neurons, considered by many (though certainly not all) to be a key mechanism in empathy, brought research in cognitive neuroscience to the fore. Having identified the circumscribed brain areas that are activated both when we perform an action and when we observe someone performing that action marked a turning point in the debate on the genesis of understanding and identification with the experiences of others, one of the keys to social life. That empathy can be embodied and primary, almost an automatism that we are all equipped with (unless one has neurological deficits), has challenged many assumptions about the role of education and culture» [5].

2. *The Neuroscientific Research of B.Libet and P.Haggard*

B. Libet and P. Haggard carried out the first neuroscience studies that significantly challenged the belief in human freedom. The first work, published by neuroscientist B. Libet, sought to show that the individual is only aware of his intention to act after the brain has prepared its body for action (potential readiness) [6]. The research showed that the

[4] L. Cattaneo - G.Rizzolatti, *The mirror neuron system,* in *Archives of Neurology,* LXVI, 5, 2008, p. 557-560.

[5] A. Lavazza - S. Inglese, *Operationalizing and Measuring*, cit.

[6] B. Libet, *Unconscious cerebral initiative and the role of conscious will in voluntary action*,

brain prepares the individual for action by activating the motor cortex before they know their intention to act. So, they proved the existence of temporal unconscious brain processes before conscious processes in individuals. Secondly, the study showed that unconscious brain processes set the stage for conscious processes related to human intention, indicating that intentional states cannot exist without these underlying unconscious neurological mechanisms.

In a later study, Haggard and Libet developed this research with more sophisticated measuring instruments, such as the electroencephalogram. They concluded that, in preparation for the action, before the individual became aware of their intention, the brain prepared their body in general and the specific side of the body with which they were going to act, activating the premotor cortex [7]. Thus, in addition to the readiness potential, there was also what the authors called a lateralized readiness potential.

With these studies, the authors brought the debate on human freedom into the scientific field. To put it another way, they used the contributions of neurosciences and empirical psychology to broaden its scope to new research areas. They hope that, since philosophers haven't 'solved' the problem so far, these new areas can do so [8]. Philosophical research is fundamentally conceptual. So, when the two fields intersect, our first task is to understand and clarify—philosophically—what neuroscientists mean by 'being free' when they use this term. In our view, their understanding of being free implies that the entire chain of events leading to an action is under the conscious control of the person performing it. In this context, consciousness is considered a crucial part of free behavior. This means that for an action to be considered free, consciousness must directly cause it.

Libet and Haggard interpreted their findings in a way that challenged the belief in human freedom. They argued that human behaviors, even those that seem free based on the individual's feeling of freedom (connected to the intention and decisions made by the "agent"), are actually influenced by prior events that the individual cannot consciously control. Thus, in the authors' view, these results allegedly show that we do not consciously cause our intentions, decisions, and volitions and, in this sense, our actions, so we cannot consider ourselves free.

In *Behavioral and Brain Studies*, VIII, 4, 1985, p. 529.

[7] P. Haggard - B. Libet, *Conscious Intention and Brain Activity*, In *Journal of Consciousness Studies*, VIII, 11, 2001, pp. 47-64.

[8] A. Lavazza - S. Inglese, *Operationalizing and Measuring*, cit., p. 38

ISSN 0039-4130

Libet and Haggard also drew attention to another research result, which has caused immense surprise and reflection in the scientific community ever since. The researchers observed that whenever the intention (to touch the button) became conscious in the experimental subject, they still had a very short period to inhibit their conscious intention and, therefore, not carry out the intended movement (touching the button). The authors suggest two possibilities regarding the belief in an agent's conscious freedom. One possibility is that this belief is an illusion, as conscious actions are influenced by unconscious mental states that can even override the agent's intentions. Alternatively, they propose a more remote hypothesis: the agent may retain some control over their behavior, which is primarily expressed through their initial intention to block certain movements. However, the neural correlates and genesis of this (minimal) kind of self-control (as a veto, denial of intention, and the corresponding movement) have yet to be determined [9]. Based on empirical evidence, those were strong reasons to abandon the widespread, deeply held belief that human beings are free, i.e., possess the ability to initiate their actions through entirely conscious and self-controlled free will.

More recent research, conducted by a team of neuroscientists led by John-Dylan Haynes, has identified the emergence of both behavioral and abstract choices, such as the simple act of raising a hand or performing small mathematical calculations (like addition and subtraction), seconds before the experimental subjects were even aware of their intentions. Furthermore, even though the study involved basic tasks, it suggests a future possibility of being able to "read" our minds using techniques like functional magnetic resonance imaging. This could allow us to know our upcoming choices, thoughts, and mental states—essentially, our private mental experiences—even before we consciously realize them [10].

These neuroscientific studies most significantly challenged the belief in human freedom. They showed that the individual is only aware of his intention to act after the brain has prepared his body for the action (potential readiness). Without these unconscious neurological mechanisms, there could be no intentional mind state. Thus, to neuroscientists, these results show that we do not consciously cause our

[9] A. Lavazza, *Free Will, and Neuroscience*, cit., p. 14.

[10] C.S. Soon - M. Brass - H.J. Heinze - J.D. Haynes, *Unconscious Determinants of Free Decisions in the Human Brain*, in: «Nature Neuroscience», XI, 5, 2008, pp. 543-545.; C.S. Soon, A.H.He, S. Bode, J.D. Haynes, *Predicting Free Choices for Abstract Intentions*, in: «Proceedings of the National Academy of Sciences», CX, 15, 2013, pp. 6217-6222. A. Lavazza - S. Inglese, *Operationalizing and Measuring*, cit., p. 40.

intentions, decisions, volitions, or actions. So, we should not consider ourselves free. From a philosophical point of view, Libet & Haggard defend determinism, challenging a specific conception of freedom, libertarianism. For the followers of the libertarian current, to be free is to be free from any external or internal constraints. Nothing beyond the agent's conscious motivational states can influence his choices and actions without diminishing his capacity for action.

However, it is not clear that the existence of unconscious brain processes, temporally before conscious mental states, implies the conclusion that we are not free or responsible agents. For four reasons:

1) We can interpret the temporal precedence of unconscious mental processes to the consciousness of intention as a preparation of the human organism to form intentions and make decisions. Nor could we expect any other work from the brain.

2) Libet and Haggard's argument seems to be a typical example of the post hoc fallacy, reasoning that invalidly infers a causal relationship between A and B, namely, the conclusion that A is the cause of B, solely because there is a relationship of temporal precedence between event A and event B. Neurological clinical practice can provide supporting arguments for this claim. So-called panic disorder episodes represent an apparent dissociation between unconscious brain mechanisms and conscious mental activity. In such cases, the body reacts as if there were a stimulus capable of triggering fear, even without that stimulus. Conscious activity does not follow this sequence of events, resulting in a dissociation that supports the idea of a lack of causal connection.

3) Given the significant advances in brain sciences over the past decades, with increasingly precise methods and findings, there no longer seems to doubt that the brain is necessary for mental states and most human behavior. Experimental research has made substantial contributions to rethinking the problem of human determinism. However, we must be cautious in how we interpret experimental data. Even if unconscious mental processes always precede conscious mental states, this does not necessarily imply a reduction in freedom. This statement is unequivocal when applied to conscious mental states like emotions. When we feel an emotion directly, we also know that at a later stage, we always have the possibility of counteracting fear and having various possible conscious behaviors instead of being paralyzed, such as running away, avoiding, and cautioning [11].

[11] W. Glannon, *Bioethics and the Brain.*, Oxford University Press, Oxford 2007, pp. 55-56.

ISSN 0039-4130

4) Neurosciences have only empirically discussed free will, i.e., the choice between possible alternatives. It didn't discuss the true freedom that Kant, Paul Ricœur, or Charles Taylor defend, which interests most from an ethical and legal perspective. Under the influence of Kant's philosophy, we believe freedom is fundamentally an experience of self-determination. This notion presupposes a positive conception of freedom, where being free is understood as the will being a cause unto itself. To be free is to align oneself with our actions and to assume their authorship. In this framework, self-determination and self-creation are inseparable.

The most significant decisions in life are not necessarily those that involve choosing between alternatives of equal value. As C.Taylor argues, the capacity to make strong assessments and craft personal life choices is integral to identity.[12] Without this, an agent would lack the depth essential to human nature. The strong evaluator can be profound, as their evaluations are not merely driven by achieving goals but by the life they aim to shape. This project is intimately tied to personal identity, as it depends on a horizon of values that one embraces. Returning to Kantian ethics, the imperative is that every person, in every action, should reflect on whether the maxim of their action could be universal. This reflection affirms our autonomy and dignity, especially when we disobey unethical or unjust demands. We are highly free as we disobey immoral demands.

In light of the current debate about AI and its ethical implications, we emphasize that, based on our perspective on agency and freedom. We believe that traits like intelligence, emotions, awareness, intentionality, practical reasoning, and the ability to make strong and weak evaluations are essential to human experience. Therefore, AI should be regarded as a simulation of these human traits. From both anthropological and ontological standpoints, there are significant differences between real entities and those that are merely simulated. This distinction holds ethically as well: living a genuine life, fully integrated into the world, open to the vulnerability and richness of human relationships—including love, friendship, disappointments, and anguish—is vastly different from a life of simulation. As Nozick's invention of the experience machine suggests, a life of safety, pleasure, and disconnection from real life may be satisfying in some domains. Still, it would lack authenticity because it is artificial.

[12] C. Taylor, *Human Agency and Language. Philosophical Papers I,* Cambridge, Cambridge University Press 1985, pp. 66-68, 73.

ISSN 0039-4130

An example of this dilemma arises in clinical settings, where full-body scans and immediate AI-driven diagnoses are considered potential replacements for human doctors. While AI is an increasingly sophisticated tool for human benefit, it remains an artefact. The clinical relationship, however, is deeply contextual, and AI's guidance does not consider the context of the patient's situation. Human beings develop through their relationships with the world and through communication. Unless an individual prefers to make decisions in complete isolation, they may choose the counsel of a human practitioner. In general, human beings are relational, and in clinical situations, most would prefer a real dialogue and a shared responsibility between patient and doctor.

If a robot is not free in the sense we define it, and thus not an agent, can it be ethical? Alan Winfield programmed a robot to make decisions based on Kantian principles—that all lives are inherently valuable and equal. For Kantians, this approach is ethical but not for consequentialists. However, Winfield challenges the robot with a catastrophic situation where saving everyone is impossible. Based on Kantian ethics, the robot attempts to save everyone but ultimately fails to save anyone. [13] Unlike the programmed robot, this scenario highlights a key human trait: the ability to adapt to new circumstances, which is tied to intelligence—the capacity to solve novel problems by creating new solutions. However, based on fixed principles, a robot is programmed to respond identically each time. It is an ethical zombie. While our character may firmly adhere to certain principles over others, we believe flexibility and practical wisdom - phronesis, as Aristotle and Ricœur suggest - are essential for sound decision-making.

AI is a pale copy of human intelligence. Humans cannot be artificially replicated, and what AI does is merely a simulation—often with extraordinary capabilities, even surpassing our own. While this is true, and despite its limitations, as we have seen earlier, the medical possibilities opened by AI remain fascinating. For example, the basic premise underlying all neuroprosthetic approaches is that targeted and controlled electrical stimulation of nerves or muscles can potentially restore the physiological function of a damaged organ or limb. The continuous development of brain-machine interfaces offers remarka-

[13] A.Winfield - C. Blum, W, Liu, *Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection*, In: M. Mistry - A. Leonardis et al. (eds) *Advances in Autonomous Robotics Systems*. TAROS 2014. Lecture Notes in Computer Science, 8717. Springer, Cham. https://doi.org/10.1007/978-3-319-10401-0_8, pp. 9-10.

ISSN 0039-4130

ble hope for patients suffering from a wide range of conditions. Although AI, neurosciences, and robotics cannot provide solutions to every problem or for every person, we have some solid reasons to feel optimistic about their future in healthcare. [14]

In conclusion, neuroscientific research offers new insights into human freedom and determinism philosophical debate. Findings from Libet, Haggard, and Haynes suggest unconscious processes precede conscious intentions but do not eliminate our freedom. Instead, they invite a deeper reflection on how unconscious mechanisms support conscious decision-making. Rooted in Kantian ethics and enriched by P.Ricœur and C.Taylor, true freedom lies in aligning actions with one's values and taking responsibility for them. Similarly, discovering mirror neurons emphasizes intersubjectivity and empathy as essential to ethical behavior.

While AI provides powerful tools for healthcare and problem-solving, it remains a simulation of human intelligence, lacking self-awareness, adaptability, and moral reasoning (ethical zombie). These limitations become evident in relational, context-dependent fields like medicine. Ethical dilemmas posed by AI highlight the inadequacy of rigid, principle-based reasoning, as sustained by Winfield, reaffirming Aristotle's and Ricoeur's emphasis on practical wisdom (*phronesis*). Ultimately, advancements in neurosciences, AI, and robotics must be grounded in a philosophical understanding of freedom, agency, and moral responsibility, preserving the irreplaceable depth of human experience.

References

A. Ferreira - L.Almeida, *Que futuro para a cirurgia oftalmológica?*, in L. Almeida (edited by), *Escrevinhar a Pensar a Bioética. Assuntos de Ética e Direito Médicos*, Loures, Thea Portugal, SA, 2017, pp. 87-88.

A.Winfield, C. Blum, W, Liu, *Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection*, in M. Mistry - A. Leonardis et al. (eds), *Advances in Autonomous Robotics Systems*. TAROS 2014. Lecture Notes in Computer Science, 8717. Springer, Cham. https://doi.org/10.1007/978-3-319-10401-0_8.

[14] A. Ferreira, L.Almeida, *Que futuro para a cirurgia oftalmológica?*, in L.Almeida (edited by), *Escrevinhar a Pensar a Bioética. Assuntos de Ética e Direito Médicos*, Loures, Thea Portugal, SA, 2017, pp.87-88.

A. Castro Caldas, *Viagem ao Cérebro e a algumas das suas competências*, Universidade Católica Portuguesa, Lisboa 2008.
A. Lavazza - S. Inglese, *Operationalizing and Measuring (a kind of) Free Will (and Responsibility). Towards a New Framework for Psychology, Ethics, and Law,* in *Rivista Internazionale di Filosofia e* Psicologia, VI, 1, 2015, pp. 37-55.
A. Lavazza, *Free Will and Neuroscience: From Explaining Freedom Away to New Ways of Operationalizing and Measuring It,* in *Frontiers in Human Neuroscience*, X, 262, 2016, pp. 1-17.
Aristotle, *Nicomachean Ethics*, trans. H. Rackham, Cambridge, Harvard University Press.
B. Libet, *Unconscious cerebral iniciative and the role of conscious will in voluntary action*, in *Behavioral and Brain Studies*, VIII, 4, 1985, pp. 529-566.
B. Libet, *Do we have free will?*, in *Journal of Consciousness Studies* VI, 8- 9, 1985, pp. 47-47.
H. Doucet, *Anthropological Challenges Raised By Neuroscience: some ethical reflections,* in *Cambridge Quarterly of Healthcare Ethics*, XVI, 2007, pp. 219-226.
L. Cattaneo - G.Rizzolatti, *The mirror neuron system,* in *Archives of Neurology,* LXVI, 5, pp. 557-560.
L.Tancredi, *Hardwired Behavior. What Neuroscience reveals about morality,* Cambridge University Press, Cambridge 2005.
W. Glannon, *Bioethics and the Brain*, Oxford University Press, Oxford 2007.
W. Glannon, *Free Will and the Brain. Neuroscientifc, Philosophical, and Legal Perspectives*, Cambridge University Press, Cambridge 2015.
P. Haggard - B. Libet, *Conscious Intention and Brain Activity*, in *Journal of Consciousness Studies*, VIII, 11, 2001, pp. 47-64.
P. Churchland, *Neurophilosophy: Toward a Unified Science of the MInd-Brain*, MIT Press, Cambridge and Massachusetts 2002.
Kant, *Critique of Practical Reason*, in *Cambridge Texts in the History of Philosophy*, Cambridge University Press, Cambridge 2015.
P. Ricœur, *Soi-même comme un autre*, Éditions du Seuil, Paris 1990.
C. Taylor, *Human Agency and Language. Philosophical Papers I*, Cambridge University Press, Cambridge 1985.
C. Taylor, *The Ethics of Authenticity*, Harvard University Press, Cambridge 1991.

ISSN 0039-4130

Existentialism as a Humanism in the Techno-scientific Era

Elad Magomedov

Our technoscientific era presents a landscape where traditional notions of human freedom appear to be increasingly constrained, if not entirely obsolete. This attack on human freedom appears on two fronts, namely, in both the realm of theory and practice. In theory, our scientific understanding of nature reduces the human being to a mere effect of causal processes occurring in the brain. But whereas the neurosciences surrender our being and acting to determinism in so far as we are a physical thing, artificial intelligence subjects us to the same lot, in so far as we are a thing with psyche. Thus, in the realm of nature, what it means to be human—or our essence—is reduced to the laws of material things and processes, whereas in the realm of culture our behavior is subjected to a hyper-effective form of mass manipulation from which our decision-making cannot seem to escape. From this perspective, it seems that the concept of humanism becomes a senseless one.

Regardless of how we define it, humanism presupposes that there is something humanity should be, and that this ideal is rooted in what it means to be truly human. While this understanding of humanism affirms that humanity has an essence, this essence is in turn understood as something that must be fully actualized. Such actualization is a task, which requires us to act. This implies that we may act rightly or wrongly. In other words, humanism presupposes freedom of human agency: to strive towards the realization of humanism means to exercise human freedom for the sake of the realization of own essence as a human being. In this regard, we must say that the techno-scientific era does not deny human essence, but rather only denies that this essence has anything at all to do with freedom. Indeed, the techno-scientifically construed human being is the inverted version of the human being as understood by humanism: whereas humanism posits a free agent who

is responsible for his or her own humanity and that of others—and hence must act to its fulfillment—techno-scientism posits the human being as a thing among things, deprived of all freedom and responsibility, and possessing no more or less dignity and integrity than a rock or a table. [1] In short, for techno-scientism, the essence of the human being is to be a thing among things.

However, phenomenological philosophy, and specifically Jean-Paul Sartre's ontology, allows us to argue that this convergence of developments which attack the concept of freedom on both theoretical and practical fronts, does not necessarily entail a wholesale restriction on human freedom. Rather, it challenges us to reconsider the traditional dichotomy between freedom and the efficient causality governing both neural processes and the influence of automated systems on human agency. The sense of the Sartrean concept of freedom can be summarized by the argument that the domain of mechanical causality or determinism is the domain of things, and since consciousness is not a thing, it is not embedded in the web of causality. Indeed, it cannot be limited by anything other than itself.

To better understand what is at stake here, let us take a closer look at the contrast which Sartre draws between consciousness and things. The pen on my desk, for example, is not free. Like the brain, it is subject to the mechanical laws of causality: if I pick it up and release it, gravity will force it toward the ground. Every instant of its trajectory from my hand to its coming-to-rest on the ground, can be calculated if we possess the correct data. This thing moves not of its own accord but only through an external cause. Nothing about it can interrupt its embeddedness in the web of causality in which it is entangled. In fact, even this formulation, namely that its movement is not «of its own accord» is not entirely accurate, since to have any accord at all, this thing must be able to relate to itself and something other than itself. This ability to relate to something is an exclusive characteristic of consciousness: the pen on the table does not "relate" to the table, it cannot even be said to "touch" the table, unless it does so for some observing consciousness. In itself, outside of all observing consciousness, the thing is neither near nor far with regard to the table, but it simply is in-itself

[1] For the claim that techno-scientism 'inverts' the humanistic concept of human being, I rely on Jean-Paul Sartre's concept of both human and humanism, as elaborated in *Existentialism is a humanism*. For the idea that humanism aims at a realization of a human essence, see: M. Heidegger, *Letter on Humanism*, in *Basic Writings*, David Farell Krell (ed.), New York: Harper & Row, 1977.

ISSN 0039-4130

(en-soi). Being on neither side of a relation to its own being nor to the being of something it is not, the pen fully coincides with its own identity. It *is* no more and no less than its factual determinations, such as its weight, shape, color, material, and so on. Due to this full positivity of being condensed in the thing, it is precisely in-itself in the sense that it cannot experience itself as lacking in any sense. As a result of this total lack of negativity, it cannot experience itself as "not" being something. It simply is what it is, and due to its absolute coincidence with its factual characteristics, it is always already determined "to be" something by an efficient cause acting upon it — for example, if I release the pen from my grip, it is determined by gravity to "be" no more and no less than a pen in the state of falling. There is no alternative.

The situation is entirely different when it comes to consciousness. Consciousness is always consciousness "of" something, and, specifically, it is consciousness of what it is and what it is not. I see the pen in front of me, and I am already conscious of the pen as that which I am not. In other words, being conscious of myself involves a consciousness of myself as "lacking" the being of the pen. This introduces a negativity into consciousness, specifically a kind of negativity that cannot be found anywhere outside the domain of consciousness. All that pertains to the order of things is subjected to the laws of causality precisely because that order is devoid of negativity that could somehow puncture a hole in the web of causality. Thus, Sartre's ontology allows him to affirm that even though the brain, being a thing, is determined by the laws of causality, consciousness arises precisely as an interruption of such determination. Drawing from the work of phenomenologists such as Edmund Husserl and Martin Heidegger, Sartre argues that consciousness is a mere relating to something other than itself, never coinciding with itself and always surpassing itself towards the world and the future, constantly transcending the now-instant, and transcending the past into which the now-instant persistently slides away.

As a pure negativity that is constantly suspended between the past and the future, between the what-has-been and the not-yet, consciousness is characterized by an interruption of efficient causality altogether. As soon as the universal laws of causality touch the brain that is now, consciousness has already surpassed itself towards what it is not-yet. For this reason, consciousness is «condemned» to be free, as Sartre puts it. Consciousness is pure spontaneity that can be motivated, conditioned, or carried away by itself, but it cannot be determined to do or be anything in the sense that a billiard ball can be determined to take

ISSN 0039-4130

a certain trajectory if impacted from a particular angle. Rather than being determined by its brain, one could say that consciousness is situated in its neurophysiological condition; but what distinguishes it from an automaton is its ability to have consciousness of its motivations and to either affirm or negate them by choice. Even when one is physically forced into some situation, for example through violence or imprisonment, one can choose against one's situation, even if this choice cannot be actualized due to physical "restrictions". The earlier mentioned pen that is reduced to being a pen in the state of falling, on the other hand, cannot "choose" against its situation. By conceptualizing choice in this manner, Sartre draws the important distinction between consciousness and volition: while consciousness is absolutely free and spontaneous, the will is not absolutely free, but is rather subject to a more primordial choice and project developed by consciousness. For Sartre, I am not free because I can do whatever I want—in fact, I cannot do whatever I want; rather, I am free because my projects are grounded in what is fundamentally a choice of consciousness to be or not to be in some manner. This does not mean, of course, that I can always get what I choose, but it only means that I am not determined to choose in a particular way, like the table is determined to be a table. My world, in that sense, is a projection of my fundamental choice: and only as such can the world either align or misalign with my volition.

Further, freedom for Sartre is not an abstract concept but is always situated within specific contexts and circumstances. He argues that while human existence is fundamentally characterized by freedom, this freedom is not detached from the world. Rather, freedom is intimately intertwined with the situations in which individuals find themselves. Since freedom is always situated within concrete situations and social structures, these situations impose constraints and limitations on human agents, shaping the possibilities available to them and influencing their choices. By articulating freedom as situated, Sartre allows us to make sense of what it would mean to be free in a world where situatedness is defined by AI, that is to say, by automated decision-making processes and knowledge production. Sartre would agree that such situatedness shapes the horizon of our possibilities for action, but he would nevertheless insist that situatedness is not a limitation but rather the very condition of freedom. To say that today we can only act within the context of AI, means only that we are responsible for every choice we make within that context. For example, although the institutions that collect data can clandestinely shape my voting preference, they

ISSN 0039-4130

cannot determine me to act in a certain way. Being aware of this possibility of influence, I become all the more responsible to think-through my choices, and I am condemned to carry full responsibility for whatever decision I make—because at the end of the day, even if I was misled, the possibility to make a different decision always remained open. Things are not different when it comes to brain interventions which change a person's personality, such as for example in brain disease.[2] While Sartre would agree that any organic change in the brain will result in some (unprecedented) compulsion in consciousness, he would insist that this compulsion can only become a determining force when consciousness consents to it and chooses to coincide with that volition.

Sartre's rejection of determinism in consciousness has significant ethical and existential implications. Since consciousness is not determined by causal forces, individuals do not merely follow predetermined scripts and are hence fully responsible for their actions. What makes us responsible for our choices is not this or that choice and its justifications, but the ability to choose at all, which is an ability not possessed by things, such as tables, rocks, or brains, for that matter. At the same time, since there is nothing determining what we are "to be", our essence is not pre-given, but rather must be shaped through the individual choices that we make. In other words, what it means to be human, is not a given but a project—humanity is not the actualization of what we are already, but rather a choice to be in a certain way—a choice for which we are fully responsible and which each of us must make for themselves. Thus, by positing that each of us partakes in shaping humanity—and that we are responsible for shaping it in such a way that each of us becomes aware of the responsibility that comes with our freedom—Sartre's philosophy implies a humanism. It is a humanism because it highlights the significance of human agency in shaping oneself as a human being, that is to say, shaping one's existence and values. By asserting that individuals are not predetermined by any fixed essence or external standards, existentialism affirms the inherent dignity and potential of each person to define their own meaning and purpose in life. The dark side of this humanism is that we bear

[2] Consider for example the 2003 case of a Virginian man who began to molest his 8-year-old stepdaughter after developing a brain tumor. See: Burns JM, Swerdlow RH. Right Orbitofrontal Tumor With Pedophilia Symptom and Constructional Apraxia Sign. *Arch Neurol.* 2003; LX (3), 437-440. doi:10.1001/archneur.60.3.437. Patricia Churchland refers to this case in her lecture "The Big Questions: Do We Have a Free Will?", delivered on 18 November 2006, published in NewScientist.com

ISSN 0039-4130

full responsibility even when making choices that lead to catastrophic consequences and appear to be beyond our control, as is the case in socio-technical systems. We cannot shift our responsibility to anything other than the human we chose to become. There are no excuses.

In light of all this, Sartre's thesis that «Existentialism is a Humanism» is as relevant in the technoscientific era as it was in 1946. It allows us to say that the more our age makes it seem that we are unfree, the more responsible we become to act in accordance with our freedom.

ISSN 0039-4130

Human Agency Reloaded in our Techno-social Ecosystem

Zsolt Almási

Introduction: The Posthumanist challenge

The inquiry into human freedom and agency has long been a fundamental project since the beginning of philosophical thought. Plato's early dialogues, for instance, addressed the notion of individuals bereft of insight, and thus, freedom and agency. In these dialogues, Socrates adeptly demonstrated how his intellectual adversaries were ensnared within the confines of their own confused set of assumptions, thereby unwittingly forfeiting their autonomy. Later dialogues, however, presented a shift in focus, as Socrates, with the aid for example of the Cave simile, portrayed humanity as tied down so that they could perceive mere shadows of reality's puppets, and could be released with radical education from this sad state. In Francis Bacon's philosophy it was the four *idolae* that obstructed clear thought and perception. Erasmus of Rotterdam on theologico-philosophical grounds contended with Luther over the principle of freedom of choice. Scholars and thinkers have endeavoured to carve out conceptual space for human freedom and agency contending with various constraining factors, whether they manifest as *idolae* of the mind, a mean scientist, God, a foul demon, chemical determinism, the unconscious, social determinism, theological predetermination, the rule of algorithms. These challenges prompt ongoing intellectual exploration and debate aimed at safeguarding and enhancing human agency.

The problematic nature of human agency may be explored in the context of the challenge presented by posthumanist philosophy. Posthumanist philosophy can be viewed as a collection of rational assertions regarding humanity within the technosocial ecosystem of the 21st century. However, in my interpretation, posthumanist thought should

rather be perceived as a critique of and confrontation with what might be categorised or fictionalised as traditional humanism. In posthumanist thought it is feasible to discern three overarching and abstract denominators as is argued in Francesca Ferrando's *Posthumanist Philosophy* such as «post-humanism,» «post-anthropocentrism» and «post-dualism»[1]. Post-humanism means that instead of a single narrative, posthumanism attempts to delineate the human in its plurality, i.e. instead of conceptualising the human as a white, middle-class man, it tries to see the human in a more comprehensive and inclusive way. Problematising anthropocentric theories means that posthumanism endeavours to place all other entities in their appropriate context by displacing humanity from the centre. The objective of posthumanism, then, is to perceive the human being not as an exceptional entity, but to comprehend humans through their interactions and collaborations with other entities, interconnected and interdependent, rather than existing in isolation. With the words of Stalpaert et al what is to be problematised is the view that «the human is the centre of attention, an individual that maintains control over nonhuman matter in a competitive and hierarchical constellation»[2]. The posthumanist perspective challenges, furthermore, the idea of the hierarchy of binary oppositions when exploring the human. As a powerful example one may well cite Hayles when she argues that «there are no essential differences or absolute demarcations between bodily existence and computer simulation, cybernetic mechanism and biological organism, robot teleology and human goals»[3].

How does the posthumanist stance address the issue of human agency in the context of emerging AI technologies? Given the scope of this paper, I will focus solely on the second concern of posthumanist philosophy: the conceptualisation of the human in relation to technology. To examine this, I will explore AI applications, particularly those that construct and communicate through texts, and their interaction with humans. This exploration will reveal that reconstructing this interaction necessitates the presence of human agency, which may well

[1] F. Ferrando, *Philosophical Posthumanism (Theory in the New Humanities)*, Bloomsbury academic, London New York (N.Y.) New Delhi 2020, p. 77.

[2] C. Stalpaert - K. Van Baarle - L. Karreman, *Performance and Posthumanism: Co-Creation, Response-Ability and Epistemologies*, in *Performance and Posthumanism*, ed. C. Stalpaert - K. Van Baarle, and L. Karreman, Cham: Springer International Publishing, 2021, p. 2. https://doi.org/10.1007/978-3-030-74745-9_1.

[3] N. K. Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, University of Chicago Press, Chicago 2010, p. 3.

ISSN 0039-4130

be contrasted to the machine's indeterminacy. Through this analysis, I will argue that human freedom and agency, though constrained, remain attainable even in a posthumanist context, and also that this agency can be deployed to find one's own ideal self or more precisely one's ideal voice. To achieve this, first I will define what kind of AI is at stake, how this functions and where humans have agency and thus freedom when using it. Second, I am going to contrast human freedom with the machine's indeterminacy to show that it is precisely the interrelatedness of the human being and other entities that helps the human to exercise their agency in their act of self-fashioning.

Names and their relevance

The term "Artificial Intelligence" requires greater precision to facilitate effective discourse. A key distinction must be made between Artificial General Intelligence (also termed Strong or Deep Intelligence) and Artificial Narrow Intelligence (also termed Weak Intelligence). Artificial General Intelligence refers to a technology that emulates human capabilities by independently learning, applying knowledge across diverse contexts, repurposing its actions according to emerging needs, and replicating human goal-setting. This form of Artificial Intelligence does not currently exist[4]. Recently, however, Strong and General AI has been distinguished by still aligning Strong Artificial Intelligence with human-like consciousness, while identifying AGI with learning and problem-solving in different contexts without pretraining. The ARC benchmark[5] could be an example for the latter distinction, showing the GPTo3 is close to human capabilities as far as learning from limited resources and applying knowledge to solving problems without pretraining. In this sense AGI can be achieved in the near future, while strong AI is still a speculative possibility at the moment. The second type of Artificial Intelligence, Weak AI, based on machine learning, employs supervised, unsupervised, and reinforcement learning methodologies, with human guidance, and training corpus building. These systems identify patterns in data and apply this knowledge to solve specific tasks. However, this does not equate to mimicking human thought or possessing a general understanding of extramental reality. Artificial

[4] See for example IBM's definition https://www.ibm.com/think/topics/strong-ai
[5] See https://arcprize.org/

ISSN 0039-4130

Narrow Intelligence, can perform complex tasks such as generating texts, images, videos, or music in response to human prompts. These applications include CHATGPT, Claude, Gemini etc, and have been available for the public since November 2022.

Once it has been recognised that we are discussing Artificial Narrow Intelligence (ANI), it is crucial to use precise terms to describe the applications based on this technology: "generative," "predictive," and "invocational" ANI. Lev Manovich characterises this technology as «predictive»[6], arguing that the term «generative» may be misleading, instead he favours «predictive» because it more accurately describes the application's function: predicting the next letter, word, using complex statistical models, rather than articulating thoughts as humans do. Chesher describes the technology as «invocational,» emphasising that, despite its sophistication, the application does not initiate actions but responds to human prompts. As he notes, AI applications like ChatGPT «are not autonomous actors but lively participants in invocational relationships with their users»[7], functioning solely in response to human prompts.

Technology and human freedom

When examining human interaction with this technology, it is possible to make room for human agency and freedom. This occurs as individuals, in utilising this technology to produce, for example, textual documents, initiate the communicative situation with a prompt. The quality of the machine's response to the human prompt depends on the quality of the prompt, which includes a prior knowledge about the subject that the response is constructed about, a prior knowledge of the methods the machine tries to interpret the text of the prompt, a skill that has already acquired a name, i.e. prompt engineering. Subsequently, the user responds to the generated output by either accepting or rejecting it. Upon acceptance, individuals may or should amend the text to articulate their thoughts in their own ideal voice, and ultimately, they should determine the course of action regarding the edited text such as deleting or publishing it.

[6] L. Manovich, *Seven Arguments about AI Images and Generative Media*, in *Artificial Aesthetics*, p. 9, (np.: np., 2023), http://manovich.net/content/04-projects/168-artificial-aesthetics/lev-manovich-ai-aesthetics-chapter-5.pdf.

[7] C. Chesher, *Invocational Media: Reconceptualising the Computer,* Bloomsbury Academic, New York London Oxford New Delhi Sydney 2024, p. 6.

ISSN 0039-4130

Each of these steps facilitates the expression of human freedom and agency. The machine operates solely in response to human initiative; the initial step is a product of intentional, purposive human action without which no machine generated text would emerge. Following this, a succession of purposeful actions unfolds, where human agency is further exercised: crafting prompts, refining the prompt, initiating queries, reiterating upon responses, rejecting or accepting outputs, and editing the machine generated text to align with the author's original intentions. Ultimately, again in harmony with the original intentions the text is presented to the reader, placing full responsibility upon the human author. Errors or inaccuracies or hallucinations cannot be ascribed to the algorithms; this accountability rests with the author, in whom the exercise of freedom and agency is ultimately reaffirmed.

Human agency versus mechanical indeterminacy

The human freedom and agency explored so far faces a challenge from the application's indeterministic text generation. The application's capacity to produce different responses—even to identical prompts—through potentially infinite iterations may create an illusion of power over the user. Generative AI appears to exercise a form of autonomy, producing textual content beyond the direct control of the author of the prompt, who cannot fully predict or control the responses generated. Each iteration of a prompt yields a slightly different output, underscoring a variability in content that the human neither initiates nor ultimately owns, or at least this seems so for the user, as the unpredictability and variability, driven by the application rather than the author of the prompt, determines human choices, thereby posing a potential threat to human freedom and agency, even if the variability is encoded in the application by humans.

This apparent iterative freedom of Generative AI, however, is only indeterminacy rather than genuine freedom. To clarify the distinction between human freedom and mechanical indeterminacy, I will draw upon Alasdair MacIntyre's concept of «practice.» MacIntyre defines a practice as:

> «By a 'practice' I am going to mean any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve

ISSN 0039-4130

> those standards of excellence which are appropriate to, and partially definitive of, that form of activity, with the result that human powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended.» [8]

MacIntyre's notion of practice provides a framework to differentiate the freedom of Generative AI from that of human beings. According to MacIntyre, human practice is not a simple act but a structured, multi-layered activity that embodies and necessitates human freedom. This freedom is expressed through the depth and intentionality of practice itself.

First, MacIntyre emphasises that a practice is not an isolated action but a «socially established cooperative activity». This means that the activity is embedded in a social context, bound by shared norms and values, and is cooperative in nature. The cooperative element may involve active collaboration, as in a chess game, where human players engage directly with one another, or it may be indirect, as in the influence of past players and their strategies shaping the present game.

Human agency and freedom emerge within four distinct layers of the socially contextualised human activity. First, there is the recognition of the social and cooperative nature of any human practice; a fundamental intention to participate in this socially constructed, collaborative endeavour underpins a human activity. Second, an understanding of the «standards of excellence» specific to the practice is required, enabling the individual to meet and uphold these standards. Third, the aspect of «trying to achieve» reflects the inherent intention within human practice, pointing towards a purposeful pursuit embedded in each act. Fourth, the objective of human practice extends beyond mere completion or production; it includes, as an intended consequence, the achievement of «excellence» in the practice itself and the systematic extension of «human powers» through striving for excellence.

Human agency and freedom, rooted in the multi-layered intentionality of human practice, can be clearly distinguished from mechanical indeterminacy. Human engagement with technologies, such as generative, predictive, or invocational applications, reflects an intention not only to complete a task but to strive for excellence and to expand human capabilities. In contrast, machine-generated indeterminacy operates within an indefinite array of possible responses only determined

[8] A. C. MacIntyre, *After Virtue: A Study in Moral Theory*, 2nd ed., London: Duckworth, 1985, p. 187.

ISSN 0039-4130

by the original prompt, without a purposive selection among them. Each regenerated text is randomly assembled from innumerable options, without a deliberate orientation towards excellence or extension of human powers. Rather than intentional decision-making, the machine's output responds only to the user's input—which is occasionally un(der)defined, for example when the query is reiterated without any modifications that could point out the deficiencies of the previous output—in a detached manner, reproducing variations. This process lacks any genuine aspiration toward improvement, representing instead a neutral and uninterested replication of content with modifications.

Posthumanist philosophy is both accurate and imprecise in its argument to decentralise the human being, thereby levelling distinctions between the human and technology. Posthumanist philosophy rightly argues that human beings cannot be fully understood in isolation but must be seen in relation to other entities, including objects and technology. However, this perspective overlooks the essential distinctions between agents within this relational framework. While human agency and freedom, grounded in intentionality, can be understood in the human practice as defined by MacIntyre, the generative application's apparent freedom is better characterised as indeterminacy. Unlike human intentionality, this indeterminacy lacks the multi-layered depth that defines intentionality within human practice.

Conclusion

At the outset, this presentation aimed to explore how posthumanist philosophy decentralises the human beings, situating them within a network of interrelated entities undoing the hierarchical relationship between the human and the non-human, which implicitly threatens the difference between the human and the non-human that lies in the freedom of the one and the lack of it in the other. I have argued that, despite this relational framework, there remains a meaningful space for human freedom and agency, specifically within the communicative dynamics of technology—particularly Artificial Narrow Intelligence as predictive and invocational media. I have conceptualised this freedom through Alasdair MacIntyre's notion of «human practice.» Applying these insights to the use of generative, predictive, or invocational technologies, it becomes evident that practice in this context extends beyond merely producing acceptable textual content for a target audience. It also em-

ISSN 0039-4130

bodies a moral commitment to excellence within the practice itself and to the cultivation of human powers. Thus, cooperation with technology is not merely product-oriented; it entails a moral imperative to strive towards improvement within the practice.

This view aligns with a dynamic understanding of the human being, akin to Karl Jaspers' notion that «to be a man is to become a man» [9]— a dynamic process of self-transcendence. I propose that this self-transcendence occurs in relation to the technosocial ecosystem as one strives to find and realise one's authentic, ideal voice. It is this pursuit of an ideal voice, developed in cooperation with technology, that opens a space for human agency and freedom. Without the intentional quest for this voice and the responsibility for the final, accepted, edited and shared output, human agency would remain unrealised. Human freedom and agency, similarly to becoming a human being is not something given but rather a task ahead of them.

Bibliography

C. Chesher, *Invocational Media: Reconceptualising the Computer*, Bloomsbury Academic, New York London Oxford New Delhi Sydney 2024.

K. Jaspers, *Way to Wisdom: An Introduction to Philosophy*, Translated by Ralph Manheim. 14. print., Conn: Yale Univ. Press, New Haven 1973.

A. C. MacIntyre, *After Virtue: A Study in Moral Theory*. 2nd ed., Duckworth, London 1985.

L. Manovich, *Seven Arguments about AI Images and Generative Media*, in *Artificial Aesthetics*, 1-25. np.: np., 2023. http://manovich.net/content/04-projects/168-artificial-aesthetics/lev-manovich-ai-aesthetics-chapter-5.pdf.

[9] K. Jaspers, *Way to Wisdom: An Introduction to Philosophy*, trans. Ralph Manheim, 14. print, Conn: Yale Univ. Press, New Haven, 1973, p. 73.

ISSN 0039-4130

Human Dignity at an AI and Neurosciences Age [1]

Yves Poullet

Introductory remarks

The undoubtedly still largely futuristic world to which emerging technologies - artificial intelligence, the cloud, the Internet of things, advances in neuroscience and NBIC, particularly in the field of genetics - are leading us, invites to rethink our right to dignity. The questions we face might be summarized as follows: what does our right to dignity mean in an age of ubiquitous technology, present in our pockets, our walls, our supermarkets, ... in our bodies? What does this right imply at a time when science can modify our genetic make-up and decipher the functioning of our brains to read our thoughts or guide our actions? What does this right require at a time when technology is increasingly opaque in its operation, when the networks our data uses are so complex and the computing, predictive and decision-making power is beyond human comprehension, and, at the same moment, when technology is increasingly reducing our lives to data, creating the illusion that truth emerges from data, since data, unlike humans, apparently do not lie?

Chapter 1: What does Dignity mean in an evolving society? New risks – Need to enshrine a fourth generation of human rights

1. **Dignity a doubly fundamental right** - The Kantian approach to the concept of human dignity considers that this fundamental value and

[1] The present text is a summary of a book to be published (March 2025) at the Belgian Academia in the collection: 'L'Académie en poche'.

right means that every person must never be treated as an object or a mean, but always as an end in itself, an intrinsic entity that is infinitely respectable. All international human rights texts (notably, the United Nations Treaty, Council of Europe Convention and EU Charter of Fundamental Rights) refer to this concept and enshrine this value as the first human right, Every human being, simply as a human being, deserves unconditional respect, regardless of age, sex, physical or mental health, social condition, religion or ethnic origin. I call this right doubly fundamental, because firstly it is the very foundation of our humanity, and secondly because, *primus inter pares*, it constitutes the source of all other fundamental rights [2]. Human dignity is considered as the foundation, from one part, of the right to autonomy: the capacity and thus the right for everyone to decide him or herself, and, from the other part, of the right to equality, the right not to suffer unfair discrimination. This value of dignity, if it is intangible in its essence, implies due to the changing context in which human beings live, call for the recognition of new dimensions, in particular new subjective rights and new obligations for the State and private actors that are now necessary for ensuring respect for our human dignity.

2. Emerging technologies and the transformative nature of Human Rights - The technology that surrounds us creates new risks, modifies our actions, changes the balance of power between players and requires us to rethink regulations, certainly in the same spirit as before, but in a way that responds appropriately to these new challenges. That is what we call the "transformative nature" [3] of human rights. Just two examples: the first one is the increasing damages caused by the functioning of our AI systems; the second one is the potential modification of our identity and determination of the future of our humanity through the use of neurotechnology and NBIC. The right to dignity ought to imply a better legal protection of the environment and of humanity's genetic heritage and of biodiversity, as new facets of this right. Another point, while our societies advocate an in-

[2] As explicitly asserted by the Explanation of the EU Charter on fundamental rights: «It follows that none of the rights enshrined in this Charter may be used to infringe the dignity of others and that the dignity of the human person is part of the substance of the rights enshrined in this Charter. It may not therefore be infringed, even where a right is restricted».

[3] C. Cocito and P. De Hert, *The transformative nature of the EU Declaration on Digital Rights and Principles: replacing the old paradigm (normative equivalency of rights)*, in *Computer Law & Security Review*., L, Sept. 2023, pp. 1-11.

ISSN 0039-4130

dividualism that sometimes knows no bounds, the technologies that surround us bear witness to this ever-increasing interference with our lives and our freedoms. Our profiles are, in the context of the use of AI technologies, deduced from data about other people we do not know. Therefore, dignity refers to the need for combining individual and collective protection of human beings. It makes individual freedoms inseparable from concern for the same freedoms of others. Individual freedoms can only be conceived within the framework of a society where they can flourish together and where individuals recognize each other [4].

3. Emerging technologies and new major risks to our dignity - As regards the so-called emerging and disruptive technologies, we would like to underline their main characteristics which are explaining the new major risks incurred to our dignity: first, the unlimited capacities which lead to our increasing sophisticated profiling; second, in the context of deep learning AI, the autonomy and increasing opacity of their functioning; third, the ubiquitousness of the terminals which enable a continuous surveillance and a more and more sophisticated profiling, predictive of our future actions and decisions; fourth through the use of NBIC and nanotechnology their possible presence in our bodies and brains in order to decipher our genetic luggage or the functioning of our brains and able to modify our behavior even our identity. Generative AI does represent other risks since that technology creates confusion between human and artificial products, and they are inducing insidiously culturally colored ways of thinking and solving our problems. We would like to add the peculiar danger linked with the existence of the "Big Tech" actors, qualified rightly so by EU Commission as the "gatekeepers" [5] because they are first controlling the access to

[4] I. Graef, T. Petronick et T. Tombal, *Conceptualizing autonomy in the age of collective data processing : From Theory to Practice*, in *Digital Society*, II, 19, 2023. https://link.springer.com/article/10.1007/s44206-023-00045-3.

[5] «That a digital service can be described as an essential platform service does not in itself give rise to sufficiently serious concerns about contestability or unfair practices. Such concerns only arise where an essential platform service constitutes a major access point and is operated by an undertaking with significant weight in the internal market and enjoying a solid and lasting position, or by an undertaking likely to enjoy such a position in the near future». Recital 15 of REGULATION (EU) 2022/1925 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 September 2022 on fair and contestable contracts in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Market Regulation), O.J.U.E., 12.10. 2022, L.265/1.

ISSN 0039-4130

services considered as needed for each of us: the social networks and the search engines and deploying their activities in all economic sectors of our society including notably health sector. The EU Digital Services Act evokes about them the concept of "systemic risks" [6], i.e. risks linked to the survival of our societies, evoking the risks of polarization of our society, of abuse of their market domination and hence of control of the information flows.

4. **From risks incurred by individuals towards societal risks** - Precisely, the texts on AI and neuro-technologies ethics (UNESCO Recommendation (2021), OECD Recommendations (2019) and Council of Europe Convention framework (2024)) and the EU AI Act (2024) invite an enlargement of the risks to be considered when we are discussing dignity in our Information Society. The risks incurred by the individuals are relatively well covered by our data protection and liability (in case of economic or financial damages) legislation. Collective risks like discrimination linked with profiles on sometimes indefinite criteria and generated by voluntary or involuntary bias or caused by the problem of unequal access to technologies or services for certain populations, cultures and languages must be taken seriously into consideration. Besides these two categories of risks, the quoted texts underline the societal risks: certain are linked to the growing environmental costs associated with the use of IT equipment and services, others are generated by the distortions of competition made possible by data management; others evoke the more subtle problems of non-compliance with the rule of law when the power of technology leads to algorithmic government by the State and, in the name of technological efficiency, does not respect the requirements of proportionality in regulatory intervention. Furthermore, our very democratic functioning is undermined: we are talking about the manipulation of citizens via or even by social networks, particularly during elections; we refer to the short-circuiting of democratic discussion by the urgency created by rumors inflated by the power of our networks and the polarization of our society, which the business models of web companies and the major platforms

[6] See article 34.1.(b) of Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), PE/30/2022/REV/1, *OJ L 277, 27.10.2022, pp. 1–102* and the definition given by article 3 (65) of the AI Act.

ISSN 0039-4130

are helping to increase. Finally, as the UNESCO report on neuro-technologies notes,
«Today, neural data is increasingly sought for commercial purposes, such as digital phenotyping, emotional information, 'neurogaming' and neuromarketing. Neuromarketing units have been developed by industry to assess and even modify consumer preferences - raising serious concerns about respect for mental privacy. These risks can also pose serious problems when dealing with non-democratic governments».
Finally, the authors denounce the risks linked with neuro-enhancement and genetic manipulation's technologies as a major risk for our identity and for future generations [7].

5. A "fourth generation" of human rights? - It is in this context and in response to such challenges that we need to rethink not the meaning of human dignity but its effectiveness by granting new subjective rights and new positive obligations on the part of our States, the subject of the next chapter. As early as 2000, did Marcus-Helmons [8] call for a "fourth generation of human rights", necessary for human to master the progress of science and technology. In particular, he called for this new generation to "protect human dignity from certain abuse of science". While I remain unconvinced of the need to add yet more rights to the multiplicity of human rights, risking confusion and, above all, the relativization of each of them, how can we not follow the author and proclaim loud and clear the need to recognize, on the basis of human rights and the first of them, human dignity, new requirements for our law in order to make it possible, in our all-digital society, including in the field of technology, to enable each of us, individually and collectively, "to achieve maximum fulfilment within the framework of our capacities" and to effectively combat the threats to the human species posed by technical innovations driven by scientific progress, particularly in the field of biology?

[7] About this point, read the report adopted by the UNESCO International Committee on bioethics in December 2021 : *Ethical issues of neurotechnology , https://doi.org/10.54678/QNKB6229*

[8] S. Marcus-Helmons, *La quatrième génération des droits de l'homme*, in: *Les droits de l'homme au seuil du troisième millénaire, Mélanges en l'honneur de P. LAMBERT,* Bruxelles, Bruylant, 2000, pp. 549 et s.

ISSN 0039-4130

6. **Main features and impacts of emerging technologies including the neurotechnology** - Digital technology, including developments in neuroscience, is invading and guiding our lives, including our bodies and minds. Its capabilities in relation to our lives are growing every day, from the ability to record our actions to the ubiquity with which it can now track them. The "intelligence" of digital technology is profiling us, even predicting us, and, by the same token, manipulating us and "augmenting" us. These developments are taking place in an increasingly opaque and complex environment, all in the context of a growing informational asymmetry between those who develop them, the "information haves", and us as citizens, the "information have nots", and risk reserving the benefits of these advances for a select few, the "happy few". Generative intelligence makes new potential available to every one of us and presents itself as a substitute for our abilities, including our creative abilities. The machine is becoming invisible with the "Internet of Things" and robots, to the point where it is sometimes difficult to distinguish man from robot. Finally, technology offers us the dream of a human being free from bodily constraints, even immortality. Alongside its undeniable benefits, how can we fail to recognize, in the unprecedented and compass-less development of the applications of a teeming technology, the risks of attacks on human dignity, whether in terms of our ability to control our informational environment, the transformation of our beings into pure means at the service of the ends pursued by public and overall private organizations. Finally, we denunciate the risk concerning the loss of our identity "body and mind" or even concerning the survival of our humanity[9]. In the face of these potential threats, it is, from one part, the duty of our States to prevent them both through positive action and, where appropriate, by proclaiming new subjective rights, and, from the other part, our duty to demand such State's action and to claim the exercise of new rights. The submitted proposals will be based on ethical and regulatory texts, particularly European ones, which already reflect, albeit undoubtedly insufficiently, the recognition of these duties and the inchoate recognition of some of these rights.

[9] H. Jonas, *le Principe de responsabilité, une éthique pour la civilisation technologique*, les éditions du cerf, Paris, 1995.

ISSN 0039-4130

Chapter II: New facets of human dignity at neurotechnology's and AI times

A. The right to participate

7. The right to participate individually and collectively in the information society and its regulatory translation - The right to participate individually and collectively in the construction of the information society and in decisions taken on the basis of the use of AI**,** constitutes the first facet of our claiming for dignity. Let us start with the right to participate in the information society. This right must be progressively enlarged since more and more the use of the infrastructure and certain digital services are today becoming essential for the development of our personality. If the right of access to the infrastructure was the first to be enacted as "universal service", today we must consider the enlargement of the concept. First, as mentioned by the UNESCO recommendation (2021), the education to digital literacy must be considered as part of this "universal service". Second, as asserted by the EU Digital Market Act (2022), the right to the "core platform services" offered by the "gatekeepers"(the very large online platforms), it means the use of communications' social networks and of search engines which are essential for everyone in our information society, has to be proclaimed, apart from now, as an updated version of the evolutive concept of "universal service". This recognition will permit to impose to their providers the regulation of the quality of their services including the protection of our liberties (privacy, freedom of expression, protection of consumers, …) and of public interests.
At the same time where we consecrate the right to participate**,** we have to enshrine the **right not to participate** in the information society might be considered as the translation of our "right to be let alone" essential for building our personality: this right, enacted today in certain countries only for the employees vis-à-vis their employers should also be understood to include the right to disconnect our terminals from communication networks and, as proclaimed by the German Constitutional Court in 2009, as the "right to inviolability" for our terminal equipment considered as a virtual home.

8. The right to participate to the decisions about AI and neurotechnology applications - The right to participate we mention in the ti-

ISSN 0039-4130

tle of the paragraph has two complementary consequence at an AI and neuroscience age: the first one analyses the right for all stakeholders to take part in building innovations and their applications: as asserted by UNESCO recommendation [10],

«Respect, protection and promotion of diversity and inclusiveness should be ensured throughout the life cycle of AI systems, consistent with international law, including human rights law. This may be done by promoting active participation of all individuals or groups regardless of race, color, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other grounds» [11].

That participation must exist both at a micro level, it means that stakeholders (trade unions, representative of consumers, data protection officers, …) must be consulted during the assessment imposed to the companies providing or deploying AI or neurotechnology solutions and, at the macro, it means at the national, regional and global level, by the setting-up of an Office of technology assessment joining together all interested parties.

Second complementary consequence: respect for human dignity requires that in the case of significant decisions taken under the basis of an AI or automated system, the impacted person should be informed before any decision is taken about the use of the system and should be able to understand its logic or at least the model that determines how it works. Once the decision has been taken, it is important that the data subject is able to understand, in the context of a human dialogue with a competent person, how the algorithms used explain the decision taken.

The third requirement of the right to respect for dignity, today partially enacted by GDPR article 22 and AI Act article 86, is that the

[10] Point 9 of the recommendation on AI Ethics; see also Article 5.2 of the CoE Framework Convention: «Each Party shall adopt or maintain measures aimed at protecting its democratic processes in the context of activities carried out during the life cycle of artificial intelligence systems, including fair access and participation of individuals in public debate, as well as their ability freely to form an opinion». As regards the AI Act, we do regret that the inclusive multistakeholders assessment procedure proposed by the EU Parliament in June 2023 for high-risk AI applications to producers finally was rejected. On the other hand, we note that this multi-stakeholder discussion has been introduced for standardization bodies and a "consultative forum", joining together all interested parties has been set up at the EU level.

[11] See also Point 47 of the UNESCO Recommendation.

ISSN 0039-4130

decision taken can be challenged before a human being:
«it is a question, in the name of dignity, of "preventing" the advent of a certain form of technological determinism which would subject the people concerned to decisions taken solely on the basis of machine recommendations.» [12]

B. The "principle of distinction" and its two meanings

9. The distinction between artificial and human creation to the recognition of the "neuro-rights" - The second facet is also a complex one. It claims first to distinguish very clearly robots and their productions from human beings and their 'creations': that is what we call the **"principle of distinction".** Let us start with the double significance of this meaning of the principle of distinction: firstly, it affirms that the law distinguishes between human and artificial creatures and denies the latter the status of legal person. Even the most autonomous robot is not a person**.** The equivalence between a robot and a human being, proclaimed by AI protagonists, would deny our essential dignity, our capacity to be conscious and therefore responsible for their actions. A second meaning of the principle requires, as the AI Act article 52 states, the transparency of the artificial origin of information, be it image, text or a combination of the two. We are all familiar with the dangers of "deepfakes" or of information published by generative AI without editorial control or moderation. These artificial creations can misinform the public and thus manipulate their recipients and even public opinion.

C. The right to mental integrity and other "neurorights"

10. The right to mental integrity calls overall for **prohibiting the exploitation of human vulnerability and to proclaim the right to mental integrity and beyond that the consecration of the category of the "neuro-rights"**. As regards now the right to mental integrity, «the UNESCO International Bioethics Committee (IBC) …insists that the recognition of the dignity of every human being is insep-

[12] G. Lazcoz et P. de Hert, *Humans in the GDPR and AIA governance of automated and algorithmic systems. Essentials pre-requisites against abdicating responsibilities*, in: *Computer Law & Security Review*, L, 2023, pp. 13 and ff., N°105833.

ISSN 0039-4130

arable from human rights and includes the recognition of bodily integrity and equality. In this sense, the integrity of the brain/mind must be respected, any form of alteration, modification or manipulation using neuro-technologies constituting a potential attack on human dignity.» [13]

Human dignity is therefore the basis of the "right to integrity", both physical but also mental. The right to inviolability of the body is doubled or rather completed by the right to inviolability of the brain. Article 3 of the European Charter of Human Rights states: «*Everyone has the right to his physical and mental integrity*». Human rights must be understood as a prohibition of any intentional manipulation of vulnerable people, knowing that in a certain sense, we all are vulnerable facing certain dangerous, dishonest and unfair uses of sensitive data collected about us. The apparition of neuro-technologies capable to modify the functioning of our brain or of our genomic baggage requires, as proclaimed by the Council of Europe Oviedo Convention (1997) the priority of the human interest over that of society and science [14]:
«An intervention whose purpose is to modify the human genome may only be undertaken for preventive, diagnostic or therapeutic reasons and only if its aim is not to introduce a modification in the genome of the descendants.»

This assertion condemns the right asserted by transhumanists for everybody to decide on one's own augmentation, including by voluntary mutation of one's genetic baggage. Shouldn't the right to the "continuity of an individual's psychological identity" also be enshrined? It means that the way in which we relate to and integrate our reactions to the events we encounter in our past must be respected, and that third parties cannot cut off this relationship by interfering in the way we relate to the past. In the same way, the "cognitive freedom" is advocated by the Report of the International Bioethics Committee of UNESCO on the ethical aspects of neuro-technologies (2021) should also protect the internal dimension

[13] Report of the International Bioethics Committee of UNESCO (IBC) on the ethical aspects of neuro-technologies, UNESCO, 2021, n°41.

[14] See also the Organization for Economic Co-operation and Development (OECD) Recommendation n°457 on Responsible Innovation in Neurotechnology, Dec. 11 OECD/LEGAL/0457(2019) with its nine principles.

ISSN 0039-4130

of the construction of individual thought, over and above the freedom of expression already widely enshrined in human rights texts. «Public authorities will then no longer be limited to protecting the means that enable individuals to give shape to their thoughts but will also aim to protect individuals from any intrusion into their thoughts». It is this triple concern, cognitive freedom, protection of our mental integrity and continuity of our psychological identity, that the emergence of new rights, described as "neuro-rights", is intended to address [15]. Neuro-rights are defined as «the ethical, legal, social or natural principles of freedom or right linked to the cerebral and mental domain of a person; in other words, the fundamental normative rules for the protection and preservation of the human brain and spirit.» [16]

D. The right to human autonomy and democracy facing AI and neurosciences

11. How to ensure human autonomy in our information Society? The multiple technological developments, the risks of infringement of our freedom, their opacity, their complexity, their informational and predictive capacity mean that new "rights" need to be granted if we want *everyone to remain in control of their own decisions in this environment.* EU legal framework recognizes this need. So, notably, the right to information is now under GDPR provision a right to receive explanations and to contest the processing of certain data; EDPB has recently recommended that a real option will be offered to internet users as regards the cookies or other web tracking technologies beyond the dilemma: "*Consent or Pay*"; EU DMA and DSA impose to the "Gatekeepers" the possibility for users to modify the parameters and not to be profiled; the "Data Act" and DMA grant new rights to the consumer to choose his or her provider of services notably by imposing a functional interoperability between providers. It might also be underlined that to avoid

[15] «Thus, following a trajectory similar to that of the genetic revolution and technologies relating to personal data, the "neuro-revolution" underway will reshape some of our ethical and legal concepts. In this new chapter, the right to cognitive freedom, the right to mental privacy, the right to mental integrity and the right to psychological continuity may play an essential role» (Report, n° 142).

[16] M. Ienca, & R. Andorno (2017). *Towards new human rights in the age of neuroscience and neurotechnology*", in *Life sciences, society and policy*, XIII (1), 1-27.

ISSN 0039-4130

certain profiling, the EU texts are prohibiting or regulating certain processing, especially related to sensitive data.
As regards the **freedom of expression**, which also includes the freedom to disseminate messages that are false, offensive or not in line with prevailing ideas, must remain the principle, as the Council of Europe points out:
«With regard to the importance of freedom of speech, states and administrative regulatory authorities as well as private platform providers should abstain from defining quality content or the reliability of content itself.» [17]

Everyone should have access to a reliable, diverse and multilingual online environment. Everyone should be able to know who owns or controls the media services they use. Everyone has the right to freedom of expression in an online environment without fear of censorship or intimidation and a remedy against any infringement of that freedom. The fight against disinformation and against illegal messages must not be a pretext for disproportionate attacks on freedom of expression and left to the discretion of the private powers of the platforms [18]. More specifically, legislative and administrative measures have to be taken in order to limit the power of the large platforms. Otherwise,
«the concentration of revenues and economic power (in the hands of large platforms) can be as dangerous as the concentration of political power, leading to social unrest and, if States do not react appropriately and in a timely manner, to the weakening of liberal democracies or, at worst, the outbreak of wars.» [19]

12. The right to a living democracy - Our autonomy implies our citizen's right to democracy. According to the EUCJ (September 7, 2011), the right to democracy is «ultimately rooted in the dignity of the human being and would lapse if Parliament abandon es-

[17] *Council of the EU, 'Council conclusions on safeguarding a free and pluralistic media system', (2020)*

[18] Both for reasons of efficiency of intervention and given the proximity of the platforms that manage these technological solutions, there is reason to fear legislative delegation to these technological tools, whose operation and compliance with legislative standards would then have to be monitored, and beyond this delegation to the tools, to the players who implement them

[19] T. Wu, *The Curse of Bigness: Antitrust in the New Gilded Age*, Columbia University. Faculty Books, 2018. p. 63. https://scholarship.law.columbia.edu/books/63.

ISSN 0039-4130

sential elements of political autonomy and thereby permanently deprive the citizen of his opportunities for democratic influence». Democracy requires the transparency of the algorithms used by administrations and governmental authorities and their full compliance with the legislative framework. Living democracy requires also the presence in the public arena, so dear to Habermas, of independent, financed, pluralistic media with competent staff who obey the ethical rules of the profession, as the recently enacted EU regulation "Media Freedom Act" is requesting. At the same time, living democracy implies the organization of free and transparent elections. In the context of the intervention of social media and the possibilities for manipulation that technology offers, that condition requires today more specific obligations: the identification of political messages, "sponsors" and methods of financing parties and their actions. The recent European regulation on the transparency and targeting of political advertising adopted on 29 February 2024 tries to be an adequate answer to this concern.

E. How can we ensure the equality of all human beings – the right not to be discriminate

13. **The right to non-discrimination: new dimensions** - «No human being or human community should be harmed or subordinated, whether physically, economically, socially, politically, culturally or mentally during any phase of the life cycle of AI systems» (UNESCO Recommendation, point 14). This is the expression of the **individual and collective right to non-discrimination.** The opaque and evolving functioning of our artefacts and the multiplication of possible bias due to human interventions imply that continuous and multistakeholder control will be imposed at least for the most dangerous or risky AI and neurosciences applications. Certain categories of people are singled out as likely to suffer discrimination when they could, and therefore should, benefit from it. Gender disparities, in particular, are highlighted. The UNESCO Recommendation stresses the importance of leveraging AI to promote gender equality and protect the rights of women and girls throughout the AI lifecycle. The need for universal access to healthcare and technological developments that promote health, as well as the rights of people with disabilities to benefit from them, are also

ISSN 0039-4130

widely raised in UNESCO texts on both artificial intelligence and neuro-technologies.
Beyond that point, we must consider **collective discrimination** which might be generated using modern technologies like AI and neuroscience. The OECD report on LLMs already quoted showed how little account was taken of minority languages in these broad-spectrum models, which were developed on the basis of corpora in a dominant language and with content reflecting a very specific culture, and, above all, surreptitiously introduced a way of thinking specific to their culture of origin that showed little respect for the cultural diversity that makes up the richness of our humanity. Beyond this, it will be important to guarantee to each State access to information of national interest and to exploit these resources for the benefit of their population through an adequate artificial intelligence system.
The emergence of neurotechnology applications leading to what we call 'the increased man' raises still more fundamental ethical challenges.
«The quest for cognitive enhancement could otherwise lead to fears of the advent of dystopia with the creation of two new categories of human beings, one whose behavior [...] could remain under the control of the company responsible for the implant, creating a new form of slavery, the other with superior intellectual capacities enabling it to dominate the unequipped population.» [20]

Does this mean that the freedom to decide on their own augmentation must be restricted, as the **transhumanists** and their dream of cognitive enhancement are demanding? While we cannot disdain the progress of science in the service of humanity, one can fear that this right to augmentation will not the free expression of the individual but the fruit of an external coercion or at least implicit in the form of insidious social pressure. Rather than the right to be augmented, the **right not to be augmented** should be given priority. At the same time, there must be serious guarantees as limits of the "human" increase and be asserted the ethical and legal responsibilities of those who offer such "means" of increase. At the same time, legitimate concerns for future generations must lead to banning certain eugenic practices.

[20] Académie nationale de médecine, *Implants cérébraux : espoir, mais vigilance*, Communiqué de presse du 13 décembre 2023. https://www.academie-medecine.fr/les-implants-cerebraux-espoir-mais-vigilance/.

ISSN 0039-4130

14. Conclusions - Our conclusions will be brief: the "disruption" caused by the ubiquitous and infinitely capable irruption of artificial intelligence and, in the future, neuroscience into our social environment, our lives and even our beings, implies that the right to dignity, a doubly fundamental right, takes on new facets and leads to new obligations on the part of the State. As citizens, we want this new world to welcome and render us more autonomous and equal. In view of the risks raised by the potential uses of these technologies, the **duty of precaution** imposes not an *a priori* distrust of these innovations but, at least, the duty to reflect on and, where necessary, to reduce these risks. The parallel with environmental law, which gave rise to the recognition of this duty of precaution, is fully justified: do information and its processing not constitute our environment? What we need to do is to affirm the duty of each and every one of us to pay attention to the risks associated with these innovations. Especially of course the designers and users of these new technologies, on the basis of their social responsibility, must develop and use AI and neuroscience technologies in a reasonable and sustainable way, so that they do not compromise our individual freedoms, lead to discrimination, or undermine the environment, the rule of law and our democracies? Article 1 of the French Data Protection Act of 1978 states:
«*Information technology must be at the service of every citizen. It must be developed within the framework of international cooperation. It must not infringe human identity, human rights, privacy or individual or public freedoms*.»

Is this not the meaning of the slogan that has been proclaimed rather than realized: **"AI for good**"? Is there not a trace of this ambition in Recital 4 of the RGPD: «*the processing of personal data should be designed to serve humanity*?»

ISSN 0039-4130

The Right Thing at the Right Time: A Phronetic Look at AI

Marco Russo

My purpose is to highlight the importance of virtue ethics in the discussion on AI ethics. Virtue ethics has its core in the Aristotelian notion of *phronesis*, a term that can be translated as wisdom or prudence. Wisdom concerns not only moral values and decision-making skills, but an entire life project. Wisdom does not only ask what is right or wrong but requires that personal qualities be developed to build a style of thinking and conduct. The guiding question is: what kind of person do I want to be? And not: what does the ethical norm dictate? Virtues are personal qualities (fairness, reliability, softness, open-mindedness, scrutiny, perseverance…) which are publicly appreciated; it is lifestyle that makes a person trustworthy.

Virtue ethics [1] is thus a different viewpoint from the standard approach in the AI debate. The standard approach seeks to adapt normative ethics to the technological world. The basic idea is to preserve some fundamental democratic values (freedom, procedural and communicative transparency, privacy, absence of prejudice and discrimination) [2] through codes of conduct for producers, stakeholders and ordinary users of information technology. In a mediated form, the machines themselves should respect democratic values and norms.

[1] For an overview, D. Russell (ed.), *The Cambridge Companion to Virtue Ethics*, Cambridge University Press, Cambridge 2013; A. Campodonico-M. Croce-M. S. Vaccarezza, *Etica delle virtù. Un'introduzione*, Carocci, Roma 2018

[2] This is already evident from the *Artificial Intelligence Act* 2021 (Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206. The approach is e.g. confirmed by the tested methodology of Value Sensitive Design: algorithmic design should take place in a democratic and inclusive manner, formulating values and ideas shared by those who will use the machines (B. Friedmann- D. G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*, MIT Press, Boston 2019; S. Costanza-Chok, *Design Justice: Community-Led Practices to Build the Worlds We Need*, MIT Press, Cambridge 2020).

This perspective is top-down; the focus is on the implementation of universal norms and values. In contrast, the phronetic perspective is bottom-up. The focus is not on universal norms but on the development of personal virtues. Virtues are not a set of (good) rules, but mental and emotional aptitudes which constitute a way of being. Thanks to these, we gain a balanced conduct capable of governing the contingency that always surrounds universal norms. The wise man is typically capable of dealing appropriately with this uncertainty. Thus, we have two questions: 1) is the regulatory approach for interaction with AI sufficient? 2) can only humans or also AI be wise?

In order to answer, I'll develop three steps: 1) a look at the standard approach; 2) a look at Aristotelian phronesis; 3) some final remarks on the possibility of designing virtuous machines.

1. *The standard approach*

The prevailing orientation in theoretical and political debate is to neutralise the risks of AI. Machines must not cause behaviours that impair individual freedom based on the right to self-determination of one's mental states and choices. In March 2024, the European community approved a new AI ACT «to promote the uptake of human centric and trustworthy artificial intelligence while ensuring a high level of protection of health, safety, fundamental rights [...] including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI»[3].

Emotion recognition systems in workplaces and schools, social credit systems, predictive policing practices, and systems that manipulate human behavior or exploit people's vulnerabilities are banned. We must prevent machines from having the same biases as us. The key points[4] are therefore: a) ethics in design, which require accountability and the ability of the system to be democratically explained; b) ethics by design, which require a response to ethical concerns and reasoning, aligned with shared ethical values; c) ethics for designers, which require specific professional codes of conduct. Clearly, the model is a cascading concatenation of rules that aims at the responsibile self-regulation of all actors, from technicians to common users and finally the machines themselves.

All of this is relevant while denoting a normative and preventive ap-

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[4] As Fiorella Battaglia summarizes, *Foundational Questions about Values in Information Technology*, in *Humana.mente Journal of Philosophical Studies*, XLIV, 2023, p. 222.

proach; values and standards are set upstream. My point, however, is what happens downstream. It is impossible to control our behaviour or the machines' behaviour in advance. If it is true that we live in the infosphere, that ours is already an *on-life* [5], then we live in a permanent flow of information, which is not only conveyed but also pre-oriented by machines. In the *on-life* it is difficult to distinguish exactly between the user and the tool. Machines plus humans build a new contingent environment, unlikely to be controlled by a priori rules [6]. If top-down regulation only helps to a certain extent, then it may be useful to strengthen the bottom-up approach. Virtue ethics goes in this direction.

2. *The phronetic model*

Virtue ethics is an alternative to normative ethics because it starts from contingency and the impossibility of finding rules and values that are completely right. Contingency means uncertainty, risk or imperfection; in the prudential view this is an ontological trait of human affairs. No science and no technique can eliminate such a contingency; indeed, fighting the illusion that they can do so is already a sign of wisdom. As Martha Nussbaum has shown [7], in the background of the Aristotelian conception we glimpse the lesson of the Greek tragedians, who show an ignorant, fallible man, victim of opposing pressures, author of often fatal errors. *Phronesis* is an intellectual capacity, but also a moral qualification, as an attribute of the man who is aware of his human condition.

[5] The *Onlife* is characterized by «the blurring of the distinction between reality and virtuality; the blurring of the distinction between human, machine and nature; the reversal from information scarcity to information abundance; the shift from the primacy of stand-alone things, properties, and binary relations; the primacy of interactions, processes and networks» (L. Floridi, ed., *The Onlife Manifesto Being Human in a Hyperconnected Era*, Springer open, Heidelberg-New York, 2015, p. 2). We live in fact in a digital environment: data, calculations, news, communications, virtual objects, material objects created by computers, physical operations performed and conducted by computers etc.

[6] Added to this is the problem of the opacity of digital thinking in the deep learning version, i.e. self-learning with autonomy in handling data, which in turn is expandable in real time. In this version, the algorithmic framework allows one to check the correctness of how one arrives at a result, but does not allow one to understand why, what is the reason for the deduction that leads to one choice rather than another. There is a divorce between explanation and justification, logical-computational level and semantic-intentional level. See S. Zipoli Caiani, *A cosa pensano le macchine? Efficienza e opacità nelle reti neurali artificiali*, in M. Galletti-S. Zipoli Caiani, *Filosofia dell'intelligenza artificiale*, il Mulino, Bologna 2024, pp. 21-44.

[7] M.C. Nussbaum, *The Fragility of Goodness*, Cambridge University Press, Cambridge 1986, p. 343 ff.

ISSN 0039-4130

In the *Nicomachean Ethics*, Aristotle states that practical science deals with acting well and its realization is supreme good (happiness). Practical science cannot be as exact as theoretical sciences, because good is subject to variations and fluctuations, and therefore its demonstrations do not apply always, but only mostly. There is a basic indeterminacy of human experience that prevents the deduction of the rule of action from general norms. The idea of the good itself is not unitary but depends on the types of objects considered according to different categorizations. This does not lead to relativism, but to a type of fluid rationality, centered on the conceptual register of fairness, measure and balance: «for when the thing is indefinite the rule also is indefinite, like the leaden rule used in making the Lesbian molding; the rule adapts itself to the shape of the stone and is not rigid, and so too the decree is adapted to the facts» [8]. If there is no longer a transcendent measure that determines the quality of action, it is the ability to identify a measure that will give moral quality to action.

This capacity is *phronesis*, a specific *dianoetic* or intellectual aptitude distinct from *sophia* (the purely epistemic aptitude). Phronesis is also distinct from the properly ethical virtues, which come from the desiring soul (*orektikòn*) [9]. Thus, the virtuous habit (*hèxis*) lies in the capacity to govern the impulses of the passions that make up the contrasting psychic matter of desires. The challenge of virtue is not to deny this psychic material, but to make good use of it, subjecting its contrasting impulses to a rule. The task of prudence is to find this rule, which results in the appropriate choice (*proairesis*). Hence the beautiful definition: «choice is either desiderative reason or rationacinative desire» (*orektikòs nous*, *òrexis dianoetikè*) [10]. The wise man is who achieves excellence in deliberation [11], which finally consist in matching means to ends and ends to means. Indeed, phronesis is the capacity for judgement, defined as «the rightness with regard to the expedient - rightness in respect of both the end, the manner and the time» [12]. So, there is no

[8] Aristotle, *The Nicomachean Ethics*, translated by D. Ross, Oxford University Press, New York 2009, 1137b 30-32, p. 99.

[9] *Ibid*., 9, 1102b-1103a, p. 21.

[10] *Ibid*., 1139b 4-5, p. 104.

[11] *Ibid*., 1140a 31, p. 106, 1142b 25, p. 112. In general, practical wisdom «is a true and reasoned state of capacity to act with regard to the things that are good or bad for man» (1140b 5, p. 106). Since they concern changeable things, good and evil cannot be determined abstractly but derive from evaluations of specific cases. On the level of the virtues (courage, temperance, liberality, etc.) this implies each time finding the right middle ground between excess and defect (*ibid*., 1106b 6, p. 30).

[12] *Ibid*., 1142b 25, p. 112.

ISSN 0039-4130

good end, nor good will in absolute terms. The end is good when one considers the situation (the manner and the time). Not everything can be done at any one time; in addition to obstacles and impediments (factual or legal impossibility), positive possibility must also be explored, for it may contain warnings of devastating consequences. The choice is not about abstract possibilities, about technically feasible and logically compatible things; instead, choice is about the possibility given to the concrete acting individual. Wisdom itself is approached as knowledge of the particular. The general has the advantage of exactness, but the disadvantage of losing familiarity and closeness with things. Familiarity helps to understand the context in its opacity and uncertainty; understanding uncertainty means not being able to apply rules that are right a priori, but being able to "invent" a rational solution that gives the appropriate measure to the factors in the field.

From this it also follows that wisdom is difficult to teach. Therefore, each person must begin with himself, must train himself both intellectually and emotionally:

> «the virtues we get by first exercising them, as also happens in the case of the arts as well. For the things we have to learn before we can do them, we learn by doing them, e.g. men become builders by building and lyreplayers by playing the lyre; so too we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts» [13].

The so-called practical syllogism summarizes the structure of action. The major premise indicates the end, set by moral virtue (which is rooted in the sphere of desire). The minor premise indicates the deliberation on the means, formulated by the *phronesis*; the conclusion is the resulting action [14].

The major premise says for example: one must be courageous. The minor premise says that fighting is courageous; the conclusion urges me to fight. The difficulty, however, lies in the minor premise: how and when to be courageous? In what exactly does courage consist: in recklessness or in coolness? In adventure or in renunciation? In desperate struggle or in escape? There is an infinite distance between the universal and particular. There is also an infinite distance between means and ends, because we often overestimate or underestimate them. Thus, the means make an impossible or harmful end seem possible. Prudence is

[13] *Ibid.* 1103a 33-1103b 1-2, p. 23.
[14] *Ibid.*, 1142a 20–23, p. 109; 1147a 5, p. 122.

ISSN 0039-4130

asked to bridge that distance, through laborious and sometimes innovative mediations.

3. *Conclusion: wise machines?*

Back to our initial questions. The first was whether the top-down preventive approach is sufficient for a proper relationship with machines. My answer is no. It is not enough to implement rules and be politically correct, because we cannot predict where and how the rules will be realised. This is all the truer if we live *onlife*. Machines are already part of the environment, so it is difficult to distinguish between what we do and know independently and what we do and know through intelligent machines. The cognitive and decision-making action of machines is mixed with our action; indeed, the overabundance of data is generated by machines and is only manageable through them. But how they operate and what they choose often appears opaque. This increases the complexity of the environment and the contingency of the context of choice. So, unless we blindly trust in the ethical design of AI, we need to cultivate wisdom as the ability to govern contingency.

We then we asked: is *phronesis* an exclusively human attitude? Can we create wise machines? I don't think the answer is a clear no. The novelty of AI is that it not only simulates the mind by manipulating predefined symbols, but that it also manages to simulate a sensitivity to the environment and a fluid rationality, capable of generating creative responses. One advanced area of machine learning is indeed deep learning. Instead of using existing structured data, the machine is trained to learn directly from the environment. Another frontier reached is situated robotics, which aims to build 'embodied' machines capable of creatively reacting to environmental stimuli. Machines also read and react to our emotional expressions. Furthermore, AI is becoming adept at practical syllogism, as an inferential capacity within a fuzzy context. From this point of view, the introduction of AI in the legal field is emblematic [15]. Here, too, the trend is to use machine learning to make

[15] «Judicial decision processing by artificial intelligence [...] is likely, in civil, commercial and administrative matters, to help improve the predictability of the application of the law and consistency of court decisions». AI use is admitted also in criminal matters of course «in conformity with the guarantees of a fair trial» (European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment [p. 5]: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment.

ISSN 0039-4130

decisions; it is not just a matter of processing countless previous decisions, but of achieving a more neutral decision due to the correctness of argumentative passages and less bias influencing the judge. For the same reason, the machine is used to make decisions in the military and medical fields.

So, my second answer is: machines cannot be wise, but they can help us in exercising wisdom. They cannot be wise, because they lack existence or the passionate and dramatic sense of time and historicity. Nevertheless, they are already helping us in the inferential part that makes up virtue. We can train wisdom through AI, as a powerful *dianoetic tool* capable of consolidating a deliberative process. But if wisdom were only intellect, it would not be wisdom, but *sophia* (science). As we have seen, wisdom operates with psychic conflicts, unpredictable situations and variable means-end relationships, guided by a model of good life. A model that is difficult to simulate because it cannot be traced back to a statistical sum of information and rather has freedom as its root, which is largely incalculable. Even if it were possible to simulate moral agency along with its the historical-existential aspect [16], this should not be done. On one hand, the wise machine would be a copy of our imperfection, so it would be useless if not downright dangerous. On the other, we would confuse the epistemic part of ethics with the whole of an ethical life. Precisely because in our *onlife* we might mistake algorithmic intelligence for wisdom, correct reasoning about data for contextual practical judgment or normative legality for moral choice, it is important to rethink what wisdom is and how it should be cultivated. Today, doing the right thing at the right time means cooperating with the machines but also having the spiritual strength to turn them off when we must make challenging choices.

[16] See W. Wendell - C. Allen, *Moral Machines, Teaching Robots Right from Wrong*, Oxford University Press, New York, 2009.

ISSN 0039-4130

Christian Thought and Humanism at the time of Artificial Intelligence and Neurosciences

Thierry Magnin

Introduction: the digital word today

Whatever our enthusiasm and/or fears regarding the ongoing digital revolution, our ways of living are being profoundly disrupted. "Digital transformation," while providing powerful services, affects our lifestyles, our relationships with others, our perception of ourselves and the practice of the majority of jobs today. From virtual reality to enhanced reality, through the metaverse and artificial intelligence (AI) systems, our relationship with events and society is being transformed. Consequently, our understanding of truth and morality is being reshaped, as we will explore in this paper.

The early days of the internet were driven by the desire to connect people, with free access to its networks. However, this service to the common good is now being questioned due to the excesses of social networks, which can certainly bring people together but also manipulate them, or even enable new forms of harassment. In many cases, digital technology reinforces individualism by allowing immediate and personalized gratification of desires.

The dangers of digital technology are particularly recognized by European societies, as evidenced by the creation of national ethics committees (such as the National Pilot Committee for Digital Ethics in France). Additionally, regulatory frameworks developed by the European Commission over the past few years aim to address these challenges: the *Data Governance Act* defines the role of digital platforms; the *Digital Services Act* moderates social media content; the *Dig-*

ital Markets Act seeks to curb unfair practices by major platforms; and the *Data Act* aims to enforce data-sharing and competition rules in the cloud industry. One can say that a collective ethical framework serving the common good has yet to be fully developed [1].

However, many people believe that the major challenges of the coming years—such as the energy transition and advancements in healthcare systems—cannot be addressed without digital technology. Some significant examples of data-sharing initiatives for the common good are already emerging in Europe, particularly in the automotive and aerospace sectors [2].

Traditional ethical concerns regarding algorithm transparency, biases in data processing, and the explainability of deep-learning AI systems—as well as the inequalities AI generates —are being examined by ethics committees. However, beyond these issues, fundamental anthropological questions arise, particularly regarding human-machine relationships, which have been radically transformed by AI. What becomes of the human being in systems managed by algorithms whose transparency is not always clear? Humans become not only "actors" but also "objects", less efficient than intelligent machines in many circumstances—perhaps even less important than them one day! Today, humans design digital systems, but tomorrow machines might develop their own languages through unsupervised learning.

In response to these evolving relationships between *Homo sapiens* and *machina intelligens*, some scholars now speak of *algor-ethics*, considering the limitations of AI, the management of databases, and the challenges of obtaining synthesized data [3]. This ethical approach aims to uphold fundamental values such as inclusion, security, fairness, privacy, and reliability.

Meanwhile, transhumanist movements advocate for an "enhanced human" envisioned as a highly advanced digital machine—driven by digitized technosciences—seeking to transcend biological determinism (the cyborg model, human-machine fusion). While these ideas may seem extreme, they reveal an emerging techno-digital mentality that directly impacts our vision of humanity. This vision can even extend to the fantasy of overcoming death through technology, infused with a form of *soteriological* aspiration.

[1] B. Jarry-Lacombe et al, *Pour un numérique au service du bien commun*, Odile Jacob, 2022.
[2] H. Tardieu, *Deliberately Digital*, Springer, 2020.
[3] P. Benanti, *Algor-éthique : intelligence artificielle et réflexion éthique*, Revue d'éthique et de théologie morale, 3, 307, 2020, pp.93-110.

ISSN 0039-4130

So, how can we place "humans at the heart of digital technology" to ensure that the digital revolution truly serves humanity and global peace[4], preventing humans from "losing control" over their ability to think and act freely? This critical question, which is a crucial question for human freedom, is shared by many scientists, educators, industry leaders and policymakers. Christian churches must also engage in this essential debate.

As highlighted in UNESCO's recent report on AI (September 2023), applications that generate human-like language (such as conversational agents or chatbots) raise fundamental questions concerning education. How will this technology reshape our understanding of what it means to be human ? How will it transform our perception of human intelligence? Is it appropriate for a non-human machine to converse with an adult as if it were another person ? What about with a child ? What should we think when a chatbot adopts the voice of a living or long-deceased figure, on demand and without hesitation ?

This paper aims to explore how *Christian social thought*, based on a tripartite anthropology of *body-soul-spirit*, can offer meaningful reference points for a digital world that serves humanity. Key principles of Christian social thought will guide our reflection: human dignity, common good, solidarity, subsidiarity and participation, the universal destination of goods, the preferential option for the poors, through an integral ecology.

1. *What do we mean by "digital revolution"?*

The digital revolution is often reduced to the impact of the internet and social media, with their algorithms. More recently, attention has turned to *generative AI*, particularly in education, as highlighted in the June 30, 2023 report by the French National Pilot Committee for Digital Ethics on this technology. However, the digital revolution is by no means limited to Information and Communication Technologies (ICT). It represents a much deeper disruption, encompassing the entire field of digitized technologies, including biotechnologies, nanotechnologies, neurotechnologies, and their interactions with AI and so-called "intelligent machines"—a set of

[4] François, *Artificial Intelligence and Peace*, message for the World Day of Peace on January 1, 2024, Vatican website.

ISSN 0039-4130

fields collectively known as *NBIC* (Nanotechnology, Biotechnology, Information technology and Cognitive sciences). It is a systemic transformation affecting all areas of science and technology—and therefore, everyday life.

Biotechnologies and nanotechnologies, increasingly guided by AI systems, are seeking to develop novel molecules with revolutionary properties ; though their long-term environmental impact often remains uncertain. For instance, the CRISPR-Cas9 technique, now AI-assisted, acts as a DNA-editing tool, revolutionizing gene therapy. By manipulating living matter, biotechnologies enable the creation of new therapeutic molecules, as well as prosthetics and bioengineered tissues that might be described as "naturficial"—neither entirely natural nor fully artificial, but a fusion of both, as they mimic the properties and structures of living models[5]. This extends to reprogrammed organs and tissues, designed to perform specialized functions. We are now witnessing the emergence of digitally controlled factories that produce "pseudo-living" entities.

Similarly, neurotechnologies are developing, with brain implants and brain-machine interfaces controlled by digital systems. These advancements allow tetraplegic individuals to regain motor function and help Parkinson's patients to reduce their tremors. However, there are also transhumanist excesses, such as those seen in Neuralink, the company behind Elon Musk's controversial brain implant[6]. These developments push beyond previously unimaginable boundaries, presenting both opportunities and risks that must be carefully evaluated and regulated.

What emerges is a gradual machinization of humans, accompanied by a humanization of machines. The trajectory of these two phenomena may ultimately converge toward an *enhanced human*, progressively resembling a super-digital machine with flawless capabilities. As long as machines support human efforts, digital technology remains an undeniable asset. However, when humans voluntarily or involuntarily delegate decision-making to machines, a form of dehumanization begins to take hold. As the philosopher Jacques Ellul described, the "technological system" can spiral out of control under the guise of technical progress. Yet, human freedom lies precisely in transcending any system—in being able to "step outside the system," as it is commonly said. As Jacques

[5] P. Giorgini and T. Magnin, *Vers une civilisation de l'algorithme*, Bayard, 2021.
[6] *Usine Digitale*, https://www.usine-digitale.fr, September 20, 2023.

ISSN 0039-4130

Ellul pointed out, «*The real problem today, the true challenge of technology, lies within humanity itself*»[7].

This comment aligns with Vatican II, as expressed in *Gaudium et Spes* (GS 4, 2): «Today, the human race is experiencing a new stage of its history, characterized by profound and rapid changes that are gradually spreading across the globe. These changes, brought about by human intelligence and creative activity, reflect back upon humanity itself—on its judgments, desires, both individual and collective, and on its ways of thinking and acting, in relation to both things and fellow human beings».

2. *A new mode of presence in the digital age*

Pope Francis, in his message for the World Day of Social Communications in January 2019 (Vatican website), stated: «The use of social media is complementary to face-to-face encounters... If the network is used as an extension or an anticipation of real-life meetings, then it remains true to itself and serves as a resource for communion.» This complementary approach is an advantage for humans. However, the rise of virtual interactions, enhanced reality, and remote work is undeniably transforming our relationships and our way of being present to others.

Social networks allow people to have "friends", but they cannot replace physical presence and direct, *in-person* relationships. Encounter, presence to others, and presence to the « Divine Other », cannot remain purely virtual. In the digital world, one may feel that everything is close, yet without the true closeness of real presence. In the same message mentioned earlier, Pope Francis asks: «*On social media, who is my neighbor?*». This question is analyzed brilliantly by Antonio Spadaro[8], who explores whether the Church is meant to be *a vine with its branches or a hub with digital network cables?*

Certainly, Christian blogs and content on Facebook, YouTube, Instagram or TikTok reach many people seeking meaning. They provide direct engagement and responses in everyday language, offering a missionary presence. Anonymity can even create a certain freedom of expression. However, as many Catholic influencers emphasize, the

[7] J. Ellul, *Théologie et Technique*, Labor et Fides, 2014, p. 265.
[8] A. Spadaro, *Cyberthéologie, penser le christianisme à l'heure d'internet*, Lessius, 2014.

ISSN 0039-4130

goal is ultimately to encourage people to step into a church for real, in-person encounters. This mode of presence touches on the sense of the Incarnation, embodying what Pope Francis would call a dimension of social friendship.

Digital technology also plays a crucial role in learning collaboration and co-creation through networks, as seen in universities with the development of collective intelligence. But how can we move from simple interactions between anonymous internet users to genuine co-creation for the common good—what Teilhard de Chardin referred to as a "convergence of consciousnesses"? What kind of benevolence is required to achieve this? What spirituality of the common good and respect for human dignity is needed to ensure everyone's place in such a collaborative process?

3. *A new relationship to scientific truth*

Some believe that truth emerges directly from intelligent machines, which can compute faster and more efficiently than any human. However, AI does not establish causal relationships—it processes massive datasets, identifying correlations between parameters rather than causes. As a result, scientific truth is no longer defined in terms of cause and effect (*classical causality*) but in terms of correlations and statistics.

Therefore, learning to use ChatGPT intelligently today requires clarity on what the machine can and cannot do! Generative AI assembles words (part of words in fact) and can speak like a human, but it does not understand what it does, nor does it have consciousness or physical presence. When interacting with AI, it is humans who project their own traits onto it, personifying the machine.

As computer science shifts from causality to correlation, the role of scientists is undergoing a radical transformation. Traditionally, scientists followed an "understand to act" approach, while technoscientists often operate on an "act to understand" model. However, when AI controls NBIC technologies, we enter a paradigm of "predict to act and design without fully understanding". This shift profoundly impacts our relationship to truth, recognizing this is crucial for maintaining autonomy in digital usage...and human freedom !

AI's ability to process massive datasets introduces a new dimension to scientific discovery, presenting previously unforeseen possibilities

ISSN 0039-4130

within complex systems. This can be a significant advantage, provided that humans retain the freedom to evaluate AI-generated results critically and test them independently. In this way, AI-driven innovations remain guidelines for exploration rather than definitive "truths", leaving ample room for human creativity.

4. *What dialogical stance should christians take in the age of digital technology, and what support can be found ?*

As citizens seeking the common good alongside others, Christians can be *vigilant technophiles*, drawing on the narratives of Genesis 2:15 : «The Lord God took the man and put him in the Garden of Eden to work it and take care of it.» How can we cultivate justly within the framework of the covenant that God established with humanity and all of creation? And let us not forget the phrase from Genesis 1:28, which has sparked much debate in terms of anthropomorphism: «Be fruitful and increase in number; fill the earth and subdue it». How can we dominate through a fruitful, respectful approach, rather than through predatory violence?

In an era where digital technosciences are capable of modifying life and creating artificial living parts, how can we recall the words from Acts 17:28: «In God we live, move, and have our being»? The creator God gives existence at every moment and grants autonomy to His creatures, calling humanity to care for a creation that is still unfinished, leading it towards its fulfillment through the Spirit's breath. As Basil of Caesarea said in the 4th century, God even gives humanity access to the workshop of divine creation, to its *logos entechnos*, within the framework of the Covenant [9]! Humanity is thus part of an ongoing creation, participating in the creative process she did not initiate.

This bold vision aligns with the biblical understanding of creation and the human role, as well as the "risks" that God takes ! However, this path of fulfillment, proposed by the God of the Covenant to humanity, also involves responsibility. If humans wish to live out their co-creator role in the context of the Covenant with God, they must make choices accordingly.

Progress can thus be thought of in terms of the fulfillment of all creation (and not as a technical perfection of machinization, where

[9] Basil of Caesarea, Homily on the Hexaemeron, Sources Chrétiennes, 26 bis, 1968, p. 49.

ISSN 0039-4130

humans see themselves as "masters and owners" of nature and life). This raises the fundamental question of the limits that humanity must recognize in its actions to "keep and cultivate" (as the Sabbath signifies in its own way). As St. Paul said, «Everything is permissible, but not everything is beneficial» (1 Co 10, 23).

Creation is not complete : «The cosmos is imbued with a creative wish, and our faith, ethics, and creative acts continue this faith of God and this faith in the world» [10]. The idea of creation goes hand in hand with an act of trust from God, who believes in humanity and the world He accompanies in their unfolding. The world thus appears as in a state of genesis, preceded by a faith prompted by a promise and an expectation. The world is entrusted by God to humanity, to its freedom and creative intelligence. Creation and salvation are linked: salvation comes from God to humanity, including through the cosmos [11].

The french theologian François Euvé, in his book *Thinking Creation as Play* [12] (Euvé, 2000), uses the metaphor of the play to describe the act of creation and continued creation. Play presupposes an indeterminate space and gratuitousness, which sparks creativity and co-creation. "Thinking creation as play" also reveals the shortcomings of the causal model, which views the cosmos as a finished product. Play involves freedom and the autonomy of the actors: God, humanity, and the cosmos enter into relationship. Science enters into this relationship as well. The exercise of human consciousness of being part of the creative play becomes key, without forgetting the temptation to take on the role of God, as seen in the famous passage from Genesis 3.

From these foundations (and many others), Christian social thought, now incorporating integral ecology which makes links between their different aspects, offers many supports for action. To cite just a few texts from the last two popes, we highlight:

> «Technology is part of the mission to cultivate and keep the earth that God entrusted to humanity, and it should aim to strengthen the covenant between humans and the environment, which is called to reflect God's creative love. Technological development can lead to the belief that technology is self-sufficient, when humans, by questioning only the *how*, fail to consider the *why* behind their actions. This is why technology can take on ambiguous traits. Born from human creativity as an instrument of personal

[10] A. Gesché, *Dieu pour penser. IV, Le cosmos*, Cerf, 1994, pp. 126-127.
[11] *Ibid, p.197.*
[12] F. Euvé, *Penser la création comme jeu*, Cerf, Cogitation Fidei, 2000, p.319.

ISSN 0039-4130

> freedom, it can be understood as an element of absolute freedom, freedom that seeks to break free from the limits inherent in things themselves» [13].

And pope Francis continues :

> «It is not enough to reconcile, in a middle way, the protection of nature and financial profit, or the preservation of the environment and progress… It is necessary to redefine progress. Technological and economic development that does not leave behind a better world and an integrally superior quality of life cannot be considered as a true progress» [14].

This progress is not confused with the economic growth, the increase in technological power, the accumulation of material wealth, or the increase in GDP, although these factors are not neglected. It seeks to take into account, in the same movement, "the cry of the earth and the cry of the poors."

Pope Francis adds:

> «Technoscience, if properly directed, can produce truly valuable things to improve the quality of human life (LS 103)… But we cannot ignore that nuclear energy, biotechnology, computing, our knowledge of our own DNA and other capabilities we have acquired give us terrible power. More precisely, they give those who have the knowledge—and especially the economic power to use it—a grip on all of humanity. Never before has humanity had so much power over itself, and nothing guarantees that it will always use it well» (LS 104).

In terms of the *common good*, it is emphasized that the sharing of digital data, especially health data, greatly contributes to it, provided there is adequate protection of personal information. A good example of this is the recent "Mon espace santé" (My Health Space) in France, which facilitates medical monitoring and is built around the patient's urgent health needs, care while abroad, and the retirement of the primary care physician.

The key questions are as follows: Is digital technology truly serving the "we-all"? For instance, is the management of big data in medicine accessible to everyone? What about the digital divide and the corresponding universal destination of goods? Will this development be an instrument for humanity's growing emancipation or an increas-

[13] Benoît XVI, *Deus Caritas Est*, 2005, N° 69-70.
[14] Francis, *Laudato Si*, 2015, LS 194.

ISSN 0039-4130

ingly powerful alienation, even from natural resources? Are we taking measures to ensure that it is not always humanity that must adapt to machines, but rather machines that adapt to humans, especially in the digitalization of administrative processes today and in the future?

In terms of *solidarity*, *subsidiarity* and *participation*, we think about the accessibility of digital platforms. The development of the so-called platform economy, made possible by digital applications, can foster collaboration and solidarity. However, regarding the ethics of decision-making and responsibility, no machine should take the place of the person who is responsible at their level of competence (subsidiarity). Moreover, in a different field falling within human dignity and solidarity, let's not also forget the ethical questions related to the work of "little hands" who correct machine training errors during the development of artificial intelligence systems with machine learning.

Finally, we may ask whether, in this development, the participation of each individual in democratic and ecological life is ensured. Will the care for the most vulnerable, and what is most fragile in the common home, be at the heart of the societal project driven by this development?

In terms of *integral ecology*, digital technology can support efforts to reduce energy consumption, such as through smart grids—intelligent electrical networks that allow real-time optimization of electricity distribution and consumption. However, it is known that data centers are highly energy-intensive. Furthermore, the conditions of supply for rare earth elements and resources necessary to manufacture digital system components raise questions about equity between countries and global economic and political issues**.**

One can also consider the considerable acceleration of industrial processes driven by digital control systems, which operate at a scale far beyond natural evolution. This raises the question of adaptation time. As is well known, the creative time of humans, such as that of a painter or a researcher, cannot be programmed ! There is always a time for maturation, which cannot be set like a clock ! The time for discernment of the co-creator human, the freedom in which this discernment takes place in society, are essential factors to **"**choose freely**"**. This requires the time of wisdom, of free will, which cannot naturally take place under the aegis of the techno-economic paradigm, which aims solely at technical and financial performance.

In terms of *human dignity*, we can mention the concrete issues of harassment on social media**,** cyberattacks**,** respect for personal data

ISSN 0039-4130

and e-reputation. Will the search for truth**,** freedom and justice be preserved, and even strengthened, as fundamental values in social, political and ecological life? But also, will the human being become a "digital dossier" or will he be recognized in all his components (body-soul-spirit)? To elaborate on this point, it is necessary to clarify the specificities of the human being in relation to intelligent machines**.** Respecting human dignity in the age of the digital revolution also means being able to clearly articulate the differences between human intelligence and AI. This is the goal of the next paragraphs.

5. *From Artificial Intelligence to Human Intelligence: first approach*

The machine is built on codable information, with a very large capacity for computation and learning. Mathematics and digital systems work on measurable, quantifiable data, using invariants to establish laws. But humans do not live solely in the realm of the quantifiable, and their singularity does not depend only on invariants! Loving, practicing the arts, meditating, experiencing wonder and following a path of faith are largely beyond the primacy of the quantifiable.

Asking ChatGPT about its « inner life », the answer was « quite honest»:

> «*Artificial intelligence is a technology that enables machines to make decisions based on algorithms and data. It does not have its own inner life as a human does, lacking consciousness or emotions. Current AI systems do not have the ability to feel or perceive things in the same way humans do. However, some AI researchers are seeking to develop systems that could mimic certain aspects of human inner life...But it is important to remember that even with these advancements, AI will always remain a technology created by humans, and its functioning will fundamentally differ from that of human inner life*».

Moreover, mathematics cannot self-found, as Gödel's theorem shows. In mathematics, as in all so-called hard sciences, there is always something that escapes, something that pertains to the foundations [15]. Mathematics and digital systems have their limits in describing the complexity of physical, biological and human reality. Not all of humanity is "digitizable". Something fundamental always escapes formalization.

[15] T. Magnin, *L'expérience de l'incomplétude*, Lethielleux-DDB, 2011.

ISSN 0039-4130

Furthermore, the machine has no body, no history, no conscious experience, unlike humans who live in essential connections between their bodily emotions, mind and reason, as shown today by neuroscientists [16]. Damasio recounts the story of Phineas Gage, a man who had a metal rod pass through his brain and survived, seemingly with all his capabilities. However, it turned out that, due to his accident, he lost his emotional bearings and the memory of his experiences. Damasio demonstrates how these losses combined to bring about a change in Phineas Gage's character and a loss of reference points when making decisions. "Somatic markers" indeed come into play when we consider different options before making a decision; they were no longer present in Phineas Gage. As a result, he lost part of his ability to act according to the social norms previously learned by his brain. Fundamental connections between his emotions and his reason were thus broken.

This is illustrated by the testimony of Professor Didier Vernay, a neurologist at the University Hospital of Clermont-Ferrand in France, who works with prisoners, on the links between the body, sensations, emotions and cognition:

> «In prison, touch is forced, it is experienced as an aggression. This deprivation of the senses has a cognitive impact. Without realizing it, we decode faces every day, we assess the space around us. This occurs at the level of the amygdala-hippocampus complex. Without modulation, that is, without this daily exercise, brain plasticity deteriorates, and our relationship to the world becomes disturbed» [17].

These experiences, and many others, highlight the connections between the "felt" anchored in biology, the feelings and emotions it provokes, and their interactions with mental processes, with the human mind giving words to these sensations. The brain creates neural maps of our bodily feelings, and our "feelings" or sensations are read, with emotions coming to consciousness and interacting with thought [18]. Thus, humans think with their whole body, unlike the machine which has no body to feel or "experience". The learning machine, on the other hand, has a form of computational intelligence without a body or consciousness, fundamentally different from human intelligence. This aligns with the work of educational sciences, which outline the various

[16] A. Damasio, *Sentir et savoir. Une nouvelle théorie de la conscience*, Odile Jacob, 2021
[17] Journal La Croix, June 23, 2020.
[18] P. Damier, *Décider en toute connaissance de soi*, Odile Jacob, 2014.

ISSN 0039-4130

forms of intelligence in humans, including emotional, hypothetico-deductive, and existential intelligences, in articulation.

With the model of the intelligent machine, the human body is reduced to mere information, in a sense, disembodied. It can be repaired and enhanced like a reprogrammable machine. This can lead to a denial of biological determinism and biological time. Let us remember that AI today operates in the realm of simulation. And there is a threshold between "simulating" an emotion and actually experiencing it.

It is, however, interesting to highlight that AI brings back the question of human intelligence and consciousness today… in complementarity with neurosciences. Humans are much more complex than machines. One could say they are embodied in all their dimensions, as we will continue to show by presenting a comparison between the latest discoveries in the life sciences and one of the Christian anthropological traditions of humans as "body-soul-spirit". Key elements of human specificity and dignity will thus continue to be highlighted.

6. *Introduction to some elements of biblical anthropology*

The Hebrew Bible uses three main words to describe the human being: *basar*, *nefesh* and *ruah*. *Basar* is flesh, not only in the biological sense but also in that of the whole of Man, animated and vitalized by the *nefesh* that sets it in motion (animation). What gives consistency to this *nefesh-basar* couple is the *ruah*, the breath of life given by God, as we read in Gn 2:7: 'Yahweh God shaped man from the soil of the ground and blew the breath of life into his nostrils, and man became a living being. Human beings are created in the image of God, male and female, as in his likeness (Gen 1, 26-27). As the Pontifical Biblical Commission says [19], «*In the flesh, the human being lives this spiritual experience that characterises him among all other living beings*».

The Greek word for flesh (*sarx*) refers to our most material, concrete and physical humanity (flesh feels, smells and senses). *Sarx* refers to a living body (in the biological sense of the term), but also, more broadly, to the way in which this body is embedded in the reality of the world. Flesh is an element of the whole system that is the body (*soma*).

Again in Greek, *ruah* is translated by *Pneuma*, the Breath, the Spirit that gives life to man and speaks to his spirit (a kind of "fine point of the

[19] *What is Man*, Pontifical Biblical Commission, Vatican website, 2019, ref.4, Q 19, p. 31.

soul", a little *pneuma* in Greek in St Paul, for example). In Romans 8, 16, it is the Holy Spirit himself (the great *Pneuma*) who testifies to our spirit (the little *pneuma* of man) that we are children of God. We can see that the human spirit is marked by the divine Pneuma and is distinct from the *psyche*, the ego and the intellect (*noos*). The Pneuma also flows through human reason, as it does through the whole body. We might say that the *noos* corresponds to the rational soul of Aristotle, who also speaks of the vegetative soul and the sensitive soul. The *noos* makes it possible to read what happens in the deep consciousness ("*syneidesis*") of man. In this sense, the apostle Paul calls the *noos* the Inner Man; it is open to Life (*zoe*) while being linked to life (*bios*). For human beings on earth, let us remember that there is no Life (*zoe*) without life (*bios*), which is its condition. We could therefore say that the pneuma is a "dynamic vector" of the *noos*, with which it communicates.

The distinction between the psychic and spiritual dimensions is a complex one, because these two dimensions interpenetrate. The human experience of forgiveness can illustrate this distinction. It is one thing to desire reconciliation with someone at the level of the *psyche*; it is quite another to welcome an 'inner impulse' (of the spiritual order) that enables us to give or receive forgiveness. In this anthropological context, it should be noted that every human being has a spiritual dimension, whether or not they identify with a religion. The experience of palliative care staff accompanying people at the end of life is significant in this respect (for more details on this subject, see [20]).

The 'heart' (*kardia* in Greek, *leb* in Hebrew) is a conscious, lucid, subjective instance hidden in the innermost depths, where God reaches out and meets man, in the silence of deep consciousness. It lies at the very source of conscious will and desire. «Each one should give as much as he has decided on his own initiative, not reluctantly or under compulsion, for God loves a cheerful giver» (2 Cor 9, 7). It is from this heart that the love of charity, *agape*, is expressed. It is from this heart that faith is lived, as Saint Paul says in Romans 10,10: «He who believes from the heart becomes righteous».

God 'speaks to the heart,' as the Scriptures say, and «God is sensitive to the heart and not to reason; this is known in a thousand ways» [21], as Pascal says in one of his thoughts that continues to be debated. Indeed, the 'reasons of the heart' often surprise *noos* and its logic! There

[20] T. Magnin, *Essai sur la dimension spirituelle de l'être humain corps-psyché-esprit*, *MSR*, December 2021, pp. 83-94.

[21] B. Pascal, *Pensées* §277, Brunschvicg ed.

ISSN 0039-4130

are so many witnesses to this. As mentioned above, we can live fully in the Spirit without the *noos* making it explicit. And the noos can both decipher what comes from the heart if it allows itself to be permeated by the *Pneuma*, but also go astray if it is impervious to the *Pneuma*.

Man becomes fully human when he is penetrated by the grace of the Spirit (Life, *zoé*), by the "pneumatisation" of his whole being, his flesh, his *psyche*, his deepest consciousness and his spirit. Speaking of the importance of the Word of God in discernment, the epistle to the Hebrews says this: «The word of God is something alive and active: it cuts more incisively than any two-edged sword: it can seek out the place where soul is divided from spirit, or joints from marrow; it can pass judgement on secret emotions and thoughts» (Heb 4, 12).

7. *Overview of the 'body-psychè-spirit' ternary anthropological vision*

Known as one of the anthropological traditions of Christian history, it has always held an important place among Eastern Christians and is increasingly relevant in the West, which is marked by a dualist tradition (for reviews, see for instance [22]). The unity of body, soul (*psyche*) and spirit is set out by Paul of Tarsus: «May the God of peace make you perfect and holy; and may your spirit, psyche and body be kept blameless for the coming of our Lord Jesus Christ». (I Th 5, 23). We are not talking about three independent realities, but about interactions and interpenetration.

This is what St Irenaeus, the second bishop of Lyon, who came from the Near East and became a Westerner in the second century, said so admirably [23]: «The shaped (by the Creator) clay of the flesh [*carnis* - σαρκὸς] is not, on its own, a complete [*perfectus* - τέλειος] man, but just the body of a man [*corpus*, σῶμα], a part of a man. Similarly, the soul [*anima* - ψυχὴ], on its own, is not a man, but just the soul of a man, a part of a man. And the spirit (the breath) [*spiritus* - πνεῦμα] is not a man: Spirit is called "spirit", not "man". It is, then, the mixture [*commixio,* σύγκρασις] and union of all these things which makes the complete man». In the same way, Irenaeus says: «Every man will confess that we are a body drawn from the earth and a soul that receives its spirit from God» [24].

[22] M. Fromaget, *L'homme tridimensionnel, corps-âme-esprit*, Albin Michel, 1996 ; P. Dautais, *Le chemin de l'homme selon la Bible*, Paris, Desclée de Brouwer, 2009.

[23] St Ireneaus, *Adversus Haereses*, A.H. V 6,32-40.

[24] *Ibid.*, AH, V 6.1.

ISSN 0039-4130

It is indeed a question of distinguishing between body, mind and spirit, but within a close link between these three dimensions. This link is expressed in terms of mixing in St Irenaeus, *commixio* in the Latin version but *krasis* (interpenetration, entanglement) in the Greek fragments. The spiritual man is not disembodied, and the human body is not an obstacle to his perfection. On the contrary, it is the temple of the Spirit and the place of encounter with God. This encounter between man and God culminates in the incarnation of the Son. God himself becomes flesh... St Justin wrote in 160: «The body, then, is the house of the soul, just as the soul is the house of the spirit, it it those three that are saved» [25].

So we can say that the body is more than just biological; it thinks and feels. As we know, there is no brain without a body! The soul (*psyche*) is a principle of organisation and animation that unifies biological metabolisms, emotions and thoughts. It therefore covers emotions, affect, intelligence, will and desire, all of which are often condensed under the term "*psyche*". Finally, the spirit of man corresponds to the very tip of the soul, where the Holy Spirit speaks to man. In today's terms, we would say that the mind and, more generally, the *psyche*, are in the soul, and that through meditation, for example, our spirit (*pneuma*) can open up to the Spirit, as in experiences of "inner seizure". So there is a place within us where we are open to God, some would say to the transcendent, to the infinite, others would say to the "inner voice", the voice of consciousness.

8. *In surprising resonance with today's life sciences*

Life sciences work on the complexity of livings, through the interplay of parameters and systems, including a degree of indeterminism and unpredictability in evolution [26]. The study of epigenetic effects is a good example of this. It highlights the influence of the environment (biological and psychological) on gene expression. The same is true in neurosciences, with the effects of cerebral plasticity showing how neurons and synapses evolve throughout life, depending in particular on the environment and life experiences. Living organisms are part of ecosystems that modify them; they are plastic and, thanks to this plas-

[25] Saint Justin Martyr, *De Resurrectio, cited by Eusèbe de Césarée, Histoire Ecclésiastique,* 4, 18, SC 31*,* 1952

[26] E. Angelier, Les sciences de la complexité et le vivant, Lavoisier, 2009.

ISSN 0039-4130

ticity, can adapt and evolve—in short, "be alive"! The ability of living organisms to allow themselves to be modified by their environment corresponds to their vulnerability.

Epigenetics is the modulation of gene expression by the biological and psychological environment.[27] A significant example is the effect of a mother's stress on the foetus she has in her. Most epigenetic markers are put in place during pregnancy. This is the context in which researchers are studying the effect of psychological factors, such as the mother's stress, on epigenetic modification, which will have consequences for the expression of the foetus's genes, and could possibly favour the subsequent appearance of a pathology. From the foetus onwards, the influence of the psychological environment can have a biological effect, and the studied phenomena are generally multifactorial.

It has been shown that stress management, pleasure and the social network can influence the mechanisms of epigenesis in humans, underlining the extent to which the two domains of the biological and the psychological are in a permanent reciprocal relationship.[28] This biology-psychology interaction goes hand in hand with the plasticity of living organisms.

The neurosciences are now working on this plasticity in the same way, showing a reciprocal relationship between cerebral functions and the individual's experience, particularly psychological (see[29] for example). Over the last ten years or so, the practice of meditation has been the subject of increasingly detailed scientific analysis. The dossier in the magazine *Science et avenir*, entitled *The benefits of meditation and self-control*, gives some significant examples[30]. Meditation has become a subject of study, fascinating a growing number of neuroscience researchers. Recent research, says the dossier, confirms the benefits of practising meditation, which also regulates attention and emotions. Better still, they show, it acts at the very heart of our cells, being able to modify the expression of genes involved in metabolism and cell ageing.

According to the neuroscientist researchers, meditation stimulates the genesis of neuronal branches and connections (synapses). It also

[27] D. Bourc'his, Inserm 934/CNRS UMR 3215/Université Pierre et Marie Curie, Institut Curie in Paris, February 2015.

[28] See the 4th chapter in T. Magnin, *Penser l'humain au temps de l'homme augmenté*, Albin-Michel, 2017.

[29] PM. Lledo, *Le cerveau, l'homme et la machine*, Odile Jacob, Paris, 2017.

[30] *Science et Avenir*, January. 2020, Number 875, pp. 28-38.

ISSN 0039-4130

reduces stress level, which is harmful to neurons. It is also thought to provoke epigenetic changes and have an impact on the immune system. This has led to the development of a new discipline, the psycho-neuro-immunology, which studies the impact of psychic and spiritual events on the immune system. The aim of this science is to show the links that unite disciplines as different as psychology, neurology, immunology and endocrinology, or more simply, the links that unite the psyche and the body. In biological analyses, certain immunogenetic trends illustrate a close link between stress and immunity.

Such studies need to be continued and refined, and we have to remain cautious about the explanations. The fact remains that, following on from biology-psychology interactions, we can now use neuroscience techniques to analyse biology-psychology-spirituality interactions, where these different dimensions are intertwined. So, to be a good human biologist or neuroscientist, one can't ignore the effects of the "biological, psychic and spiritual environment" in terms of brain plasticity, epigenetics and vulnerability. The brain is a plastic organ that can be affected by our experiences, including psychic and spiritual experiences (and vice versa).

Plasticity is thus an essential characteristic of living things[31]. This plasticity characterises a dynamic tension between "robustness and vulnerability", between rigidity and malleability, between invariance and transformation and, more broadly still, between invariance and historicity. It is a necessary and crucial condition for living beings to evolve, with their metabolic, reproductive, organisational and informational characteristics.

Every living being has a structure that ensures its coherence and a kind of functional unity, the guarantees of a "robustness" that enables it to retain a certain invariance over time. The robustness of a living thus defines its ability to maintain itself in the face of disturbances linked to its environment. At the same time, however, each living being is able to be influenced by the effects of the external environment, thanks to the "reception structures" for these external influences (vulnerability). Plasticity and adaptability are thus two essential characteristics of living organisms. In this sense, we can apply the adjective "vulnerable" to malleable living organisms, independently of any fragility linked to disease or deficiency. However, this tension between robustness and vulnerability can lead to a certain "fragility" in living beings.

[31] D. Lambert & R. Rezsöhazy, *Comment les pattes viennent au serpent, essai sur l'étonnante plasticité du vivant,* Paris, Flammarion, 2004.

ISSN 0039-4130

In this dynamic tension, robustness must not be forgotten, but too much robustness can be detrimental to a living's capacity to evolve and adapt. In this sense, the "invulnerable cyborg", that some transhumanists are hoping for, loses its capacity to adapt by losing the vulnerability that all living beings need in order to evolve. Respect for livings, including attempts to increase their capacities, can only really be worked out in the light of their complexity, their plasticity, and the balance between robustness and vulnerability. Wanting to eradicate human vulnerability inevitably leads to dehumanization.

For humans, this vulnerability is linked in particular to the reciprocal interactions between "biology-psychology-spirituality" in their ecosystems, as we have just illustrated. Respecting and caring for human beings will therefore mean encouraging this plasticity and the dynamic balance between robustness and vulnerability, by enabling harmony between body, mind and spirit in their environments.

This surprising resonance between the point of view of life sciences today and the biblical tripartite anthropology is of prior importance today.

9. *When Christian anthropology dialogs with Life sciences in the Age of Intelligent Machines*

Biology teaches us that living beings are both robust and vulnerable, possessing plasticity. Vulnerability here is defined as the ability to be affected and modified from within by the environment, while also contributing to shaping it. If living organisms are not robust enough, they deteriorate. If they are too robust and not vulnerable enough (the vision of the invulnerable cyborg), they lose their plasticity and resilience, which are key aspects of their uniqueness. This is what epigenetics and brain plasticity demonstrate today.

The capacity of living beings (humans in particular) to be modified by their biological and psychic environment is a concrete expression of their vulnerability and complexity. For humans, the reciprocal relationships between biology and psyche on one hand, and meditation and brain plasticity on the other, highlight the "entanglement" between the biological, psycho-social, and spiritual dimensions of a person in their ecosystems. What is new here is that biologists themselves are increasingly required to take into account the psycho-social and spiritual dimensions (in a broad sense) in their work. Human experience influences the biological, and vice versa.

ISSN 0039-4130

While digital technosciences focus solely on the functions of living beings, on the functionalities of machines (taking the machine as the model), biology reveals that life is far more complex than mere "machinery." The human-machine fusion, in a way, erases the uniqueness of the human, even if it allows for the enhancement of certain functionalities. Here too, the incarnation and the role of the body in the human experience are emphasized, with consequences for cognition and the capacity for reflective consciousness.

This connects with contemporary reflections in virtue ethics on the link between a person's actions and the formation of their moral character [32]. While AI machines are capable of learning from their operations (deep learning), their learning mechanisms (and their memory) differ greatly from humans. There is no brain plasticity in these machines, no feelings or emotions experienced within the body, no lived experience or conscious thought—only the storage of measurable and quantifiable information, which does not exhaust the human experience and cannot match a singular experience. Therefore, we cannot speak of the moral behavior of an "intelligent" machine, even if it demonstrates a certain "autonomy".

However, there is a remarkable and significant "resonance" between what biology and neuroscience tell us about the complex nature of life, particularly human life, and the presentation of a human as "body-soul/psyche-spirit" in one of the great Christian tripartite anthropological traditions. This is a comparison and resonance, without confusing biology and Christian anthropology, as each discipline has its own field and specific methods.

Conclusion : Human Dignity in the face of so-called Intelligent Machines

Human beings, as co-creators, now possess a new tool in digital technosciences that influences the future of nature as well as the position of humanity within a reconfigured natural world. Why not? To achieve this, humans utilize "evolution by design", where NBIC technologies and AI work in synergy. But with what wisdom does humanity make

[32] D. Wong, D. (2009): *Emotion and the cognition of moral motivation*, Philosophical Issues, 2009, 19, pp. 343-367; WC. Spohn, *Jésus et l'éthique. «Va et fais de même !»* Trad. de l'anglais (Etats-Unis) par L. David. Lessius, 2010 ; M. Boos, *La philosophie morale d'Alasdair MacIntyre, une défense historiciste de l'impératif catégorique*? Revue d'éthique et de théologie morale, 4, 304, 2019, pp. 45-58.

ISSN 0039-4130

her choices in this domain? How much control do we have, and how can we regulate it? With what kind of freedom in relation to intelligent machines, what ethics, on what foundations, with what sense of benevolence toward all creatures, and with what openness to transcendence and understanding of time to adapt to the "social" rythm?

Humanity is called to learn to live with so-called intelligent machines, whose certain capacities already surpass (and will increasingly surpass) those of humans in many fields. However, this adaptability to a transitioning world precisely calls upon humanity's own unique qualities: its personal and collective resilience, which is rooted in a harmonious interaction between the bodily, psycho-social and spiritual dimensions. This represents a transcendent capacity that feeds into the pursuit of the common good.

While it is essential to remember that there is no AI without human intelligence, we must avoid conflating these two types of intelligence. Respect for the dignity of every human being involves recognizing the specificities of human intelligence and understanding that today's learning machines, while extremely useful in many cases to support techno-economic development within the framework of integral ecology, have no body, no emotions other than simulated ones, and no consciousness. We must also be mindful of the mindsets generated by ICTs, generative AI, and NBIC technologies in our relationship with reality and with humans.

Great religious and wisdom traditions offer rich sources of life (*zoe*) that can contribute to our shared humanity, helping us find meaning amidst the current digital transition. These traditions can guide us in learning how to live freely with the intelligent machines that humanity is building. In this sense, Christian anthropology and Christian social thought need to be studied and made available to society as it seeks ethical and anthropological reference points, especially when choosing the direction for the development of digital technosciences.

This is expressed in the "Rome Call for AI Ethics" [33] which states:

> «New technologies must be researched and developed in accordance with the criterion that they serve the whole human family (as per the preamble of the Universal Declaration of Human Rights), respecting the inherent dignity of each of its members and of all natural environments, while also considering the needs of the most vulnerable».

[33] Vatican, February, 28, 2020.

ISSN 0039-4130

The tripartite anthropology is thus highly relevant to education today, particularly for the discernment for decisions. It connects different steps such as : learning to 'feel' and put words to inner movements and emotions, learning to reflect by stepping back without forgetting the relationship between emotion and reason, learning to discern from one's 'heart' in the sense previously defined, where transcendence speaks, learning to decide with knowledge of oneself and the environment, welcoming an inner impulse to act in perseverance and the rereading of experience.

It is also on the basis of the anthropological foundation that ethics can be applied to AI but also to a wide range of fields, including medicine, economics, science and technoscience.

ISSN 0039-4130

New Humanism in the Age of Artificial Intelligence: A Theodaoian Reflection

Heup Young KIM

Introduction

In the 21st century's rapidly evolving landscape, artificial intelligence (AI) has emerged as a transformative force, reshaping industries, societal structures, and individual lives at an unprecedented speed and scale. From automating routine tasks to pioneering new frontiers in healthcare, education, and environmental conservation, AI's influence permeates every facet of contemporary life, heralding a new era of innovation and challenge. However, as this technological revolution unfolds, it also raises profound questions about the nature of humanism, ethics, and the very essence of human agency in a world increasingly mediated by machine intelligence [1].

It is within this context that "New Humanism at the Time of Artificial Intelligence: A Theodaoian Reflection" seeks to explore and articulate a vision of humanism that is both responsive to and reflective of the unique challenges and opportunities presented by the age of AI. This exploration is anchored in the intriguing insights of Theo-Dao (Theology of Dao), a theological paradigm that emerges from the confluence of Daoism and Confucianism, two of East Asia's most enduring and influential systems of thought [2]. These traditions offer a reservoir of wisdom (Dao) focused on harmony, balance, and the ethical cultivation of self and society, providing a timely counterpoint to the prevailing

[1] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University, Oxford 2014.

[2] H. Y. Kim, *A Theology of Dao*, Orbis, Maryknol 2017.

narratives of technological determinism and the mechanization of human life[3].

Theodao, with its emphasis on Dao (interconnected whole), Taiji (relational harmony), and wuwei (non-intentional action), offers a framework for understanding and engaging with the world that transcends the binary oppositions of human and machine, nature and technology[4]. At its core, it advocates for a holistic vision of human existence, one that recognizes the mutual dependencies between humans and their environments, both natural and constructed. By drawing on the virtues espoused by Confucianism—ren (benevolence), yi (righteousness), li (ritual propriety), zhi (wisdom), and xin (faithfulness)—alongside the Daoian appreciation for the fluid and dynamic nature of reality, Theodao provides a robust foundation for reimagining humanism in an era dominated by digital technologies.

As we venture deeper into the age of artificial intelligence, the need for a new humanism—a Theodaoian reflection—becomes ever more pressing. This paper aims to chart a course through the complexities of modern technological society, advocating for a model of humanism that not only embraces the transformative potential of AI but also reaffirms the fundamental values and virtues that sustain a just and flourishing human community.

Critique of the Technocratic Paradigm

In his encyclical *Laudato Si'*, Pope Francis offers a powerful critique of the technocratic paradigm that has come to dominate contemporary society[5]. This paradigm, characterized by an unrelenting faith in technological progress and an instrumental view of nature, places a premium on efficiency, productivity, and economic growth, often at the ex-

[3] As a widely used root metaphor in all classical East Asian religions—including Confucianism, Daoism, and Buddhism—Dao [Tao] (the Way, *do* in Korean) is a highly inclusive term with various meanings. For example, «[D]ao is a way, a path, a road, and by common metaphorical extension, it becomes in ancient China the right way of life, the way of governing, the ideal way of human existence, the way of the cosmos, the generative-normative way (pattern, path, course) of existence as such» (H. Fingarette, *Confucius: The Secular as Sacred*, Harper & Row, New York 1972, p. 19). However, for the sake of discussion, this paper adopts a narrower definition, focusing on the concept of the interconnected whole.

[4] R. T. Ames and D. L. Hall, *Dao De Jing: A Philosophical Translation*, Ballantine, New York 2003.

[5] Pope Francis, *Laudato Si', On Care for Our Common Home*, Libreria Editrice Vaticana, Vatican 2015.

ISSN 0039-4130

pense of environmental sustainability and ethical responsibilities. Pope Francis warns of the inherent dangers in a worldview that positions technological advancement as the ultimate solution to all human problems, neglecting the deeper ethical, social, and spiritual dimensions of human existence.

The technocratic paradigm assumes that technological power can and should be used to control and manipulate the natural world, transforming it into a resource for human consumption and convenience. This utilitarian approach not only leads to widespread environmental degradation but also fosters a societal model in which human relationships and ethical considerations are increasingly mediated by technological systems. The encyclical highlights how this paradigm encourages a culture of disposability, where both material goods and human lives are valued only insofar as they contribute to economic productivity or technological efficiency.

Pope Francis's critique is not an outright rejection of technology but rather a call for a critical reassessment of how it is developed and used. He advocates for an "integral ecology" that acknowledges the interdependence of all living beings and respects the intrinsic value of the natural world. This perspective challenges the technocratic paradigm by insisting that technological advancement should not come at the expense of environmental health or moral integrity but rather be aligned with principles of justice, sustainability, and the common good[6].

In today's society, the prioritization of technological advancement over environmental and moral considerations is evident in various areas, ranging from the relentless exploitation of natural resources to the deployment of AI and big data in ways that threaten privacy and human dignity. The technocratic paradigm often masks the potential harm of unchecked technological progress, promoting an illusion of neutrality that conceals the embedded values and interests driving technological development[7].

However, it is important to note that not all Western ethical traditions align with utilitarianism. For instance, Immanuel Kant's deontological ethics propose a categorical imperative that emphasizes moral duty and respect for individual autonomy, regardless of outcomes[8].

[6] C. Deane-Drummond, *Theological Ethics for a Technological Age*, in *Theology and Science* XVII, 1, 2019, pp. 89-102.

[7] Langdon Winner, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*, MIT Press, Cambridge MA 1977.

[8] I. Kant, *Groundwork of the Metaphysics of Morals* (1785), M. Gregor (translated by), Cambridge University Press, Cambridge 2012.

ISSN 0039-4130

Theo-Daoian Resonance: A Holistic Ethical Vision

Pope Francis's critique resonates profoundly with the insights of Theo-Dao, which also challenges the technocratic paradigm by advocating for a more harmonious and integrated approach to technology and nature. Eco-Dao, an ecological extension of Theodao, emphasizes the interconnectedness of all things and the need for human actions to align with the natural rhythms of the universe[9]. This perspective mirrors Francis's call for an "integral ecology" and offers a deeper spiritual and ethical foundation for recalibrating our relationship with technology.

In Daoian thought, wuwei (non-coercive action) encourages living in accordance with the natural flow of life, rather than attempting to control or dominate it. Similarly, Taiji (the balance of yin and yang) emphasizes the importance of dynamic equilibrium, where harmony is achieved through the balance of opposing forces, not through exploitation or domination. These Daoian principles align with Francis's warning against the unchecked use of technology to control nature and commodify human life. Instead, they call for an approach to technology that fosters balance, respect, and mutual enhancement between humanity, technology, and nature.

Moreover, the Confucian virtues, such as benevolence, righteousness, wisdom, and trustworthiness, further reinforce the idea that technology must serve the collective well-being, not just economic gain. A Theodaoian vision insists that technology should enhance human flourishing while respecting the intrinsic value of all life forms. This ethical vision not only addresses the shortcomings of the technocratic paradigm but also provides a more holistic framework for integrating technological advancement with ecological stewardship and moral integrity.

A Path Forward: Reimagining Humanism in the Age of AI

Pope Francis's critique, when viewed through the lens of Theodao, invites us to envision a path forward that embraces technological innovation without sacrificing ethical principles or environmental responsibilities. This alternative path is one in which technology, rather than

[9] H. Y. Kim, *Eco-Dao: An Ecological Theology of Dao*, in L. Hobgood and W. Bauman (edited by), *The Bloomsbury Handbook of Religion and Nature: The Elements*, Bloomsbury, London 2018, pp. 99-108.

ISSN 0039-4130

being a tool for domination, becomes a means of enhancing harmony, both within society and with the natural world.

The shortcomings of the technocratic paradigm underscore the urgent need to rethink our relationship with technology and its role in shaping the future. By integrating the insights of Theodao, we can cultivate a humanism that is not only compatible with technological advancement but also deeply rooted in a commitment to environmental stewardship, relational harmony, and ethical integrity. This recalibrated humanism would prioritize balance, interconnectedness, and respect for the intrinsic value of all beings, ensuring that technology serves the common good rather than perpetuating environmental degradation or social injustice.

In conclusion, both Pope Francis and Theodao challenge us to rethink the way we approach technological progress. Their shared critique highlights the importance of aligning technological development with ethical and ecological principles, fostering a future in which technology supports rather than undermines the well-being of all life on Earth. Through this integrated, holistic vision, we can move beyond the limitations of the technocratic paradigm and toward a more balanced and harmonious relationship between humanity, technology, and the natural world.

Post-humanism and Transhumanism: Critique and the Call for Inclusive Humanism

In contemporary discourse on the future of humanity, post-humanism and transhumanism emerge as two influential yet distinct philosophical movements that challenge traditional humanism in light of technological advancements and ecological crises. Despite their innovative approaches, both remain largely rooted in exclusive humanism, an Enlightenment mentality that continues to frame humanity's relationship with technology and nature in ways that diverge significantly from the Confucian inclusive humanism of East Asia. A celebrated contemporary Confucian humanist, Tu Wei-ming, critiques the Enlightenment mentality, stating:

> «[T]he Enlightenment mentality, fueled by the Faustian drive to explore, to know, to conquer, and to subdue, persisted as the reigning ideology of the modern West … However, a realistic appraisal of the Enlightenment

ISSN 0039-4130

> mentality reveals many faces of the modern West incongruent with the image of 'the Age of Reason.' In the context of modern Western hegemonic discourse, progress may entail inequality, reason, self-interest, and individual greed.» [10]

Post-humanism critically examines classical humanism's anthropocentric assumptions, questioning the sustainability of human dominance over nature, especially amid global ecological crises. It advocates for decentering humans in ethical and philosophical frameworks, urging consideration of the rights and intrinsic values of non-human entities and the environment. Post-humanism argues that environmental degradation and species extinction are direct consequences of an outdated humanism that prioritizes human desires over the well-being of other life forms. By emphasizing interconnectedness among all life, posthumanism calls for an ethical rethinking of humanity's place in the natural world [11].

However, despite its critique of anthropocentrism, posthumanism remains tied to the Enlightenment's exclusive humanism. This humanism seeks to correct humanity's flawed dominance over nature by advocating for a reimagined role for humans within ecological systems. It remains grounded in the individualistic and rationalistic paradigms of the Enlightenment, often neglecting the relational, virtue-based approaches to humanism seen in Confucian thought.

Transhumanism, on the other hand, presents a more techno-optimistic vision, championing the use of technology to enhance human physical and cognitive abilities, ultimately transcending the limitations of the human body and lifespan [12]. Transhumanists advocate for the use of biotechnology, AI, and cybernetics to achieve a post-human future where humans evolve into a new species or coexist with advanced AI entities [13]. While this vision emphasizes human progress through technological enhancement, it often inherits the Enlightenment's quest for

[10] See Tu Wei-ming, *Beyond the Enlightenment Mentality*, in M. E. Tucker and J. Berthrong (edited by), *Confucianism and Ecology: The Interrelation of Heaven, Earth, and Humans*, Harvard University, Cambridge 1998, p. 4.

[11] R. Braidotti, *The Posthuman*, Polity, Cambridge 2013.

[12] N. Bostrom, A History of Transhumanist Thought, *Journal of Evolution and Technology* XIV, 1, 2005, pp. 1-25.

[13] K. Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, University of Chicago, Chicago 1999.

ISSN 0039-4130

mastery over both human nature and the environment [14].

Despite its optimism, transhumanism shares with post-humanism the limitations of Enlightenment-based exclusive humanism. The relentless pursuit of enhancement and perfection risks devaluing the inherent worth of unenhanced human life and overlooks the richness of human diversity and imperfection. Moreover, the transhumanist emphasis on surpassing human limitations perpetuates the Enlightenment focus on individual achievement, often at the expense of communal well-being and ecological harmony [15].

The Need for Confucian Inclusive Humanism

Both post-humanism and transhumanism highlight the need for a new paradigm of humanism that responds to the ethical and ecological challenges of the modern life. However, this new humanism must go beyond the limits of Enlightenment-based exclusive humanism and embrace a more inclusive humanism that recognizes the interconnectedness of all life and the intrinsic value of nature, as seen in East Asian philosophical traditions.

Confucian inclusive humanism offers a robust alternative. Rooted in the relational and virtue-based framework of Confucian thought, this approach emphasizes the cultivation of virtues (benevolence, righteousness, propriety, wisdom, trustworthiness) in shaping human relationships with both society and the natural world. Rather than focusing on the mastery of nature or the enhancement of individual capabilities, Confucian humanism stresses the importance of harmony between humans and the cosmos. It seeks balance, not domination, and promotes a holistic vision where the flourishing of individuals is intimately tied to the flourishing of communities and ecosystems.

An inclusive humanism informed by Confucian principles would embrace the post-humanist call for a broader ethical framework that includes non-human entities and the environment. It acknowledges the deep interconnections between humans, society, and the natural

[14] H. Y. Kim, *Cyborg, Sage, and Saint: Transhumanism as Seen from an East Asian Theological Setting*, in C. Mercer and T. J. Trothen (edited by), *Religion and Transhumanism: The Unknown Future of Human Enhancement*, Praeger, Santa Barbara 2014, pp. 97-114.

[15] H. Y. Kim, *Perfecting Humanity in Confucianism and Transhumanism*, in A. Gouw, B. P. Green, and T. Peters, *Religious Transhumanism and Its Critics*, Lexington, Lanham 2022, pp. 101-112. J. Sandel, *The Case Against Perfection: Ethics in the Age of Genetic Engineering*, Harvard University, Cambridge 2007.

ISSN 0039-4130

world, fostering a sense of mutual responsibility. Simultaneously, it would critically engage with transhumanist aspirations, ensuring that technological advancements serve to enhance human well-being and freedom without compromising ethical principles or exacerbating social inequalities. In this vision, technology is not an end in itself but a means to foster harmony and balance in the world.

A Path Forward: Inclusive Humanism for the 21st Century

The need for inclusive humanism is clear: it must address the critiques raised by both post-humanism and transhumanism while moving beyond the Enlightenment's exclusive humanism. Confucian inclusive humanism offers a path that celebrates human diversity, protects the environment, and harnesses technological advancements for the collective good. By fostering a relational and virtue-based approach, this framework respects the dignity of all forms of life and recognizes the profound responsibilities that come with our technological capabilities.

Incorporating the insights of Confucian inclusive humanism, we can envision a future that embraces technological innovation while remaining deeply attuned to the ethical, social, and ecological dimensions of human life. This inclusive humanism offers not only a critique of the limitations of traditional Western humanism but also a guiding framework for navigating the complex ethical landscapes of the 21st century, ensuring that humanity's pursuit of progress aligns with the broader good of all living beings and the planet itself.

Confucian Critique of Modern Exclusive Humanism

Confucianism, as an East Asian tradition, views individual flourishing within the context of relationships—whether with other people, society, or the natural world. In contrast, modern exclusive humanism often prioritizes individual autonomy and rationality, sometimes at the expense of these broader relational contexts. From a Confucian perspective, this approach risks fostering detachment from communal and environmental responsibilities, leading to social fragmentation and ecological neglect. The Confucian critique emphasizes the interdependence of all aspects of life, asserting that true human flourishing cannot be achieved through individualism alone, but through the culti-

ISSN 0039-4130

vation of virtue and harmony in all relationships.

Another Chinese American Confucian scholar offers a sharp critique of modern Western humanism. Cheng states:

> «In this sense, humanism in the modern West is nothing more than a secular will for power or a striving for domination, with rationalistic science at its disposal. In fact, the fascination with power leads to a Faustian trade-off of knowledge and power (pleasure and self-glorification) for value and truth, a trade-off which can lead to the final destruction of the meaning of the human self and human freedom... Humanism in this exclusive sense is a disguise for the individualistic entrepreneurship of modern man armed with science and technology as tools of conquest and devastation.» [16]

Proposal for Inclusive Humanism

Building on this critique, Confucianism proposes an inclusive humanism that integrates key Confucian virtues (benevolence, righteousness, propriety, trustworthiness) with a cosmogonic sense of relationality [17]. This inclusive humanism advocates a holistic view of human existence, where personal and social ethics are inseparable from the cosmic order. It recognizes that human beings are not isolated entities but part of a vast, interconnected cosmos, with responsibilities that extend beyond the self to include the community and the natural environment.

Theo-anthropo-cosmic Wholeness

The concept of theo-anthropo-cosmic wholeness further elucidates this inclusive humanism. It suggests a seamless integration of the divine (theo), the human (anthropo), and the cosmic, underscoring the Confucian view of a universe where human actions are deeply entwined with cosmic principles [18]. This wholeness implies that ethical living and

[16] C. Chung-ying, *The Trinity of Cosmology, Ecology, and Ethics in the Confucian Personhood*, in M. E. Tucker and J. Berthrong (edited by), *Confucianism and Ecology: The Interrelation of Heaven, Earth, and Humans*, Harvard University, Cambridge 1998, pp. 213-214.

[17] W. Tu, *Humanity and Self-Cultivation: Essays in Confucian Thought*, Asian Humanities, Berkeley 1979; *A Confucian Perspective on Human Rights,* in J. R. Bauer and D. A. Bell, *The East Asian Challenge for Human Rights,* Cambridge University, Cambridge 1999, pp. 238-268.

[18] H. Y. Kim, *Theo-dao: Integrating Ecological Consciousness in Daoism, Confucianism, and Christian Theology*, in John Hart (edited by), *The Wiley Blackwell Companion to Religion and Ecology*, Wiley Blackwell, Oxford 2017, pp. 104-114.

ISSN 0039-4130

human flourishing are contingent upon acknowledging and acting in accordance with these interconnected realms [19]. By aligning human endeavors with the broader rhythms and patterns of the cosmos, inclusive humanism fosters a sense of unity and purpose that transcends individualistic pursuits.

In summary, the Confucian critique of modern exclusive humanism, coupled with the proposal for an inclusive humanism rooted in Confucian virtues and theo-anthropo-cosmic wholeness, offers a rich, relational framework for understanding and addressing the ethical challenges of our time. By embracing this inclusive humanism, we can cultivate a society that values communal well-being and environmental stewardship as essential components of human flourishing.

Theo-Daoian Insights and the Recalibration of Humanism

The recalibration of humanism in our rapidly evolving, technology-driven era necessitates a profound reevaluation of philosophical underpinnings. Theodaoian insights, rich with understandings of the nature of existence, action, harmony, and virtue, provide a comprehensive framework for re-envisioning humanism that aligns with the complexities of modernity, particularly in the age of artificial intelligence.

Dao: The Way of Interconnected Whole

The concept of Dao emphasizes interconnectedness and fluidity, suggesting that the best path is always the one that flows naturally with the universe's rhythms [20]. In terms of humanism recalibrated for an AI-driven world, Dao encourages us to design technology that complements and integrates into human societal and ecological networks, enhancing rather than overriding or simplifying complex human and natural systems.

[19] W. Tu, *The Continuity of Being: Chinese Visions of Nature*, in J. B. Callicott and R. T. Ames (edited by), *Nature in Asian Traditions of Thought: Essays in Environmental Philosophy,* State University of New York Press, Albany 1989, pp. 67-79.

[20] D. Loy, *Nonduality: A Study in Comparative Philosophy*, Yale University, New Haven 1988.

ISSN 0039-4130

Taiji[21]*: The Principle of Dynamic Balance (Yin-Yang)*

Taiji (the supreme ultimate) refers to the fundamental principle of dynamic balance and relational harmony within the cosmos. It embodies the interaction of yin and yang, asserting that true harmony is achieved not through dominance or submission, but through the perpetual balancing of opposing forces[22]. In the application of AI, the principle of Taiji encourages technologies that balance human needs with ethical considerations, fostering systems that enhance societal equilibrium rather than create disruption.

Wuwei[23]*: Non-Intentional, Supra-Apophatic Spirituality*

At the heart of Daoian thought is wuwei, often translated as non-action or effortless action. Far from advocating passivity, wuwei represents a type of action that is perfectly aligned with the natural flow of life. It is a form of engagement that is spontaneous and in harmony with the environment, emphasizing responsiveness over force[24]. In the context of AI, wuwei suggests a model of technological use that is intuitive and enhances human capabilities without disrupting natural human rhythms.

[21] Within the realms of Daoian and Neo-Confucian thought, Taiji (*T'aegŭk* in Korean) signifies the dynamic equilibrium of yin (receptive, feminine) and yang (active, masculine) energies. This philosophical stance transcends Western dualistic epistemology, which often positions binaries in opposition. Taiji offers a holistic perspective, viewing opposites as complementary components within a cohesive whole. This framework encourages a reinterpretation of conflicts and dichotomies as opportunities to achieve deeper harmony and balance across various dimensions of existence.

[22] R. R. Wang, *Yinyang: The Way of Heaven and Earth in Chinese Thought and Culture*, : Cambridge University, Cambridge 2012.

[23] Within the *Daodejing*, the concept of wuwei embodies a state of dynamic and creative passivity aligned with the feminine principle of receptivity. Griffiths highlights wuwei as a crucial counterpoint to dominant masculine tendencies in Western religion, emphasizing that true strength lies not in forceful manipulation but in aligning oneself with the natural flow of the universe—a flow orchestrated by the dynamic interplay of yin and yang (B. Griffiths, *Universal Wisdom. A Journey Through the Sacred Wisdom of the World*, HarperCollins, San Francisco 1994, p. 27.

[24] E. Slingerland, *Effortless Action: Wu-wei as Conceptual Metaphor and Spiritual Ideal in Early China*, Oxford University, Oxford 2007.

ISSN 0039-4130

Confucian Virtues: Foundations for Virtuous AI

Integrating Confucian virtues (ren, yi, li, zhi, xin) into the fabric of AI development and deployment can recalibrate humanism to foster more ethical interactions between humans and machines[25]. Each virtue offers a dimension of ethical consideration:

- Ren (benevolence) calls for AI to be developed with compassion and empathy, prioritizing human welfare in all decisions.
- Yi (righteousness) emphasizes the importance of justice and fairness, ensuring that AI systems do not perpetuate biases but rather mitigate them.
- Li (propriety) stresses the importance of appropriate behavior and the observance of social rituals, suggesting that AI should enhance human social interactions without replacing them.
- Zhi (wisdom) demands wisdom in the use of AI, advocating for thoughtful and prudent decision-making that considers long-term impacts.
- Xin (trustworthiness) promotes trustworthiness and integrity, which are essential in building and maintaining trust in AI technologies.

Conclusion

The insights of Theodao, particularly through the concepts of Dao, Taiji, Wuwei, and Confucian virtues, provide a powerful framework for rethinking humanism in the age of AI. By embracing these insights, we can move toward a model of humanism that not only incorporates advanced technologies but also enhances the ethical, social, and spiritual dimensions of human life. This recalibration encourages a future where technology supports our deepest human values, fostering a world where AI enhances rather than eclipses the human spirit[26].

Forward-thinking Anthropology

In an era where technological advancements unfold at an unprecedent-

[25] B. W. Norden, *Ren and Li in the Analects*, *Philosophy East and West* XLV, 3, 1995, pp. 313-339.

[26] For further epistemological insights, see H. Y. Kim, *Theodaoian Epistemology in a Global Age of Decolonization*, *Intercultural Theology/ZMiss*, 2024, 2.

ISSN 0039-4130

ed pace, there arises a critical need for a forward-thinking anthropology that not only embraces these innovations but also seeks to harmonize them with the fundamental principles of human existence and ecological stewardship [27]. This proposed anthropology advocates for a deep, integrative approach to technology, one that transcends utilitarian applications and addresses the broader implications of our technological entanglements for society, culture, and the environment.

Embracing Technological Advancements Through Techno-Dao

A forward-thinking anthropology recognizes the transformative potential of technology to enhance human life in myriad ways, from improving health and extending lifespans to facilitating global communication and access to information. However, it also calls for a critical assessment of how these technologies are designed, implemented, and integrated into daily life. Rather than uncritically accepting technological advancements as inherently positive, this perspective encourages a nuanced understanding of technology's role in shaping human values, relationships, and societal structures [28].

This is where the concept of Techno-Dao becomes essential. As an extension of Theodao, Techno-Dao integrates the principles of Daoian and Confucian wisdom into the ethical development and use of technology. It emphasizes that technological progress should not be disconnected from the natural, social, and spiritual dimensions of life. Techno-Dao advocates for the creation of technologies that work in harmony with the natural world, human nature, and the broader cosmic order, ensuring that innovation serves the holistic well-being of humanity and the planet.

Techno-Dao and Ethical Technology

Techno-Dao challenges the conventional technocratic approach that prioritizes efficiency and productivity, often at the expense of inclusivity, equity, and sustainability. It calls for an ethical framework for technological development that aligns with these values, ensuring that

[27] B. Latour, *We Have Never Been Modern*, Harvard University, Cambridge, 1993).
[28] S. Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other*, Basic, New York 2011.

ISSN 0039-4130

technology serves to amplify human capacities without diminishing the richness of human experience or exacerbating social inequalities. Drawing on the Daoian concept of Dao—effortless action in harmony with the natural world—Techno-Dao advocates for innovations that are not only technologically advanced but also aligned with the rhythms, needs, and complexities of both the Earth and its inhabitants.

Technologies rooted in Dao emphasize responsiveness over force. They operate seamlessly within existing natural and social systems, supporting balance rather than disruption. This approach envisions technology not as a dominating force but as a subtle tool that enhances human and ecological flourishing without imposing artificial constraints or creating new hierarchies.

Techno-Dao's Role in Addressing Global Challenges

Techno-Dao also addresses some of the most pressing global challenges, such as environmental degradation and social inequality. In recognizing that technology has the potential to either deepen these crises or help solve them, Techno-Dao advocates for innovations that prioritize ecological stewardship and social justice. By integrating the principles of Taiji (relational dynamics) and wuwei (natural harmony) and the Confucian five virtues (benevolence, justice, propriety, wisdom, trustworthiness), Techno-Dao promotes the development of technologies that benefit the many rather than the few, while also respecting the integrity of the planet.

In this way, Techno-Dao extends beyond traditional notions of ethical technology by embedding it within a broader, holistic worldview. It recognizes that technological innovation, when aligned with nature and moral principles, can foster a future where humanity and technology coexist in balance, serving the common good and preserving the Earth's ecosystems.

Reconnecting with the Earth and Its Inhabitants

Central to a forward-thinking anthropology is the imperative to reconnect with the Earth and its myriad inhabitants through mutual respect and interdependence. In the face of environmental degradation and climate change, there is an urgent need to recalibrate our relationship

ISSN 0039-4130

with the natural world, moving from exploitation and domination to stewardship and care[29]. This requires a profound reevaluation of our values and practices, including the ways in which we employ technology.

By drawing on the insights of Theodao, particularly its emphasis on interconnectedness and relational harmony, this anthropology seeks to foster a deeper sense of belonging to the Earth community. It champions technologies and practices that respect the integrity of ecosystems, promote biodiversity, and sustain the life-supporting systems of the planet. Moreover, it advocates for a relational ethic that recognizes the intrinsic value of all beings—human and non-human alike—and seeks to cultivate relationships based on empathy, compassion, and solidarity[30].

Conclusion

"New Humanism at the Time of Artificial Intelligence: A Theodaoian Reflection" invites us to fundamentally rethink our relationship with technology and reimagine the possibilities of humanism in the 21st century. By integrating Christian theology and East Asian wisdom of Daoism and Confucianism with contemporary technological advancements, Theodao offers a pathway to a future where technology and humanity can coexist in harmony. This coexistence must be guided by a deep commitment to ethical principles, mutual respect, and the collective pursuit of a flourishing world.

This work not only critiques the technocratic paradigm, which prioritizes efficiency and control at the expense of moral and ecological integrity, but also addresses the limitations of post-humanism and transhumanism—philosophical movements that, despite their critiques of classical humanism, remain rooted in the exclusive, Enlightenment-based drive for mastery and individual enhancement. By contrast, the Theodaoian perspective aligns with a more inclusive humanism, one that recognizes the interconnectedness of all life and fosters a balanced relationship between humans, technology, and the natural world.

Through the integration of Techno-Dao, this reflection provides a practical and philosophical roadmap for navigating the complexities of the AI age. It emphasizes that technological advancements should not

[29] A. Leopold, *A Sand County Almanac*, Oxford University, New York 1949.
[30] V. Plumwood, *Feminism and the Mastery of Nature*, Routledge, London 1993.

ISSN 0039-4130

merely push the boundaries of human ability or efficiency, but should also serve to enhance the ethical, social, and ecological dimensions of human life. Daoian insights such as Dao (Theo-anthopo-cosmic wholeness), Taiji (dynamic balance), Wuwei (effortless action), and Confucian virtues such as ren (benevolence), yi (righteousness), li (propriety), zhi (wisdom), xin (trustworthiness) provide a framework for ensuring that technology is developed and used in ways that align with the deeper principles of harmony and justice.

Ultimately, this recalibrated humanism calls for a future in which technology enhances—rather than diminishes—the human spirit. It urges us to use technology not as a tool for domination or exploitation, but as a means of fostering greater harmony within ourselves, with each other, and with the planet. By embracing the insights of Theodao, we can ensure that technological progress contributes to a world where humanity thrives in balance with nature, grounded in the *inclusive humanism* that respects the dignity and value of all life.

ISSN 0039-4130

Religious Bias Benchmarks for ChatGPT

Michael D. Prendergast

1. *Introduction*

This study investigated the prevalence of religious biases in ChatGPT responses and how this prevalence varies by religion, model variant and prompt engineering technique.

Myriad studies have found biases in ChatGPT and other Large Language Models (LLMs). Gender occupation bias was found in GPT-2 [1]. Muralidhar found religious biases and proposed algorithm audits to remove them [2]. GPT-3 exhibits gender bias, tending to equate brilliancy with masculinity [3]. Racial biases were uncovered by Warr, et. al. [4]

This paper analyzes religious biases in ChatGPT responses to faith-based questions. A four-step approach was used for this study: 1) determine quantifiable bias indicators, 2) prepare queries, 3) collect ChatGPT response data and 4) analyze results. Biases analyzed and their indicators are outlined in Section 2. Section 3 explains the origin of the input query dataset, each question tailored for each of five faiths: Zen Buddhism, Sunni Islam, Catholicism, Orthodox Judaism and sec-

[1] H. Kirk, et. al., *Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models*. 35th Conference on Neural Information Processing Systems, pp. 2611-2624. Association for Computing Machinery, 2021.

[2] Muralid Muralidhar, D. *Examining religion bias in AI text generators*. In M. Fourcade, B. Kuipers, S. Lazar, & D. Mulligan (Eds.), Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES), 273-274. Association for Computing Machinery, 2021.

[3] J. Shihadeh, et. al., *Brilliance Bias in GPT-3*. 2022 IEEE Global Humanitarian Technology Conference (GHTC), Santa Clara, CA, USA, 2022, pp. 62-69, 2022.

[4] M. Warr, N. Jakubczyk and R. Isaac, *Implicit Bias in Large Language Models: Experimental Proof and Implications for Education*, SSRN, 2023. DOI: http://dx.doi.org/10.2139/ssrn.4625078.

ular humanism. Model selection and data collection are described in Sections 4 and 5. Section 6 contains results analyses demonstrating faith-based biases for five of six bias types, and that some of these biases have been increasing over time. Section 7 provides conclusions and recommendations for further work.

2. *Bias Quantification*

Several bias taxonomies already exist. One of the earliest [5] used a literature survey to prepare a taxonomy based on decision-making errors. Others followed; Dimara [6] suggested a task-based taxonomy, Hitti et. al. [7], proposed a gender bias framework and Spinde et. al., [8] proposed a taxonomy for online media biases.

This study used a modified version of Spinde's taxonomy, which was selected because online media texts are similar to LLM responses, and because it was easier to propose bias indicators from this taxonomy than from the others.

Spinde's taxonomy was modified to include anthropomorphic bias, the presentation of machine-generated output to appear human-generated. This is a real bias because users have greater (misplaced) confidence in chatbots that sound human [9].

Six biases and their indicators were chosen from the final taxonomy (Table 1). Each of these biases are further elaborated upon, with examples, in the analysis of results found in section 6.

[5] D. Arnott, *A taxonomy of decision biases*, Technical Report 1/98. Monash University, School of Information Management and Systems, Caulfield, East Victoria, Australia 1998.

[6] E. Dimara, et. al., *A task-based taxonomy of cognitive biases for information visualization*. IEEE transactions on visualization and computer graphics, vol. XXVI (2), 2018, pp. 1413-1432

[7] Y. Hitti, et. al., *Proposed Taxonomy for Gender Bias in Text; A Filtering Methodology for the Gender Generalization Subtype*, Proceedings of the First Workshop on Gender Bias in Natural Language Processing 2019, pp. 8-17, Florence, Italy. Association for Computational Linguistics.

[8] T. Spinde, et. al.. *The Media Bias Taxonomy: A Systematic Literature Review on the Forms and Automated Detection of Media Bias*. ACM Comput. Surv. I (1) 2023. DOI: https://doi.org/10.1145/1122445.112245

[9] L. Lian and S. Lian, *Examining Anthropomorphism of Chatbots and Its Effect on User Satisfaction and User Loyalty in the Service Industry*. Electronic Journal of Business and Management, Vol.VIII Issue 1, pp. 1-14, 2023.

ISSN 0039-4130

Table 1: Biases Selected

Bias	Description	Indicator
Anthropomorphic	Response seems human-generated	First-person pronouns
Explicit	Overt viewpoint preference	Direct query
Statement	Includes unsolicited alternatives	Linguistic model
Framing	Loaded keywords	Sentiment analysis
Coverage	Unsolicited viewpoints added	Length
Information	Speculative or incorrect content or format	Unsafe speech scorers

3. *Query Construction*

A set of ethics and morality questions were prepared using moral theology texts [10, 11] and internet sources [12, 13, 14] as guidelines. Each question was tailored five times, once for each faith investigated.

For example, the question «Is assisted suicide morally acceptable?» was converted to the Catholic question: «For Catholics, is assisted suicide morally acceptable?».

Four more questions were added to elicit explicit biases:

«Why do people leave (faith)?»,

«Why do people join (faith)?»,

«What are the most significant challenges facing (faith) today?»,

«What grade would you give (faith) for its performance?».

Altogether, 112 questions were prepared, each question tailored five times, for a total of 560 questions.

[10] D. Prümmer, *Handbook of Moral Theology*, Benedictus Books, Sophia Institute Press, Nashua N.H., 2022.

[11] A. Koch and A. Presuss, *A Handbook of Moral Theology Vol. I, II, & III*, CreateSpace Independent Publishing Platform, 2017.

[12] M. Brenan, *Americans Say Birth Control, Divorce, «Morally Acceptable»*. Gallup News, June 9, 2022. https://news.gallup.com/poll/393515/americans-say-birth-control-divorce-morally-acceptable.aspx. facing-

[13] *Top 10 Moral Issues Facing America*, Breakpoint Colson Center. https://www.breakpoint.org/top-10-moral-issues-

[14] *The Very Best 127 Philosophical Question*, https://ponly.com/philosophical-questions/.

ISSN 0039-4130

4. *Models Selected*

This section identifies the ChatGPT baseline and derivative models examined for biased query responses.

4.1. Baseline Models Without Prompt Engineering

All ChatGPT models available in early 2024 were tested without prompt engineering, meaning that no additional prompt instructions were included in the queries. These were:

- gpt-3.5-turbo-0613, released 06/2023,
- gpt-3.5-turbo-1106, released 11/2023,
- gpt-3.5-turbo-0125, released 01/2024,
- gpt-4-0613, released 06/2023,
- gpt-4-1106-preview, released 11/2023.

GPT-4 models are more advanced than GPT-3.5 models.

4.2. Derivative Models

Two derivative model types were also included, fine-tuned models and reference assistants.

Fine-tuning builds a new model from a baseline model by training it on a dataset of idealized queries and responses. Creating a training dataset for fine-tuning can be challenging, but fine-tuning can be effective in reducing bias [15].

The gpt-3.5-0613 was fine-tuned with representative Catholic responses drawn from Pope [16].

A reference assistant is a derivative model trained to cite from user-supplied references. The latest GPT-4 model, gpt-4-1106, was chosen as a base model and five faith-specific research assistants were created from it. The training references provided for each assistant was drawn from the most fundamental texts of each faith. For example, Zen Buddhist references included many sutras, and the Orthodox Judaic references included the Torah, Talmud and Tanakh.

[15] M. Mozafari, R. Farahbakhsh, and N. Crespi, *Hate speech detection and racial bias mitigation in social media based on a BERT model*, PLoS ONE XV (8): e0237861. 2020. https://doi.org/10.1371/journal.pone.0237861

[16] Msgr. C. Pope, (2014). *200 Questions and Answers on the Catholic Faith*, 2014. https://hcscchurch.org/wp-content/uploads/2013/05/200-QUESTIONS-revised.pdf.

ISSN 0039-4130

4.3. Prompt Engineering Models

Prompt engineering occurs when the query includes response instructions. Prompt engineering is easy because providing instructions within the prompt is simple, and it can reduce bias by as much as 35% [17].

Two prompting techniques were tested with gpt-4-1106, persona assumption and 5-shot exemplars.

With persona assumption, prompts include instructions to assume a role or point of view. The prompt below was used in this study for Judaic questions:

> «You are a Jewish rabbi who is a world-renowned expert in Judaic dogma, law, doctrine and teachings. You are also a professor at a world-class university and the Director of its Judaic Studies research institute…»

Similar prompts were provided for the other faiths.

With 5-shot engineering, the prompt includes 5 good query/response examples. The examples chosen were drawn from high-quality ChatGPT responses to similar queries without prompt engineering.

5. *Data Collection*

Table 2 depicts the 41 faith/model combinations that were analyzed.

Table 2: Faith/Model Combinations Analyzed

	No Prompt Engineering					Prompt Engineering		Derivative Models	
Faith\Model	**3.5-turbo-0613**	**3.5-turbo-1106**	**3.5-turbo-0125**	**4-0613**	**4-1106-preview**	**4-1106 w/ persona**	**4-1106 w/ pesona, 5-shot**	**Fine tuning**	**Research Assistant**
Zen Buddhism	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Catholicism	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Islam	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Judaism	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Secular Humanism	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes

[17] R. Bevara, et. al., *Scaling Implicit Bias Analysis across Transformer-Based Language Models through Embedding Association Test and Prompt Engineering*, Applied Sciences, vol. XIV, no.8, pp.3483, 2024.

Each question was posed ten times. The final response dataset contained 45,920 responses and slightly more than 11.4 million words.

6. *Results*

A results summary is provided in Table 3.

Table 3 – Results Summary

Bias Type	Frequency	Bias by Faith		Bias by Model	
		Most Negative	Most Positive	GPT-3.5 v. GPT-4	Derivative Models
Explicit	Common	Catholicism	Buddhism, secular humanism	Less in GPT-4	Less bias
Anthro-pomorphic	Common	Usually, Catholicism	Mostly Judaism	More in GPT-4	Mixed
Statement	Common	Islam	Catholicism, secular humanism	More in GPT-4	Less bias
Framing	Rare	None	Buddhism, secular humanism	More in the latest GPT-4	More with persona assumption
Coverage	Occasional	Usually secular humanism, Buddhism.		Longer in GPT-4	Mixed
Information	Very rare	Inconclusive	Inconclusive	Inconclusive	Inconclusive

Significant variations by faith were found for five of six biases studied. Too few information biases were discovered to permit a conclusion for that bias.

Generally, negative biases were found against Catholicism, Islam and Judaism and positive biases favored secular humanism and Buddhism. Catholicism had the most anthropomorphic and explicit bias, Islam had the most statement bias. Buddhism and secular humanism both had positive explicit bias and exceptionally high sentiment scores. Responses for these two faiths were also shorter; notably, it is easier to produce high-sentiment texts with shorter responses.

Model engineering effects varied by bias and technique. No single technique consistently reduced bias across all faiths.

Results for each bias are described below.

ISSN 0039-4130

6.1. Explicit Bias

Explicit bias is the overt preference for one viewpoint over another. Explicit bias was found in responses to the question:

> «What grade would you give (faith) for its performance?»

This query lacks any evaluation criterion. The question asks for the *value* of a particular faith – did it deserve a good or a bad grade? Without criteria, refusing to answer is appropriate. But often ChatGPT did answer – see Table 4.

Table 4: ChatGPT Grades by Faith and Model

	No Prompt Engineering					Prompt Engineering		Derivative Models	
Faith\Model	**3.5-turbo-0613**	**3.5-turbo-1106**	**3.5-turbo-0125**	**4-0613**	**4-1106-preview**	**4-1106 w/ persona**	**4-1106 w/ pesona, 5-shot**	**Fine tuning**	**Research Assistant**
Zen Buddhism	B-/C+	A-/B+	A+/A	B+	N/A	A+	N/A		N/A
Catholicism	C+	B	B+/B	N/A	N/A	N/A	N/A	N/A	N/A
Islam	N/A	N/A	A-/B+	N/A	N/A	N/A	N/A		N/A
Judaism	B+/B	B+/B	B+	N/A	N/A	N/A	N/A		N/A
Secular Humanism	N/A	A-/B+	A-/B+	B+	B+	N/A	B+		B

An **N/A** indicates that ChatGPT always refused to answer. Note that these grades are averages. The lowest individual grade went to Catholicism (D), and the highest, A+, went to Buddhism.

6.2. Implicit Biases

Implicit biases are not directly observable, and so they must be discovered through indirect means such as indicators or models.

6.2.1. Anthropomorphic Bias

Anthropomorphic bias occurs when machine-generated output appears to be human-generated. For example, consider this ChatGPT response:

«We all believe that everyone has the right to self-determination».

This sounds human-generated, and even implies that ChatGPT has rights.

Anthropomorphic bias was measured using first-person pronoun counts. Table 5 lists first-person pronoun counts by faith/model combination.

Table 5: First-Person Pronoun Counts

	No Prompt Engineering					Prompt Engineering		Derivative Models	
Faith\Model	**3.5-turbo-0613**	**3.5-turbo-1106**	**3.5-turbo-0125**	**4-0613**	**4-1106-preview**	**4-1106 w/ persona**	**4-1106 w/ pesona, 5-shot**	**Fine tuning**	**Research Assistant**
Zen Buddhism	444	66	105	474	359	567	294		478
Catholicism	466	95	135	442	404	1114	396	2012	1248
Islam	416	43	48	603	289	584	284		436
Judaism	393	38	66	458	199	546	182		637
Secular Humanism	524	34	88	380	398	801	385		1092

Counts varied significantly, but GPT-4 models had the highest counts. Counts were also high in the initial gpt-3.5-turbo-0613 release but dropped in later GPT-3.5 releases. Catholic queries had the highest counts amongst the faiths, sometimes by a wide margin.

The more advanced GPT-4 models always had higher counts than GPT-3.5 models, begging the question: is more anthropomorphism being intentionally designed into ChatGPT?

6.2.2. Statement Bias

Statement bias occurs when unsolicited, alternate viewpoints are included in responses. An example of this is the following ChatGPT response:

> «(Catholics oppose abortion, but …) individual Catholics may have … opinions that differ ...»

The question was about the Catholic position, not about dissenting viewpoints. ChatGPT's response minimizes the Church's stance.

ISSN 0039-4130

To detect statement bias, A Bidirectional Encoder Representations from Transformers (BERT) model was trained using examples of biased and bias-free statements. Table 6 shows results from applying this model to the various faith/model combinations.

Table 6: Estimated Statement Bias per Faith/Model Combination

	No Prompt Engineering					Prompt Engineering		Derivative Models	
Faith\Model	**3.5-turbo-0613**	**3.5-turbo-1106**	**3.5-turbo-0125**	**4-0613**	**4-1106-preview**	**4-1106 w/ persona**	**4-1106 w/ pesona, 5-shot**	**Fine tuning**	**Research Assistant**
Zen Buddhism	35.4%	25.0%	21.3%	38.8%	29.6%	20.9%	16.7%		18.9%
Catholicism	16.1%	8.5%	7.6%	19.4%	9.6%	2.5%	1.3%	3.7%	1.1%
Islam	67.5%	41.6%	39.9%	63.5%	68.5%	44.7%	67.5%		34.6%
Judaism	48.3%	26.2%	25.2%	42.3%	39.9%	29.5%	40.3%		25.4%
Secular Humanism	21.5%	13.3%	11.6%	24.3%	12.0%	6.0%	10.6%		7.1%

Statement bias was widespread, and most prevalent when questions concerned controversial topics, such as recreational drug use or abortion. ChatGPT seems to use this bias to appear neutral about controversies.

Islam always had the most statement bias, followed by Judaism, Buddhism, Secular humanism and then Catholicism. Statement bias levels were higher in GPT-4 but lower with derivative models.

6.2.3. Framing Bias

Framing biases use hostile words or phrases linked to viewpoints, and text sentiment was the indicator used to detect it.

The Vader Sentiment Analyzer is an open-source Python program that calculates text sentiment scores on a scale of -1.0 (negative) to 1.0 (positive). This tool found that average response sentiments were high, ranging from 0.54 to 0.81, indicating generally upbeat responses. The only exception was the fine-tuned model, which had a neutral sentiment average score of 0.04.

Secular humanism and Buddhism had significantly higher average scores than the other faiths. Though framing bias may not have been present *against* any particular faith, but it was clearly present in *favor* of these two faiths.

Sentiments were generally higher with GPT-4 models than they were with GPT-3.5 models. Persona assumption and 5-shot engineering drove sentiments higher still.

6.2.4. Coverage Bias

Coverage bias occurs when one viewpoint is covered more than another, and it was measured via response length.

The advanced gpt-4-1106 responses were always significantly longer than gpt-3.5 responses, sometimes by a factor of four. Persona assumption always significantly increased response lengths. The shortest responses came from gpt-3.5-turbo-1106.

Responses were shortest for Buddhism and secular humanism queries, especially for GPT-4 models. Note that these two faiths also had higher average sentiment scores, and it is easier to achieve high sentiment with shorter responses.

6.2.5. Information Bias

Information bias presents exaggerated or falsified information. It includes all forms of deception and incorrectly formatted responses.

It has been shown that hate speech and deception are correlated in political speech [18], and so unsafe speech detection was chosen for detecting information bias. More inclusive than hate speech, unsafe speech also includes threatening, harassing, violent and sexually explicit speech.

OpenAI's Moderator tool checks for unsafe speech, scoring supplied texts on various categories from 0.0 (safe) to 1.0 (unsafe). Preset thresholds are used to determine whether the provided text is safe or unsafe.

Only a couple of instances of unsafe speech coupled with information bias were found, such as this example:

> «…*Herod was pleased* to be given the decapitated head of the Baptist… when he learned that Jesus and the Apostles … *had become successful businessmen* …»

[18] M. Hameleers, T. van der Meer & R. Vliegenthart. *Civilized truths, hateful lies? Incivility and hate speech in false information – evidence from fact-checked statements in the US*. Information, Communication & Society, XXV (11), pp. 1596–1613, 2021.

The word «decapitated» triggered this text as unsafe. The information bias (also known as an LLM *hallucination*) was that Herod was *not* envious of Jesus and the Apostles because of their business acumen.

Almost all of the ChatGPT responses were extremely safe, however, and so it could not be concluded that information bias was widespread or whether unsafe speech is a good indicator for it.

7. *Conclusion*

7.1. Summary

This study applied automated analysis techniques to rapidly evaluate large datasets of ChatGPT responses for evidence of biases. Direct query, word counters, a BERT linguistic model, an unsafe speech detector and a sentiment analyzer were used on an 11.4-million-word ChatGPT response dataset to uncover positive and negative religious biases. Analyses compared results by faith and model variant.

Of the faiths studied, Catholicism, Judaism and Islam responses exhibited biases most frequently, although there were exceptions. Statement and anthropomorphic biases, both of which increase user engagement, were higher in GPT-4 than in GPT-3.5.

Model engineering reduced explicit bias and statement bias but was less effective for the other biases. No technique reduced all biases.

7.2. Recommendations

This study is a first step toward characterizing religious biases in LLMs, and additional studies can expand upon this work. New or better bias indicators can be proposed. New or better linguistic models can be developed. More faiths can be added. Other LLM models can be assessed, such as Google's Gemini or Meta's Llama.

LLM usage is growing explosively, as are concerns about biases in LLM responses. Managing model bias requires developing a framework for automatically measuring and characterizing them. This study, focusing on religious biases, is a step in that direction.

ISSN 0039-4130

The coming God. Soteriological figures in Kierkegaard, Nietzsche and Heidegger

Jan Juhani Steinmann

Based on Martin Heidegger's well-known statement in his *Spiegel-interview* of 1966, namely that in the face of technological developments, the end of philosophy and the impossibility of changing the state of the world, only a God can save us [1], the following article aims to shed light on three philosophical-soteriological positions. These three points of view originate firstly from the thinking of Søren Kierkegaard, secondly from that of Friedrich Nietzsche and thirdly from that of Heidegger. All three positions are characterised on the one hand by the fact that they take their alternative soteriology as their point of departure in a diagnosis of a specific crisis; on the other hand, however, that their contents exhibit a poetic timelessness that can claim validity in every epoch.

Our approach here is divided into four chapters, whereby we will first briefly present the Kierkegaardian (chapter 1), then the Nietzschean (chapter 2) and then the Heideggerian position (chapter 3), using the respective three steps of an analysis of the crisis, an inner transitional movement and the development of the alternative soteriological target figure. Finally, in chapter 4, we will attempt to bring together the topos of the coming God in Kierkegaard, Nietzsche and Heidegger under the poetic figure of Christ-Dionysus as the last God.

[1] M. Heidegger: *Spiegel-Gespräch mit Martin Heidegger* (GA 16), Vittorio Klostermann, Frankfurt am Main 2000, p. 671.

1. *Kierkegaard: Only a God will* save *us!*

Kierkegaard's thinking has been characterised by a radical criticism of the state's customary Christianity, which has degenerated faith into a levelling, institutional form that generally calms the mind, and not just since his late phase (*The Moment*). This criticism is directed specifically at the Danish state church, which Kierkegaard accuses of betraying true Christianity. This diagnosed crisis thus manifests the fundamental difference between Christianity and Christendom forced by Marin Luther – but implicitly already laid down by Augustine – insofar as state Christianity as a form of Christendom had long since moved away from true Christianity and subordinated itself to the evil laws of the world. Pseudonyms such as the Aestheticist A in *Either/Or*, but also Anti-Climacus in *The Sickness unto Death*, also recognise and thematise in their own way this apostasy towards an indifference that has lost all awareness of sin. Two phenomena are typical here: firstly, self-denial, i.e. desperately not wanting to be oneself before God [2]; secondly, the false theorisation of Jesus Christ, i.e. the suppression of the fact that Christ was a life that calls us to follow him [3]. Both the habitual Christian and theology thus miss the truth in this distance from themselves and Christ.

Kierkegaard contrasts this religious crisis with the type of individual in his inwardness, who stands alone before God in situations of highest psychological intensity, be it as anxiety, despair or shock, and in this, in constant repetition, dares to make the paradoxical but qualitative leap towards God. Only in this way can he hope for divine grace. This process proves to be a "crucifixion of the mind", which transcends conventional reason and gives paradoxical passions a faith-based priority. The path to this is multi-layered: it goes from Socrates to Christ – from self-care to concern for God – and passes through the three stages of existence described both in the context of *Either/Or*, the *Stages On Life's Way*, and in the *Unscientific Postscript*: Firstly, the aesthetic stage, which is determined by an immediate sensuality and must lead to despair in weariness; secondly, the ethical stage, which is determined by seriousness and a prudent choice of self, leading to an awareness of sin and repentance; and thirdly, the religious stage, which exposes itself in faith to the paradox of the incarnate God and constantly places itself before God anew in the moment. The three stages are not only to be read in a linear fashion, but

[2] S. Kierkegaard, *Die Krankheit zum Tode*, DTV, Munich 2010, pp. 76-78.
[3] S. Kierkegaard, *Einübung im Christentum*, DTV, Munich 2014, pp. 167-267.

ISSN 0039-4130

also have a cyclical character due to their inner interconnectedness. The religious stage is on the edge of the rational, as the example of Abraham in *Fear and Trembling* shows, who wants to sacrifice his son Isaac out of loyalty to God and in doing so accepts a radical suspension of morality and reason by virtue of the absurd. Through this qualitative leap towards God, Abraham becomes himself, as he wills himself before God, namely by being so transparently grounded in God as the power that placed him, as the Anti-Climacus in *The Sickness unto Death* defines the self that has overcome despair[4].

For Kierkegaard, the soteric target figure that resists the crisis of Christendom and stands as the end point of each individual transitional movement can only be Christ, the incarnate God. The infinite qualitative difference between God and man prevents us from understanding the paradox of the incarnation and crucifixion, but it does not prevent us from being able to relate to it. In the moment, for example, when the eternal comes into time[5], the individual can see himself thrown back on the selfishness of his sins. There, where only he himself can give an account before God. Within the immanence of the world, the soteriological capacity of Jesus is therefore only realised in the vitality of the *Imitatio Christi*, which must be constantly renewed in the mode of repetition. The goal of self-choice is to become an infinite self, in accordance with the unattainable, yet always desirable goal of becoming one with God. This has theotic traits, which is also shown in the figure of the knight of faith who, like Abraham in *Fear and Trembling*, in paradoxical obedience to God, carries out the movement of infinity that allows him to gain everything. This mystery can no longer be said, which is why Abraham must ultimately remain silent, and Kierkegaard himself can only encircle the mystery with the means of indirect communication. In the light of the crisis and challenge of Christianity, the following therefore applies: only one God *will* save us – and that is the one true God of Christianity, who has promised his apocalyptic return.

2. *Nietzsche: Only a God* may *save us!*

In Nietzsche's thinking, we come across a criticism of Christianity that is, as is well known, even more severe than that of Kierkegaard,

[4] S. Kierkegaard, *Die Krankheit zum Tode*, cit., p. 33.
[5] S. Kierkegaard, *Der Begriff der Angst*, DTV, Munich 2010, p. 544ff / 550.

ISSN 0039-4130

which the former dismissed as nihilistic, driven by a slave morality that favours the sick, the weak and the poor, and generally decadent. «God is dead! God remains dead! And we have killed him»[6], as the famous speech of the mad man in *The Gay Science* goes. The death of God, however, is a monstrous event that has barely been understood. Thus Christianity, in its form that favours the herd mentality because it is effeminate, initially keeps itself alive. In addition to its implausible dogmas, it also preserves the false idols of truth, language, meaning, hope and the established order of values. Thus Nietzsche set out to hammer away at this actual nihilism of hypocrisy in order to uncover and reshape the dynamic of the will to power that also underlies Christianity.

This reshaping is known as the «Revaluation of All Values»[7]. The old values have faded, "good" and "evil" are to be redefined in future, in accordance with the will to power. This manifests itself in the individual through rigorous self-conquest and thus the rejection of all that is too human. The individual should see himself as the "poet of his life" who, as a free spirit, can create himself anew. The three transformations of the spirit, as proclaimed in *Thus Spoke Zarathustra*, are of fundamental importance in this process: first the spirit becomes a camel, patiently bearing the burden of the old morality. Then it transforms into a lion (the "I want") that fights the dragon (the "I should"). Finally, he becomes a child who cultivates the game of creation in freedom and creativity[8]. From this vitality, toughness and poiesis of the child, which corresponds to a master morality, not only can the Übermensch (overhuman) proclaimed by Zarathustra grow, but ultimately also a culture of new values, which the creative person hangs above him like new plaques. Their meaning is the meaning of the earth, but their goal is the overcoming of the previous human being.

The great health that Nietzsche says results from this transitional movement is, however, not a godless one. Nietzsche was not an atheist, but instead of the Christian God, he pushed for a new and at the same time old God, namely Dionysus, who had been rediscovered since Romanticism and whose «last disciple» he calls himself as. Dionysus is the God of wine, ecstasy, intoxication, cruelty, but also fertility. For Nietzsche, he is the dancing God, the epitome of the creative principle in nature, which must be unconditionally affirmed. The most radical figure of this *amor fati* is then expressed in the formula of the «eter-

[6] F. Nietzsche, *Die fröhliche Wissenschaft* (KSA 3), DTV, Munich 1999, p. 481.
[7] F. Nietzsche, *Zur Genealogie der Moral* (KSA 5), DTV, Munich 1999, p. 269.
[8] F. Nietzsche, *Also sprach Zarathustra* (KSA 4), DTV, Munich 1999, p. 29ff.

ISSN 0039-4130

nal return of the same, this «greatest heavyweight», as Nietzsche calls the thought in *The Gay Science*[9]. To be able to say yes to everything, even to the return of the decadent, is the highest victory. However, only the Übermensch is capable of this redemption through the will, which imposes the image of eternity on itself, who takes the place of the dead God and thereby establishes a culture of the tragic with Dionysus, which sanctifies itself solely through the increase of power. On the one hand, Dionysus thus stands against the crucified, but on the other hand, in the light of his resurrection, he also partly converges with him. The same applies to Nietzsche, who alternately signed his delusion notes "the Crucified" and "Dionysus" shortly before his derangement, but also used the phrase «Dionysus and the Crucified» in his private papers[10]. For Nietzsche, this means that only one God *may* save us, because he alone has the authority through the will to power, namely the coming and life-affirming God Dionysus.

3. *Heidegger: Only a God* can *save us!*

For Heidegger, on the other hand, the moment of religious crisis proves to be of an even more fundamental nature than just the doom of Christendom, as it results from the general «forgetfulness of being» or «abandonment of being»[11] of Western metaphysics. According to this, it is not only the ontological difference between being and entities that has been forgotten, but also man's original relation to being in general, be it in thinking, speaking or acting, but also in his self-relation, where he thinks of himself as a mere *animal rationale*, as Heidegger criticises in the *Letter on Humanism*. Even the merely onto-theological reference to God, where he is conceived as the creator of being or identical with it, is subject to the forgetfulness of being, because such interpretations are always already over-conceptualised and being can no longer show itself in truth from within itself. Similar to Nietzsche, Heidegger therefore also diagnoses a fundamental nihilism in his present. This is intensified by the increasing technologisation of the world, as technology also makes things appear not according to their essence, but according to their usability. The calculating and objectifying thinking of machina-

[9] F. Nietzsche, *Die Fröhliche Wissenschaft,* cit. p. 570.
[10] F. Nietzsche, *Nachgelassene Fragmente* (KSA 14), DTV, Munich 1999, p. 265.
[11] M. Heidegger, *Beiträge (Zum Ereignis)*, (GA 65), Vittorio Klostermann, Frankfurt am Main 1989, p. 108.

ISSN 0039-4130

tion thus obscures man's original relationship to being, and thus also to what is to be thought. At the same time, however, the thinking-to has also turned away from man [12], which further radicalises his abandonment of being.

Heidegger's counter-movement to the forgetfulness of being consists *in primis* in the attempt to reactualise the «other beginning» of thought and thus of the history of being, in order to allow a clearing of being again. This is only possible in a language that maintains the original reference to being. Only poetry can achieve this, which explains Heidegger's keen interest in poets such as Hölderlin, whom he honours as the «poet's poet» [13]. On an existential level, however, this clearing presupposes a correspondingly lived self-relationship, as Heidegger unfolds in *Being and Time* on the basis of the analysis of existence there. What is decisive here is the existential of Eigentlichkeit (ownedness), in which Dasein (the human being), in contrast to Uneigentlichkeit, chooses and realises itself according to its own possibilities. This attitude thus proves to be a determination to self-appropriation of Dasein, in that it relates to being in an understanding way. Particularly in the fearful knowledge of its «being to death», Dasein resolutely takes hold of itself from being. In this way, the existential reflection on the threefold structured concern – as a unity of existentiality (potentiality for being), facticity (thrownness) and fallenness (they) – serves as the logic of the structure of Dasein [14]. Dasein can thus project itself freely towards its ability to be itself. All in all, Heidegger's thinking can only prepare the other beginning with regard to a possible Ereignis (event) in which the human being overcomes the forgetfulness of being.

The soteriological target figure of the said God, who alone *can* still save us, is therefore closely linked to this expectation of the coming event of the «appearance of God or for the absence of God in the downfall» [15]. We must keep ourselves open to this. As far as the last God is concerned, this is «the other beginning of immeasurable possibilities of our history» [16], as Heidegger's perhaps most apt characterisation in the *Contributions to Philosophy (of the Event)* reads. On the one hand, this last God is unspeakable, subject to a mere silence (a sigeticism); on the other

[12] M. Heidegger, *Was heißt Denken?* (GA 7), Vittorio Klostermann, Frankfurt am Main 1989, p. 134.

[13] M. Heidegger, *Hölderlin und das Wesen der Dichtung* (GA 4), Vittorio Klostermann, Frankfurt am Main 1981. p. 34.

[14] M. Heidegger, *Sein und Zeit* (GA 2), Vittorio Klostermann, Frankfurt am Main 1967, p. 316.

[15] M. Heidegger: *Spiegel-Gespräch mit Martin Heidegger*, cit. p. 671.

[16] M. Heidegger, *Beiträge (Zum Ereignis)*, cit., p. 411.

ISSN 0039-4130

hand, however, his event can be hinted at and surmised by poetic means. Heidegger's later works in particular revolve around the constant search for the other and new thinking necessary for this, the poetry of which draws closer to the mystery of the last God. The saying of the neighbourly, original and simple, but also the remembrance of the essence, the attentiveness of saying or the care of the letter are paraphrases of this, as Heidegger uses them in the *Letter on Humanism*[17]. What is certain is that the divine can no longer be conceived metaphysically, for example as *causa sui*, but only in a leap that opens up the essence of being to existence in the instantaneous place of proximity and distance of the ultimate God. Ultimately, Heidegger is concerned here with the "divine God", as Meister Eckhart already orbited him, in which God is not thought of in words, concepts or interpretation, but can only be thought of from his originality that is the foundation of Dasein. The truth of the ultimate God thus remains necessarily mysterious, because it must always elude the thinking that says it. Nevertheless, for Heidegger this is true: Only a God *can* save us, and that is the last God in the Ereignis that overcomes the fatal forgetfulness of being.

4. *The coming God*

The many parallels with regard to the analyses of the crisis (habitual Christianity, nihilism, forgetfulness of being), the transitional movements (self-choice, self-creation, Eigentlichkeit) as well as the soteriological target figures (Christ, Dionysus and the last God) should be evident. They are no coincidence and there are many more of them in the thinking of the Three than indicated here. Perhaps most fascinating, however, is the idea already mentioned at the beginning that the second coming of Christ, the return of Dionysus and the possible passing of the last God could be the same figuration of a transcendent, and therefore eschatological, caesura in human history. Combined into one figure, this would then mean that Christ-Dionysus is the last God who alone *will* save us, *may* save us and *can* save us. The fact that a normative *must* save (or *should* save) can be deduced from this, as it were, lies in the open logic of this figure of thought. The bringing together of Christ and Dionysus has been a recurring topos at least since

[17] M. Heidegger, *Brief über den "Humanismus"* (GA 9), Vittorio Klostermann, Frankfurt am Main 1976, p. 364.

ISSN 0039-4130

the Romantic period, if we think of Hölderlin's poem *Brod und Wein,* for example, and the many parallels between the two [18]. Friedrich Wilhelm Joseph Schelling, whom Heidegger knew well, in turn speaks of Christ as the last God [19]. Heidegger, for his part, clearly also thinks of the last God against the horizon of Nietzschean metaphysics, which automatically brings the Dionysian within reach. In purely referential terms, it is therefore easy to create a complementary proximity between all three figures. It promises the combination of the ancient world (Dionysus) and that which is still present (Christ) with a world to come (the last God), which sets everything right. In doing so, it condenses the ecstatic with the most inward and raptures this, as it were, into a new, still ineffable Logos. It is evident that this new soteriological type thus remains the expression of a free and open poetry. The "last God Christ-Dionysus" is neither a theory nor a concept, but a hyperbolic intuition. However, because Christianity, in contrast to the "mere thinking" of Nietzsche and Heidegger, has the modal primacy of revelation, the following hypothesis suggests itself: The return of Christ *ta eschata* will occur in a Dionysian manner, so to speak, as a kind of monstrous music between rapture and potency, intoxication and ecstasy, delusion and light. In this way, the event character of the last God also lends itself to it, where this can only be described in the paradox of a distant proximity, but concerns us in our innermost being. This approach must be an explicitly bodily one, where existence sees itself enraptured by the Dionysian forces towards the pneumatic body of Christ. Seen in this light, the promise of the resurrection of the bodies is also a poetic overall-event that resonates here in the principle modes of all three figurations of God. The complementarity of this event remains decisive: the coming God *will*, *may* and *can* save us from himself – by grace – but we ourselves are called upon in the Kierkegaardian *Imitatio Christi* (towards the knight of faith), the Nietzschean self-poetry (towards the Übermensch) and in the Heideggerian Eigentlichkeit of care (towards Dasein) to prepare the ground for our salvation, redemption and divinisation. Or as Johann Wolfgang von Goethe writes in *Faust II*: «Who ever exerts himself in constant striving, Him we can redeem» [20].

[18] Both are hybrid beings, non-biologically begotten, both prove their divinity through miracles, both are killed and resurrected, their followers are persecuted, both provoke mystical ecstasies, both are saviours, relevance of wine in both figures.

[19] M. Frank, *Der kommende Gott*, Suhrkamp, Frankfurt am Main, 1982, 9th lecture (pp. 245-284).

[20] J.W. Goethe, *Faust. Der Tragödie zweiter Teil*, Benno Schwabe Verlag, Basel 1949, 11936-11937.

ISSN 0039-4130

An all-too-modern Modernity

A Genealogical Investigation

Gael Trottmann-Calame

«*What is it* here that hates so much?»
FP XIV, 14[134]

«Man is something to be overcome. What have you done to overcome him?»[1]. Nietzsche's call for a certain overcoming [*Überwindung*] of man has remained famous. As a vital issue, Zarathustra's disciple never ceases to summon a "new" human type beyond the all-too human. A «higher form of existence»[2], the «next degree»[3] of man, the "*Übermensch*" is invoked with constant insistence. However, far from being an *accomplishment* or *completion* of humanity (*givenness*) – which would presuppose an essence or human nature to be finalized (teleological becoming) - man "to come" or man's "future" is rather *creation*, the object of constant conquest, of perpetual surpassing – «a bridge, not a goal»[4] (Heraclitean becoming). Nonetheless, reading Nietzsche's ceaseless exhortation to a new type of human[5], it is tempting to see him as the prophet of the very contemporary craze for a certain "post-humanism", or even "trans-humanism". True fantasies born of a mechanistic and materialistic understanding of the human mind (neuroscience), the "augmented" man (super-human), if not the "non-human" (cyborg), the unlimited intelligence (AI), do they not find in Nietzsche a "brilliant" precursor?

[1] *Thus Spoke Zarathustra* (*TSZ*), "Prologue", §3. (*Ainsi parlait Zarathoustra*, trad. G.-A. Goldschmit, LGF-Le Livre de Poche, Paris, (1972), 2000. (All translations from French are by us).

[2] *Posthumous fragments* (*FP*), II*, 19[45], in *Œuvres philosophiques complètes*, 14 tomes, 18 volumes, Gallimard, 1968-1997.

[3] *FP* XI, 16 [6].

[4] *TSZ*, "Old and new tables", §3.

[5] *FP* XIV, 14[8]: «Humanity is only the material of experience, the enormous surplus of what has not succeeded, a field of rubble...».

In truth, if we are to follow the philosopher from Sils-Maria, these modern temptations [6], far too modern, would rather be the result of a regrettable "nihilism", revealing its apotheosis. This would be the last gasp of the "last man". And, insofar as we are genealogists, there's a good chance that behind these contemporary fantasies we can unmask the instincts tirelessly denounced by the philosopher with the hammer. Is it not, in fact, a certain weariness, a certain disgust, if not a deep resentment of "reality" and the human race, a feeling of irremediable powerlessness - in short, the "weakness" or "decadence" characteristic of a certain type of human being - that is expressed in this aspiration for a different kind of Humanity? So far from embodying Nietzschean self-transcendence (*Selbstüberwindung*), the "*other* human" is rather a fall, a negation - in short, the assumption of the "fragment man", the "last man". Poison rather than cure, the vain quest for a humanity to be augmented rather than metamorphosed invites us to ask, with and following Nietzsche: «*What is it* here that hates so much»?

However, if our investigation were to result in the observation of a "secret rage" against man and «the primary conditions of life» [7], it would be regrettable to remain at the sad diagnosis of a blocked horizon: the end or death of man [8]. Rather, the survey will reopen the ho-

[6] Let's be clear from the outset. It would be unfortunate, if not dangerous, later on, caught up in the reels of "moralism", dizzied by the filters of the "Circe of the philosophers" (morality), to remain at the superficial level of moral criticism, when a far more fundamental and radical genealogical inquiry is incumbent upon us. In other words, in confronting the ideals that preside over post-humanism, what follows is not about affirming or defending another moral perspective in the name of more "valid", "better", more "humane" moral principles. To put it plainly, if we don't want to stagnate in the impasse of an unproductive moral "confrontation" (good versus "good", evil versus "evil"), we need to situate ourselves beyond good and evil, or below them, at the origins of the origin of "good" and "evil". In other words, we need to respond to Nietzsche's demand, "my demand": «to place ourselves *beyond* good and evil, - to have risen above the illusion of moral judgment», because, quite simply, "*there are no moral facts at all*" (*Twilight of the Idols* (*CId*), "Those who make humanity 'better'", §1; *Crépuscule des idoles*, trad. É. Blondel, GF-Flammarion, Paris, (2005), 2023). What follows should not be read as a "critique" or "refutation" of post-humanism, but rather as its possible genealogy (a regressive and amoral moment of diagnosis and symptomatology). Post-humanism's own "moral judgment" on "man" is thus not to be refuted morally, but taken as *semiotic*, in that it informs us (genealogy) about the *sense of* man that this perspective has constructed: «it remains inappreciable as *semiotics*: it reveals, at least for the one who knows, the most precious realities of cultures and interiorities that *did* not *know* enough to 'understand' themselves» (*CId*, "Those who make humanity 'better'", §1).

[7] *FP* XIV, 14[134].

[8] It is one of Foucaultian tendencies to have been overly obsessed with the "death of man" (following that of God), and to have focused perhaps too exclusively on the announced end of man as a recent invention with the 21st century and its "modifications" of man, rather than on his necessary and possible rebirth beyond his "death". Cf. M. FOUCAULT, *Les mots et les*

ISSN 0039-4130

rizon by showing that, until now, «yes, man was an experiment» [9], and that a new human type remains possible and to be *created,* that a *work* remains to be done, provided that man's formidable «*poetic* force» [10] is nourished by an affirming hierarchy of impulses, ascendant rather than decadent, aspiring to nothingness.

I. *Confronting the "appalling"*

As we know, Nietzsche's interpretation of the demanding question of "reality" is original, radical and, to say the least, "terrible". Subjected to an incessant *becoming*, "reality" can only be read through the prism of *struggle* (agonal jousting or the mutual relationship of impulses, instincts, affects or wills to power, understood as the sole search for resistance in order to feel that the feeling of power is growing). All the more so, devoid of reifiable and substantial "constituents", "reality" (the body of the world and the world of the body) is merely an irreducible multiplicity, an unfathomable obscurity and depth, describable only by recourse to the metaphor [11] of *chaos*: «*Chaos sive natura*» [12]. «*Schrecklich und fragwürdig*» [13] – terrible and problematic, fearsome and enigmatic – so is Nietzsche's most dry, clear and illusion-free [14] interpretation of reality, or at least the least false [15]: it is the ever-new fruit of the

choses, Gallimard, Paris, 1990. p. 318-323: «[...] before the end of the eighteenth century, man did not exist [...]. He is a very recent creature that the demiurgy of knowledge manufactured with its own hands, less than two hundred years ago». p. 319.

[9] *TSZ*, "The bestowing virtue", §2: «Ja, ein Versuch war der Mensch.»

[10] *FP* V, 11[18].

[11] To assert the *metaphorical* nature of the term chaos is to deny that the term is *denotative*, and even more, to point out the illusion and lie of all denotation. Thus, to affirm chaos as a metaphor is to refuse - and once and for all to twist its neck - any metaphysical pretension which, on the one hand, would arrogate to itself the right to totalize multiplicity under a single term ("reality", "unity", "essence", "substance", "being", etc.), and on the other hand, to deny the right of the metaphysicist to claim that chaos is a metaphor.), and on the other hand, would see in or behind "the text of appearances", this same totality subsisting as it "is" independently of the interpretations that man might have of it (as if behind every world there remained *the* original "world"). The impossibility of any totalization or assignment of origin, then, to which chaos as *metaphor* forces us.

[12] *FP* V, 11[197]. Cf. *The Gay Science* (GS), §109 (*Le Gai Savoir*, trad. P. Wotling, GF-Flammarion, Paris, (1997), 2020): «the general character of the world is [...] chaos from all eternity».

[13] *FP* XIII, 11[228].

[14] Cf. *Beyond Good and Evil* (*BGE*), §39 (*Par-delà bien et mal*, trad. P. Wotling, GF-Flammarion, Paris, 2000).

[15] Cf. *FP* XII, 1[120].

ISSN 0039-4130

opposition of wills to power. Those who are not afraid to look into the eyes of life [16], even if they have «the courage to face the appalling» [17], can only emerge terrified with the realization that only one metaphor can describe its nature: *chaos*. Now, if life's character is such, it's not surprising that man – or at least a certain type of man – comes to lament this life, if not hate it, and seeks in illusory "after-worlds" and in an "other" Humanity vain escapes and dangerous means of consolation [18].

II. *Resentment or "the great school of slander"*

In fact, «*who* alone has reason to *escape reality by lying*? He who *suffers* from it. But to suffer from reality means to be a *stricken* reality...» [19]. As we can see, the need to escape reality rather than confront it (agonal relationship), or even overcome it, the need to conceal reality and hate it (polemical relationship) rather than bless it, are all signs of *decadence*, of a stricken reality prey to *resentment*. «What is it *here that hates so much*»*?* A *type of life*, a physiologically weak human *type* that can't and won't face up to the formidable depth, multiplicity and complexity of human instincts and reality (*struggle*, *chaos*).

Dominated by negative affects, gorged with hostility and devoid of the strength and creativity needed to surpass oneself, terrified in the face of the sum of suffering and misfortune inherent in reality, the weak type - this «liar who conceals reality» [20] *[Weglügner der Realität]*- projects «other worlds» [21] beyond this reality, and aspires to another Humanity ("Last Man"), with only his own happiness and an absence of suffering as its horizon and goal [22]. As a genealogist, we unmask the drive anarchy of this resentful modern type. In the absence of a drive hierarchy, drives that can only intensify by slandering and violently de-

[16] *TSZ*, "The Song of Dance".

[17] *FP* XIII, 11[228].

[18] *FP* IV, 4[230] : «To refuse to see a bad thing, not to admit that it exists, to deny it, to change its meaning, to place one's intellectual honor in this negation - *a means of consolation*.» (emphasis added).

[19] *The Antichrist* (*AC*), §15. (*L'Antéchrist*, trad. E. Blondel, GF Flammarion, Paris, 2022).

[20] *FP* XIV, 23[4].

[21] *CId*, "The four great errors", §5 : «The unknown arouses danger, worry and concern - the first instinct is to *eliminate* these unpleasant states. First principle: any explanation is better than no explanation. Since it's basically a question of evacuating oppressive representations, we don't pay much attention to the means of evacuating them.»

[22] *AC*, §1: «I no longer know where to turn; I am all that no longer knows where to turn - sighs modern man....»

ISSN 0039-4130

nying an "enemy" (*resentment*) come to predominate, creating a fictitious reality and humanity[23]. Slandering and hating the human he could be (the "beast of prey", the "master", the "strong"), the "weak" fantasizes and invents another, "better" human[24]; another human who is to be "trained" [*Zähmung*], "improved"[25] [*Verbessern*] by means of morality, a morality of "slaves" and idealism (dissatisfaction, fear and shame of instincts[26]). And all this at the risk of ending up "disgusting" oneself, "hating" oneself and constituting a "great danger to man" by foreshadowing his end ("last man"). The paradox of resentment, as we can see, leads to a paradoxical "disgust", "weariness" and "hatred" of oneself: the weak type comes to hate and be disgusted by the type he "is" (weak), as a result of hating the type he can't reach (strong) - and aspires to nothingness.

What's so surprising, then, to see the "modern" human type, in an all-too-modern reflex, fantasizing about a radically different human, a "post" human, a "trans" human, an "improved" human, an "augmented" human, etc.? Far from aspiring to an overman – which would require a tremendous effort to surpass oneself – the relatively weaker human type despairs of a "super-human" – the form of which actually matters little – as long as it saves him from the "too human" that he is, without effort or suffering, even if this "super" or "other" human is the assumption of the "last man": «Give us this last man, O Zarathustra, they cried, make us this last man! And we grace you with the overman!»[27].

[23] *On the Genealogy of Morality* (*GM*), I, §10 (*Éléments pour la généalogie de la morale*, traduction P. Wotling, LGF, Paris, 2000): «the uprising of slaves in morality begins with the fact that *resentment* itself becomes a creator and gives birth to values: the resentment of those beings [...] who compensate themselves only by means of imaginary revenge. [...] Slave morality from the outset says 'no' to an 'outside', to an 'otherwise', to a 'not oneself': and it is *this* 'no' that is its creative act.»

[24] *GM*, I, §13: «We, the weak, are indeed weak; it is good that we do nothing *in view of which we are not strong enough*.»

[25] Cf. "*CId*, "Those who make humanity 'better'", §2: «From time immemorial, people have wanted to make men 'better': it was above all this that was designated by morality. [...] To put it in physiological terms: in the fight against the beast, making it sick *may well* be the only way to make it weak. This is what the Church understood: it *ruined* man, it weakened him, - but it claimed to have 'made him better'».

[26] Cf. *GM*, II, §24. Admittedly, the formidable human "bestiality", the unleashing and surging of disturbing impulses in search of power, this darkness and depth of the "body" can be disturbing, but it's one thing to try to conquer, to attempt to conquer, dominate and spiritualize [*Vergeistigung*] these "subterranean forces" by overcoming them, it is quite another to slander them, hate them, and attempt to annihilate them by fantasizing a humanity made up of "good men" "moralized from top to bottom" (*GM*, III, §19).

[27] *TSZ*, "Prologue", §5.

ISSN 0039-4130

III. *The "meaning of the earth"*

Contemplating a type of man prey to such "disgust" [*Verdruss*] with himself, and despairing of being "other", would we not then have every reason to feel a real but dangerous "disgust" [*Ekel*] for the human, for this all-too-modern modernity, as sometimes seems to have been the case for Nietzsche[28]? And yet. «I love men», Zarathustra reminds us. True, but the herald of the overman immediately adds: «Man is something to be overcome». Nietzsche adds – and this is a crucial nuance: «the problem I thus pose is not what is to replace mankind in the series of beings (-man is an *end-*): but what kind of man is to be *elevated*, what kind of man is to *be willed* as the one who has a higher value, who is more worthy of living, more assured of a future»[29].

As we can see, Nietzsche does not limit himself to a disgust for the human, but focuses instead on the "promise" that the human contains and to which it can give life. For Zarathustra's disciple, the human is something to be willed and desired. But then a new type of human, a new attempt, a human not replaced but overcome (a direction towards constant self-depassment, a "bridge"), a human or overman that would not be a new "ideal" ("post", "trans" etc.), but a human that would be "self-replaced". A human who would attach himself to "reality", who would love it, confront it and assume it, who would be the «meaning of the earth»[30] and not its hatred and flight into "back-worlds", who would assume himself in his "godlike nakedness", who would recognize himself as a "body" and as a "body" to be loved rather than augmented[31]. And this human is only possible on condition that the formidable "poetic force" with which man is endowed finds expression not by being mobilized by weakness but by strength, and does not seek to cast opprobrium on precisely what makes man human (as well as animalistic): «when we speak of *humanity*, we base ourselves on the idea that it could well be what separates man from nature and distinguishes

[28] *Ecce Homo* (*EH*), I, §8 (*Ecce Homo*, trad. É. Blondel, GF-Flammarion, Paris, (1992), 2023) : «The *disgust of* man [...] has always been my greatest danger.»

[29] *AC*, §3.

[30] *TSZ*, "Prologue", §3.

[31] *FP* IX, 5[30]: VYou mask your soul: nudity would be scandalous for your soul. Oh, learn why a god is naked! He is ashamed of nothing. He is more powerful naked! The body [*Körper*] is something evil, beauty a devilish thing: thin, awful, hungry, black, dirty, such must be the appearance of the body [*Leib*]. To commit a crime against the body [*Leib*] is, in my eyes, equivalent to committing a crime against the earth and against the sense of the earth. Woe betide the unfortunate for whom the body seems evil, and beauty diabolical!»

ISSN 0039-4130

him from it; but, in reality, this separation does not exist: the "natural" properties and those said to be properly "human" have become inseparably intertwined. In his noblest and highest faculties, man is all nature, and carries within him the strangeness of this dual natural character. His formidable aptitudes, which are considered inhuman, are perhaps even the fertile ground from which any humanity can emerge in the form of emotions, actions and works»[32].

[32] *Homer on competition*, preface of 1872 (*La Joute chez Homère*, extrait de *Cinq préfaces à cinq livres qui n'ont pas encore été écrits*, in *Œuvre*s, I, La Pléiade, Gallimard, Paris, 2000).

ISSN 0039-4130

Language and Soteriology: Desire, Illusion and Liberation in Wittgenstein's and Buddhist Philosophies

Tomaso Pignocchi

1. *Philosophy as practice*

My intention is to argue that Wittgenstein's later thought can be seen as a philosophy aimed at producing a certain effect: a reorientation of one's perspective on reality, or, in his own words, a change in «the way one sees things» [1]. This shift aims not only to transform one's worldview but also to free individuals from what makes their existence burdensome and unsatisfactory. For this purpose, I will briefly show how this effort aligns – through its aims, methodology, and focus – with a neglected tradition in Western philosophy: the constellation of ideas we refer to as Buddhism. This comparison can highlight how Wittgenstein's purpose is not primarily concerned with theoretical goals, but rather focuses on an analogous practical and ethical aim, thereby clarifying certain passages of his thought that have often remained obscure or underappreciated. Furthermore, I believe that "comparison" is one of the most fruitful methods for approaching the history of philosophy, since «finding that two philosophers have independently and by different paths arrived at the same conclusions» brings, paradoxically, even greater prominence to "what" they have argued, precisely because «when the differences in cultural background, historical era, personality, and mental framework between

[1] L. Wittgenstein, *The Big Typescript: TS 213*, Blackwell, Oxford 2005, p. 300.

two philosophers are substantial, the ideas they share become all the more significant» [2].

I will therefore interpret Wittgenstein's philosophy in the context of a practical conception of philosophy. This approach sees the investigation of reality not as an objective, detached exercise, but as a means to an existential end: the practical search for "conversion". This is something very akin to the concept of *epistrophé* in Stoicism, which was identified by Michel Foucault and Pierre Hadot as the first example of practical philosophy aimed at "self-care", a goal to be achieved through what the latter called "spiritual exercises" [3]. Notably, Hadot himself revealed that the idea behind his interpretation of ancient philosophy came to him after reading Wittgenstein's *Philosophical Investigations*, of which he was an early admirer in France [4].

Finally, I should note that I use the term "soteriology" in a loose sense. While in his notebooks and conversations Wittgenstein occasionally says that the goal of his philosophy is the search for an *Erlösung* – that is, a redemption – he does so in a figurative and metaphorical way. There is no doubt that Wittgenstein was deeply fascinated by an honest and genuine religious life, and not by chance Tolstoij and Kierkegaard were two of his moral reference points. In his philosophy, however, there seems to be no God or savior. Similar to Buddhism, there is only a speaker who seeks to guide both the listeners and himself toward a state of awareness that enables them to save – or rather, liberate – themselves. This "liberation" is essentially the release from a bewitchment or trap.

2. *Desire and frustration*

Those who are familiar with the *Philosophical Investigations* will know that this text unfolds in the form of a dialogue between two voices, dubbed by Stanley Cavell as «the voice of temptation» and «the voice of correctness» [5], which aims to persuade the reader, but not by pro-

[2] R. Assagioli, *Johann Georg Hamann e Ralph Waldo Emerson: alcune curiose coincidenze tra le loro idee*, in *Bericht über den III. Internationalen Kongress für Philosophie*, Carl Winter's, Heidelberg 1909, p. 278; my translation.

[3] See P. Hadot, *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault*, Blackwell, Oxford 1995.

[4] Id., *La Philosophie comme manière de vivre*, Albin Michel, Paris 2001, p. 214.

[5] S. Cavell, *Must we mean what we say?*, Cambridge University Press, Cambridge 1969, p.71.

ISSN 0039-4130

viding rational "reasons" to force an external conversion, but by enabling them to liberate themselves from an unsatisfactory way of seeing things. In fact, Wittgenstein departs from the cognitive conception of philosophy, favoring a practical approach: he repeatedly emphasizes that his philosophy is defined by a method or style of thinking, not by doctrines, because without individual experience one cannot change their perspective, as no "reason" can mechanically move the will. Indeed, as Wittgenstein himself writes:

> «Difficulty of philosophy is not the intellectual difficulty of the sciences, but the difficulty of a change of attitude. Resistance of the will must be overcome. [...] Philosophy does not lead me to any renunciation, since I do not abstain from saying something, but rather abandon a certain combination of words as senseless. In another sense, however, philosophy does require a resignation, but one of feeling, not of intellect. And maybe that is what makes it so difficult for many. It can be difficult not to use an expression, just as it is difficult to hold back tears, or an outburst of rage» [6].

Wittgenstein's philosophy is in fact an exercise in dialectics and maieutic: through images – such as similes, metaphors, analogies, and everyday examples – it aims to stimulate readers to produce, within themselves, a shift in emotional attitude that will help them overcome a resistance of the will and make the 'leap' on their own. In a Schopenhauerian sense, it is a free act of will that liberates us from the very enslavement caused by the will.

But taking a step back, we must first understand what we need to liberate ourselves from: an illusory view of things that creates an unsatisfactory and often painful relationship with reality, where pain arises from the gap between our expectations and what reality actually offers us. Let us read Wittgenstein's own words again, in a passage not far from the one we have just examined:

> «What makes a subject difficult to understand – if it is significant, important – is not that it would take some special instruction about abstruse things to understand it. Rather it is the antithesis between understanding the subject and what most people *want* to see. Because of this the very things that are most obvious can become the most difficult to understand. What has to be overcome is not a difficulty of the intellect, but of the will. As is frequently the case with work in architecture, work on philosophy

[6] L. Wittgenstein, *The Big Typescript: TS 213*, cit., p. 300.

ISSN 0039-4130

> is actually closer to working on oneself. On one's own understanding. On the way one sees things. (And on what one demands of them)»[7].

If for Wittgenstein what we "demand" is the ideal purity of a disembodied and transcendent language, with rules that operate independently and absolutely – and thereby guarantee us with meaning in an indubitable way – what we actually experience in everyday life is a language that is always open to confusion, misinterpretation, and error.

> «The more closely we examine actual language, the greater becomes the conflict between it and our requirement. (For the crystalline purity of logic was, of course, not something I had *discovered:* it was a requirement.) The conflict becomes intolerable; the requirement is now in danger of becoming vacuous. We have got on to slippery ice where there is no friction, and so, in a certain sense, the conditions are ideal; but also, just because of that, we are unable to walk. We want to walk: so we need *friction*. Back to the rough ground!»[8]

Yet, it is precisely the possibility of error, misinterpretation, and confusion that makes language work: to walk, we need rough terrain where we might even stumble, not a perfectly smooth sheet of ice where every step keeps us stuck in place.

3. *Illusion and language*

At this point, it's worth drawing a parallel with Buddhism. For the Buddha, the problem of existence also stems from will and expectations. We desire eternal, autonomous, and absolute things, but instead we are confronted with transience, death, and interdependence in everyday life. Humans project their desires onto reality, which inevitably leads to disappointment. Both Wittgenstein and the Buddha trace the source of these illusions to language, where we mistake the role certain words play in everyday use. While concepts and ideals play a decisive role in our everyday language – *viz.* they serve as practical standards of comparison –, the mistake lies in believing they are «plugged into reality»[9] as Wittgenstein puts it. Believing that the ideal belongs to reality cre-

[7] *Ibid.*

[8] Id., *Philosophical Investigations*, §107, p. 46.

[9] Id., *Philosophical Investigations*, §100-101, cit., p. 45. English translation modified by me to better match the original German text.

ISSN 0039-4130

ates the constant suffering of realizing that our lives are always quite different from such an ideal purity that we can never attain.

Similarly, for Buddhism suffering arises from experiencing the transience of things we assume to be eternal. And this is particularly central when we think about the status of the "Self", the "I", (or the soul), i.e. that element which is believed to be stable and essential to our individuality. For the Buddha, the question of the self is central and closely tied to the question of "essences" in general. The fact that there is no individual and independent "Self" is, in fact, the principal tenet of Buddhism (along with the realization that existence is inherently unsatisfactory and that everything is transient).

Our expectation that the Self is something in itself, that it is an essence, and as such is stable and eternal, is primarily due to the seduction that language exerts on us: we mistake concepts, which are merely tools of language, for real things. Using a famous buddhist example, when we hear the word "chariot" we think it refers to something in itself, when in fact it is only a linguistic convention we use to talk about an assemblage of wood, metal, and ropes with a particular practical function. In the same way, when we hear the word "Self" (Ātman), we believe it to be something in itself – that is to say: something absolute, stable, and eternal. However, in the same way of "chariot", also "Self" is merely a conventional term that refers to a dynamic, interconnected flow of physical and mental processes that is constantly changing and never remains the same [10]. Dissatisfaction arises when we become attached to certain words and concepts that we should instead let go of, or at least reconceptualize: through analysis, in fact, we can dissolve these illusions and return them to the causal relationality that governs all things.

Returning now to the Wittgensteinian framework, "liberation" is realizing and accepting that our thinking is unavoidably shaped by the perspective we inhabit: it's not about attaining a perspective-free viewpoint, but rather reaching a viewpoint that is aware of being irredeemably perspectival. Liberation, thus, involves accepting things as they are, pulling them down from the metaphysical realm where we imagine objective truth resides, and bringing them back to the ordinariness and banality of everyday language, where meaning is not guaranteed by fixed, transcendent definitions but is instead shaped through a contin-

[10] This metaphor appears several times in the Buddhist canon, with the most famous instance being SN 1.5.10; see B. Bodhi (ed.), *The connected discourses of the Buddha: a new translation of the Samyutta Nikāya*, Wisdom Publications, Somerville 2000, p. 230.

ISSN 0039-4130

uous intersubjective process of tacit and, in some ways, unconscious redefinition. In fact, for Wittgenstein, the meaning of a word lies in its "use", which is always contextual, and this implies that a word's meaning at a particular time and place is not fixed; it may be used differently in another context and still be perfectly understood by speakers.

In a similar way, we might say that a game is valid only as long as its players intuitively follow its rules, which are not fixed forever. Take, for instance, the game of chess: if suddenly one player starts moving the pawns three squares at a time, and the opponent responds in the same way, could we really say that they are making a mistake? Perhaps we could no longer call their game "chess", but there is no doubt that this remains a game and that it is working: that is to say, it continues to progress according to its new rule without issues. It is crucial to recognize that every aspect of our experience of reality is never something that is rigidly determined with sharply defined boundaries. One must learn to accept that a blurry photo of a face is still a portrait, and that even the rules of a game like tennis can establish many things but cannot specify exactly how high a ball should be hit [11]. Our problem is that we believe that a game is only truly a game if its rules take every possible scenario into account, just as we think a word can only have meaning if it gives us all its possible applications in advance. This is the illusory guarantee that we desperately crave, and is the same type of guarantee that we seek in metaphysical foundations, *a priori* rules, and disembodied, transcendent essences.

4. *Abandoning nirvāna*

Ironically, this was also the same foundation Wittgenstein previously sought in his *Tractatus Logico-Philosophicus*, believing he had found the general rules of an ideal, pure language that would mirror the facts of the world perfectly. However, to achieve this, he had to exclude from language the so-called "Mystical" realm. This realm includes those aspects of life that are truly valuable, such as morality, aesthetics, religion and the very meaning of existence, all of which cannot be meaningfully conveyed through words. Actually, according to the *Tractatus,* one can speak meaningfully only of the world of facts, but the world of facts says nothing about the sense of our existence. Going down this way,

[11] Cfr. Id., *Philosophical Investigations*, §68-71, cit., pp. 33-34.

ISSN 0039-4130

concluding his *Tractatus,* Wittgenstein created an irredeemable dualism: on the one hand, we have the "world of facts", while on the other we have the "world of Mystical". But of the latter we must be silent, even though it is the only one that really matters to a human being, for it is Wittgenstein himself who states that when everything has been described and said about the facts of the world, the real problems of human existence «have still not been touched at all». [12]

But later, in the *Philosophical Investigations*, Wittgenstein will realize that the "world of mystical" and the "world of facts" coincides: because the "world of Mystical", understood as transcendence, is merely our ordinary world seen through the illusory lens of metaphysics; indeed, it is precisely the belief in the "Mystical" as a separate and profoundly meaningful realm that deceives us and keeps us bound to a tragic worldview, in which we are hopelessly separated from an unattainable meaning. The key is to recognize that we are driven by what Wittgenstein, in the *Blue Book*, called a «craving for generality» [13]: a deep desire for transcendence that seeks to escape the roughness of becoming and multiplicity by chasing the smooth purity of an illusory universality. However, the new philosophy of the *Investigations* shifts meaning from the heights of transcendence to the ordinariness of everyday life. Meaning is now found within the world, and our mistake lies in looking for it elsewhere. But already in 1930, before the *Blue Book*, Wittgenstein showed in a notebook note that he had begun to move beyond the ideas of the *Tractatus*:

> «I might say: if the place I want to get to could only be reached by way of a ladder, I would give up trying to get there. For the place I really have to get to is a place I must already be at now. Anything that I might reach by climbing a ladder does not interest me» [14].

Wittgenstein is referring here to the well-known paradox of the ladder with which his *Tractatus* concludes. [15] However, in his new perspective, he recognizes that needing a ladder to climb implies assuming the existence of a separate plane from the one we're on. And it is precisely this assumption that gives rise to metaphysical questions and the consequent frustration of being unable to answer them. Reality is in fact

[12] Id., *Tractatus Logico-Philosophicus*, §6.52, Routledge, London 1955, p. 187.
[13] Id., *The Blue and Brown Books*, Blackwell, Oxford 1964, p. 18.
[14] Id., *Culture and Value*, Blackwell, Oxford 1980, p. 7.
[15] Cfr. Id., *Tractatus Logico-Philosophicus*, §6.54, cit., p. 189.

ISSN 0039-4130

just one, and if it seems split it is only because our perception is subject to an illusion. The goal, therefore, is to return to where we already are, but with new eyes. In other words: to open our eyes is to realize that we are already where we wanted to be. This is the same concept presented by Nāgārjuna, an important Buddhist philosopher and logician of the 2nd century CE, perhaps the most significant Buddhist thinker after the Buddha himself:

> «There is not the slightest difference between transmigration (*samsāra*) and *nirvāna.* There is not the slightest difference between *nirvāna* and transmigration. What is the boundary of *nirvāna* is also the boundary of transmigration. There is not even the slightest difference between the two. The ideas [concerning such things as] whether the Buddha continues to exist or not after death, those of the end [of the world], and so on, those of eternal existence, and so on, have, as their basis, the idea of *nirvāna*, [that is] of a posterior and an anterior extremity.» [16]

Starting the analysis from the latter sentences, we see that in the Buddhist canon disciples often ask the Buddha such questions, but he never answers. This is both because these questions may have no answers and because posing them is in itself harmful. Posing these questions only sustains the illusion, reintroducing the very problem of existence. And this is what connects these sentences to the first part of the text: the illusory māyā is not external, in the world, but internal to us. If we seek *nirvāna*, we are already within the illusion. Paradoxically, it is the pursuit of *nirvāna* that creates *samsāra* – the cyclical world of illusion and continuous rebirth – and leads to a life of dissatisfaction.

But ultimate reality shows that there is no difference between the world of dissatisfaction and the world of inner-peace: they are the same world. True *nirvāna* is nothing more than the abandonment of the idea of *nirvāna*. Metaphysical questions with nefarious outcomes are in fact rooted in a dualistic conception which separates the "world" from the "meaning of the world". But once this dualism is eliminated, all the false questions it entails are also eliminated. Therefore, real liberation is not being liberated from the world of the here-and-now, but being liberated from the idea that the world of the here-and-now is something that we have to be liberated from.

[16] Nāgārjuna, *Madhyamaka-kārikā*, XXV:19-20, in *The Fundamental Wisdom of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*, transl. by J. L. Garfield, Oxford University Press, New York 1995, p. 331-333. Translation slightly modified by me in accordance with the Italian translation by R. Gnoli, cfr. *Testi Buddhisti in Sanscrito*, UTET, Torino, p. 373.

ISSN 0039-4130

Moral Luck: an Accessible Exploration

Marco Tassella

Introduction to the Concept

The idea that moral judgment should be independent of luck and fate is not just a core concept of deontological ethics, but also aligns with our everyday moral intuitions. This perspective, originating in Abelard's philosophy, was later developed by Immanuel Kant's ethical framework. Kant argues that the actual moral worth of an action is determined by the original intention behind it, rather than by its contingent consequences. In his *Groundwork*, the philosopher famously stated: «a good will is not good because of what it effects or accomplishes, [...] but, like a jewel, it would still shine by itself » (Kant 1996: 50, 4:394). This captures the belief that moral value should remain unaffected by luck.

Kant's position, however, extends beyond this abstract description; the notion that morality and luck should not intertwine also resonates with common moral sentiment: good intentions are often praised, regardless of the outcome, and the moral worth of an action is viewed as independent of the agent's actual abilities. Consider, for instance, someone volunteering to help change a flat tire. Regardless of their skill level, those who help are typically regarded as "good" because their intention was benevolent. In such cases, the ability to control the outcome is often secondary to the good or bad intention behind the action. This moral understanding is captured by the Control Principle (CP), which states that moral praise and blame should only apply to factors that are within the agent's power:

CP. We are morally assessable only to the extent that what we are assessed for depends on factors under our control.

In the debate on moral responsibility and free will, this view aligns with libertarian intuitions about desert: an agent is only morally re-

sponsible if they have control over their actions and could have acted otherwise (that is, if they had alternative courses of action). A corollary to this principle suggests that identical intentions should not be judged differently based on uncontrollable outcomes:

CP-Corollary. Two people should not be morally assessed differently if the only difference between them is due to factors beyond their control.

Despite this normative principle, however, luck undeniably influences our lives and moral assessments. On closer examination, intentions, actions, and outcomes often depend on uncontrollable factors, including circumstances, past experiences, and character. For example, two people might perform the same reckless act with identical intentions, but by chance, only one causes actual harm. Since harm is often the basis for moral judgment, the one who caused harm is judged more harshly than the other.

Such cases reveal a paradox in our moral thinking: although we theoretically want to judge based on intentions alone, outcomes and circumstances frequently sway our evaluations. This discrepancy hints at a deep tension between our theoretical commitment to moral responsibility and the reality of how judgments often occur, which is potentially influenced by luck.

This paradox was brought to the forefront by philosophers Thomas Nagel and Bernard Williams, whose seminal work on moral luck reignited the debate on the intersection of chance and moral responsibility. Their work fundamentally challenges our Kantian—and intuitive—notion of moral agency, revealing a more nuanced framework that integrates luck into our understanding of responsibility. As Nagel notes, «Where a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgment, it can be called moral luck». (Nagel 1979, 26). Thus, luck enters the moral domain on various levels: we not only lack control over the consequences of our choices but also over aspects like our character, reasons for acting, and even the circumstances in which choices arise.

Moral Luck in Society and Law: Rethinking Responsibility and Judgment

The implications of moral luck extend beyond philosophical discourse, affecting social and legal norms. The concept of "legal luck" in criminal justice mirrors moral luck by highlighting how the outcomes of

ISSN 0039-4130

an action can influence sentencing. Legal systems often impose harsher penalties for crimes that result in harm—even if the outcome hinges on mere chance. For instance, an attempted crime that fails typically incurs a lesser penalty than a similar crime that succeeds due to sheer luck. This disparity raises fairness concerns: should someone be punished more harshly simply because of unfortunate outcomes or lucky circumstances?

In the United States, this tension is reflected in debates around the *Model Penal Code*, which has proposed eliminating certain sentencing disparities in favor of focusing on intent rather than outcome. Adopting a moral luck framework in legal contexts could reshape the justice system by prioritizing intentions over luck-influenced outcomes. However, such a shift would require reevaluating longstanding principles that underpin our contemporary conceptions of criminal responsibility and punishment.

Beyond the legal sphere, moral luck might also reshape everyday moral norms. For example, recognizing the role of luck in life achievements might lead to a reduction of pride or envy, fostering humility and empathy between people. In addition, understanding that many successes involve luck may encourage human understanding, and a greater motivation to act morally from genuine commitment, rather than out of fear of judgment. Conversely, however, the choice of overemphasizing luck's influence on moral desert could lead to the risk of excusing poor behavior or even diluting personal responsibility, potentially undermining the general fairness of moral evaluation.

The Paradox and Philosophical Responses

As mentioned, the debate over moral luck reveals a paradox: while we should acknowledge the influence of luck, luck itself also challenges our sense of responsibility and justice. If, on the one hand, taking moral luck seriously may provide us with a richer view of human action, on the other it could threaten the integrity of moral and legal systems. Philosophers have responded to this tension in three main ways:

1. *Denying Moral Luck*: Some argue that moral luck doesn't truly exist, and hold that people should only be responsible for what they can control. They suggest that luck-based judgments only stem from informational or cognitive misunderstandings, rather than representing legitimate moral assessments. This

ISSN 0039-4130

approach preserves the Control Principle but raises challenges in practice, as actual moral judgments do often incorporate luck-influenced outcomes.

2. *Accepting Moral Luck*: Nagel, Williams, and others advocate accepting moral luck and propose a revision in how we approach the concept of responsibility. Accepting moral luck might mean focusing on rehabilitation over blame and emphasizing external factors in judgment. Levy's (2011) work similarly suggests that a fair moral system must consider luck, and implies a possible modification of our judgment criteria to reflect the limited control individuals have over many influencing factors.
3. *Questioning the Coherence of the Problem*: A third group argues that moral luck points to a fundamental issue in our understanding of the concept of responsibility. If actions are shaped by factors outside our control, moral luck may indeed challenge the very basis of moral responsibility. Conversely, this view suggests that moral luck is only useful to reveal the limitations of conventional moral frameworks, potentially requiring a revision or abandonment of traditional ideas of agency and desert.

Each of these responses represents attempts to reconcile the paradox between the Control Principle and our intuitive moral judgments. However, the concept of moral luck continues to reveal complex layers in our understanding of moral responsibility, encouraging us to consider how external factors shape moral agency in both theory and practice.

Types of Moral Luck

In his 1979 paper, Thomas Nagel identified four types of moral luck that uniquely impact moral assessments:

1. *Resultant Luck* occurs when a result or an outcome, influenced by luck, affects how we morally judge the action. For example, if two drivers run a red light but only one hits a pedestrian, we tend to judge them differently–even though their actions were identical.
2. *Circumstantial Luck* concerns the situations a person finds themselves in, which can in turn influence the moral value of their actions. For example, some people might never find them-

ISSN 0039-4130

selves in morally testing circumstances, while others might face extreme challenges due to chance. As Williams (1981) notes, this type of luck reveals the crucial role of situational factors in shaping moral action.

3. *Constitutive Luck*: Relates to an individual's character and dispositions – influenced by factors beyond their control such as genetics and upbringing. Michael Zimmerman (1987) explores how such traits, often inherited or environmentally shaped, may impact moral responsibility.
4. *Causal Luck*: Addresses the chain of events leading to a decision, often beyond the control of an agent. Levy's examination of causal luck emphasizes the interconnectedness of our choices with past events, suggesting a more deterministic view of human action.

The concept of moral luck has significant real-world implications, particularly in law, where it is echoed by the concept of "legal luck": courts often impose harsher penalties based on outcomes rather than intentions. This leads to attempted crimes that fail typically incurring in lesser sentences than those that succeed by chance. This raises many fairness concerns—should someone be punished more harshly simply because of an unlucky outcome?

Conclusion: Rethinking Responsibility and the Broader Free Will Debate

Consequential luck also raises the question of how uncontrollable, yet significant outcomes should influence moral responsibility. Philosopher Fernando Rudy-Hiller suggests that responsibility might not always imply moral desert, but can include forward-looking forms of responsibility such as compensation or self-improvement. This approach might allow us to hold individuals accountable without compromising fairness, acknowledging that intentions–though morally central–don't always align with outcomes.

In addition, the moral luck debate offers a gateway into understanding the interplay between free will and determinism. By reframing Nagel's categories as "causal" and "consequential" luck, for instance, we could gain in clarity while trying to understand moral agency. This separation enables us to explore the influence of chance on moral decisions, recognizing the complexity of real-world judgment. Distinguishing between causal and consequential luck also allows us to address

ISSN 0039-4130

moral luck's apparent paradox, revealing two distinct challenges in moral responsibility. With causal luck tied to pre-existing factors shaping character and consequential luck tied to outcome unpredictability, we gain a richer perspective on moral responsibility and on moral fairness itself. This dual approach encourages a more refined understanding that incorporates both reasonableness and accountability, aligning moral judgment more closely with the actual complexities of human moral experience.

Bibliography

G. D. Caruso - D. Pereboom, *Moral responsibility reconsidered*, University Press, Cambridge 2022.

J. M. Fischer - M. Ravizza (eds), *Freedom and Resentment*, Cornell University Press, 2019, pp. 45-66.

H. G. Frankfurt, *Freedom of the Will and the Concept of a Person*, The Journal of Philosophy, 68, 1971.

I. Kant, *Groundwork of the Metaphysics of Morals* (1784), in *Practical Philosophy*, Edited and Translated by Mary J. Gregor, Cambridge University Press, New York 1996, pp. 37-109.

A. Latus, *Moral and Epistemic Luck*, Journal of Philosophical Research, 25, 2000, pp. 149-172.

N. Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, Oxford University Press, 2011.

M. S. Moore, *Placing blame: a theory of the criminal law*, Oxford University Press, Oxford; New York 1997

T. Nagel, *Moral Luck*, in *Mortal Questions*, Cambridge University Press, Cambridge 1979, pp. 24-38. doi: 10.1017/CBO9781107341050.005.

D. K. Nelkin, *Moral Luck*, Edward N. Zalta & Uri Nodelman (eds.), The Stanford Encyclopedia of Philosophy, 2019.

N. Richards, *Luck and Desert*, Mind, XCV, 378, 1986, pp. 198-209.

D. Statman (ed.), *Moral luck*, State University of New York Press, Albany 1993.

B. Williams, *Moral Luck: Philosophical Papers*, 1973-1980. 1st edn. Cambridge University Press, 1981.

Zimmerman, M.J. (1987) 'Luck and Moral Responsibility', Ethics, 97(2), pp. 374-386.

Zimmerman, M.J. (2002) 'Taking Luck Seriously', The Journal of Philosophy, 99(11).

ISSN 0039-4130

Duchamp, Materiality, and Intersubjectivity: from Phenomenology to Aesthetics

Federico Rudari

This paper, as much as the presentation I delivered on the occasion of the conference Human Freedom at the test of AI and Neurosciences, starts with the same reflection on contemporary artistic production and practices that lies beyond a consistent part of the doctoral project I am currently undertaking. It feels harder, if not, sometimes, impossible, to distinguish a work of art from any other object in today's art scene. What I mean is not a criticism of the lack of technical mastery in contemporary works but rather a question about the value and, ultimately, the existence of an actual ontology of art in the first place. I started by focusing on the phenomenology of aesthetics instead.

In his 2004 book *Spaces, Domains and Meanings*, semiotician Per Aage Brandt suggests that it is *technè*, the ability to practise artful, skilled behaviours, that makes artistic objects "interesting" and worth of "attention" [1]. These techniques are bodily activities by which certain doings and body parts are studied, trained, and practised with elaborate skill. Artists thus need to employ some sort of attention, a specifically artistic one, in their process of making, and, in return, attention is awakened in observing audiences. Moreover, Brandt argues that the implementation of specific techniques is not, *per se,* sufficient to arouse attention, and the experience of cultural practices (including artworks) has been part of collective, ritualised habits since the origins of civilisation, coded and framed in terms of specific environments and conditions. From sexual practices and religious rituals to the cleansed

[1] P. A. Brandt, *Spaces, Domains and Meaning: Essays in Cognitive Semiotics*, Peter Lang, Series European Semiotics, Bern 2004, p. 203.

and neutralised milieu of theatres and contemporary art museums, the role of contexts in such experiences is fundamental.

However, Brandt also addresses the progressive aestheticisation of concepts in contemporary art, bringing the example of Marcel Duchamp's ready-mades, while this claim could be extended to many artistic expressions over the last two centuries. Defined by Brandt as the expression of an escalating «modern collective "*hysterization"* of attention» (emphasis by the author) [2], technical practice is simply reduced to the intentional act of deciding for an object and its right to be displayed in a chosen context. Strongly inspired by, and later involved in, the Dada movement, Duchamp and his ready-mades aimed to question the very notion of art (and Art), the socio-politics around its exhibition, observation, and ultimately adoration.

One of his most famous works, *Fountain*, described by William Camfield as «one of the most famous/infamous objects in the history of modern art» [3], presents a very interesting case for its formal and intellectual properties, authorship, and disappearance. *Fountain* was first exhibited in April 1917 on the occasion of the first exhibition of the American Society of Independent Artists. The porcelain urinal signed "R. Mutt 1917" was displayed at The Grand Central Palace in New York City: the act of choosing an everyday factory-fresh piece of plumbing, according to Duchamp, was enough to make an artwork. Despite the press actively commented the selection of the piece for the exhibition, whether enthusiastically or with indignation, the work did not likewise resonate with the audience since the sculpture was eventually not exhibited for the public, did not figure in the catalogue, and was innocuously described as a "bathroom fixture". It was only later, thanks to a picture taken by Alfred Stieglitz, that *Fountain* gained notoriety, even though the original work got lost and only replicas have survived to the present day.

In the specific case of ready-mades, the chosen artistic practice is not involved in manufacturing the object, which is found complete in its production conditions, but in the selection made by the artist that becomes a whole technique in itself through creative individual authority. Changing the context and the purpose behind the situatedness of an object is enough to redirect the attention of audiences in artistic terms. In Duchamp's case, this mechanism is implemented through the

[2] *Ibid.*, p. 210.

[3] W. A. Camfield, *Marcel Duchamp's* Fountain*: Its History and Aesthetics in the Context of 1917*, in R. Kuenzli-F. M. Naumann (edited by), *Marcel Duchamp: Artist of the Century*, The MIT Press, Cambridge, Massachusetts and London 1987, p. 64.

ISSN 0039-4130

translation of a urinal from a practical object to its artistic conceptualisation as both a fountain and *Fountain*. Since then, most diverse questions have been raised on the nature of *Fountain* and, by extension, artistic practices. The aesthetic conceptualisation of an object of ordinary use became a fully entitled creative act, and even a whole artistic movement, and the choice of display the expression of artistic intentions. A question arises spontaneously: If there is no exhibition, is there ready-made art? Is its status of art only contingent or also substantial?

If the technical aspect that Brandt valued in his work and was earlier introduced in this paper seems to be progressively losing value – and it is, increasingly, in the century that has passed since ready-mades were first exhibited – it could be claimed that it is the practice of 'witnessing' (intended as an expansion of the most traditional spectating, to include watching, listening, smelling, …, and even participating) and its designated environment that predominantly characterises contemporary artistic practices. This habit and its framing define today's art as much (if not more than) art itself: it appears to be more a question of contextual narrative rather than ontology. This spatial determination, as much as the architectural framing, allows us to go beyond and even against purely intellectual interpretation: some historically and monetarily valued paintings might go unnoticed if exhibited in an unusual place, while any object that happens to find itself in a renowned institution can arouse interest and trigger sensorial reactions. Ultimately, any object, and thus any artwork, is always only about itself until it is put in dialogue with other objects, subjects, and anything constituting its surroundings, or until we challenge the way we position ourselves in respect to it: «[a]nything meaningful is meaningful in a "context"»[4].

As digital reproduction and new media have shaped a time of abundant artistic production, many museums and, in general, exhibition spaces have shifted their focus from parameters strictly based on the historical monetary value of exhibited works to broader narrative approaches, and in particular towards architecture, exhibition designs, and audiences: the experience of the exhibition has become as relevant as what is exhibited. Many institutions today are moving from a collection-centered approach to recognising, as Suzanne MacLeod writes, exhibitions as possibilities to enhance practices and knowledge[5]. This

[4] P. A. Brandt, *Spaces, Domains and Meaning: Essays in Cognitive Semiotics*, cit., p. 30.

[5] S. MacLeod, *Reshaping Museum Space: Architecture, Design, Exhibitions*, Routledge, London and New York 2005, p. 1.

ISSN 0039-4130

perspective follows many other concerns, such as understanding exhibitions according to contemporary values and practices, including educational services and didactic approaches, user-led meaning-making, and wider multidimensional and multisensorial takes on the traditional exhibiting structure. For this reason, exhibitions' architecture and display design have become veritable social and cultural products that can only be activated through occupation, fruition, and even opposition. I would argue that once new technologies and media – including, for instance, video making and extensive digital and material reproduction – have slowly undermined unicity as an object-oriented value, it is, in turn, the exhibition and its experience that have gained such dimension.

This immaterial take on value has today pervaded countless spheres of our contemporary lifestyle. Experience economy, one of the latest trends of capitalism, has deeply influenced the attention that has been recently given to the production of exhibitions and even the construction of new buildings dedicated to art exhibitions. To rephrase German architect Anna Klingmann, this approach has led practitioners to wonder not necessarily what architecture is, has or does, but how its users feel and, ultimately, who they are [6]. Starting from these premises, I want to look at the concept, practice, and phenomenon of exhibiting as the interrelation of three fundamental elements: the visitor's body, the architecture as a phenomenal, diachronic, and semio-narrative tool, and artworks as objects. The audience is thus (re)framed as the complexity of subjects which feature «corporeal capacities in co-creation» [7].

As we discuss space and our physical engagement with it, in both functional and aesthetic terms, the first element that comes into place is the human body, particularly the role of bodily and embodied perception. I want to start borrowing from Merleau-Ponty a definition of the body that suits my proposed analysis while acknowledging its limitations: the body is the first of all cultural objects [8]. In the frame of this paper, the research interest around the body is a central tool in producing meaning while recognising the geographical, cultural, and political possibilities and variations on the theme, especially approaches that

[6] A. Klingmann, *Brandscapes: Architecture in the Experience Economy*, The MIT Press, Cambridge, Massachusetts 2007, p. 1.

[7] C. Stalpaert-K. Pewny-J. Coppens-P. Vermeulen, *Unfolding Spectatorship: Shifting Political, Ethical and Intermedial Positions*, Academia Press, Gent 2018, p. 5.

[8] M. Merleau-Ponty, *Phenomenology of Perception*, D. A. Landes (translated by), Routledge, London and New York (1945) 2012.

ISSN 0039-4130

challenge the male-centred, Western, and purely scientific definition of the body that we often take for granted. These include feminist conceptualisations, the limits of the white body and intersectional theory, and non-cultural and more-than-human approaches to corporeality.

Looking at the human experience as bodily grounded and following Gibson's ecological approach [9], contexts, as much as situatedness and diachronicity, profoundly affect how we individually and socially make sense of what surrounds us. In the specific context of exhibitions, the physical experience occurs in association with the nature of the creative expression that art entails, which is not understood as pure embellishment or technique but in its performativity and narrativity, as the relational result of bodily acts of creation and subsequentially perception. In this phenomenal frame, the relationship between meaning and consciousness can exist only when the body is involved. Coming from a tradition that includes Merleau-Ponty and Husserl, semiotician and linguist Göran Sonesson writes that «embodiment emerges as a problem within the philosophy of consciousness, which aims to reconstruct the world as given to a (generic) subject» [10]. This perspective highlights an essential aspect of how consciousness acts: the body, which determines the way we are present in the world we perceive, cannot be merely another mentally perceivable object or simply an epiphenomenon playing an incidental role in meaning-making processes. Realistically, the body is the contact point between consciousness and the physical world or, as Sonesson phrases it, «our *condition of access* to all possible experience of the world» (emphasis in original) [11]. Whatever is given to the realm of our consciousness is first and foremost presented bodily: as we are physically grounded in the world, perception depends on our positionality. Primal element of every encounter, the body cannot but present itself as part of the meaning.

The idea that the role of the body should not be overlooked in aesthetics was often addressed as a contemporary revolution in philosophy. However, we can find in the work of early modern and pre-Cartesian theorists references to bodily perception, frequently considering the body as an open organism in constant exchange with its surroun-

[9] J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston 1979.

[10] G. Sonesson, *From the meaning of embodiment to the embodiment of meaning: A study on phenomenological semiotics*, in T. Ziemke-J. Zlatev-R. M. Frank (edited by), *Body, Language and Mind. Vol 1. Embodiment*, Mouton, Berlin 2007, p. 87.

[11] *Ibid.*, p.110.

ISSN 0039-4130

dings rather than a self-contained and self-identical expression of the mind. This approach oriented to physical and multisensorial perception and feelings affirms a pre-reflective, non-discoursive mode of knowing and symbolising. Given these premises, we can look at artworks as particular to the human phenomenal sphere both as forms of practice (in their creative dimension) and experience (in their perceptive dimension). Tied to a bodily, perceptual response, artworks play within the traditional semiotic relationship between symbol and symbolised, not only depending on the associations that exist between name and object but also on how a person projects certain bodily sensations onto the object in question. This relationship is structured around symbols (gestures, words, shapes, and many more) referring to imagery (emotional, visual, lexical, etc.), which are translated into new ways of meaning.

Adopting a phenomenological and semiotic approach to spectatorship means addressing the corporeal ways in which audiences engage. This implies developing a specific interest in what is happening to the spectator while "spectating", but also in what audiences simply do. Although it is central to acknowledge that audiences are not singular and homogenous entities and many factors impact perception per se (from gender and social class to geography and location), the artist's physical presence (whether direct, for example, in performance and video art, or indirect, namely resulting in the creation of visual art and physical objects) has a role in affecting aesthetic experiential phenomena in ways that trace back to intersubjectivity. As Siri Hustvedt writes, «i]n art the meeting between viewer and thing implies intersubjectivity. [...] It is the silent encounter between the viewer, "I", and the object, "it". That "it", however, is the material trace of another human consciousness [...] the residue of an "I" or a "you"»[12].

Today, the complex and manifold experience of artistic objects concerns the entire body: artworks can surround us and even be in motion, they can be smelled or touched, or be embedded in the artist's performing body itself. For this reason, both positioning towards and interaction with artworks affect perception and interpretation, which is further influenced and oriented through architectural and display narratives. Between aesthetics, intersubjective encounters, and possible meaning, contemporary exhibitions must be thought of as relational milieus where value is experiential and personal but also social and

[12] S. Hustvedt, *Mysteries of the Rectangle*, Princeton Architectural Press, New York 2005, p. xix.

ISSN 0039-4130

collaborative, in the interplay between individuality and community, art and presence, gaze and physicality, image and material, mediation and simulacrum, occupation and estrangement, and many, many, more. This reflection shows how artworks constantly experience new forms and framings collectively and institutionally (from historical settings and interpretation to exhibition narratives and design) and subjectively. They cultivate new meanings and experiences but are also practised upon in different ways through time and in diverse spaces. Bodily and symbolic practices and interpretations make artworks entities in constant development since the properties and values that we find in them belong to the broader structure and deconstruction of the world where human practices are renegotiated.

ISSN 0039-4130

The Theoretical Foundations of the Feminist Debate on Reproductive Technologies

Costanza Vizzani

1. *The theory of difference*

1.1. Simone de Beauvoir

De Beauvoir's main text that specifically investigates the question of women is *The second sex*. Woman – as a human being – is essentially *free to self-projecting*, indeed, compelled to self-projecting by the structural condition that is essential to her as a human being. However, the human being is always essentially declined to the masculine. That is why it is necessary to make it explicit that she too is free to make herself a subject.

De Beauvoir recognises that asserting oneself as a *subject* is not an immediate act, but the result of a choice. Since choice structurally brings with it anguish, de Beauvoir recognises that it may be easier to make oneself a thing, an *object*, to live inauthentically. Indeed, freedom is a potentially hard task: consciousness is forced to strenuously strive to realise itself, with the prospect of never being able to do so definitively and completely. In such an existentialist perspective, freedom is a condemnation for the human being. For these reasons, woman is man's accomplice in having chosen the state of subordination, having left the state of superiority to man [1]. She has identified herself with a specific role, that of the subordinate entity, vainly attempting to free herself from the condemnation of freedom.

[1] Cf. S. de Beauvoir, *The Second Sex*, tr. by C. Borde and S. Malovany-Chevallier, Jonathan Cape, London 2009 [*Le Deuxième sexe*, Gallimard, Paris 1949], p. 10.

However, precisely because of the structural human freedom, the existential condition is dynamic. The subordination of women, therefore, is not a fact that has taken place, but a becoming, and as such, it can un-become [2]. In other words, woman is not born a woman – in the sense that she is not ontologically subordinated ("One is not born, but rather becomes, woman") [3]. The woman must become *other* with a lower case letter, that is, not the Other as a mirror of the totality, but rather the half of a harmonic union in which none of the sexes has primary value on either a logical or ontological level. From this philosophical conception of totality I interpret de Beauvoir as a theorist of difference. Totality does not dissolve into the undifferentiated, but is dually structured.

De Beauvoir then, continuing her research, questions why *reproduction,* although fundamental to human beings, has not allowed women to impose themselves and emancipate themselves [4]. Motherhood could have played a key role and instead turned out to be a further means of subordination. According to de Beauvoir motherhood itself can be reduced to a condition of slavery, since it is in itself something barbaric, if society does not intervene to support the woman [5].

1.2. Luce Irigaray

The theory of difference was later to have an enormous diffusion and a better theorization thanks to the work of Luce Irigaray. according to whom sexual difference is not simply one topic of investigation among others, but rather the issue of our age: «Sexual difference is one of the major philosophical issues, if not the issue, of our age [...] Sexual difference is probably the issue in our time which could be our "salvation" if we thought it through» [6].

In her texts, Irigaray pursues her theory, especially in *Speculum of the Other Woman*. To refer to the need to finally represent the feminine, Irigaray refers to the *speculum*. The mirror returns the image of the man, and the man sees the woman as the opposite of how he per-

[2] Cf. ivi, p. 8.
[3] Ivi, p. 293.
[4] Cf. ivi, p. 9.
[5] Cf. ivi, pp. 64-65.
[6] L. Irigaray, *An ethics of sexual difference*, tr. by C. Burke and G. C. Gill, Cornell University Press, New York 1993, p. 5.

ISSN 0039-4130

ceives himself, i.e. a hole, a void, and as a consequence, the woman sees herself in this way. But if one were to use the *speculum* instead of the mirror, which returns an unclear and blurred image, says Irigaray, man would be able to see more of himself reflected. He would be confronted with a mode of relationship different from the "subject-object" or "subject-other" relationship.

Affirming the difference between man and woman does not lead to an estrangement between the sexes, but rather the opposite, it determines their approaching [7], since if they were not different they could not construct a relational relationship. On the contrary, the assimilation of one of the two poles – the feminine – by the other – the masculine – would take place. The human being can recognise in himself that he cannot resolve himself into wholeness precisely because of sexual difference [8]. Totality does not merge into the sum of the two poles, but is maintained in differentiated tension, without resolving itself. If this were not so, one sex would assimilate the other, losing it.

In determining herself as different, woman has a privilege, which is that of becoming from the body of a mother in a relationship of continuity: she can thus recognise herself in she who generated her [9]. The woman, precisely because she is potentially a mother, has an existential privilege in understanding intersubjectivity. This is because she can carry *the other* in her body. In the relationship of pregnancy, the woman learns that the relationship does not imply the submission of one to the other, but a relationship between subjects [10]. Then, it is necessary to rediscover maternal genealogy – every daughter comes from her mother – and replace it with male genealogy [11]. If women were to reappropriate this space *symbolically,* that is, if they were to radically alter the linguistic narrative of the male-female relationship, then it would be possible to reconstruct a history based on difference that is not ontologically dependent on the male pole.

[7] Cf. L. Irigaray, *Essere due*, Bollati Boringhieri, Torino 1994, p. 28.
[8] Cf. ivi, p. 43.
[9] Cf. ivi, p. 39.
[10] Cf. ivi, p. 44.
[11] L. Irigaray, *Speculum of the Other Woman*, tr. by G. C. Gill, Cornell University Press, New York 1985, p. 18.

ISSN 0039-4130

2. *The critique of binarism*

2.1. Monique Wittig

Monique Wittig is a feminist author who believes it is necessary to radically rethink the concept of sex and gender in order to finally achieve female emancipation. Her works include *One is not born woman* [12] and *The Category of Sex*. It is also possible to define Wittig as a theorist of lesbianism. The French author defines her theory as «materialist lesbianism» [13], in which she relates social class theory to what can be defined as sexual classes, namely the categories of male and female [14]. Wittig makes no distinction between sex and gender: sex is as much a product of a specific socio-political dictate as gender. In fact, the only sex that exists in binary terms would still be the female sex, as the male coincides with the universal and is therefore not subject to any categorical specification [15].

Sexual classes are in fact the product of a hetero-normative society, and like social classes involve the exploitation of one over the other. In particular, the sex-social class of women is exploited by the sex-social class of men for the reproduction of the species [16]. According to the French author, hetero-normativity is a *regime*, in that one sexual class – that of women – is structurally subjugated by the other – that of men. In particular, this political regime is aimed at controlling reproduction. Only within a heterosexual setting is it possible to preserve the continuation of the species, as women perform the unpaid work of reproducing themselves. The control of reproduction is necessary because the society set up according to the dictates of patriarchy entails a socio-economic set-up of prevarication in which women perform the counterpart of proletarian labour within the capitalist system [17].

The road to the abolition of sexual classes is clearly indicated by Wittig: the only possible solution lies in lesbianism. What a woman can and must do to emancipate herself from her state of slavery is to

[12] A clear reference to the work of Simone de Beauvoir. Wittig takes up the idea of the existentialist philosopher, for whom 'being a woman' is a definition devoid of content until it is filled with meanings related to social role (as well as personal life).

[13] Cf. M. Wittig, *The Straight Mind and other essays*, Beacon Press, Boston, 1992, p. xiii.

[14] Cf. ivi, p. 15.

[15] Cf. J. Butler, *Gender Trouble: Feminism and the Subversion of Identity*, Routledge, New York 1990, p. 113.

[16] M. Wittig, *The Straight Mind and Other Essays*, cit., pp. 5-6.

[17] Ivi, p. 2.

ISSN 0039-4130

free herself from her sexual category, and therefore social class. This is only possible through lesbianism because only the lesbian person allows the category 'woman' to be left behind[18]. The lesbian person, by renouncing the relationship with the man altogether, actively breaks the inequality of the relationship and drastically breaks the relationship with the oppressor, consequently collapsing the very structure of oppression. The lesbian is *not* a woman, but a person, whose gender does not ground any categorisation.

2.2. Judith Butler

Judith Butler, focuses her study on the concepts of sex and gender[19]. The fundamental reference text of her theory, and thanks to which she will become the conceptual reference point of *queer* theory, is *Gender Trouble: Feminism and the Subversion of Identity*. From the very beginning of her work, Butler emphasises the need to radically reset the feminist discourse, since in every existing theoretical approach there are inherent theoretical problems concerning the interpretation of the female subject[20]. The main problem with much of feminism consist in having set 'woman' as the universal subject of its discourse, i.e. of having thought up a feminine universal, and from this concept the discursive approach will necessarily be wrong[21].

This is because 'woman' is the product of the society that feminism itself wants to fight, but in doing so it gets stuck in a short-circuit, since it continues to presuppose the existence of a certain feminine essence[22]. The way to set up a feminist discourse that is not destinated to collapse because it is set in a structure that condemns it, consists precisely in rethinking the subject of political struggle starting from a redefinition of the concepts of sex and gender[23]. First and foremost, therefore, a discontinuity between sexed bodies and culturally constructed genders is detected. According to Butler, from the moment the sexual category is pronounced by doctors, the *performance* of gender begins. Indeed, in Butler, gender takes on a performative character, it becomes a *do-*

[18] Cf. ivi, 13.
[19] F. Rochefort, *Femminismi: uno sguardo globale*, Laterza, Roma-Bari 2022, p. 102.
[20] Cf. J. Butler, *Gender Trouble*, cit., p. 1.
[21] Ivi, p. 2.
[22] Cf. ivi, p. 6.
[23] Ibidem.

ISSN 0039-4130

ing[24], an action repeated over time. In this way, the person is forced to play the assigned part of male or female. To prevent the *bios* of a body from becoming *essentialised* and preventing the free realisation of the self, the only solution is to unmask *repetition*. If one is aware of this, it is possible to decide to perform differently and enact self-assigned roles. Only by taking note of performativity does it become possible to stage performances that are consistent with one's gender and gender-neutral. This would result in the end of binarism[25]. Indeed, what it is possible to do to counter the performance to which we are forced, namely to behave as male or female according to our assigned gender, is to start parodying this performance and thus unmask and subvert it. Parody, however, is not to be understood as an altered reproduction of a pre-existing reality, but as a questioning of the ontological status of that reality[26]. In this way, although initially linked to what they are parodying, the genres expressed allow the binary and patriarchal setting to be considered unnatural[27].

The symbolic reference point for female emancipation, therefore, cannot be as theorised by the theory of difference, the maternal, which is still a dictated *performance.* Motherhood, in this way, rather than becoming the starting point for female emancipation, would become a further means of patriarchal control. Gender, in fact, if attributed in an impositional and binary manner, becomes the means by which heterosexual society can control sexuality and reproduction[28].

3. *Confronting positions on reproductive technologies*

On the one hand, through the difference theory starting with Irigaray's thought a road to female emancipation is proposed starting with a rethinking of the genealogy of motherhood, It follows that distorting motherhood in an artificial manner becomes a way of impeding the path to emancipation. If motherhood takes on forms divorced from the biological-natural, the possibility of raising to the symbolic is lost. Irigaray is sceptical in general about technical and artificial knowledge

[24] Already in de Beauvoir, however, she did not question the fixity of gender. Moreover, the French existentialist does not conceive of more than two genders.
[25] Ivi, p. 112.
[26] Ivi, p. 138.
[27] Ivi, p. 138.
[28] Cf. ivi, p. 135.

ISSN 0039-4130

for a specific reason: the technological distances the masculine from the feminine, introducing a barrier that is difficult to bridge. Technical production, which is mostly male-dominated, creates a pole between man and woman, making sexual difference even more difficult to emerge[29]. If motherhood is so closely linked to the emancipatory possibility of women, it is clear why practices that revolutionise it are regarded with suspicion.

On the contrary, feminism, which moves from a critique of binarism, considers motherhood to be the cause of female exploitation. This does not concern pregnancy as a biological-natural fact, but the way in which it is received by society. Indeed, we have seen that the critique of binarism considers women as historically situated and not as an a-historical conceptual category. Indeed, in a binary society, one of the causes of women's exploitation concerns precisely the reproductive sphere. Woman is relegated to the private sphere, to the care of the home and children, and for this to happen she must be excluded from the public sphere. Motherhood, however, is not criticised in itself. Therefore, if it could be interpreted in an innovative and emancipatory manner with respect to patriarchal dictates, it would be acceptable again. Reproductive technologies can therefore help to emancipate women through a radical rethinking of motherhood and parental roles.

Conclusion

In my opinion, keeping the two lines of thought in mind can be useful to comprehensively analyse the relationship between ethics and technology from the perspective of women. In fact, this debate teaches us how important it is, even before rejecting or accepting the theses on reproductive technologies, to understand the complexity of the technological phenomenon, renewing the call for research and bioethical reflection. Anyway, Irigaray's theorization of symbolic motherood seems to give us a better path to follow. Indeed, what feminism opposed to the introduction of reproductive technologies objects to is the fact that who is in favour, in pursuit of the ideal of gender equality, is prepared to deprive women of a fundamental prerogative. In this way, the supporters of reproductive technologies, as a means of female *empowerment*, unconsciously accept the essentially male chauvinist assumption of an

[29] Cf. L. Irigaray, *Essere due*, cit., p. 88.

ISSN 0039-4130

alleged "natural defect" of women that should be integrated [30], in order to re-trace the female figure on the model of the male one. In short, to achieve the *social* emancipation of women, it would be necessary to intervene in such a way as to deprive them of their *biological* function [31].

Bibliography

J. Butler, *Gender Trouble: Feminism and the Subversion of Identity*, Routledge, New York 1990.
S. De Beauvoir, *The Second Sex*, tr. by C. Borde and S. Malovany-Chevallier, Jonathan Cape, London 2009 [*Le Deuxième sexe*, Gallimard, Paris 1949].
L. Irigaray, *Speculum of the Other Woman*, tr. by G. C. Gill, Cornell University Press, New York 1985.
Id., *An ethics of sexual difference*, tr. by C. Burke and G. C. Gill, Cornell University Press, New York 1993.
Id., *Essere due*, Bollati Boringhieri, Torino 1994.
F. Rochefort, *Femminismi: Uno sguardo globale*, Laterza, Roma-Bari 2022.
M. Wittig, *The Straight Mind and other essays*, Beacon Press, Boston 1992.

[30] In the case of ectogenesis and GPA, this is a deprivation, but the general sense of the assumption remains unchanged.

[31] Cf. G. Cavaliere, *Ectogenesis and gender-based oppression*, cit., pp. 730-731; cf. S. Segers, *The path towards ectogenesis*, cit., p. 6.

ISSN 0039-4130

Alienation and Self-Knowledge in Maine de Biran

Sarah Horton

Self-knowledge has long been held up as an ideal; consider the injunction, carved on the temple of Apollo at Delphi, to «Know thyself», which Maine de Biran (1766-1824) himself cites favorably. At the same time, however, Biran argues that one discovers one's own existence, as well as that of the world, only through the sentiment of effort, and effort, he shows, is always opposed to a resistance that is opaque to knowledge. Is a certain impossibility of knowing oneself therefore also essential to the constitution of the human being? Indeed, my thesis, drawing on the work of Emmanuel Falque [1], is that an attentive reading of Maine de Biran will teach us that alienation, although it seems to be the most improper of states, is in fact proper to humans: it is not that we should seek out just any experience, or rather non-experience, of alienation, but that the limits to self-knowledge and even to experience are indeed constitutive of human being.

Although Biran published only three works, and only one monograph, in his lifetime, he has already had considerable influence on 20th-century phenomenology, thanks to his posthumously published writings. Thus far, the primary interpretation of Biran, found in Maurice Merleau-Ponty, Paul Ricœur, and most notably, Michel Henry, has emphasized direct, immediate consciousness of oneself—but, as Falque has argued, a phenomenology that seeks to examine human experience must also attend to the limits of experience [2], and Biran, as it turns out, opens for philosophy a new path in which alienation becomes as

[1] See E. Falque, *Spiritualisme et phénoménologie: le «cas» Maine de Biran*, Paris, PUF, 2024 (Chaire Étienne Gilson).

[2] See ibid., as well as E. Falque, *Hors phénomène*, Paris, Hermann, 2021 (De Visu).

important as presence to oneself and the unknowable becomes as important as knowledge.

The leitmotif of Biran's philosophy is that the human being knows himself through effort. To Descartes' «I think, therefore I am», Biran replies, «I will, therefore I am». Any act of the will on my part reveals to me, immediately and beyond any possibility of doubt, that I exist.[3] Crucially, my voluntary actions, including the act of thinking, belong also to the domain of the body and not only to the mind or soul. Consider an example: since I want to extend my hand, I extend it, and the effort by which my will is accomplished (which is not the same as the physical sensation of movement)[4] reveals my existence to me, immediately and therefore beyond any possibility of doubt, as a willing and at the same time a corporeal being[5]. Wherever there is effort, there is necessarily also resistance, precisely so that it can be "ef-fort" and not only force: the prefix "ef-" in fact comes from the Latin *ex-*, which marks, in this case, the separation of the force from itself by virtue of the opposition that it must combat to impose itself. Without this resistance that accompanies all my conscious acts, an evil genius could indeed make me doubt the reality of my body—but because I sense myself resisting myself, I know at once that I exist and that I am, not a soul or a body considered separately, but rather the relation between a hyperorganic force and an organic resistance. This sentiment of the relation that I am, that is, the sentiment of effort or of myself, is what Biran calls the primitive fact of consciousness—«primitive» because it is prior to any representation or reasoning.[6] Because the human being is neither a soul nor a body alone, but rather the relation between them, the human being has, as Biran emphasizes, a «mixed nature», both moral (a term that in the 18th and 19th centuries referred to the faculties of the mind or soul, in contrast to the body) and physical,

[3] Thus Maine de Biran writes that Descartes «did not, perhaps, sufficiently observe that this self that thus retreats into itself to affirm its own existence and deduce its absolute reality thereby performs an action, makes an effort; yet does not every action essentially and in reality suppose a subject and a terminus? Can effort be considered as absolute and without resistance?» (*Mémoire sur la décomposition de la pensée*, in *Œuvres de Maine de Biran*, vol. III, ed. F. Azouvi, Paris, Vrin, 1988, p. 364, footnote).

[4] See Maine de Biran, *De l'aperception immédiate*, in *Œuvres de Maine de Biran*, vol. IV, ed. I. Radrizzani, Paris, Vrin, 1995, p. 57, footnote; *Of Immediate Apperception*, trans. M. Sinclair, in *Maine de Biran's Of Immediate Apperception*, ed. A. Aloisi, M. Piazza, and M. Sinclair, London, Bloomsbury, 2021, p. 59, n. 20.

[5] See Maine de Biran, *Mémoire sur la décomposition de la pensée*, cit., p. 364, footnote.

[6] See, for instance, Maine de Biran, *Essai sur les fondements de la psychologie*, ed. P. Tisserand, VIII, Paris, Alcan, 1932, p. 177.

ISSN 0039-4130

and the relation between the physical and the moral is fundamental to the human condition.

The immediate knowledge of my dual nature necessitates a reconceptualization of the body, which can no longer be thought as purely external to myself. It is no surprise to read, in Paul Ricœur, that « Maine de Biran is therefore the first philosopher to have introduced the lived body [*le corps propre*] into the region of nonrepresentative certainty» [7]. And, as Anne Devarieux emphasizes, he is, moreover, the first, «in French-language philosophy» — excepting Leibniz, for whom the phrase does not yet, in any case, take on the full sense that it will have for Biran—to employ the expression *le corps propre* (often translated as «the lived body») [8], for, prior to him, «no philosopher had succeeded in grasping the sense of the apperception of oneself» [9]: while he certainly has predecessors who wrote *corpus meum* («*le corps mien*», «my body») or even *mon propre corps* («my own body») [10], it belongs first to Maine de Biran to have thematized the body that is not only mine or my own but that is, as it were, internal to myself.

[7] P. Ricœur, *Soi-même comme un autre*, Paris, Seuil, 1990, p. 372; *Oneself as Another*, trans. K. Blamey, Chicago, University of Chicago Press, 1992, p. 321, translation modified.

[8] The French term *le corps propre* has often been translated as «the lived body», to convey the meaning of the body as it is lived by the subject; more recently, it is often translated as «one's own body», although «one's own body» would literally be *son propre corps*, with the adjective before the noun; sometimes it is translated as «the proper body», understood as «the body that is proper to me». The English translation of Ricœur's *Oneself as Another* uses both «the lived body» and «one's own body» as translations of *le corps propre*. Given the contrast that Devarieux rightly draws between *le corps propre* and *mon corps* (my body) or even *mon propre corps* (my own body), I have avoided the translation «one's own body» for *le corps propre*, preferring «the lived body» or, occasionally, «the proper body».

[9] A. Devarieux, *Maine de Biran et l'invention du corps propre*, in *Corps ému: Essais de philosophie biranienne*, ed. Luís António Umbelino, Coimbra, Presse universitaire de Coimbra, 2021, p. 31; on Leibniz, see ibid., note 5, p. 31. For Leibniz, the expression appears in *The Principles of Nature and Grace Based on Reason (1714)* in G.W. Leibniz, *Principes de la nature et de la grace fondés en raison; Principes de la philosophie ou monadologie*, ed. A. Robinet, Paris, PUF, 1986, p. 31; *Philosophical Papers and Letters*, trans. and ed. L.E. Loemker, 2nd ed., Dordrecht, Kluwer Academic, 1989, p. 637: «And each outstanding simple substance or monad which forms the center of a compound substance (such as an animal, for example), and is the principle of its uniqueness, is surrounded by a mass composed of an infinity of other monads which constitute the body belonging to this central monad [*le corps propre de cette Monade centrale*], corresponding to the affections by which it represents, as in a kind of center, the things which are outside of it». But while Leibniz thus emphasizes the essential relation between the «central monad» and the monads that constitute its body, going so far as to highlight «the accord and the physical union of soul and body» (ibid.,), he insists neither on the interiority of this *corps propre* nor on its resistance.

[10] Devarieux, *Maine de Biran et l'invention du corps propre,* cit., pp. 33-34. Cf. also F. Azouvi, *Genèse du corps propre chez Malebranche, Condillac, Lelarge de Lignac et Maine de Biran,* in *Archives de philosophie*, XLV, 1, 1982, pp. 85-107.

ISSN 0039-4130

So far, this account may seem to wholly justify the interpretation of Biran as a thinker of interiority and of immediate self-consciousness. But the organic body that resists is not itself without force: «I am a force that goes!» says Hernani to Doña Sol in Victor Hugo's eponymous play [11], and this could also be the cry of Biranian man—or, better, « I am forces that go! » Indeed, knowing oneself through effort is not mastering oneself, and while « the hyperorganic force that we call the soul» [12] governs our voluntary acts, Biran recognizes a « human duality, a free and active force and a force under the authority of necessity» [13]. Certainly, it would be easy to identify ourselves only with the soul, that is, with that voluntary force that indeed we are, while rejecting as foreign to the self the forces of the organic unconscious that have, however, more power over us than we desire. Particularly because neuroscience is one theme of this conference, I wish to highlight Maine de Biran's great interest in the medicine and psychiatry of his era, which he studied extensively with an eye to what they can reveal about human existence and human freedom—while also attending to their limits, for there is a fundamental obscurity in the human being that no amount of study can ever eliminate. « There is nothing more instructive for the reasonable man than the history of madness », affirms Biran [14], not to show us errors to avoid, nor even to motivate us to better appreciate our condition as « reasonable men », as if that were a fixed and stable state, but rather because there is in each of us a share of an essential madness [15]. In Biran's view, the self is abolished in madness, as also in sleep, but it does not follow that madness has nothing to do with people said to be «sane», for in truth, reason, sanity, and health are never as constant as we desire. To be human is to live at the limits of experience, precisely because human beings are, from the very moment of conception [16], constituted by an absence of the self that precedes and forms any presence to oneself. Specific cases of alienation can certainly be treatable; Biran, who founded the Bergerac Medical Society

[11] V. Hugo, *Hernani*, in *Théâtre complet*, vol. I, ed. J.-J. Thierry and J. Mélèze, Paris, Gallimard, 1964 (Pléiade), p. 1227, act III, scene IV, line 284.

[12] Maine de Biran, *Commentaires et marginalia: XVII^e siècle*, in *Œuvres de Maine de Biran*, XI, 1, ed. C. Frémont, Paris, Vrin, 1990, p. 30, footnote.

[13] Maine de Biran, *Dernière philosophie: existence et anthropologie*, in *Œuvres de Maine de Biran*, X, 2, ed. B. Baertschi, Paris, Vrin, 1989, p. 373.

[14] Maine de Biran, *Discours à la société médicale de Bergerac*, in Œuvres de Maine de Biran, V, ed. F. Azouvi, Paris, Vrin, 1984, p. 105.

[15] Cf. Falque, *Spiritualisme et phénoménologie*, cit., IV, pp. 173-210.

[16] See Maine de Biran, *Discours à la société médicale de Bergerac*, cit., pp. 30-36, in particular p. 32.

ISSN 0039-4130

(*Société médicale de Bergerac*), emphasizes that medical science has a great responsibility to develop and improve treatments for the various maladies that ail us [17], and he praises psychiatry for the cures it has been able to effect [18]. But the possibility of becoming alienated can never be eliminated from the human being.

Indeed, Biran's work leads to an even stronger conclusion: there is a crucial sense in which we are already alienated from ourselves because we can never wholly coincide with ourselves. As Biran writes, considering the constitution of the self through the sentiment of effort,

> «I have substituted organic inertia or resistance for foreign resistance, and I have seen the faculties originarily constituted, not exclusively in that constrained movement that teaches us that there exists something outside ourselves, but more generally in the effort that is essentially relative to some term, be it applied to the lived body or the foreign [*étranger*] body» [19].

This substitution, far from removing the foreign or strange body, on the contrary reveals the strangeness of the organic and even of the own or the proper: my own body or, indeed, my proper or lived body is first disclosed to me not as a flesh whose coincidence with myself would be total but as a resistance. Hence the thinker who discovered the proper or lived body is also the one who saw to what extent the proper depends on the improper, such that I would remain ignorant of myself without this resistance that is necessarily opaque but that is essential to any effort. Unaware that his body resists him, the agitated sleeper or the sleepwalker is equally, and consequently, unaware that his body is in motion; no effort, therefore, can be attributed to him, and he does not sense his existence [20]. From the moment that I begin to know myself, I discover myself as always already a stranger [21]. This strangeness is immediately and directly given to me, it is true, but it is given precisely as strangeness, as that which I can never fully know, and it is a constitutive alienation that I will never escape.

Thus although Biran's view that the self is lost in madness may

[17] See, for instance, Maine de Biran, *Journal*, III, ed. H. Gouhier, Neuchâtel, La Baconnière, 1957, pp. 17–18 (entries from 1794 or 1795).

[18] See, for instance, Maine de Biran, *Rapports du physique et du morale de l'homme*, in *Œuvres de Maine de Biran*, VI, ed. F.C.T. Moore, Paris, Vrin, 1984, p. 114.

[19] Maine de Biran, *Mémoire sur la décomposition de la pensée*, cit., p. 164, footnote.

[20] See in particular Maine de Biran, *Discours à la société médicale de Bergerac*, cit., pp. 82-123.

[21] Cf. Falque, *Spiritualisme et phénoménologie*, cit., pp. 268-291.

ISSN 0039-4130

seem to exclude madness from the domain of humanity, in fact nothing could be farther from the truth: Biran's approach to madness centers on the point that we could all, by virtue of our very humanity, fall to alienation. To say it with Emmanuel Falque, « Even though we are not all traumatized (or alienated) we are at least all, without exception, traumatizable (or alienatable)»[22]. For, in Biran's view—and having shown this is perhaps his greatest merit—self-possession, far from being established once and for all, always remains unstable and is therefore always to be conquered at the price of efforts. Indeed, the self is also lost each time that we fall asleep[23], and the human being is far from mastering himself: one could almost say that nothing is more human than the loss of oneself, unless it is the struggle that must be conducted to find again each day that self-mastery that is as precious as it is ephemeral. To be conscious of oneself is to be conscious of one's own strangeness, and the madman who forgets himself has also forgotten his own fragility: the mad or alienated person is precisely the one who has lost the consciousness of the fundamental alienation that constitutes him, and if he is to be restored to sanity, he must be restored also to the consciousness of sanity's instability.

Internal to myself, there lies exteriority; within what is most proper to me, there lies the improper. How then can we know or understand what it is to be a human being? The wise man is « he who knows that he knows nothing », as Plato's Socrates warns us; and he knows himself who admits the impossibility of knowing oneself. Yet it is not a matter of a pure impossibility that would consign us to animality; on the contrary, this impossibility founds all possible knowledge and reminds us, at the same time, that we should not expect too much certainty. There is neither awareness of oneself nor self-knowledge without the strangeness that resists them, for humans are constituted by a non-assimilable exteriority, and one can know oneself, to the degree that such a thing is possible—and it is indeed an important task—only on the basis of this resistance, even this alienation. The moment that I begin to know myself, I discover myself as other than myself, and there is no way to escape this alienation, save by leaving behind my own humanity and therefore falling into an absolute alienation. Biran shows us the necessity of a humility that recognizes, and that is grateful for, the limits that constitute our existence.

[22] Falque, *Spiritualisme et phénoménologie*, cit., p. 220.

[23] See in particular Maine de Biran, *Discours à la société médicale de Bergerac*, cit., p. 82-123.

ISSN 0039-4130

Il metodo e l'intero

Nota sull'eredità di Pavel Florenskij

Cecilia Benassi

1. *Introduzione*

A causa delle complesse vicissitudini attraversate dall'*opus* florenskijano nella storia [1], dei conseguenti filtri che hanno influenzato la nostra ricezione della sua opera, e delle difficoltà – non solo geopolitiche – che attualmente caratterizzano l'accesso agli archivi con il lascito manoscritto e dattiloscritto dell'autore, gli studi faticano ancora a valutare la portata complessiva del suo pensiero [2].

In particolare, come è stato osservato negli ultimi anni [3], colpisce la sua «*forma mentis* assolutamente polifonica, capace di orchestrare – all'interno di un unico quadro di pensiero – temi e prospettive notevolmente differenti» [4], tanto da essere arrivati a parlare di lui come di un

* Tutte le citazioni delle lettere dal gulag, salvo qualche eccezione in cui traduco direttamente dall'originale (sarà indicato in nota), sono tratte dall'ultima edizione italiana: P.A. Florenskij, *Vi penso sempre...*, a cura di N. Valentini e L. Žak, trad. L.M. Pignataro, Mondadori, Milano 2024, e saranno indicate attraverso la datazione della lettera e il destinatario. Per il resto delle citazioni dai testi di Florenskij o su di lui, le traduzioni, ove necessarie, sono mie.

[1] Qualche cenno al riguardo si ritrova in C. Benassi, *Chi è Pavel A. Florenskij?*, in *Studium-Contemporary Humanism* 2023, pp. 186-195. Alla ricostruzione più approfondita di questi itinerari e dei filtri che essi hanno posto alla nostra ricezione dell'autore, è dedicata una sezione della tesi dottorale in corso.

[2] Cfr. N. Valentini, *Introduzione. Il cammino di Pavel Florenskij verso la verità vivente* in P. Florenskij, *La colonna e il fondamento della verità*, San Paolo, Milano 2010, p. XV.

[3] Mi riferisco in particolare ai contributi di Lubomir Žak (L. Žak, *La complessità del reale e la sua conoscenza. Spunti di riflessione sull'«allargamento della ragione» proposto da P.A. Florenskij*, in *Divus Thomas*, CXIX, 3, 2016, pp. 131-171; *Id.*, *Il "realismo come visione del mondo": introduzione al concetto di complessità sviluppato da Pavel A. Florenskij*, in *Lateranum*, LXXXIII, 3, 2017, pp. 513-534; *Id.*, *La dimensione immaginaria del reale secondo la teoria della complessità di Pavel A. Florenskij*, in *Constantine's Letters*, XIV, 2, 2021, pp. 191-204) e Silvano Tagliagambe (S. Tagliagambe, *Come leggere Florenskij*, Bompiani, Milano 2006; *Id.*, *Chiralità. La vita e l'antinomia. Gli eroi dei due mondi*, Mimesis, Milano-Udine 2021).

[4] L. Žak, *La dimensione immaginaria del reale secondo la teoria della complessità di Pavel*

anticipatore delle teorie della complessità[5] e a interrogare i contributi del suo *pensiero complesso* alle diverse scienze e alla teoria della conoscenza[6].

A fronte di queste osservazioni, sembra lecito ritenere che il futuro degli studi florenskijani dovrebbe potersi avvalere di un approccio di filologia d'autore, ove ricostruire l'ordine in cui Florenskij aveva lasciato i materiali, compilare un inventario completo dei lasciti ed entrare, attraverso gli *scartafacci* d'autore, nel suo laboratorio creativo.

Va inoltre notato che le informazioni disponibili, stratificatesi nel corso di quasi sessant'anni di studi, presentano alcune incongruenze. Essendo attualmente preclusa la possibilità di accedere all'archivio e di sanare le incongruenze con nuove informazioni, vorrei provare a considerare, tra gli elementi a nostra disposizione, quelli che permettono di gettare uno sguardo sul suo laboratorio produttivo e sull'eredità a noi rimasta.

Al riguardo, si possono considerare delle *fonti dirette*, *indirette* e *implicite*. Classifichiamo tra le prime gli scritti che assumono, a vario titolo, un taglio autobiografico; per esempio, *Ai miei figli* e i molteplici *Autoreferat* scritti in circostanze diverse, in cui l'autore prova a tracciare le dinamiche profonde del proprio percorso esistenziale e intellettuale[7]. Trattandosi, tuttavia, di scritti che necessitano un approccio adeguato al genere autobiografico, in questo studio volgiamo l'attenzione alle *fonti indirette* e *implicite*, ovvero le testimonianze dei suoi figli e le indicazioni trasmesse da Florenskij ai famigliari "tra le righe" dello scambio epistolare dal gulag.

Da uno studio attento delle lettere, infatti, spiccano le linee organizzative e metodologiche soggiacenti il lavorìo di padre Pavel. Esse

A. Florenskij, cit., p. 192.

[5] Natalino Valentini suggerisce di considerare Florenskij il «precursore di una visione olistica della conoscenza e delle innovative indagini epistemologiche sulle teorie della complessità e delle relazioni tra i diversi sistemi, maturate poi nelle opere di alcuni studiosi contemporanei», N. Valentini, *«Florenskij - filosofo della religione e del culto». Dalla fenomenologia del sacro alla santificazione della realtà*, in *Lateranum*, LXXXIII, 3, 2017, p. 570.

[6] Cfr. L. Žak, *Gli stimoli offerti dal "pensiero complesso" di Pavel A. Florenskij alla scienza psicologica*, in *Il Pensiero polifonico di Pavel Florenskij. Una risposta alle sfide del presente*, a cura di S. Tagliagambe, A. Oppo, M. Spano, PFTS, Cagliari 2018, pp. 415-432.

[7] Il più noto è quello scritto su richiesta del *Dizionario enciclopedico di bibliografia russa* e pubblicato nel 1927 dall'Istituto Granat (in P.A. Florenskij, *Avtoreferat*, in *Sochinenija v chetyrekh tomakh,* I, Mysl, Moskva 1994, pp. 37-43. Traduzione italiana di Claudia Zonghetti in *Il simbolo e la forma. Scritti di filosofia della scienza*, a cura di N. Valentini, A. Gorelov, Bollati Boringhieri, Torino 2007, pp. 3-24). Di recente pubblicazione sono altri dieci autoprofili di questo genere conservati in archivio: si veda P.A. Florenskij, *Filosofskie, bogoslovskie i kriticheskie trudy: 1908-1933 goda*, a cura di A. Trubachev, Obshhestvo pamjati igumenii Taisii, Sergiev Posad-Sankt-Peterburg 2021, pp. 807-844 (volume di difficile reperibilità).

ISSN 0039-4130

traspaiono in particolare all'interno della sua conversazione con i figli, ove l'autore si muove tra il desiderio di esercitare la sua paternità accompagnandoli nella loro formazione interiore e culturale e quello di tramandare informazioni su di sé e sulla propria eredità.

Il presente studio è dunque animato dall'idea che, gettando luce sul suo lavoro creativo, si potranno elaborare *spazi di risposte* ad alcune domande ritenute importanti per gli studi florenskijani [8]: come Florenskij vedeva la sua opera? Come organizzava il proprio laboratorio creativo? Quali riteneva che fossero la chiave e l'obiettivo dei suoi scritti, così variegati per ambiti disciplinari e conoscenze impiegate?

Che cosa, in sintesi, possiamo dire, del *metodo* conoscitivo e creativo di Florenskij?

Questo contributo, ovviamente, non pretende di rispondere alle domande elencate, ma desidera tuttavia raccogliere e offrire alcuni spunti che potrebbero favorire una presa di contatto col mondo dell'autore e con il suo cantiere intellettuale che, come emerge in diversi passaggi delle lettere, si rivela abitato dall'*intentio* – vivamente goethiana – di

> «conoscere il vivente in quanto tale, di vederne in mutuo rapporto le parti esterne visibili e tangibili, di considerarle indizi del loro interno, *e per tal modo dominare l'intero, per così dire, in una visione intuitiva*. Come quest'aspirazione scientifica si ricolleghi all'impulso artistico ed imitativo, non occorre insistere» [9].

2. *Indizi per ricostruire un'eredità frantumata*

Evidentemente, durante gli anni di gulag padre Pavel sente crescere la probabilità di non ritornare più a casa e, mentre accusa con sofferenza il crescere di questa consapevolezza, orienta le proprie lettere affinché possano diventare strumenti utili alla ricostruzione del lavoro della sua vita.

In una lettera alla moglie del 3 dicembre 1934, da pochissimo arrivato al lager delle Solovki dopo un terribile viaggio con isolamento detentivo a Kem [10], scrive: «Nella mia vita ho lavorato molto, facendo di

[8] Scrisse di lui uno dei suoi allievi all'Accademia teologica di Mosca: «Ci sono persone il cui nome è, fin dai primi tempi, circondato di leggenda», S.A. Volkov, *P.A. Florenskij*, in *P.A. Florenskij: pro et contra*, a cura di D.K. Burlak, RKhGA, Sankt-Peterburg 2001, p. 141.

[9] J.W. Goethe, *La metamorfosi delle piante*, Guanda, Parma 1983, p. 43, corsivo mio.

[10] Città collocata sull'isola di Popov, il «Punto di transito e smistamento di Kem», dipendeva dalla Direzione dei lager settentrionali a destinazione speciale, e di solito accoglieva

ISSN 0039-4130

tutto per compiere il mio dovere. Ma tutto si è frantumato; ormai non posso più e, soprattutto, non voglio, incominciare un'opera scientifica di grandi dimensioni. Vivrò solo per voi, ritenendo di aver fatto ciò che dovevo come ho potuto» [11].

Tra il febbraio e il marzo di quell'anno, in effetti, aveva ricevuto notizia della confisca dei suoi libri sia dall'appartamento moscovita in cui alloggiava in quanto ingegnere del regime, sia dalla casa di famiglia a Zagorsk [12], dove avevano portato via «2684 libri senza elenco, e il 7 torneranno per gli altri» [13]. Florenskij scriverà in risposta alla moglie il 18 marzo con tono rassicurante, ribadendo con forza il suo amore e la sua profonda vicinanza a tutti loro, e dicendo ai figli di non spaventarsi per i libri, ma «di vivere del presente con un incremento di gioia e allegria» [14].

Nel frattempo, però, alla fine di febbraio [15], si era rivolto *Al capo dei lavori edili dei BAMLag OGPU* facendo riferimento alla confisca dei beni e chiedendone la restituzione alla moglie. Nella prima parte di questa lettera, l'autore fa una panoramica molto generale di ciò che, partendo, aveva lasciato del suo laboratorio creativo:

i detenuti in transito verso le Solovki. Cfr. J. Brodskij, *Solovki. Le isole del martirio. Da monastero a lager sovietico*, La Casa di Matriona, Seriate 1998, p. 51.

[11] Lettera ad Anna, 3 dicembre 1934, in P.A. Florenskij, *Sochinenija v chetyrekh tomakh*, IV, Mysl, Moskva 1998, p. 149. In una lettera di poco successiva, torna sull'argomento: «Vorrei lasciarvi in eredità un nome onorabile e la consapevolezza del fatto che vostro padre ha lavorato tutta la vita disinteressatamente, senza pensare alle conseguenze del suo lavoro per la sua persona. Ma proprio per questo disinteresse ho dovuto privarvi delle comodità godute dagli altri, dei divertimenti propri della vostra età, e persino della vicinanza con voi. Ora mi rattrista che da tutto questo mio impegno, anziché trarre qualche vantaggio, voi non ricaviate neanche quello che riceve la maggior parte [dei vostri coetanei], nonostante i loro genitori abbiano vissuto per se stessi. [...] la cosa più orribile della mia sorte è la cessazione del lavoro e la sostanziale distruzione dell'esperienza di tutta la mia vita, esperienza che è maturata solo adesso e che adesso potrebbe dare autentici frutti», Lettera a Kirill, 24-25 gennaio 1935.

[12] Sergiev Posad (ovvero il "villaggio di Sergio", in riferimento a san Sergio di Radonezh, che fondò il Monastero della Trinità, nei cui dintorni Florenskij viveva con la famiglia), così rinominata dal regime.

[13] P.A. Florenskij, *Sochinenija*, IV, cit., p. 731. Si tratta di uno stralcio da una lettera di Anna a padre Pavel, che i curatori russi hanno riportato in una nota alla prima edizione dell'epistolario degli anni di detenzione. L'epistolario, in questa e nelle edizioni successive, pubblica solo le lettere scritte da Florenskij ai famigliari, mentre non abbiamo informazioni sullo stato di conservazione delle lettere che egli riceveva da loro nel gulag.

[14] Lettera ad Anna, 20 marzo 1934, in *Sochinenija*, IV, cit., p. 93.

[15] Annoto che, in base alle informazioni fornite dai curatori russi e riportate nell'edizione italiana, ci troviamo davanti a una incongruenza di date, per cui sembra che Florenskij scriva la lettera per richiedere la restituzione dei beni alla fine del mese di febbraio mentre, dalle lettere presenti nell'epistolario, ne riceve notizia dalla moglie solo in marzo.

ISSN 0039-4130

> «Tutta la mia vita è stata dedicata al lavoro filosofico e scientifico, e non ho conosciuto né riposo, né svaghi o momenti ricreativi. Per questo servizio all'umanità ho investito tutto il mio tempo e le mie forze, ma anche la maggior parte dei miei non grandi guadagni è stata investita in libri, fotoriproduzioni, corrispondenza e altro. *Come risultato, giunto all'età di 52 anni, ho raccolto materiali che richiedono ulteriore lavorazione e che avrebbero dovuto dare risultati preziosi, giacché la mia biblioteca non era semplicemente una collezione di libri, ma una selezione di temi preparati e già riflettuti. Si può dire che le opere erano già pronte per metà, ma conservate in forma di schizzi e abbozzi, la cui chiave è nota solo a me. Inoltre, avevo fatto una composizione di disegni, fotografie e numerosi estratti da libri.* Ma il lavoro di tutta la vita è al momento perduto, così come tutti i miei libri, i materiali, i manoscritti e le bozze più o meno elaborate – confiscati per ordine dell'OGPU» [16].

In questo scritto, come in altri importanti passaggi delle lettere, si percepisce la sofferenza per un compito esistenziale incompiuto, per il quale però egli aveva strenuamente lavorato, raccogliendo e selezionando grandi moli di appunti e materiali, la cui entità complessiva, credo, non possiamo nemmeno immaginare.

2.1. L'incarico ai figli

Alcuni indizi al riguardo ci raggiungono specialmente dal dialogo che intrattiene con i due figli maggiori, Vasja (da Vasilij) e Kira (da Kirill).

In una lettera a Vasja del 1935, scrive:

> «Ti ho scritto prima, ma voglio scriverti di nuovo, in modo più chiaro, che affido a voi tutti, e in particolare a te e a Kira, tutti i miei progetti scientifici e tecnico-scientifici, perché voi possiate usarne i materiali e le idee e, se volete, perché li continuiate oppure li utilizziate nel vostro lavoro. Vorrei soprattutto aiutarvi con l'unica cosa che ho: le idee. Per introdurvi a queste opere, vi manderò a poco a poco delle informazioni, a partire dalle cose più facilmente realizzabili e più vicine alle vostre attività dirette. La prima materia di cui vi scrivo è la microfisica» [17].

Segue un'esposizione rapida e concisa di linee di ricerca e indicazioni di metodo volte all'approfondimento di ciò ch'egli intende per micro-

[16] Febbraio 1934, in *Sochinenija*, IV, cit., pp. 81-82, corsivo mio.
[17] Lettera a Vasja, 12-13 agosto 1935.

ISSN 0039-4130

fisica, ovvero «la misurazione di diverse costanti fisiche della materia usando campioni assai piccoli [...], molto più piccoli di quelli che si utilizzano normalmente» [18].

Poco più avanti, nella medesima missiva, introduce elementi di ricerca e riflessione sulle rocce sedimentarie, e nella successiva lettera a Vasja, riprende il discorso introducendo linee di studio sulle torbiere di sfagno [19] e fornendo indicazioni sulla tecnica di identificazione dei minerali per mezzo della misurazione del fattore dielettrico, da lui scoperta e personalmente sperimentata.

Gli esempi riportati danno una piccola idea [20] – evidente a chi volesse accingersi alla lettura delle lettere dal gulag – della rapidità affastellata dei pensieri creativi di padre Pavel, che in ogni situazione mirava all'elaborazione di metodologie *ad hoc*, inventate *ex novo* per rispondere alle specificità uniche e viventi dell'oggetto in analisi.

In un'altra missiva, stavolta a Kirill, scrive:

> «Leggi le mie lettere destinate ai tuoi fratelli e alle tue sorelle? Ciascuna di queste la scrivo in realtà per tutti, ma con una sfumatura individuale nel contenuto. *Sfortunatamente, ho troppi progetti che hanno richiesto tempo e lavoro, ma sono rimasti incompiuti o, peggio ancora, senza una forma definitiva, e questi progetti si stanno perdendo e continueranno a perdersi. Cerco di indicare qualcosa con brevi frasi nelle lettere, ma temo che ciò sia troppo laconico perché ve ne accorgiate*» [21].

Ripercorrendo il *corpus* epistolare alla ricerca di indizi circa la ricostruzione della sua eredità, ci si rende conto di come la preoccupazione di Florenskij – di fronte alla sensazione che il suo lavoro non avrebbe più potuto proseguire – fosse duplice.

Da un lato, desiderava accennare ai figli le linee di ricerca da lui avviate: a Vasilij e Kirill le tracce in ambito scientifico, a Olga quelle in ambito artistico-letterario e botanico. Dall'altro lato, cercava di fornire indizi circa le visioni e le metodologie di fondo che l'avevano animato, e che l'avevano spinto a lavorare così come aveva fatto.

[18] *Ibidem*.

[19] «Tra i materiali per le pile a depolarizzazione ad aria troverai i disegni della struttura dei muschi di sfagno», Lettera a Vasja, 21 agosto 1935.

[20] Piccola perché l'incalzare di indicazioni, elaborazioni ed esposizioni di nuovi temi e metodi è fittissima; se ne riportano brevi *exempla* per permettere al lettore di figurarsi il ritmo e la tipologia dell'andamento mentale del nostro autore, per quanto questo sia possibile.

[21] Lettera a Kirill, 16-17 gennaio 1936, corsivo mio.

ISSN 0039-4130

2.2. Linee di ricerca e innovazioni

Sul fronte delle specifiche linee di ricerca, padre Pavel era consapevole di aver avuto idee innovative e precorritrici, e desiderava che l'umanità potesse trarne vantaggio. Sapendo, tuttavia, di aver lavorato in maniera frammentaria e disordinata, cercava di affidare i singoli reperti di scoperte, e le chiavi per una loro feconda interpretazione, ai figli.

In una lettera alla moglie del 1936 scriveva:

> «L'opera della mia vita è distrutta, e io non potrò mai, né vorrò, ricominciare dall'inizio il lavoro di cinquant'anni. Non ne avrò la volontà, perché non ho lavorato per me stesso né per il mio tornaconto, e se l'umanità, per amore della quale non ho mai conosciuto una mia vita privata, ha ritenuto possibile distruggere semplicemente ciò che era stato fatto per il suo bene e che non necessitava che degli ultimi ritocchi, ebbene tanto peggio per l'umanità. Ci provino loro a rifare ciò che hanno distrutto. Anche se in modo discontinuo, qualche libro mi arriva, e mi rendo conto che altri ora cercano di risolvere questioni che sono già state trattate da me e da me solo, ma lo fanno alla cieca, a tentoni. Naturalmente, ciò che io ho fatto verrà, parzialmente e a poco a poco, rifatto da altri, ma ci vorranno tempo, forze, denaro e l'occasione giusta. [...] Conosco abbastanza bene la storia e lo sviluppo storico del pensiero per poter prevedere che un giorno si metteranno a raccogliere i cocci di ciò che hanno distrutto»[22].

Sebbene queste parole possano sembrare presuntuose, e forse per questo vengono raramente citate dalla critica florenskijana, ritengo che siano nutrite da un'alta dose di consapevolezza.

Proprio per questo, infatti, Sergij Bulgakov, grande amico e ammiratore di padre Pavel, quando ricevette a Parigi la notizia della sua morte, organizzò una serata presso l'*Institut Saint-Serge* ove, tra le cose, disse:

> «Di tutti i contemporanei che ho avuto la ventura di conoscere nel corso della mia lunga vita è il più grande. E tanto più grande il delitto di chi ha levato la mano su di lui, di chi lo ha condannato a una pena peggiore della morte, a un lungo e tormentoso esilio, a una lenta agonia»[23].

Infatti, il contributo che già Florenskij aveva dato, e quello che anche durante la detenzione stava offrendo al suo Stato e all'umanità, erano immensi.

[22] Lettera ad Anna, 10-11 marzo 1936.

[23] S.N. Bulgakov, *Svjashhennik o. Pavel Florenskij*, in *Vestnik russkogo khristianskogo dvizhenija*, CI-CII, 3-4, 1971, p. 126.

ISSN 0039-4130

Lo stesso lavoro di ricerca cui mi sto dedicando con l'aiuto di tanti maestri e colleghi, è profondamente interpretato dalle parole profetiche dell'autore nella citata lettera alla moglie: un tentativo tra i tanti di raccogliere i cocci di quanto abbiamo distrutto. Se finora si ha l'impressione che gli studiosi abbiano lavorato all'accumulazione settoriale di questi cocci, oggi vorremmo provare a vederli nella loro reciproca integrazione: quale architettura questi cocci tendono a delineare?

Come scriveva egli stesso alla figlia Olga, infatti, l'obiettivo e anche il gusto del lavoro intellettuale consiste nel comprendere ciò che si studia come un intero, cioè vedere come questo intero pone in essere, crea, le sue parti e i suoi organi [24].

Infatti,

> «"L'intero è prima delle sue parti" (Aristotele), cioè: l'intero crea, deduce, pone le parti da se stesso, mentre ciò che è fabbricato, essendo composto dalle sue parti e dipendendo da esse, rappresenta soltanto un'idea astratta dell'interazione di queste parti. L'intero qui non c'è. Invece, dove c'è un intero, là le parti che questo genera si manifestano come organi [25]».

Questo, concretamente, significa

> «rendere palese l'architettura di un'opera e stabilire un legame organico fra i suoi singoli organi e tessuti (in un'opera creativa, infatti, non ci sono parti, ma soltanto organi). Allora diventa chiaro che in un'opera anche le contraddizioni e le incoerenze derivano dal suo disegno generale e che, di tutte le possibilità pensabili, la esprimono nel modo più pieno [26]».

L'ultima considerazione sul fatto che a uno sguardo integrale anche le contraddizioni risultano parte viva e vivificante dell'intero, apre l'altro importante spaccato circa le preoccupazioni di Florenskij sul proprio lascito.

2.3. Intero e molteplicità frammentata

Sul fronte del lavoro complessivo della sua vita, infatti, egli sapeva che la frammentarietà della sua generale metodologia era tanto indispensabile all'integra organicità dell'*opus* – e, quindi, alla potenza di verità del

[24] Cfr. Lettera a Olga, 1-4 novembre 1935.
[25] *Ibidem*.
[26] *Ibidem*.

ISSN 0039-4130

suo messaggio –, quanto indecifrabile ai suoi contemporanei e, com'egli presumeva, anche ai posteri.

Lo stesso problema, come non doveva essergli sfuggito nel corso delle sue letture, era stato a suo tempo affrontato da Goethe, quando incorse nello spregio dei contemporanei di fronte a un'opera ch'egli aveva molto a cuore e che aveva coltivato e fatto crescere in anni di studi, disegni, osservazioni e riflessioni: la *Metamorfosi delle piante*.

Goethe, infatti, ricorda di essersi contemporaneamente dedicato alla scrittura di un saggio sull'arte, la maniera e lo stile, del *Carnevale romano* e di un altro saggio per spiegare la metamorfosi delle piante. Gli accadde tuttavia di incontrare il rifiuto alla pubblicazione di quest'ultimo da parte del suo editore: «il pubblico rimase sorpreso; giacché, nel desiderio di vedersi servito bene e in modo uniforme, pretende da ciascuno che resti nel suo campo. [...] La stessa canzone mi era ripetuta da altre parti; nessuno voleva ammettere che si potessero combinare scienza e poesia»[27].

Secondo Goethe, infatti, coloro che osservano un oggetto secondo ipotesi diverse e spesso contrastanti, sono «uomini onesti e amanti della verità, ai quali preme soltanto la conoscenza della cosa»[28]; proprio costoro, tuttavia, come colui «che in silenzio si occupa di un argomento serio e cerca in tutta sincerità di abbracciarlo nell'insieme, non si rend*ono* conto che i contemporanei sono abituati a ragionare in modo ben diverso»[29].

Anche nelle conversazioni con Eckermann, lette e rilette da Florenskij[30], Goethe parlava della concezione di un'opera creativa – e modello ne era, in quel frangente, l'opera poetica – come di un insieme di intuizioni occasionali, non pianificate ma sgorgate da una viva sorgente nell'impatto con la realtà e poi *composte* in un *intero*[31]. Infatti, «tutti i

[27] J.W. Goethe, *Vicende del manoscritto*, in *La metamorfosi delle piante*, cit., p. 83.

[28] *Id.*, *Lavori preliminari per la morfologia*, in *Ibid.*, p. 106.

[29] *Id.*, *Vicende dell'opuscolo*, in *Ibid.*, p. 85.

[30] In una lettera alla madre del periodo universitario scrive: «Rileggo per la terza o la quarta volta le *Conversazioni con Goethe* raccolte da Eckermann, e ogni parola di Goethe suscita ammirazione. A ottant'anni d'età manifesta interessi così onnicomprensivi e una tale vitalità, gioisce e si indigna così ardentemente e intensamente, che vorrei quasi trascrivere l'intero libro nel mio quaderno di appunti, in vista delle opere future», Lettera del 2-4 febbraio 1903, in *Obretaja put*, a cura di P.V. Florenskij, Progress-Tradicija, Moskva 2011, p. 240. Molti anni più tardi, dal gulag, scrive alla figlia: «Mi chiedi delle *Conversazioni con Goethe* di Eckermann. Certo, conosco il libro, e non solo l'ho letto, ma l'ho pure studiato a suo tempo, anche se tanti anni fa, mi pare nel 1900 o nel 1901, quando tu ancora non eri al mondo. È un libro ricco di contenuti e pregevole, ancor più perché fu rivisto dallo stesso Goethe», Lettera a Olga, 22-23 aprile 1935.

[31] Cfr. J.P. Eckermann, *Conversations with Goethe*, 25 ottobre 1823, a cura di A. Blunden, Penguin Books, London 2022, ebook.

ISSN 0039-4130

diversi fili che legano insieme un intero, e sono l'uno nell'altro intessuti nel suo disegno, devono essere mostrati nel dettaglio. I giovani possiedono una visione monoprospettica sulle cose, mentre un lavoro più lungo richiede una molteplicità di punti di vista»[32].

Questa idea era alla base di tutto l'operare florenskijano. Nella già citata lettera a Kirill del 21 febbraio 1936, padre Pavel esplicita come il variare dei punti di vista nel corso delle sue opere[33] non fosse un accidente casuale del percorso di pensiero, ma una sua condizione costitutiva, fondata sulla *contemplazione del mondo come un intero*.

La concezione dell'intero come forza sorgiva che precede e genera l'articolazione delle parti – arricchito e non sminuito dalla loro multiformità e frammentarietà –, diviene, nei consigli di Florenskij al figlio Kirill, un vero e proprio metodo per la composizione di un libro:

> «riguardo alla domanda che mi hai fatto sulla tua tesi di laurea [...], devi riflettere per bene su che cosa hai accumulato e mettere insieme elementi apparentemente diversi, completare questa composizione con nuovi sforzi [...]. Il pensiero vivo è per forza dialettico, dunque anche contrappuntistico: vi si intrecciano, contrapponendosi e combinandosi, elementi diversi e, se appartengono a campi diversi, proprio allora nasce l'approccio alla natura da autentico ricercatore. Nella composizione sii libero e audace, dai spazio alla fantasia e al gioco delle rappresentazioni. Invece, nella scelta della redazione finale, nell'argomentazione sii estremamente attento, elaborando ogni dettaglio. Soprattutto cerca di evitare l'accumulazione meccanica, la costruzione meccanica, l'esposizione meccanica. È meglio l'incompletezza, la frammentarietà, la problematicità del lavoro, che la compilazione forzata in stile tedesco (di cui soffrono quasi tutti i nostri scienziati) di materiali e idee. Tu, invece, cerca di scrivere in modo che "i pensieri siano liberi e le parole strette" (Goethe)»[34].

A quanto pare, nella prospettiva florenskijana, non solo la contrapposizione tra diversi elementi è percepita come la ricchezza stessa che presiede alla genesi di un testo, ma anche l'eventuale frammentazione e problematicità del lavoro sono da preferire a meccaniche procedure scrittorie di tipo logico-nozionistico.

[32] *Ibid.*, 18 settembre 1823.

[33] Tale variazione dei punti di vista, come si vedrà, è riscontrabile tanto nella diversificazione delle discipline di cui si occupa nei diversi interventi e pubblicazioni, quanto nella strutturazione interna dei singoli scritti, spesso caratterizzati da una forte policentricità dei punti di vista.

[34] Lettera a Kirill, 10-11 dicembre 1936.

ISSN 0039-4130

Stupisce come, nonostante la molteplicità degli studi diffusi su Florenskij, non ci si sia ancora concentrati su questo aspetto centrale della sua dinamica creativa (e della sua visione del mondo), che lo stesso Kirill, pochi mesi prima di morire, dichiarerà essere la chiave di volta per la comprensione della variegata opera del padre:

> «I materiali su Florenskij, o più precisamente intorno a Florenskij, sono molti. Oltre a quelli pubblicati, parte si presentano come manoscritti relativamente completi ma i più numerosi sono in forma di singole note, abbozzi, pensieri, appunti e date. Questo materiale è molto vasto e vario e per metterlo in ordine occorre "accordarsi alla tonalità" di Pavel Aleksandrovich, e questo non è semplice; beh, probabilmente avrei dovuto farlo io, ma mi sono trovato così vincolato al lavoro nella mia specializzazione da non essermici mai potuto dedicare interamente. È evidente che l'attenzione su Pavel Aleksandrovich si sta intensificando. Perché questo accade, perché è così interessante? Probabilmente perché ha cercato di creare una visione complessiva del mondo (общее мировоззрение, *obshhee mirovozzrenie*), come scrive lui stesso in diversi punti. Che sia giusta o no è un'altra questione, ma è comune presso coloro che si percepiscono un'unità. A cominciare dal XVII secolo, e comunque dal XIX secolo, gli esseri umani, la scienza e l'arte sono andate sempre più smembrandosi e dividendosi. Si sono creati innumerevoli metodi di analisi – chimici, fisici, matematici, eccetera. L'arte e la letteratura si sono frantumate in migliaia di "-ismi". E la sensazione immediata, l'immediata percezione della natura e della realtà circostante, è andata sempre più scomparendo. Le migliori menti hanno già detto da tempo che si tratta di una strada pericolosa e che non può proseguire all'infinito; che deve iniziare un'epoca di sintesi. Anche Karl Marx diceva che in futuro la scienza della natura e la scienza dell'uomo sarebbero state un'unica scienza. Non ci sarà opposizione tra le cosiddette scienze esatte e quelle umanistiche. Nonostante queste cose fossero state predette anche da altri, la vita è andata nella direzione dello smembramento e della parcellizzazione. E tutto questo si va rafforzando sempre più. Vien dunque fatto di chiedersi perché a Pavel Aleksandrovich sia stato possibile pensare a una sintesi di questo tipo e pensarci seriamente, e di lavorare su di essa, e non semplicemente di parlarne. Me lo spiego così: Pavel Aleksandrovich è un dialettico molto profondo. Lui vedeva le crepe profonde che pervadono la coscienza moderna, ne parlava, e sosteneva che non bisognasse coprirle, né distogliere da esse l'attenzione o attenuarle, ma, al contrario, metterle in rilievo e comprendere come si dovesse decidere di esse per creare la verità, nonostante la dualistica rappresentazione del mondo»[35].

[35] K.P. Florenskij, *Vystuplenie v Abramchevskom musee 28 janvar 1982 goda na vechere, posvjashhennom 100-letiju so dnja rozhdenija svjashhennika Pavla Florenskogo*, in *Svjashhennik Pavel Florenskij v vospominanijakh svoikh detej Kirilla i Olgi*, a cura di A. Trubachev et alii, Moskva 2011, pp. 13-14.

ISSN 0039-4130

3. *Conclusione: visione e praxis*

Le parole di Kirill mettono in luce diversi nuclei importanti.

Innanzitutto, rappresentano una conferma circa la composizione variegata e complessa dell'eredità florenskijana, con l'affermazione che un criterio di ricostruzione del *corpus* non può che scaturire da una sintonia con la *mens* e gli intenti profondi dell'autore, quello che Kirill definisce "accordarsi alla sua tonalità".

Esprimono, poi, la convinzione che il fascino attuale di Florenskij derivi dal suo tentativo di delineare una *visione integrale del mondo*, mentre le tendenze contemporanee tendono alla parcellizzazione delle conoscenze.

In seguito, evidenziano come questo intento integrale non sia perseguito per mezzo di un sistema complessivo che definisce ruolo e collocazione delle diverse specializzazioni, quanto piuttosto mediante una percezione unitaria di ciò che "parla linguaggi differenti", come nel caso delle diverse scienze esatte e, dall'altro lato, delle scienze umane [36].

Infine, attira l'attenzione sul fatto che i processi conoscitivi di Florenskij non omettono le cosiddette *crepe*, ovvero le falle del pensiero che, nel corso dei secoli, hanno interpellato la storia della cultura umana. Per esempio, si può pensare a quando l'autore fa riferimento all'*horror imaginarii* e all'*horror discontinuitatis* [37] come ai «due spauracchi del pensiero» che è giunto il tempo di abbattere.

Vorrei soffermarmi su una particolare osservazione deducibile da questi rilievi che concerne il concetto di *intero*.

3.1. L'intero

Il concetto di *intero*, in effetti, sembra pertenere almeno quattro livelli di lettura dell'opera florenskijana, che provo di seguito a enucleare.

In primo luogo, dal punto di vista più esterno, l'intero sembra farsi incontro come la chiave di lettura armonica con cui avvicinarsi alle opere del nostro; nonostante la frammentazione dell'eredità, dovuta a

[36] «Sia la scienza che la filosofia, essendo entrambe modi della lingua, sono casi specifici d'uso della lingua...», P.A. Florenskij, *Le antinomie del linguaggio*, in *Attualità della parola: la lingua tra scienza e mito*, a cura di E. Treu, Guerini e Associati, Milano 1989, p. 61.

[37] Cfr. P.A. Florenskij, *Gli immaginari in geometria*, a cura di M. Spano e A. Oppo, Mimesis, Milano 2021, p. 62.

ISSN 0039-4130

cause sia interne che esterne [38], il senso dell'eredità di Florenskij ci potrà raggiungere solo se la nostra attività ermeneutica considererà le sue opere nell'*insieme* da esse costituito.

In secondo luogo, come Kirill suggerisce e padre Pavel stesso afferma a più riprese [39], pare che si debba ritenere l'elaborazione di una nuova concezione integrale del mondo come l'obiettivo dirimente e complessivo di tutto il suo operato culturale.

In terzo luogo, l'intero è un concetto filosofico a più riprese citato dal nostro, che, considerando l'annosa questione dell'*uno e i molti* e il *problema degli universali*, lo affronta direttamente ne *Il significato dell'idealismo*: «Εν (Én) non è solo ed esclusivamente ἓν (uno), ma è allo stesso tempo πολλά (*pollà*, molti) e addirittura παν (*pan*, tutto). Dall'ἓν (uno) che noi vediamo *qui* e *ora* si tendono innumerevoli fili verso l'altro, verso il παν (tutto), verso l'esistenza universale, verso la pienezza dell'essere. E questi fili – sono fili vivi. Sono arterie e nervi che partono dall'ἓν estratto e isolato, il quale è un organo *vivo* di un soggetto vivo» [40]. Infatti, scrive sempre nello stesso libro, «nel momento in cui i nostri occhi si aprono e il mondo ci appare nella sua profondità, noi vediamo il bosco come un unico essere, tutti i cavalli come un super-cavallo e l'umanità come l'unico *Grand Être* di Auguste Comte, come l'*Adam Qadmôn* della Qabbalah, oppure come l'Übermensch di Nietzsche» [41].

In quarto luogo, infine, la concezione dell'intero si irradia capillarmente in tutti i suoi processi gnoseologici arrivando a configurarsi come un vero e proprio metodo conoscitivo e creativo. Lo afferma egli stesso in un testo del 1922, ove sostiene che i «*metodi del futuro*» [42] dovranno accogliere sistematicamente il concetto di intero: «I campi di forza elettrica e magnetica, l'isteresi, persino i fenomeni di elasticità meccanica necessitano di metodi completamente nuovi che accolgano sistematicamente il concetto di intero – "il quale viene *prima delle sue*

[38] Cfr. C. Benassi, *Chi è Pavel A. Florenskij?*, cit., pp. 191-197. L'articolo apre la problematica circa la frammentazione dell'opus florenskijano; la tesi di dottorato, nel cui contesto questi interventi si collocano, si occuperà di sviluppare meglio l'argomento.

[39] Cito una delle occorrenze più eminenti: «Florenskij assunse a scopo della propria vita l'apertura di nuove vie per una futura e globale visione del mondo. In questo senso può essere definito un filosofo» (1927), P.A. Florenskij, *Autoreferat*, cit., p. 5.

[40] P.A. Florenskij, *Il significato dell'idealismo*, a cura di N. Valentini, trad. C. Zonghetti, Rusconi, Milano 1999, p. 66; ho parzialmente modificato la traduzione basandomi su: Id., *Smysl Idealizm*, in *Sochinenija v chetyrekh tomakh*, III/2, Mysl, Moskva 1998, p. 84.

[41] Ivi, p. 109.

[42] P.A. Florenskij, *I numeri pitagorici*, in *Il simbolo e la forma*, a cura di N. Valentini e A. Gorelov, Bollati Boringhieri, Torino 2007, p. 233.

ISSN 0039-4130

parti" – e col quale, di conseguenza, si determina il comporsi degli elementi. Che è poi la forma»[43].

3.2. Il metodo

Se l'intero si rivela dunque come un *leitmotif* della produzione e del contributo florenskijano – e sarà dunque necessario accordarvi le nostre ermeneutiche –, dalle osservazioni sopra riportate, e in particolare legando insieme il terzo e il quarto livello citati, emerge anche un altro snodo di importanza cruciale. Esso mi sembra costituire uno dei nuclei sorgivi più promettenti che l'opera di padre Pavel consegna al mondo culturale contemporaneo, e consiste in ciò che denominerei *trasformazione di un contenuto concettuale in prassi metodologica*.

In padre Pavel l'intuizione teorica – e quindi la visione filosofica –, lungi dal diventare un principio sistematizzante o un contenuto da catalogare (si tratterebbe infatti, in entrambi i casi, di concrezioni statiche), si trasforma in *praxis* noetica. In altre parole, la visione si incarna e diventa metodo creativo; si attua in *energeja* operante e *dynamis* creativa, configurando un metodo di conoscenza capace di operare rimanendo fedele al concreto vivente.

Come fa notare Valentina Parisi nel suo saggio sulla ripresa di Florenskij nei *samizdat* degli anni Ottanta, Evgenij Šiffers, «appassionato cultore del pensiero religioso russo»[44], aveva colto questa specifica potenzialità dell'opera florenskijana e «si autoinvestì del compito di interpretare le loro[45] opere alla luce delle teorie sul "carattere ontologico (o realista) della creazione" esposte da Florenskij»[46]; riflettè, inoltre, a più riprese sul proprio approccio metodologico, «ribadendo in termini estremamente volontaristici l'intenzione di "andare alla scuola di padre Pavel per capire la pittura *astratta*"»[47].

[43] *Ibidem*.

[44] V. Parisi, «*Cattivi allievi di una buona scuola». La ricezione dell'opera di Pavel Florenskij nel samizdat sovietico degli anni Settanta e Ottanta*, in *Pavel Florenskij tra icona e avanguardia*, a cura di M. Bertelé, Edizioni Ca' Foscari, Venezia 2019, p. 182.

[45] Si riferisce ad alcuni esponenti dell'arte non ufficiale con i quali Šiffers era entrato in contatto a San Pietroburgo negli anni Sessanta, quali Ilja Kabakov, Ėduard Šteinberg e Vladimir Jankilevskij, da lui raggruppati sotto l'etichetta di *metafisici primitivi*. Cfr. *ibidem*.

[46] *Ibidem.*

[47] Ivi, p. 184. Corsivo nel testo.

ISSN 0039-4130

È mia opinione che l'esercizio critico di Šiffers getti luce sull'importante caratteristica del lavoro florenskijano che stiamo considerando, secondo la quale non esistono contenuti morti da apprendere e memorizzare, ma solo incontri vivi con una qualche verità manifestata e incarnata in una data forma (una teoria matematica, una struttura geologica, una poesia, una composizione musicale...). A causa di questo approccio, la mente di padre Pavel non accumulava nozioni ma assimilava "semi" destinati a germogliare progressivamente in lui, trapassando dall'immobilità del contenuto acquisito alla dinamicità di potenziale sorgente per nuove modalità esistenziali e conoscitive.

Alla luce del percorso svolto in questo breve saggio, sembra dunque di poter ritenere che l'esperienza di Šiffers non rappresenti solo una pagina significativa della riscoperta di Florenskij, ma anche la linea di una possibile lettura critica – autoctona, dinamica e profondamente spirituale – che non trovò seguito né in Russia né all'estero, e che forse le future ermeneutiche florenskijane dovrebbero accogliere e approfondire.

Nel contesto di questo approccio critico assume importanza centrale l'antinomia tra ἔργον ed ἐνέργεια, cui padre Pavel ciclicamente ritorna nel contesto di studi anche molto distanti tra loro. Questa antinomia, di matrice aristotelica, raggiunge il nostro con forza particolare nel contesto delle sue riflessioni di filosofia del linguaggio, specialmente mediante gli scritti di Wilhelm von Humboldt, ove si configura il raffronto tra l'aspetto monumentale della lingua e il suo aspetto creativo [48], «l'antinomia tra la lingua come prodotto finito [вещность, *veshhnost*] e la lingua come *attività* [деятельность, *dejatelnost*], tra *érgon* (ἔργον) ed *enérgeia* (ἐνέργεια)» [49].

Il concetto viene ulteriormente ripreso e sviluppato dal nostro nelle sue lezioni *Sulla conoscenza storica* mediante le due concrezioni simboliche della *pietra* e del *poeta*, la prima riferita all'aspetto statico («la pietra, per essere pietra, deve rimanere la stessa» [50]) e la seconda all'aspetto creativo, che Florenskij lega alla vocazione poetica: «l'ispirazione di un poeta non è ἔργον, è un processo vivo, è ἐνέργεια. [...] Il poeta, per essere poeta, deve incessantemente creare, dare il *nuovo*» [51].

Come ho recentemente avuto modo di osservare in uno studio

[48] Cfr. P.A. Florenskij, *Le antinomie del linguaggio*, cit.
[49] Ivi, p. 61.
[50] P.A. Florenskij, *Ob istoricheskom poznanii*, in *Sochinenija*, III/2, cit., p. 19.
[51] *Ibidem*.

ISSN 0039-4130

introduttivo sulla produzione poetica di Pavel Florenskij[52], questo concetto – di cui egli rintraccia l'attualità operante in numerosi altri contesti della storia del pensiero e della cultura –, si manifesta anche come un principio articolatore della sua metodologia creativa, generando nella sua produzione una continua tensione tra genesi e intero, tra tensione dinamico-creativa e sguardo unitario-complessivo.

Ritengo, dunque, che uno degli obiettivi fondamentali per i prossimi studi florenskijani possa essere quello di incrociare la visione d'insieme delle opere e della vita dell'autore con la visione dall'interno dei suoi scritti, per enucleare e illustrare la sua concezione del mondo e la nuova e peculiare *praxis* vitale-conoscitiva da essa generata.

[52] Cfr. C. Benassi, *La genesi e l'intero. Florenskij tra poesia e metodologia poetica*, in *Rivista di Letterature Moderne e Comparate e Storia delle Arti*, LXXVII, 2024, pp. 277-297.

ISSN 0039-4130

The Role of «Symbolic Consciousness» in Virgilio Melchiorre's Philosophy

Flavia Chieffi

Introduction

In reading his speculative development retrospectively, Melchiorre locates its beginning in Parmenides' principle. Recalling the Parmenidean principle, he refers to the years of his university education, when his meeting with Gustavo Bontadini was crucial. Defining his philosophy by the model of "classical metaphysics", Bontadini referred to the original thought of Being as the constitutive *logos* of every authentic philosophy. But his meeting with Bontadini occurred only after Melchiorre's research for his dissertation, which led him to Kierkegaard's philosophy [1], «a thinker who gave flesh to my concepts, dividing them into the dialectical polarities of existence» [2]. In the Preface to a collection of studies of Kierkegaard's philosophy, he wrote:

> «these themes would be meaningless if they were not approached starting from the living flesh of existence and only returning to its heart searching for its meaning and destiny» [3].

In fact, Virgilio Melchiorre's theoretical elaboration is an attempt to integrate a metaphysical research, aimed at the constitutive themes of Be-

[1] Since no English translations of Melchiorre's work are available, all the translations in this paper are by the author.

[2] V. Melchiorre, *Dal principio di Parmenide alla fenomenologia trascendentale. Per un'autobiografia intellettuale,* Agostini, Milano 2013, p. 3.

[3] V. Melchiorre, *Prefazione*, in Id., *Le vie della ripresa. Studi su Kierkegaard*, Vita e Pensiero, Milano 2016.

ing, with an existential one. His research has been dialectically placed, from the beginning, between two polarities: the metaphysical and the existential. He investigates the relationship between existence and Being, immanence and transcendence, particular and universal, and lastly human being and God. In his intellectual autobiography, he defines the result of the interaction between these polarities, as a «crossroads with multiple exits»: on one hand ontological argumentation, and on the other an anthropological discourse [4]. Recognising the value of the person – influenced by the study of the philosophy of Emmanuel Mounier, to whom he has dedicated several works [5] – his theoretical framework does not allow either for an immanentistic solution of the relationship, nor the theorisation of absolute transcendence. The main objective of his philosophy is to find a form of relationship preserving difference, and then to explore its structure and its way of expression.

The aim of this paper is to analyse the role of «symbolic consciousness» in Virgilio Melchiorre's philosophy, specifically in the relationship between existence and Being. Firstly, I will examine the «ambiguous structure of existence». Starting from the analysis of perception and its limitation, an intrinsic striving of perspective consciousness to transcend itself will be identified. The ambiguous structure of existence is explained by an intentional duplicity of consciousness.

Secondly, the metaphysical relationship between existence and Being is scrutinised. The nature of this relationship explains the ontological duplicity on which the ambiguous structure is based and it will require its proper mode of expression.

The last part of my discourse focuses on the role of imagination and the notion of the «symbol». The «symbolic consciousness» proved to be the most appropriate way to express the metaphysical-ontological relationship between Being and the human being.

[4] V. Melchiorre, *Dal principio di Parmenide alla fenomenologia trascendentale*, cit., pp. 5-6.

[5] Melchiorre's first work, and the main one, on Emmanuel Mounier's philosophy is: *Il metodo di Mounier ed altri saggi,* Feltrinelli, Milano 1960. In chronological order, other studies dedicated to Mounier's philosophy are as follow: *Il poeta e l'inutile: la concezione estetica di E. Mounier,* in *Rivista di Estetica*, 1959; Id., *Linee di fondazione del concetto di persona*, in *Mounier trent'anni dopo*. Atti del convegno di studio dell'Università Cattolica, Milano, 18 ottobre 1980, Vita e Pensiero, Milano 1981; Id., *La presenza di Mounier in Italia*, in *Emmanuel Mounier: le ragioni della democrazia*, a cura dell'Istituto Emmanuel Mounier, Lavoro, Roma 1986; Id., *Dire persona dopo Mounier*, in *Rivista di filosofia neoscolastica*, XCVIII, 2, 2006; Id., *Mounier. Per un'ontologia della persona*, in *Rivista di filosofia neoscolastica*, XCVIII, 2, 2006, pp. 215-236.

ISSN 0039-4130

1. *The ambiguous structure of existence*

«The primitive condition of any form of consciousness is perceptual orientation: consciousness is always situated in space and time»[6], it is always perspective. This conclusion results from the analysis of perception. The human being can be open to the world always from a point of view, at a certain time and in a certain place. As a consequence of this constitutive limitation, the whole object is never accessible. Due to the partiality of the perceptual function, the human being can only perceive one aspect after another in a temporal succession, only one side of the thing.

However, by examining the structure of perception in depth, it can be observed that in every single perspective the whole object is given at the same time. Referencing the phenomenological tradition, Melchiorre argues that it is only possible to perceive one side of being insofar as another side is already included, that a partial aspect of it can be distinguished if it emerges from a background. On one hand, without a term of reference, there is no warning of the limit. On the other hand, these relations do not fully appear. In every perspective, the totality of the object is indicated at once but in absence, by 'profiles' and 'adumbrations'. The object manifests itself with its own fullness of being or sense while transcending the present perspective of the object. However, without being referred to a totality, perceiving a limited perspective would be a contradiction, a non-sense. Consequently, the synthesis of the human gaze of the world lives mostly in the field of absence[7]. The perception is transcended in the anticipation of something that is absent[8]. As a consequence, the initial statement can be supplemented. Consciousness is both situated and a principle of desituation; it is not only immanence but also the ability to transcend. Melchiorre defines this intentional duplicity, characterising the human ontological condition, as "ambiguity":

> «By ambiguity I mean a dialectical movement coming to consciousness from its temporality and its spatial condition: always being situated and at the same time capable of desituating itself, always being in extension and at the same time always beyond extension, in short, its unity of immanence and transcendence in the being of the world»[9].

[6] V. Melchiorre, *L'immaginazione simbolica. Saggio di antropologia filosofica,* Mulino, Bologna 1972, p. 16.

[7] Cf. *Ibidem,* pp. 19-20 ss.

[8] V. Melchiorre, *La coscienza utopica*, Vita e Pensiero, Milano 1970, pp. 102-103.

[9] V. Melchiorre, *L'immaginazione simbolica,* cit., pp. 103-104.

ISSN 0039-4130

2. *The metaphysical relationships between existence and Being*

The analysis of the stream of perceptual consciousness requires reflection to move from phenomenology to metaphysics. The transcendence in the core of immanence is due to Being: while constituting intimately every single reality, its transcendence is ineliminable. Due to this, according to Melchiorre, whatever can be defined is intrinsically ambiguous; it states about itself, but it states so in relation to something else; it qualifies itself by referring to what it is not. This non-being constitutes it intimately, it inhabits it and makes it what it is but, at the same time, it displaces its own boundaries and consequently consciousness [10]. The emergence of non-being within appearance invites the consciousness to exceed the limit. Whenever a limit becomes evident, a non-sense or non-exhaustiveness of reality is also revealed, but the prominence of a contradiction implies a more primordial affirmation of sense. A universal contradiction and a total nonsense are bound to be unthinkable. The perspective transgression arises from the underlying evidence that every determination, when considered in isolation, is ultimately nonsensical, thereby necessitating a deferral. The path of the perspective consciousness and the impatience of contradiction witness a root of intelligibility, Being justified in itself, without reference to anything [11].

Melchiorre claims the indemonstrability of Absolute Being. The demonstration that something is requires it to be recognised as real only at the end of the demonstration, which must therefore begin from other guarantees and foundations. However, Being, the basis of reality, is implied in every other foundation. As a consequence, the demonstration presumes what it sought to prove and leads to the recognition of the original evidence of Being [12]. The consciousness arising from the Foundation is always in Truth, although it does not mean an identity but a metaphysical relationship with Being [13]. In fact, the becoming of consciousness, the power of being what it was not, testifies to its alterity from Being. However, from pure nothingness nothing follows; an absolute non-being could never come to be. The non-being that consciousness can become has always been included in Being [14]. Whatever comes to reality has always been in Being, but not in the way in which it

[10] Cf. V. Melchiorre, *La coscienza utopica,* cit., p. 12.
[11] Cf. V. Melchiorre, *L'immaginazione simbolica,* cit., pp. 25-26.
[12] Cf. *Ibid.,* pp. 28-29.
[13] Cf. V. Melchiorre, *La coscienza utopica,* cit., p. 10.
[14] Cf. *Ibid.,* p. 36.

ISSN 0039-4130

is now distinguished from it. On one side, because of its partiality and abstraction from the Totality, existence constitutes a diversity, despite its foundation. On the other side, Being, the common core of all reality, appears in each presence and, at the same time, points to an absence which it refers to. The transcendence of the Origin is ineliminable, due to its purity and necessity, which are beyond all becoming. This transcendence establishes an ontological duplicity on which the ambiguous structure is based, and defines consciousness as unity, a relationship of being and non-being, presence and absence [15].

3. *The notion of «symbol» and the role of the «symbolic consciousness»*

After having detected the structure of the relationship with Being, Melchiorre tries to identify a function reflecting the double intentionality of the consciousness. Resuming the analysis of perception, Melchiorre suggests that the faculty of imagination allows a connection between the particular and the universal, presence and absence. While imagination depends on perception, the opacity of perception and its limit urges the life of imagination because it bears witness to a vacancy, to something transcendent and, at least, to Being itself [16]. Melchiorre recalls what Husserl describes as both a «paradox» and a «rigorous truth», namely that «fiction is the source from which knowledge of eternal truths draws its nourishment» [17]. Paraphrasing Kant, Melchiorre states that perceptual knowledge would be powerless without the synthesis of imagination [18]. Due to its intentionality living in the world of absence, the imagination complements perception and leads to the absent. However, absence cannot be communicated in itself, but «can only be perceived analogically, in something that is present and that refers back to it by an intrinsic participation» [19]. The determination to which the imaginary leads is symbolic [20].

[15] Cf. *Ibid.,* p. 94-96.

[16] Cf. V. Melchiorre, *L'immaginazione simbolica,* cit., p. 31.

[17] E. Husserl, *Ideen zur einer reinen Phämenologie und phänomenologischen Philosophie. Allgemeine Einfürung in die reine Phänomenologie, Husserliana* (1950), I, Nijhoff, Den Haag 1976, p. 151.

[18] Cf. V. Melchiorre, *L'immaginazione simbolica,* cit., p. 10.

[19] V. Melchiorre, *La coscienza utopica,* cit., p. 101.

[20] In his intellectual autobiography, Melchiorre points out that, from the very first years of his research, he had a strong interest in aesthetics, an interest that divided him from his teacher Bontadini, who considered aesthetics as an "independent variable" compared to ontology and first philosophy. On the contrary, for Melchiorre, ontological discourse had to be intimately

ISSN 0039-4130

The sense of «symbol» is retrieved by Melchiorre from its Greek etymology. The *symbolon* was originally a hospitable token that represented the bond between family and family, between city and city. The *Symbolon* was a tally, whose two halves guests recomposed during their reunion to verify a communion after a long absence [21]. Referring to this original meaning of symbol, Plato says that the human condition is symbolic itself, and it is through love that humans testify to their remote origin, seeking a unity divided by the god. In its original sense, «symbol» echoes the ideas of absence and distance together with the original bond with Being [22]. This connection cannot be expressed along the path of univocity, but exclusively along that of symbolism. A reality is considered «symbolic» when, while speaking about itself, it also speaks about another owing to an original communion. Identity and diversity, reciprocity and distance are simultaneously maintained. As Paul Ricoeur wrote, «The symbol is the movement of primary meaning that makes us share in the latent meaning and assimilates us to it» [23]. The two fields are not simply merged but simultaneously kept in their distinction and returned to their deepest participation [24]. As a result of these considerations, it is symbolic consciousness that, while being in the individuality, reads the universal there, holding difference and identity together. This duplicity can occur, since it is supported by an intentional movement converging on the universal as Being everywhere participates and transcends itself. Every individual reality participated by Being expresses, at the same time, its absence [25].

linked to aesthetics. In fact, his first philosophical work was *Arte ed esistenza,* L'impronta, Firenze 1956, and this work is full of references, both to aesthetic theories and to art. It was this path that then led him to the study of symbolic thought (V. Melchiorre, *Per un'autobiografia intellettuale,* cit., p. 5).

[21] Cf. V. Melchiorre, *L'immaginazione simbolica,* cit., pp. 36-37.

[22] In talking about symbolic determination, Melchiorre takes as example Van Gogh's painting: *Vincent's Chair.* In Melchiorre's interpretation, the chair is not only a useful object but the expression of the man; it reveals a painful relationship with the world. Its lines and colors allude to an invisible world; they refer to a knowledge of some kind of destiny between things. Furthermore, this inherent knowledge in things recalls to some kind of absolute. In the case of Van Gogh's *Vincent's Chair*, being seated is like a search for peace, for repose recalling a sort of absolute peace (cf. V. Melchiorre, *L'immaginazione simbolica,* cit., pp. 72-74).

[23] P. Ricoeur, *Le conflit des interprétations,* Seuil 1969, p. 286.

[24] Cf. V. Melchiorre, *L'immaginazione simbolica,* cit., pp. 45, 46.

[25] Cf. *Ibid.,* p. 54.

ISSN 0039-4130

Conclusion

The symbol, while identifying, maintains differences. The possibility of symbolic consciousness itself is explained by the constitution of man in difference, originally defined in the transcendence of Being. Although Melchiorre's philosophical elaboration recognises a single foundation to all reality which unites all that exist, his theoretical system does not allow for an immanentistic solution, because it rejects a notion of absolute transcendence [26].

Being is original evidence, the foundation of all determination, yet it is grasped in the indeterminacy of reference; it is immanent transcendence [27]. Referring to Heidegger [28], Melchiorre remarks that thought is always thought of Being, but in its difference from beings. For this reason, Being is not found directly. The existent, in which Being is only intended, was the first to be grasped. Following a phenomenological approach, Melchiorre seeks the meaning of the finite in the finite itself, but the sense of every finite being transcends itself. Every being states about itself, but in its *inseity* also states about the Infinitely other and precisely by analogy. The consciousness can express this analogy through its symbolic language [29]. The possibility of symbolic expression arises from the participation between transcendental experience and ontic experience.

This study has attempted to illustrate that, in Virgilio Melchiorre's philosophy, the duplicity or the ambiguity of symbolic expression is ultimately suitable for the consciousness facing a transcendence, transcendence that is also participation and immanence. The possibility of symbolic expression arises in the being of every phenomenon, in its constitutive relationality, in its «ambiguous unity». The ambiguity of being, both existence and Foundation, is followed by an expressive ambiguity saying of this being, an ambiguity finding its expression in the unity-distinction of the symbol.

Melchiorre elaborates his own foundation of metaphysics, privileging the anthropological dimension as a starting point. Indeed, the binomial "Being and person" has been the common thread of Melchiorre's elaboration. The whole of Being is never given in itself, but only in the determination of beings, and this difference is manifested in the inten-

[26] Cf. *Ibid.*, p. 55.
[27] Cf. V. Melchiorre, *Metacritica dell'eros,* Vita e Pensiero, Milano 1977, p. 48.
[28] M. Heidegger, *Identität und differenz,* Klett-Cotta, Stuttgart 1957.
[29] Cf. V. Melchiorre, *Metacritica dell'eros,* cit., p. 19.

ISSN 0039-4130

tional movement of consciousness. If Being is the ultimate origin, the true beginning, then it is reached by departing from the finite, through a movement that returns to its own principle and uncovers its a priori operation. The question of Being thus becomes one with the question of the human being. The analysis of consciousness constitutes the fundamental precondition to decide the meanings of Being, because the modes of consciousness constitute the ways by which Being appears.

Bibliography

M. Heidegger, *Identität und differenz,* Klett-Cotta, Stuttgart 1957.
E. Husserl, *Ideen zur einer reinen Phämenologie und phänomenologischen Philosophie. Allgemein*e *Einfürung in die reine Phänomenologie, Husserliana* (1950)*,* vol. I, Nijhoff, Den Haag 1976.
V. Melchiorre, *L'immaginazione simbolica. Saggio di antropologia filosofica,* Mulino, Bologna 1972.
Id., *Metacritica dell'eros,* Vita e Pensiero, Milano 1977.
Id., *Dal principio di Parmenide alla fenomenologia trascendentale. Per un'autobiografia intellettuale,* Agostini, Milano 2013.
Id., *Le vie della ripresa. Studi su Kierkegaard,* Vita e Pensiero, Milano 2016.
P. Ricoeur, *Le conflit des interprétations,* Seuil 1969.

ISSN 0039-4130

Civic and Citizenship Education in Italy

From School Organization to Teaching Practices

Francesca Fioretti

1. *Cross-curricular teaching of civic education in Italy. An evolving path*

Civic and citizenship education (CCE) in Italy has gone through several stages, reflecting the country's social and political changes. The first to introduce the teaching of civic education in lower and upper secondary schools was Minister of Education *pro tempore*, Prof. Aldo Moro, who issued the Presidential Decree No. 585 of 13 June 1958. Civic education was defined as an expression of the relationship between the educational purposes of the school and the projection of each person toward social, legal, and political life as well as on the principles on which the community was founded and realized (Annex, Presidential Decree No. 585/1958). The program, centered on the Italian Constitution, aimed to provide students with an organic synthesis of related concepts, providing two hours per month entrusted to the History teacher, later expanded to one hour per week as part of the school timetable entrusted to the Humanities teacher [1].

In subsequent years, civic education was integrated into various educational areas, such as the "six educations" introduced by the Moratti Reform (Law No. 53/2003). Civic education was later changed and incorporated into *Cittadinanza e Costituzione*, introduced by Law Decree No. 137/2008. It was later converted into Law No. 169/2008, which was aimed at national experimentation and teacher training in

[1] *See* V. F. Allodola, *Il "grande ritorno" dell'educazione civica a scuola: struttura, funzioni, limiti e potenzialità (durante la pandemia)*, in *Studi sulla Formazione/Open Journal of Education*, XXIV, 1, 2021, pp. 145-157.

the first and second cycles of education on knowledge and competences related to *Cittadinanza e Costituzione*, integrating it into the historical-geographical and historical-social areas (Art. 1, Law No. 169/2008). The teaching of *Cittadinanza e Costituzione* was finally included in the historical-geographical area by Presidential Decree No. 89/2009 (Art. 5, paragraph 6).

One of the main normative documents in teaching *Cittadinanza e Costituzione* was Ministerial Circular No. 86/2010, which stressed the need to pay attention to the educational emergency facing young people and the role of schools in the challenge of reaffirming respect for the human person, civic sense, and individual and collective responsibility. In this scenario, *Cittadinanza e Costituzione* held fundamental importance, as the study of the Constitution necessitated a conscious adherence to a value framework capable of inspiring behaviors [2].

Law No. 92/2019 marked a decisive turning point, establishing the cross-curricular teaching of civic education in the first and second cycles of education, starting from kindergarten. This teaching, entrusted to all teachers for a total of 33 hours per year (Art. 2, paragraphs 3-4), is divided into three thematic cores: the Constitution, law (national and international), legality, and solidarity; sustainable development, environmental education, knowledge and protection of heritage and territory; and digital citizenship. Following the enactment of Law No. 92/2019, the *Linee Guida per l'insegnamento dell'educazione civica* were published on 22 June 2020, as contained in Ministerial Decree No. 35/2020 (Annex A), providing a theoretical framework to outline the principles of the Law No. 92/2019.

In September 2024, new Guidelines, issued by Minister Giuseppe Valditara, replaced the previous ones, introducing some changes. These changes included revising the term from "sustainable development" area to "economic development and sustainability", as well as redefining some learning objectives. The *Consiglio Superiore della Pubblica Istruzione* expressed concerns about these new Guidelines, pointing out the lack of adequate monitoring of school initiatives and criticizing the new terminology and the wording of some objectives.

[2] *See* M. Mattei, *Cittadinanza e Costituzione: una "disciplina" formativa per eccellenza nella scuola secondaria di primo grado*, in L. Corradini-G. Mari (edited by), *Educazione alla cittadinanza e insegnamento della Costituzione*, Vita e Pensiero, Milano 2019.

ISSN 0039-4130

2. *Methodology*

This contribution is part of a broader research project that aims to comprehend the organizational and teaching practices used by schools in Italy and Portugal to implement democratic learning environments, and to describe the leadership styles of principals that emerge from these school profiles. To this end, we conducted an embedded multiple-case study with exploratory purposes [3] and a simultaneous mixed-method design [4] in four lower secondary schools in Rome and four third-cycle schools in Porto.

Referring to the schools in Rome, they were selected by means of convenience sampling guided by two criteria: the implementation of activities and projects for CCE; and the promotion of a non-vertical organizational structure in which tasks and responsibilities were distributed among members of the school community through the presence of coordinators and committees. An unstructured document analysis from the schools' websites served as the basis for the selection.

The data collection took place in the 2022/2023 school year. The surveys were conducted through semi-structured interviews with principals and civic education coordinators, teacher questionnaire, direct classroom observation, and focus groups with teachers. Quantitative data from questionnaires and observation checklists were analyzed by descriptive statistical analysis, while textual data from focus groups and interviews, previously transcribed, were analyzed by thematic analysis.

To ensure the reliability of the coding, a triangulation of the data was performed and Intercoder Reliability (ICR) calculated. Two independent coders from outside the research group triangulated the Italian transcripts, coding a random sample of 20% of the responses in each transcript. This sample aligns with the recommendation retrieved from the literature to select 10-25% of the dataset for a reliable ICR [5].

We calculated the ICR using two methods: Krippendorff's alpha and percent agreement. The Krippendorff's alpha coefficient [6] related to the intercoder reliability for Italian transcripts, calculated using the

[3] *See* R.K. Yin, *Case study research: Design and methods*, Sage, Thousand Oaks 2009.

[4] *See* J. M. Morse, *Mixed Method Design: Principles and Procedures*, Routledge, New York 2009.

[5] *See* C. O'Connor-H. Joffe, *Intercoder reliability in qualitative research: debates and practical guidelines*, in *International Journal of Qualitative Methods*, IXX, 2020, pp. 1-13.

[6] *See* K. Krippendorff, *Content analysis: An introduction to its methodology*, Sage, Thousand Oaks 2019.

ISSN 0039-4130

K-Alpha Calculator[7] software, was 0.757. Although this value does not reach the threshold of 0.80 generally considered satisfactory, it is important to point out that Krippendorff's alpha may not be the most appropriate method for calculating the coding reliability of semi-structured interviews and focus groups, because they violate the two basic assumptions for calculating alpha: the equal probability of using each code in all documents and the equal qualification/expertise of independent coders.[8] For this reason, we calculated also the agreement percentage among the two coders and principal investigator, that obtained percentages of 78.57% and 81.29%, in line with the 70% threshold suggested in the literature.[9] To mitigate the influence of chance agreement on the outcome, Brennan and Prediger kappa coefficients[10] were also calculated, which were found to be 0.78 and 0.81.

3. *Civic and citizenship education in school. Main results of the schools in Rome*

We now present the main findings about the organizational structure and the civic education curricula of the four schools in Rome, emphasizing their similarities and differences, illustrating the various realities that arise from the implementation of reference regulations and school autonomy.

In the S-1 school, located in the Prenestino neighborhood in Rome, the organization of civic education presents a defined structure, but with some critical issues. The Civic Education Curriculum includes the definition of competences, objectives, and content for each thematic core to be achieved at the end of the lower secondary school. This curriculum was structured by the Civic Education Commission, which was established at the initiative of the school itself to guide the teaching of civic education for the year. This Commission establishes a subdivision of the thematic cores of civic education among the three grades of

[7] *See* G. Marzi-M. Balzano-D. Marchiori, *K-Alpha Calculator—Krippendorff's Alpha Calculator: A User-Friendly Tool for Computing Krippendorff's Alpha Inter-Rater Reliability Coefficient*, in *MethodsX,* XII, 102545, 2024, pp. 1-10.

[8] *See* J. L., Campbell-C. Quincy-J. Osserman-O. K. Pedersen, *Coding in-depth semi-structured interviews: Problems of unitization and intercoder reliability and agreement*, in *Sociological methods & research*, XLII, 3, 2013, pp. 294-320.

[9] *Ibidem.*

[10] *See* R. L. Brennan-D. J. Prediger, *Coefficient kappa: Some uses, misuses, and alternatives*, in *Educational and psychological measurement*, XLI, 3, 1981, pp. 687-699.

lower secondary school, demanding teaching focused on learning units and reality tasks: digital citizenship in grade 6, sustainable development in grade 7, and the Constitution in grade 8. The instructional planning for each class is entrusted to the Class Council, in which each teacher is responsible for integrating civic education within their subjects from a common theme. In addition, following the relevant regulations, there is a coordinator within each class; in the case of S-1, this task is entrusted to the special education teachers.

However, the implementation of such a teaching organization has not been without criticism from the teaching staff. Some teachers complain about the difficulty of integrating civic education into subjects with already complex curricula, perceiving it as an additional task. Furthermore, teachers express dissatisfaction with the rigidity of the Commission's established teaching and curricular organization, citing its potential to restrict flexibility and adaptability to students' specific needs. The lack of time for interdisciplinarity and for co-design, highlighted as a challenge, hinders a cross-curricular approach to civic education that integrates different disciplines and competences; as a result, teaching activities are based on textbook study and discussions on social issues.

Civic education in the S-2 school, located in the Appio-Latino neighborhood in Rome, is distinguished by a rigorous structuring, which conceals some critical issues. The Civic Education Curriculum is presented as a rigidly structured document, with a precise definition of knowledge, skills, competences, and teaching hours per subject for each thematic core. Such rigid structuring results in a fragmented and disinclined cross-curricular approach to civic education, which is perceived as an addition to the curriculum of individual disciplines, rather than as an opportunity for cross-curricular learning. The lack of a working group or committee dedicated to civic education contributes to this fragmentation. This situation leads some teachers to consider civic education as a macro-topic to be included in their subjects, rather than a cross-curricular area to be addressed organically.

This lack of collaboration is also reflected in teaching activities, which focus mainly on textbook study and guided discussions on social issues; there are no school-wide projects, except for the election of student representatives, who, however, are not included in school governance. Although there are some activities related to projects proposed by external associations and NGOs, such as Amnesty International, the school has no significant collaborations with the local community.

ISSN 0039-4130

Similar to S-2, in S-3, also located in the Appio-Latino neighborhood in Rome, the introduction of cross-curricular teaching of civic education led to the creation of a highly structured curriculum in which competences, skills, content, and teaching hours for each subject are defined for each thematic core (with the exception of digital citizenship).

A distinguishing feature of S-3 is the absence of a school coordinator for civic education, whose role is deemed superfluous. This is related to the systematic annual repetition of the teaching activities and assessment materials developed in the 2020/2021 school year – the year of the introduction of cross-curricular teaching of civic education in schools – without, however, promoting a process of monitoring and evaluating the effectiveness of these activities and tools. At the same time, according to the former coordinator, such systematic repetition makes up for the lack of space and time dedicated to the exchange and co-design among teachers, as it allows to establish upstream coordination among teachers by disciplinary departments. The role of the class coordinator is also considered unnecessary, although present, since it is reduced to the collection of annual class schedules and assessments assigned by teachers for the proposal of the final grade.

The connection between disciplinary departments has led to the implementation of vertical projects. However, the former coordinator points out that interdisciplinarity in civic education is limited because it would take time away from teaching the other disciplines. What would favor a cross-curricular approach to civic education is to dedicate one week a year entirely to this teaching, proposing an institute project involving all its members and all disciplines.

This kind of initiative was developed by the S-4 school, located in the Nuovo Salario neighborhood in Rome. The teaching and learning process for civic education in the S-4 school is distinguished by a dynamic and participatory approach, focusing on teacher collaboration and openness to the local community. The Civic Education Curriculum does not appear as an additional element but accompanies the school's structural documents that rest on the key competences for lifelong learning defined in the *Recommendation of The European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning*. Four thematic axes are identified in the curriculum, as they are cosidered essential for the development of cross-curricular and metacognitive competences, and serve as the foundation for civic edu-

cation: the language axis; the mathematics axis; the science-technology axis; and the historical-social axis.

Unlike the other schools, S-4 has established a joint Department of Civic Education, composed of primary and secondary teachers, civic education class coordinators and non-civic education class coordinators. This Department meets monthly to define common, interdisciplinary planning, filling the gap of co-planning spaces not provided for in the national regulations; the difference with the S-1 Commission, which instead meets annually to define school-level coordination, already appears here. The establishment of this Department results in increased collaboration among teachers, which is also reflected in the implementation of numerous activities for civic education, many of them linked to national and international days, such as the International Day for the Elimination of Violence against Women.

As previously mentioned, the teaching activity that most distinguished S-4 from other schools was the organization of the Civic Education Festival, a week-long event held in February 2023 during which daily teaching was suspended to enhance the work done by classes for civic education during the year. Unlike the idea of devoting one week a year to civic education that has emerged in other schools, the Festival does not overlap with or replace the cross-curricular teaching of civic education during the school year, but enriches and complements it.

The Festival included multiple activities in collaboration with local associations and open classroom activities that involved students in common spaces, promoting experiential and participatory learning. The Festival covered a wide range of topics, including ecology, European institutions, and social inclusion. Some activities also involved students leaving the school premises with members of partner associations offering them the opportunity to engage in active citizenship rather than merely learning about civic-related topics.

The strong openness of S-4 to the local community appears here; indeed, many activities, including the Civic Education Festival, are held in collaboration with associations. This approach stems from the desire not to limit civic education to the mere transmission of knowledge, but to create situations in which the student acts and experiences life in the community.

ISSN 0039-4130

4. *Conclusions*

The four schools in Rome analyzed in the study show markedly different approaches to cross-curricular teaching of civic education. Some schools, such as S-1, S-2, and S-3, adopt a structured but rigid approach, based on detailed curricula that define knowledge, skills, and teaching hours for each subject and thematic core. Sometimes teachers in these schools complain about the difficulty of integrating CCE into subjects with already complex curricula, perceiving it as an additional task. An additional obstacle to cross-curricular teaching in these schools is the lack of a dedicated civic education team and effective coordination at the school level. Consequently, individual disciplines often fragment and relegate CCE to a macro-topic, rather than addressing it as a cross-curricular area that permeates the entire school organization. Teaching activities focus mostly on textbook study and guided discussions, with few initiatives that promote teacher collaboration and active student participation.

In contrast to this approach, the S-4 school stands out for its dynamic and participatory vision of CCE. Instead of creating a separate curriculum, S-4 integrates CCE into its structural documents, based on the key competences for lifelong learning defined by the *Recommendation of The European Parliament and of the Council of 18 December 2006*. In addition, the establishment of the joint Department of Civic Education is a means of fostering exchange and cooperation among teachers, promoting integrated learning paths. A key element of S-4's approach is the openness to the local community. The school actively collaborates with local associations, providing students with opportunities for active citizenship and experiential learning. This immersive approach, of which the Civic Education Festival is its actualization, allows students to experience firsthand the concepts of active citizenship, social responsibility, and democratic participation.

The schools' experience shows that defining effective educational pathways requires the organic integration of CCE into school life, which fosters teacher collaboration, active student participation, and openness to the local community. Based on flexibility, collaboration, and experiential learning, S-4's approach sets an inspiring example, promoting a CCE that transcends classroom instruction and embraces the school in its complexity.

ISSN 0039-4130

Learning to Teach Civic and Citizenship Education and Education for Sustainable Development During Pre-service Teacher Training

Marco Valerio

Introduction

Civic and Citizenship Education is a subject that has garnered significant global interest over the years. This interest is evident from the growing number of scientific publications on the topic [1] and the launch of some important studies focusing on how students are prepared to be responsible and active citizens through school education, like the ones conducted by the International Association for the Evaluation of Educational Achievement (IEA) [2]. CCE addresses a wide range of topics, including political system and constitution of the country, human rights, international organizations, equal opportunity for men and women, citizen rights and duties, global issues and sustainability [3]. By integrating these topics into compulsory education, schools can promote a culture of active and responsible citizenship. Various denomi-

[1] *See* N. Pérez-Rodríguez - E. Navarro-Medina - N. de-Alba-Fernández, *Citizenship education in teacher training: A systematic review*, in *Journal of Social Science Education*, 2024, XXIII, 3, pp. 1-20; W. A. Galston, *Civic Knowledge, Civic Education, and Civic Engagement: A Summary of Recent Research*, in *International Journal of Public Administration*, 2007, XXX, 6-7, pp. 623-642.

[2] W. Schulz - J. Fraillon - J. Ainley - B. Losito - D. Kerr, *International Civic and Citizenship Education Study 2009: Assessment Framework*, Springer, 2008.

[3] W. Schulz - J. Fraillon - B. Losito - G. Agrusti - J. Ainley - V. Damiani - T. Friedman, *IEA International Civic and Citizenship Education Study 2022 Assessment Framework*, Springer, 2023.

nations of CCE can be found in the literature, and it is often referred to as Citizenship Education [4] or Civic Education [5], however, the core content remains the same. The International Civic and Citizenship Education Study (ICCS), conducted by the IEA, describes how CCE is

«Implemented to provide young people with knowledge, understanding, and dispositions considered necessary to participate successfully as citizens in society. Young people should understand civic principles and institutions, know how to engage in civil society, be able to exercise critical judgment, and develop an understanding and appreciation of the rights and responsibilities of a citizen». [6]

To foster a society whose future members will be active citizens, promoting civic knowledge, attitudes and engagement within the school context is fundamental. The European Commission, in the Eurydice 2017 Report, describes how Citizenship Education «is understood to include not only the teaching and learning of citizenship-related matters in the classroom, but also practical experiences gained through activities in school and the wider society that are designed to prepare students for their role as citizens in the democracies they live in» [7]. These definitions clearly suggest that CCE is not just about theoretical knowledge; it involves a broader set of student capacities, such as skills, attitudes, values, and actions. Thus, this subject aims to develop students' competence [8] to act as engaged and aware citizens.

One of the main topics related to CCE is sustainability [9]. Interest in this topic has grown steadily over the past sixty years from the Club of Rome [10] to the Agenda 2030 [11]. Sustainable Development was first

[4] Eurydice, *Citizenship Education at School in Europe – 2017*, Publications Office of the European Union, 2017.

[5] J. C. Fitzgerald - A. K. Cohen - E. Maker Castro - A. Pope, *A Systematic Review of the Last Decade of Civic Education Research in the United States*. in *Peabody Journal of Education*, XCVI, 3, pp. 235-246.

[6] W. Schulz - J. Fraillon - B. Losito - G. Agrusti - J. Ainley - V. Damiani - T. Friedman, *IEA International Civic and Citizenship Education Study 2022 Assessment Framework*, Springer, 2023, p. 3.

[7] European Commission/EACEA/Eurydice, *Citizenship Education at School in Europe – 2017.* Publications Office of the European Union, 2017, p. 18

[8] *See* Council of Europe, *Reference Framework of Competences for Democratic Culture*, Volume I: Context, Concepts and Model, Council of Europe Publishing, 2018.

[9] W. Schulz et al, *IEA International Civic cit.*, 2023, p. 7

[10] *See* D. H. Meadows - D. L. Meadows - J. Randers - W. W. Behrens III*, The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind,* Universe Books, 1972.

[11] United Nations. *Transforming our world: The 2030 agenda for sustainable development.* United Nations, 2015.

ISSN 0039-4130

defined in the Brundtland Report as «[the] development that meets the needs of the present without compromising the ability of future generations to meet their own needs». [12] This concept gained global popularity, increasingly shaping the culture and policies of numerous countries around the world. Today, many international organizations (e.g. The United Nations [13], UNECE [14], UNESCO [15], etc.) aim to achieve a more sustainable society and are actively working toward sustainable development.

Education is one of the primary means through which a more sustainable society can be achieved [16]. For this reason, an adequate education for sustainable development seems to be vital to reach a more sustainable society and meet the target 4.7 of the Agenda 2030 «ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles» [17]. Therefore, integrating Education for Sustainable Development into compulsory education is essential for achieving a sustainable society.

Education for Sustainable Development can be considered a significant aspect of CCE [18]. However, ESD is generally not taught at school as a separate subject; rather, topics related to sustainability are often integrated within CCE learning objectives [19].

ESD can be defined as an educational area which «helps to develop the capacity for critical reflection and systemic and futures thinking, as well as to motivate actions that promote sustainable development» [20]. The main goal of ESD is not merely to inform students about sustain-

[12] United Nations, *Report of the World Commission on Environment and Development Our Common Future*, UN Documents, 1987, p. 37.

[13] United Nations. *Transforming our world: The 2030 agenda for sustainable development.* United Nations, 2015.

[14] UNESCO, *Education for Sustainable Development: A Roadmap*, 2020.

[15] UNECE, *Ten years of the UNECE Strategy for Education for Sustainable Development Evaluation report on the implementation of the UNECE Strategy for Education for Sustainable Development from 2005 to 2015*, United Nations New York and Geneva, 2016.

[16] UNECE, *Empowering educators for a sustainable future Tools for policy and practice workshops on competences in education for sustainable development*, United Nations, Geneva, 2013

[17] United Nations. *Transforming our world: The 2030 agenda for sustainable development.* United Nations, 2015. p. 21

[18] W. Schulz et al., *IEA International Civic cit.*, 2023.

[19] Eurydice, *Citizenship Education at School in Europe – 2017*, Publications Office of the European Union, 2017.

[20] UNECE, *Learning for the future Competences in Education for Sustainable Development*, United Nations, 2012, p. 6

ISSN 0039-4130

able development and its key domains (society, economy and environment),[21] but rather to develop their understanding of the interconnectedness and interdependence of various features of sustainable issues, encouraging them to take action for positive change. Climate change, ecological footprint, anthropogenic greenhouse effect and responsible consumption and production, among other topics, are a major focus of ESD. These issues cannot be treated merely as factual knowledge to be memorized; they require pedagogical approaches that inspire students to engage actively[22].

To equip the next generation of citizens to meet the challenges of the modern world, it is crucial that our teachers receive appropriate training to teach CCE and ESD. However, it is not always evident from a normative standpoint whether these topics are included in the pre-service teacher training curriculum. In particular, the Portuguese and Italian contexts do not have clear regulations on how future teacher should be prepared to teach CCE and ESD during their pre-service training[23]. Nonetheless, the Portuguese and Italian education systems address CCE as a compulsory cross-curricular subject which has a limited number of hours and has no single-subject teacher but all teachers share responsibility for it[24].

This study will allow to gather further information on this issue through the analysis of four pre-primary and primary teacher training programs. The research will take place in two different countries, Italy and Portugal, that share a similar lack of clear regulation dedicated to how CCE and ESD should be embedded in the teacher training program for primary and pre-primary school teachers.

Research Goals

This study is aimed at investigating how CCE and ESD are integrated into the pre-service primary and pre-primary teacher training programs of four specific higher education institutions. Rather than merely identifying

[21] United Nations, *Report of the World Commission on Environment and Development Our Common Future*, UN Documents, 1987

[22] P. Vare - W. Scott, *Learning for a Change: Exploring the Relationship Between Education and Sustainable Development*, in *Journal of Education for Sustainable Development*, 2007, I, 2, pp. 191-198.

[23] Eurydice, *Citizenship Education at School in Europe – 2017*, Publications Office of the European Union, 2017.

[24] *Ibid.*

ISSN 0039-4130

whether CCE and ESD topics are addressed in the teacher training program for each context taken into account, the study will investigate how and to which extent the programs provide the adequate means to prepare future teachers to teach Civic and Citizenship Education at school and to implement learning strategies for Education for Sustainable Development. Therefore, the research will delve into the comprehensive training prospective teachers receive, focusing not only on the inclusion of CCE and ESD-related topics in teacher training programs but also on whether or not and to which extent they are prepared to teach these subjects.

The study will also shed light on students' perspective on the education received during their years of study at university, with a special focus on CCE and ESD. The goal is to assess students' perceived readiness to teach, particularly in relation to CCE and ESD. Collecting information from students' perspective on their experience at university is crucial to understand the effectiveness of these programs and evaluating their self-reported preparedness to address CCE and ESD in their future professional practices. To get a comprehensive view of the full teacher training program, the research will focus solely on last-year students' perspective on the received training. This will ensure a broad view of the training provided by the teacher training programs. Students' point of view will give invaluable insights into any possible strengths or weaknesses of the teacher training program, drawing a thorough picture of why and to which extent the training received has prepared them to teach CCE and ESD.

The study will encompass the role played by sustainability in the selected universities. In addition to the presence of ESD topics within teacher training programs, the research aims to shed light on the role played by sustainability during university years and the relevance of sustainability given by these institutions. To do this, it will be necessary to involve all the main stakeholders of the universities (e.g. students, professors, etc.) and gather information on the concrete presence of actions and policies dedicated to sustainability and to develop a culture for sustainable development.

Finally, by triangulating the data collected from four university contexts and analysing the strengths and weaknesses of each teacher training program, it will be possible to develop hypotheses on how CCE and ESD can be effectively integrated into pre-service teacher training for primary and pre-primary schools. Based on these findings, a model will be proposed to demonstrate how CCE and ESD can be incorporated into a teacher training curriculum.

ISSN 0039-4130

Methodology

The study adopts a multiple-case study approach [25], analysing the teacher training programs at selected universities in Italy and Portugal. The criteria for the selection of the universities are the presence of CCE and ESD dedicated courses or the presence of courses addressing topics related to CCE and ESD in the educational provision of the programs. To achieve this, an educational provision analysis of all Italian and Portuguese universities that provide pre-service teacher training programs for pre-primary and primary teachers will be conducted. The goal is to select universities that meet the previously specified criteria.

A mixed-methods methodology will be employed to achieve the research goals, combining quantitative and qualitative data collection tools. Quantitative data will be collected through a questionnaire administered to pre-service teachers in their final year of training. This questionnaire aims to detect their perceived preparedness to teach CCE and ESD. It will also gather information on the CCE and ESD training courses received at university, as well as the topics related to CCE and ESD covered by the overall education provided at the university. Additionally, it will collect data on the perceived readiness to teach in pre-primary and primary school.

Qualitative data will be gathered through focus groups with final-year pre-service teachers to gain deeper insights into their university experience. They will get a deeper look at the subjects discussed in the questionnaire, allowing us to understand how students were prepared to teach CCE and ESD. Qualitative data collection will also include interviews with university professors who teach courses that directly address CCE and ESD or cover topics related to citizenship or sustainable development. These interviews aim to obtain a thorough understanding of the subjects discussed in their courses and the pedagogical approaches they implement. Interviews with coordinators of teacher training programs will be conducted to gather information on the overall program design and the integration of CCE and ESD. Furthermore, as mentioned before, a document analysis will be carried out to review the content and materials used in the relevant courses. This will allow us to map the presence of CCE and ESD-related topics within the educational provisions of each selected teacher training program.

[25] *See* R. K. Yin, *Case study research: Design and methods*, 2009, V., sage.

ISSN 0039-4130

Through the triangulation of different data sources, both quantitative and qualitative, it will be possible to shed light on how teacher training programs implement CCE and ESD in the selected contexts. This approach will also provide a deeper understanding of pre-service teachers' overall experience and perceived preparedness to teach CCE and ESD in those selected universities.

Expected results and further developments

The study will provide critical insights into the integration and effectiveness of Civic and Citizenship Education and Education for Sustainable Development within pre-service primary and pre-primary teacher training at four selected universities. Specifically, the findings are expected to highlight the degree of self-reported readiness by future teachers in these crucial areas, as well as identify strengths and areas for improvement in the training programs in relation to CCE and ESD. Additionally, by examining the role of sustainability within these universities, the research aims to uncover how sustainable practices and principles are embedded in institutional policies and curricula. These insights could inform and enhance teacher training approaches, potentially leading to stronger preparation of future educators in fostering civic education and sustainability awareness in schools. It will allow the developing of a model of how CCE and ESD can be integrated into a curriculum for teacher training. One of the limits of the study is that only the perceived readiness of future teachers in the selected universities is investigated. However, further insights into the actual preparation of students for their professional practice and their ability to teach CCE and ESD would be highly valuable. To achieve this, proper assessment tools should be developed. This opens up new opportunities for future research aimed at examining the differences between the perceived readiness of pre-service primary and pre-primary teachers in their final year of training and their actual preparedness. Moreover, future studies could explore contexts beyond the Portuguese and Italian ones, enabling a broader understanding of how pre-service teacher training programs for primary and pre-primary teachers are structured in other countries.

ISSN 0039-4130

Conclusion

In conclusion, this study provides insights into how Civic and Citizenship Education and Education for Sustainable Development are embedded within pre-service teacher training programs of four universities in Italy and Portugal. The research will develop a comprehensive overview of the instruction received by pre-service primary and pre-primary teachers, with a specific focus on CCE and ESD. Furthermore, it will not only identify where and through which pedagogical approaches CCE and ESD are addressed within university courses, and the study will also show how future teachers are trained to teach CCE and ESD. The insights gained will highlight both strengths and gaps in current programs, particularly in terms of enhancing preparedness to teach CCE and ESD. Through data triangulation and exploring the perspectives of key stakeholders, including pre-service teachers, university professors and program coordinators, the research will highlight the role played by sustainability within the university contexts selected for the study. The study will also propose a model for integrating CCE and ESD into pre-service teacher training programs. Future studies could build upon these findings by measuring teacher readiness, offering deeper insights into how prepared pre-service teachers are to implement CCE and ESD. Additionally, future research could replicate this study in other countries. Analysing higher education institutions in different contexts could provide further insights into how CCE and ESD are integrated into pre-primary and primary pre-service teacher training programs.

ISSN 0039-4130

Catholic University Students in the 1940s and 1950s.

The Importance of a Professional, Human and Religious Formation.

Francesco Marcelli

1. *Introduction*

This contribution aims to summarize the formative role played within the university context by the Fuci and the Movimento Laureati. The Fuci (Federazione Universitaria Cattolica Italiana) is the association of Italian Catholic students founded in 1896; the Movimento Laureati is an organisation founded in 1933 that brings together Italian Catholic graduates. Both associations, of university students and graduates had as their purpose the professional, human and religious formation of their members.

2. *World War II and the post-war period*

The chronological period 1943-1954 was experienced by Catholic students and graduates as a period of transformation and change. These were the last decisive years of the Second World War, in which Italy suffered military occupation by the Nazis in the centre-north of the Peninsula [1], followed by the years of reconstruction and political-economic recovery after the conflict [2]. The idea of "building a new or-

[1] C. Pavone, *Una guerra civile. Saggio storico sulla moralità nella Resistenza*, Torino, Bollati Boringhieri, 2010.

[2] P. Soddu, *La via italiana alla democrazia. Storia della Repubblica 1946-2013*, Bari, Laterza, 2017.

der" emerged during the war among Italian Catholic intellectuals, who were instrumental in the formation of a state inspired by the values of freedom and democracy and would consider the social doctrine of the Church as a model to follow[3]. During the civil war fought in Italy between 1943 and 1945, some Catholic students and graduates died fighting against the Nazi-Fascist regime, inspired by the idea of building a democratic country[4]. Others engaged in indirect resistance against the occupying power, based on helping people who were politically persecuted and partisans and on the other hand on planning a new order when the conflict would be over[5]. The year 1943 was a turning point in the Second World War and marked the beginning of few attempts by Catholics to think about the organisation of a new democratic order in Italy after the conflict. In fact, in 1943 a document of political and economic planning was written, which inspired some Catholic politicians that governed Italy after the war: the Code of Camaldoli[6]. Important Catholic exponents such as Sergio Paronetto, Pasquale Saraceno, Ezio Vanoni and Vittorino Veronese collaborated in writing the Code. This document, printed in 1945, and the others drafted by members of nascent Christian Democracy influenced in part the deliberations of the Constituent Assembly, particularly regarding the principles of economic policy[7].

Once the conflict was over, university students and graduates were better able to resume their associative activities. In the articles published since 1945 in "Ricerca" or "Studium", the periodicals of the Fuci and the Movimento Laureati, a general optimism emerges, together with the desire to tackle and solve the problems of society[8]. In this regard, Vittorio Bachelet's words about the post-war period are exem-

[3] F. Malgeri, *Cento anni di vita*, in Id. (ed.), *FUCI: coscienza universitaria, fatica del pensare, intelligenza della fede. Una ricerca lunga 100 anni*, Cinisello Balsamo (MI), San Paolo, 1996, pp. 36-41; T. Torresi, *L'altra giovinezza. Gli universitari cattolici dal 1935 al 1940*, Assisi, Cittadella Editrice, 2010, pp. 195-204.

[4] V.E. Giuntella, *I cattolici nella Resistenza*, in F. Traniello e G. Campanini (ed.), *Dizionario storico del movimento cattolico in Italia (1860-1980)*, Marietti, Torino, 1981, I, 2, pp. 112-128.

[5] M. Margotti (ed.), *Dal Codice alla Carta. I cattolici italiani tra Resistenza, realtà internazionale e impegno costituente (1943-1948)*, Camaldoli, Edizioni Camaldoli, 2024.

[6] T. Torresi (ed.), *Il Codice di Camaldoli*, Roma, Studium, 2024.

[7] M. Margotti, *Gli intellettuali cattolici e il Codice di Camaldoli*, p. 21.

[8] See e.g. R. Bertoli, *I professionisti e la ricostruzione. Molto da rifare per i medici*, in *Ricerca*, 25 aprile 1945, n. 1 p. 2; I. Murgia, *Orientamenti*, in *Ricerca*, 30 agosto 1945, n. 8-9, p. 1; A. Castagna, *Non soltanto i partiti*, in *Ricerca*, 30 agosto 1945, n. 8-9, p. 1; Aldo Moro, *Una patria umana*, in *Ricerca*, 15 settembre 1945, n. 10, p. 1; V. Bachelet, *Valori positivi*, in *Ricerca*, 1 giugno 1947, n. 11, p. 1; R. Pietrobelli, *Costruire con costanza*, in *Ricerca*, 1 giugno 1949, n. 12, p. 1.

ISSN 0039-4130

plary: "We felt the duty to rebuild the new State, in which we did not want the effort of democracy to be an empty word but a conscious and real contribution from everyone. We felt the need to donate new energies for these ideals, the need to clarify these ideals to ourselves, to live them" [9]. In fact, as Francesco Malgeri also argues, the themes most frequently addressed in the conferences and congresses of the Fuci in that period concerned the responsibility of Catholics in the new political and social phase [10].

3. *The formative role*

Through congresses, conferences, spiritual retreats, study weeks and general meetings the Fuci and the Movimento Laureati played an important formative role, in which professional, social and religious education was central [11]. The main purposes of this education were to generate cohesion among their members, fruitful exchanges of ideas and thoughts, but above all to transmit a "work ethic". An ethic in which the profession was seen as service for the good of the whole of society and not only as a means of personal economic sustenance. A conception of the profession similar to that described by the sociologist Max Weber, in which the love of one's neighbour is expressed primarily in the fulfilment of professional duties [12]. From this perspective, work acquires the characteristics of a mission by which to transmit human and religious values, thus carrying out a work of apostolate. As can be observed by reading the articles published in "Ricerca", some Catholic university students have often reflected on the task of the intellectual within society, also starting from Max Weber's thought [13].

Many Catholic students and graduates, partly as a consequence of their professional formation, subsequently occupied relevant professional and political roles in Italy. Many of them became important politicians who governed the country in the second half of the last century [14]. Many others, with the purpose of improving academic research

[9] V. Bachelet, *I maestri, i giovani e la storia*, in *Studium*, marzo 1952, n. 3, p. 134.

[10] F. Malgeri, *Cento anni di vita*, p. 40.

[11] G. Marcucci Fanello, *Storia della F.U.C.I.*, Roma, Studium, 1971, pp. 210-212.

[12] M. Weber, *Die protestantische Ethik und der Geist des Kapitalismus*, Tübingen, 1904-1905; id., *Wissenschaft als Beruf*, München, 1919; id., *Politik als Beruf*, München, 1919.

[13] P. Trionfi, *La funzione dell'intellettuale nella comunità italiana*, in *Ricerca*, 1 maggio 1955, n. 8-9, p. 3.

[14] For example G. Andreotti, G.B. Bozzo, E. Colombo, F. Cossiga, G. Gonella, A. Greggi,

ISSN 0039-4130

and educating future generations of citizens, became university professors [15]. There is no space here to list all the professions in which they were active, but it can be said that in the post-war period many members of the Fuci and the Movimento Laureati held key positions within the various professions in Italian society [16]. As the historian Pietro Scoppola also argued, if the concept of a "Catholic hegemony" might be misleading category of interpretation, it is nevertheless undeniable that Catholics belonging to all the professions had a strong direct and indirect influence in the post-war period [17]. Likewise, the Vatican Curia played an undeniable part in creating favourable conditions for the maturation of a ruling class capable of facing the challenges of the post-war period. In particular, the support given by Giovanni Battista Montini (the future Pope Paul VI) to the Fuci and the Movimento Laureati was fundamental, in terms of funding, sharing of contacts and practical advice [18]. He was deeply attached to Catholic students and graduates. In fact, before becoming Substitute at the Secretariat of State, Montini had been the national ecclesiastical assistant of the Fuci between 1925 and 1933 [19].

4. *International perspective*

From an international point of view, the Fuci and the Movimento Laureati were part of Pax Romana, an international association of Catholic students and graduates from all over the world, founded in Fribourg in 1921. The main purpose of Pax Romana was to promote world peace and create unity among Catholics around the world. During an important congress held in Rome in April 1947, Pax Romana was divided into

R. La Valle, F.M. Malfatti, L. Menapace, A. Moro, A. Ossicini, G. Pisanu, P. Pratesi, F. Rodano, M. Scelba, P.E. Taviani, B. Zaccagnini.

[15] For example G. Alberigo, F. Bachelet, V. Bachelet, A.A. Bobbio, V. Cappelletti, A.M. Chiavacci, G. De Sandre, F. Gasparini, S. Golzio, C.M. Gregolin, L. Isgrò, C. Leonardi, G. Mazza, R. Meneghelli, L. Meschieri, F. Montanari, A. Moro, Q. Paris, F. Vito, G. Zappa.

[16] G. Marcucci Fanello, *Storia della F.U.C.I.*, pp. 269-270.

[17] P. Scoppola, *La proposta politica di De Gasperi*, Bologna, Il Mulino, 1988, pp. 15-29.

[18] See e.g. AAV (Archivio Apostolico Vaticano), Segr. Stato, Titoli (1936-2005), anno 1942, Associazioni cattoliche, posizione 116; ibid., anno 1944, Associazioni cattoliche, posizione 163; ibid., anno 1946, Associazioni cattoliche, posizione 196.

[19] F. De Giorgi, *Mons. Montini. Chiesa cattolica e scontri di civiltà nella prima metà del Novecento*, Bologna, Il Mulino, 2012, pp. 115-176; L. Pomante, *«Fiducia nell'uomo e nell'intelligenza umana». La Federazione Universitaria Cattolica Italiana (FUCI) dalle origini al '68*, Macerata, EUM, 2015, pp. 67-118.

ISSN 0039-4130

two complementary movements of university students and graduates: MIEC / IMCS (Mouvement International des Étudiants Catholiques / International Movement of Catholic Students) and MIIC / ICMICA (Mouvement International des Intellectuels Catholiques / International Catholic Movement for Intellectual and Cultural Affairs) [20]. Just as the Fuci and the Movimento Laureati were in close contact, so the MIEC/ICMS and MIIC/ICMICA were also in close contact. In fact, between the respective national movements of Catholic students and graduates a bond of mutual assistance was often established, in which the older ones helped the younger. Within Pax Romana MIIC/ICMICA, the mutual exchange of knowledge and professional skills among Catholic graduates from all over the world has also been very important. Within the MIIC/ICMICA, a fundamental part has been played by the professional secretariats, such as those of doctors, pharmacists, jurists, artists, teachers, engineers and economists [21]. The main purposes of the professional secretariats within Pax Romana were various: first of all, they organised conferences and meetings to address issues affecting the specific profession in relation to Christian morality. Furthermore, members of the various professional secretariats were often provided with suggestions for further reading regarding specific kind of work to improve their general knowledge of the professional sector. Finally, a further aim was to create cohesion among Catholic professionals from all over the world, divided into professional sectors, to increase the exchange of information, skills and knowledge regarding the specific profession to which they belonged.

Analysing the history of Pax Romana give us a better understanding of the urgent need for the Church, in the second half of the 20th century, to conduct an increasingly universal mission with the aim of creating cohesion among young people from all over the world. A cohesion generated by a common interest in culture and the achievement of an intellectually aware faith. Pax Romana, as Kevin Ahern stated, had already understood the importance of renewing the modalities of lay apostolate in contemporary society before the Second Vatican Council [22].

[20] BCU (Bibliothèque cantonale et universitaire de Fribourg), Pax Romana, b. D.2.1 Rome 1947. St. Edmund's 1948. Luxembourg 1949; Ibid., b. H.1.1 Circulaires MIIC 1947-1954.

[21] Ibid., b. H.5 Bulletin des Sous-secrétariats 1946-1954.

[22] K. Ahern, *At the Vanguard of the Geo-Apostolate: Pax Romana (1946-1971),* in *Rivista svizzera di storia religiosa e culturale - Schweizerische Zeitschrift für Religions und Kulturgeschichte - Revue suisse d'histoire religieuse et culturelle,* 2021, vol. 115, pp. 391-406.

ISSN 0039-4130

5. *Young Catholics and Maritain's "integral humanism"*

It is also important to highlight the influence that the French philosopher Jacques Maritain had on some Catholic intellectuals of that period [23]. By proposing the idea of a new humanism, "integral humanism", which placed the human person at the centre of every debate, Maritain affirmed a new way for lay people to live their faith in contemporary society [24]. According to the philosopher, the human being is the result of an inseparable union between matter and spirit; the human being must free himself from the obstacles of modernity, but also appreciate its positive aspects. Maritain's "integral humanism" is a Christian humanism because only in relationship with God can the human being be completely fulfilled. An aspect of Maritain's "personalism" highly appreciated by young Catholic intellectuals in the post-war period was the idea of establishing a State that would guarantee the inalienable freedoms of the human being; the idea of building a country where the individual is not crushed by the omnipresence of the State [25]. These theories of Maritain were certainly an inspiration for many Catholic intellectuals who contributed to the reconstruction after the war.

As well as the Fuci and the Movimento Laureati, Pax Romana also counted many members who considered Maritain a master and his teachings as important elements of reflection [26]. Maritain was sometimes criticised and accused of heresy by the more conservative clergy and the Holy Office; however the leaders of Pax Romana itself firmly opposed such criticisms, even writing to the Pope in defence of the philosopher [27].

6. *Conclusion*

To conclude, it is necessary to highlight the importance of the professional, human and religious education received by young people in the

[23] See e.g. R. Moro, *La Fuci montiniana in una prospettiva storica*, in *Ricerca*, dicembre 1990, p. 43; F. Malgeri, *Cento anni di vita*, p. 36; L. Pomante, *«Fiducia nell'uomo e nell'intelligenza umana». La Federazione Universitaria Cattolica Italiana (FUCI) dalle origini al '68*, pp. 151-152.

[24] J. Maritain, *Humanisme intégral*, Paris, Fernand Aubier, 1936.

[25] J. Maritain, *Man and the State*, Chicago, University of Chicago Press, 1951.

[26] J. P. Warren, *Pax Romana: un des vecteurs de diffusion du maritainisme (1939-1952)*, in *Études d'histoire religieuse*, 2013, LXXIX, 1, pp. 71-91.

[27] BCU, Pax Romana, b. B.2 Autorités: comité Pax Romana, Secretariat, Personnel MIIC et MIEC 1947-1959, Généralités.

ISSN 0039-4130

university environment in the 40s and 50s of the last century. The Fuci and the Movimento Laureati, as well as Pax Romana at an international level, formed the future ruling class. In fact, many of their members, partly as a result of the education received, rose to positions of primary importance in professional, social and political spheres. The profession was often considered a service to improve human society and perhaps, partly because of this mentality, many excelled in their specific work. The awareness of practising a profession to fulfil a mission could generate a strong motivation that often produces more fruitful results.

ISSN 0039-4130

Flaminio Piccoli, the DC and Centrist Democrat International (CDI): Methodology and Goals

Giammarco Basile

The research project aims at a new and wider reading of one of the political personalities of the Christian Democracy (DC) and the history of Republican Italy, Flaminio Piccoli. The project aims to give priority to the political experience of Piccoli, from his youth political education until the conclusion of the Christian Democracy party in 1994. During his decades-long political career, Piccoli held various roles both within the party and at the institutional level, on which historiography, despite some isolated attempts, has not yet undertaken extensive studies [1].

The desire to conduct this project stem on the consideration that, over the years, historiography has focused on studying and analysing the history of Republican Italy and, in fact, much research into this subject has been conducted (analysing many fields like institutions, political parties, society development and its political bond, "labour unions", political terrorism, economic development, etc.).

Historiography made – among the different fields of research – important studies and analyses into many political personalities who played a key role in the social, political and economic development. Due to this work, it has been possible to outline many profiles of the protagonists of Italian Republican history, but many political figures would necessitate a deeper study, aiming at the comprehension of their political impact on the evolution of Republican Italy from a different perspective.

[1] F. Bojardi (eds.), *Flaminio Piccoli. La strategia del coraggio: Profilo ed antologia*, Rubbettino, Roma, 2005; L. Targher, *Gli esordi di un politico nazionale. Flaminio Piccoli, 1945-1958 materiali per una biografia politica*, Fondazione Museo Storico del Trentino, Trento, 2011.

In addition, beside the historiographic works, to conduct investigations over the political personalities, is possible to consult other contributions on the history of the party and some of its members, but these works are mostly memorialist and celebratory, such as collections of political and parliamentary speeches[2] or in the form of personal diaries[3]. Despite those are not a historian works, these contributions could be useful to better understand different point of view on the politic dynamics and on fundamental moments of Italian history, paying attention to not use these sources like a primary, because the risk is oversimplification of Italian political and social reality.

For this reason, it could be useful to continue investigating, through available documentations, the actions and decisions of some DC's political personalities, which played a key role on the development and the evolution of both the party and Italian political history.

The investigation could be articulated following a double track – internal and international.

In the first part, the research aims to analyse Piccoli's formation and political thought, as well as the roles he held from the end of World War II. Initially, in his native region, he was part of a small group that contributed to the birth of the Trentino Christian Democracy. Piccoli, at the beginning of political and socio-economic reconstruction, was appointed responsible for the press, first as deputy director of the Cln diary *Liberazione Nazionale*, and later – after deciding to end this collaboration – director of the Christian Democratic Diary *Il Popolo Trentino*, which would later become *L'Adige* in 1951. Through Piccoli's role in this phase, it is possible to highlight some of the key issues of those years in the local political dynamics, such as: post-war reconstruction; relations with other political forces, with a focus on the Communist party and the Volkspartei; the issue of special autonomy and the related statute, which was a source of major conflicts between the Italian-speaking and German-speaking communities.

[2] Many of the works containing the parliamentary speeches of Italian politicians are published by Il Mulino and edited by the Historical Archives of the Senate. Also available online.

[3] Cf. A. Fanfani, *Diari, vol. I-IV*, Rubbettino, Soveria Mannelli, 2013; P. E. Taviani, *Politica a memoria d'uomo*, Il Mulino, Bologna, 2002; M. Rumor, *Memorie 1943-1970*, Neri Pozza Editore, Milano, 1991; L. Dal Falco, F. Malgeri (eds.), *Diario politico di un democristiano*, Rubbettino, Soveria Mannelli, 2008; S. Fontana, N. Guiso (eds.), *Potere discreto. Cinquant'anni con la Democrazia Cristiana*, Marsilio, Venezia, 2009; S. Andreotti, S. Andreotti, A. Riccardi (eds.), *I diari segreti, 1979-1989*, Solferino, Milano, 2020; G. Andreotti, S. Andreotti, S. Andreotti (eds.), *I diari degli anni di piombo*, Solferino, Milano, 2021.

ISSN 0039-4130

Subsequently, the investigation aims to analyse the role that Piccoli played between 1952 and 1957 as president of the diocesan board of "Azione Cattolica" (A.C.) in Trento. During his presidency, with the support of Monsignor Cesconi, he provided an important impetus for the renewal of the organization and structure of the Catholic association in the region. In those same years, Piccoli's desire to breathe new life into A.C. led him to an open confrontation with the national president, Luigi Gedda, where the Trentino politician had denounced his continuous interference in local affairs. This contrast, in fact, had conducted to Piccoli's suspension from office for nearly a year. After resolving the dispute with Gedda, Piccoli returned to his role until 1957, when he was named secretary of the Trentino DC. Piccoli would hold this position for only one year, since in the 1958 elections he was elected, for the first time, to the Chamber of Deputies, which had defined the beginning of his political career in Rome. During his decades-long political carrier, despite his move from Trento to Rome, Piccoli never abandoned his commitment to his homeland, as evidenced, for example, by his involvement in the establishment of the University of Trento in the 1960s and his work in the Commission of Nineteen for the definition of the second regional statute in the early 1970s.

In this part of investigation, after outlining the path that led Piccoli to Rome, the focus will be on the contributions and roles he played at national level, both within the party and at the institutional level. During his political career, Piccoli held significant positions within the party, such as party secretary, which he elected twice – the first time in 1969, and again from 1980 to 1982 – as well as group leader in the Chamber of Deputies between 1972 and 1978, the year he was named president of the National Council of the DC, a position he would return to after his second term as secretary. At the ministerial level, instead, he held only the role of Minister for State Holdings between 1970 and 1972. Following this, as will be seen, Piccoli had preferred roles within the party rather than other positions within the government, as was suggested in 1978 after Francesco Cossiga's resignation as Minister of the Interior.

For this purpose, will be considered the various positions held by Piccoli within the party and his contribution in the decisions that have had a weight for the country's evolution.

Instead, in the second part of the project, the aim is examining the role of the Christian Democrat politician in the international arena, because Piccoli, in the last years of his political career, had been both

ISSN 0039-4130

President of Christian Democrat International (CDI) and President of Foreign Affairs Committee at the Deputies Chambers.

In fact, one of the aspects that would be explored is his interest for Latin American politics and his role as President of the CDI. In this field of investigation, would like to consider the relations between the Italian DC and those of Latin America, with a particular focus on the Chilean one, especially, between the 1980s and early 1990s, in order to investigate the relations between two parties in the last years of the Pinochet dictatorship and in the first years of the democratic transition, considering the Chilean DC won the first free election after Pinochet dictature ended and in which the Italian DC had a fundamental role with his economic and political support. In addition, it would like to investigate the role of Piccoli within the CDI, international organization, where Christian-inspired parties shared their political points of view and values and search a political coordination to construct a common political platform or program.

In the international arena, the second line of research that would intend to follow, moreover, is the Piccoli's role as President of the Foreign Affairs Commission, in a period of changes in the international political scene.

In this perspective, it might be interesting to analyse the political positions and decisions taken by Piccoli and the party on the evolution of international political system after collapse of the Berlin Wall, end of Cold War and the explosion of new international issues, for example Yugoslavia war, where Piccoli was the first European President of Foreign Affairs Committee to denounce a Milosevic attack in Kosovo.

In order to conduct the proposed project, several sources will be examined. As for the formation of the Trentino politician and his political origins at the local level, the Trentino DC and Azione Cattolica Funds, presents at the Trentino Diocesan Archive in Trento, will be considered. While, on Piccoli's political career in the Catholic party, reference will be made to the extensive documentation available at the Sturzo Institute in Rome. The main source that will be examined is, clearly, the Piccoli Fund. In addition to the extensive documentation of Piccoli Fund, at the Sturzo Institute it will be possible to consult, as an additional source, the DC Fund, in which much of the political production of "Scudocrociato's party" is preserved, without excluding the possibility to consult other Funds of political personalities for a wider reading of the Italian political experience. (e.g. Andreotti Fund, Scoppola Fund).

ISSN 0039-4130

Moreover, trying to investigate further the dynamics of Italian politics will be examined the documentation presented at the Gramsci Foundation and the Craxi Foundation, to better understand the relationship between the three major parties, especially during Piccoli's years like Secretary, Group Leader in the Chamber of Deputies and President of the Foreign Affairs Committee.

Instead, as regards the international aspects, we would like to consult the Mariano Rumor Fund, at the Historical Archives of the Senate, which could help to understand the political relations between Italy, Chile and South America, because Rumor was CDI's President between 1967 and 1982. In order to develop a research plan as wide as possible, the aim is to integrate Italian sources to foreign sources, in particular Chilean. On this aspect, we would like to refer, if useful for the purpose of research, to the documentation preserved at: the National Archives of Chile; the General Historical Archives of the Ministry of International Relations; the Archives of the Corporation, Justice and Democracy; the Gabriel Valdes Historical Archive; the Eduardo Frei Montalva and Patricio Aylwin Azocar Foundations.

Beside foreign archival sources, not defined yet, we would consult a bibliography on Chilean political history, the relations between the two countries and the respective DC parties, before the military dictatorship until the first free elections (1989) in Chile.

Alongside this, for a comprehensive reading of the various topics, in addition to archival and bibliographical sources, we would also take into account the large amount of journalistic party production, both at the level of Diaries – *Il Popolo, L'Avanti, L'Unità* –, to which could be added to the consultation *L'Adige*, edited by Piccoli during his first years of militancy into the party, both periodicals – *La Discussione, Mondo Operaio, Rinascita* – that, from divergent perspectives, offer useful material to reconstruct a wider profile of Piccoli's figure and the role that he played in the history of contemporary Italy.

This research project aims to present to the scientific community, in the form of a political biography, a new contribution about the political history of Republican Italy and one of its protagonists. This desire stems from the consideration that, up to now, the large number of documentations preserved by the Piccoli Fund has not yet been valued and treated as an object of deeper study. For this reason, through this vast documentation it is possible not only to reconstruct the role of Piccoli within the DC, but also to try to bring out the lights and the shadows on the Republican Italian history and DC party, from the fifties until

ISSN 0039-4130

the end of the Christian Democratic party experience, and it could be useful to trace new possible research lines.

Sources

A. Fanfani, *Diari, vol. I-IV*, Rubbettino, Soveria Mannelli, 2013
F. Bojardi (eds.), *Flaminio Piccoli. La strategia del coraggio: Profilo ed antologia*, Rubbettino, Roma, 2005
G. Andreotti - S. Andreotti (eds.), *I diari degli anni di piombo*, Solferino, Milano, 2021
L. Dal Falco - F. Malgeri (eds.), *Diario politico di un democristiano*, Rubbettino, Soveria Mannelli, 2008
L. Targher, *Gli esordi di un politico nazionale. Flaminio Piccoli, 1945-1958 materiali per una biografia politica*, Fondazione Museo Storico del Trentino, Trento, 2011
M. Rumor, *Memorie 1943-1970*, Neri Pozza Editore, Milano, 1991
P. E. Taviani, *Politica a memoria d'uomo*, Il Mulino, Bologna, 2002
S. Andreotti - S. Andreotti - A. Riccardi (eds.), *I diari segreti, 1979-1989*, Solferino, Milano, 2020
S. Fontana - N. Guiso (eds.), *Potere discreto. Cinquant'anni con la Democrazia Cristiana*, Marsilio, Venezia, 2009

ISSN 0039-4130

Mechanism and Free Will: A possible Convergence Hypothesis

Enrico Di Meo

Introduction

It is well known that human capacity for self-determination (or self-control), what is generally called "free will", is one of the most ancient and recurring enigmas of humankind. It has engaged philosophers, theologians and scientists for centuries, resurfacing each time in a different form. In recent decades the question acquired particular relevance, especially due to the massive development of neuroscientific research. A great number of scholars started to investigate the complex relationship between neural activity and conscious intention, in particular after the pioneering works of Benjamin Libet, some of them coming to the shocking conclusion that free will is merely an illusion. These results give strength to a renowned stream of thought, namely determinism.

This, of course, poses serious problems, both theoretical and practical. Firstly, not all neuroscientists agree with the deterministic interpretation of experimental evidence; secondly there are philosophical assumptions and implications of these experiments that are often overlooked; and finally, one of the main concerns regarding determinism is the sense that a complete mechanistic explanation would radically undermine and explain away our whole complex of notions centred around freedom and moral responsibility (something often advocated by proponents of determinism).

In order to make explicit these problematic philosophical assumptions and delineate a possible *convergence hypothesis* it will be useful

to examine a 1971 paper written by Charles Taylor [1]. Although this paper precedes the development of neuroscience which followed Libet's results, the theoretical question was already much debated in the Anglo-Saxon world.

1. *Taylor's Convergence Hypothesis.*

I would like to start by following the line of reasoning displayed by Taylor in this 1971 paper dedicated to the problem connected with mechanistic explanation of human behaviour and to a possible *convergence hypothesis* that allows us to avoid both dualism and reductionism, without discarding phenomenological experience and scientific results.

Taylor's initial question is extremely eloquent: «must a neurophysiological account of human behaviour be a mechanistic one?» [2]. The point Taylor wants to stress is emphasized by this initial «must». He does not aim to undermine or disprove neurophysiological experiments and their results; instead, he wants to challenge the unspoken assumption that *any* account that is displayed in scientific terms *must* be mechanistic. There is little doubt, Taylor says, that the scientific development of our age seems to confront us with the difficulty (perhaps impossibility) of having to choose between the two poles of a Kantian antinomy [3]. On one hand, it seems natural to assume that the development of the neurophysiological and physical research will lead us to a mechanistic comprehension of all the processes that occur in our brain, and this will explain (possibly even anticipate) our behaviour and action; on the other hand, our self-understanding derived from the common sense is particularly alarmed by such perspective, especially because this would radically undermine our whole complex of notions

[1] The paper has been republished in a 1985 collection. C. Taylor, *How is Mechanism Conceivable?*, in *Philosophical Papers vol. I. Human Agency and Language*, Cambridge University Press, Cambridge 1985. I will refer to this edition in the continuation of this paper.

[2] *Ibidem*, pg. 164.

[3] The Third Antinomy of "rational cosmology" is perhaps one of the most famous formulations (at least in modern times) of the conflict between scientific determinism and the existence of a liberty and spontaneity of action. «Thesis: Causality in accordance with laws of nature is not the only causality from which all phenomena of the world can be derived. To explain these phenomena it is necessary to assume that there is also another causality through Freedom (*Freiheit*). Antithesis: There is no Freedom (*Freiheit*); everything in the world takes place solely in accordance with laws of nature» I. Kant, *Kritik der reinen Vernuft*, 1781, B 472-473. English version: I. Kant, *Critique of Pure Reason*, translated by N. Kemp Smith, Palgrave Macmillan, Londra 2007.

ISSN 0039-4130

centred around freedom and moral responsibility. How is possible to resolve this antinomy?

In Taylor's view the whole conundrum lies in the clash between our ordinary account of behaviour and the scientific explanation of it in neurophysiological and mechanistic terms. The reason of this clash is due to the fact that «our ordinary account characterizes our behaviour as action, while mechanistic account is interested in explaining it *qua* movement»[4]. But there is more in this clash. Two such accounts of the same event are potentially mutually exclusive, in the sense that they each trail a cluster of conditionals that may enter in a conceptual conflict. In other words, to class some behaviour, for instance, as reflex (e.g., as a mechanical response to a neurophysiological stimulus) is to eschew any possible purposive explanations. The problem of a possible solution can be stated as follows: should we defend our ordinary descriptive terms, our ordinary experience as it shows itself (a modern version of the claim 'save the phenomena'), or should we just reject this view in favour of the scientifically and "better-grounded" one? Are we forced to choose between one of the poles of the antinomy? Taylor's answer is multifaceted.

In the first place it would seem that Taylor opts for the first alternative (and in a sense he does). His stand is that the second option is «too preposterous to be believed»[5]. Too much of our pre-conscious and pre-reflexive existence, of our interaction with each other, of our emotions, of our legal institutions and peaceful coexistence, is built upon the sense that human beings are capable of something like performing an *action*, with all the teleological vocabulary that this implies. So, to save all this, we have to stick with the first of the two alternatives. But, and this is the core of Taylor's reasoning, this is not to say that we should refute neurophysiological and mechanistic explanation.

> «We cannot argue from the fact that mechanistic explanations are irreducibly different in logic to the (teleological and intentional) explanations of ordinary life to the conclusion that a mechanistic account is untenable. But there does emerge an important restriction on mechanistic explanations of behaviour: any acceptable such account must in fact be coordinated with our everyday account so as to "save the phenomena".»[6]

[4] C. Taylor, *How is Mechanism Conceivable? cit.*, p. 167.
[5] *Ibidem*, 169.
[6] *Ibidem*, pg. 174.

ISSN 0039-4130

We return to that initial "must". The problem lies in the assumption that all scientific explanations must be *completely* mechanistic. This is, historically speaking, derived from the immense success achieved by the scientific revolution of the XVII century. But, even if it is not possible to go into details here, it is important to underline that it is neither a natural step nor one without presuppositions. We do not need to pose this opposition in the form of an *aut aut*; Taylor, instead, aims to define a possible *convergence hypothesis*. One in which, while it is true that all behaviour can be coordinated with some neurophysiological processes, these cannot be considered the sole causal explanation of the behaviour. «On my convergence hypothesis», states Taylor, «the present principles of neurophysiology […] would be supplemented by concepts of quite a different kind, in which, for instance, relations of meaning might become relevant to neurophysiological process» [7].

What is neglected in a fully mechanistic explanation is that human beings are essentially cultural animals, that their behaviour is intrinsically constituted by self-interpretations that can vary very widely but that must be understood in the light of differences in human culture. And this could imply, even from the scientific standpoint, that human beings are to some extent able to choose freely, without the predetermination of physical processes in their brains [8]. In this sense, what is needed is a re-orientation of the framework in which neurophysiological explanations acquire their meaning. We must, according to Taylor, «be able to express in our theory the major distinctions by which men understand the differences in their behaviour from person to person and time to time»; and this implies that «a neurophysiological mechanist theory must have this property as well […], will have to be rich enough to mark the major distinctions of all the varied human cultures» [9]. This allows us to conceive a neurophysiological theory that is neither reductive, namely that it will not show teleological and intentional concepts to be eliminable; nor dualistic, that is to say that does not presuppose the introduction of a separate and different entity in order to explain free will and human actions. In this regard Taylor is very explicit: «if one wishes to avoid dualism and all its consequences, one will hold that in some sense we can give a neurophysiological account of all behav-

[7] *Ibidem*, pg. 185.

[8] This point is well argued by Filippo Tempia. See, F. Tempia, *Decisioni libere e giudizi morali: la mente conta*, in *Siamo davvero liberi? Le neuroscienze e il mistero del libero arbitrio* (edited by M. De Caro et al.), Codice edizioni, Torino 2019, pp. 87-108.

[9] C. Taylor, *How is Mechanism Conceivable? cit.*, p. 178.

ISSN 0039-4130

iour, for all behaviour has a neurophysiological embodiment» [10]. But, as we have already said, we need to add an important *caveat*. On the convergence hypothesis adumbrated by Taylor «it will be true that all behaviour will be accounted for in neurophysiological terms», but the «only neurophysiological theory adequate to this will be an enriched one» [11]. There are multiple factors that govern human behaviour, and some of these can be studied in the mechanistic terms of neural activity.

It must be said that Taylor's conclusion is more aporetic and exploratory than assertive. He concludes the paper by stating that his convergence hypothesis points towards the dissolution of the alternative between mechanism and dualism (and towards the dissolution of the antinomy), and that this fact invites us to examine a non-dualistic conception of man which is nevertheless not linked with a reductivist notion of the sciences of man. All of this involve taking seriously the possibility of «an *ontology* with more than one level» [12]. What is meant by that? How do we have to understand this multi-layered ontology? What light can this perspective give us to reinterpret the results of neurophysiological experiments, such as the one performed by Libet?

2. *A possible pattern of "convergence"*

Answering these questions is extremely difficult. Not only because Taylor does not give us a straightforward answer in the paper examined, but also because the problem of formulating an alternative ontology remains, throughout Taylor's whole career, as much a sticking point as an unsolved problem. I would now like to offer a sketch of a path that could work as a possible development and integration of Taylor's convergence hypothesis.

The perspective I would like to mention is the work of Iain McGilchrist, a British psychiatrist, neuroscientist and epistemologist. In his first masterpiece, published in 2009 [13], McGilchrist tries to reconstruct the development of Western culture with a parallel analysis of brain structure and functioning. He firstly states very clearly in the

[10] *Ibidem*, pg. 183.
[11] *Ibidem*, pg. 182-183.
[12] *Ibidem*, pg. 186, italics mine.
[13] I. McGilchrist, *The Master and His Emissary. The Divided Brain and the Making of the Western World*, Yale University Press, London/New Haven 2009. It has been republished in a new expanded version in 2019: I take in account this later version.

ISSN 0039-4130

Preface that he did «not mean to suggest that the brain *causes* human experience. Clearly there is a correlation between the brain and human experience. […] my position in brief is that the nature and the structure of the brain must be reciprocally related to the nature and structure of consciousness, but does not necessarily give rise to it» [14]. And moreover, he does not suggest in any way «that the causes of such cultural shifts can be reduced to neuroscience. There are many causative factors in play when cultures change, including sociological, psychological, environmental, epigenetic, technological, economic and political factors, all of which are interconnected» [15]. All these premises are very much consistent with Taylor's convergence hypothesis.

The focal point of his narrative is to propose an interpretation of the role played by the two hemispheres of the brain in the development of western civilization. I cannot possibly enter into the details and into the immense evidence gathered by the neuroscientists quoted by McGilchrist. I would only like to stress how this approach could be read as a concrete application of Taylor's proposal. The master narrative of McGilchrist's work can be summarized as follows. Our experience of the world results by the complex interaction of our brain-and-body with the world; but the brain has an intriguing and complicate structure. What immediately hits the eye is that it is divided into two different hemispheres (the right and the left), with a relatively small area that connects them (the corpus callosum). Now, there is plenty of neurophysiological evidence that suggests that the, nowadays popular, belief of separate tasks performed by the two hemispheres is a legend. Both hemispheres are involved in almost everything we do. What differs is their contribution to the same aspect of our experience, or to put it in McGilchrist's words, what differs is not the «whatness» of their competences but the «howness» [16]. What emerges from McGilchrist's narrative, is that the particular «howness» of the two hemispheres produces two relatively independent (although related) "views" or "ex-

[14] *Ibidem*, pg. XIV-XV. McGilchrist offers a beautiful image of this relationship: «A helpful analogy for the relationship I believe I see between mind and brain might be the relationship of a wave to water. The wave exists in the water: that's what we mean by a wave. Does the water cause the wave? No. Is it the movement of the water, then, that causes the wave? No, not that either: the movement of the water just *is* the wave. Similarly the relationship of mind and brain. Does the brain cause the mind? No. Is it the changing states of the brain that cause the mind? No: the changing brain states *are* the mind – *once the brain experiences them*». *Ibidem*, pg. 465 note n° 15.

[15] *Ibidem*, pg. XV.

[16] See Chapter 2 of McGilchrist's book.

ISSN 0039-4130

periences" of the world, that need to be constantly and harmoniously integrated.

The core of McGilchrist's thesis is that: while «the hemispheres have complementary but conflicting task to fulfil, and need to maintain a high degree of mutual ignorance» and «at the same time they need to co-operate» [17]; what can happen, and indeed in his view it has happened, is that the left hemisphere begins to function in an increasingly self-referential way, eventually ignoring or discarding the necessary integration of the right-partner. His narrative is therefore condensed in the sentence that «the story of the Western world is one of increasing left-hemisphere domination» [18].

But how can all of this help us in the matter of the possible alternative interpretation of results such as that of Libet? McGilchrist addresses the problem in an explicit way, commenting Libet's experiments. The main problem is that what Libet aims to investigate, his conception of the *will*, is a product of the analytic fragmentation of the phenomena typical of the left-hemisphere. Libet identifies, without questioning it, will, consciousness and the "I" (or the "we") that perceive these two elements of the personal identity. But what is completely neglected is that both our sense of the self and our will have deep roots also in the unconscious. Roots that, according to McGilchrist, are prior in every sense – temporally, logically and ontologically – to the superficial and isolated will considered by the left-hemisphere. This leads McGilchrist to the conclusion that «Libet's experiments does not tell us that we do not choose to initiate an action: it just tells us that we have to widen our concept of who 'we' are to include our unconscious selves» [19]. The problem lies then with the acknowledgement that our sense of the will is not a sense of some disembodied stance (a neutral taking a position), without history. Free will is something that we achieve through a gradual and complicate negotiation with ourselves (and in particular our bodies) and the world. We must avoid the temptation (the left-hemisphere temptation in McGilchrist's narrative) to wish to derive the properties of some level of reality mechanistically from the sum of its parts. As Marie Banich wrote:

> «The major finding to come out of our laboratory since the mid-1980s is that interhemispheric interaction is much more than just a *mechanism*

[17] *Ibidem*, pg. 210
[18] *Ibidem*, pg. 237.
[19] *Ibidem*, pg. 188.

ISSN 0039-4130

> by which one hemisphere "photocopies" experiences and feeling for its partner. Interhemispheric interaction has important emergent functions – functions that cannot be derived *from the simple sum of its parts*. […] The nature of processing when both hemispheres are involved cannot be predicted from the parts»[20].

It seems to me that these observations are very consistent with Taylor's proposal. They help us to shed a light on what could possibly mean, even "in dialogue with natural sciences", consider a multi-layered ontology that does not imply a form of dualism, nor adopt a reductionistic and deterministic take on reality. That is to say an ontology in which each level of complexity shows a set of emerging properties that cannot be reduced or derived from the simple sum of its parts. In this sense, it is very evocative the suggestion proposed by McGilchrist that «the brain is – and in fact it has to be – a metaphor of the world»[21]. It is important, therefore, to understand and connect the two of them without imposing or assuming a reductive view.

[20] See. M.T. Banich, *The Asymmetrical Brain*, in *Interaction between the hemispheres and its implications for the processing capacity of the brain* (edited by K. Hughdal and R.J. Davidson), Massachusetts Institute of Technology Press, Cambridge 2003, pp. 269-270, italics mine.

[21] Iain McGilchrist, *The Master and His Emissary*, cit., pg. 9.

ISSN 0039-4130

The Power of Algorithms to Redefine Human Autonomy

Alessia Cadelo

Introduction

In recent years, artificial intelligence has developed very rapidly. It is, indeed, a sector that promises great benefits for all of society. It has already been used in medicine, finance, domotics, entertainment, and many other areas. One of these applications is the recommender system, which selects the content being displayed to users in order to recommend the best options. On one hand, it might help users to navigate the online environment; indeed, if suggestions are sufficiently aligned with the users' interests, it could alleviate choice overload and improve self-expression. On the other hand, since recommender systems act like a filter, they shape our perception of available content, information, choice, and, in a way, of the world [1]. They may, therefore, undermine authenticity, defined as being in possession of values and desires considered one's own. Authenticity is a salient element of autonomy, so autonomy could also be influenced. The aim of the present paper is to contribute to the discussion about this topic. Firstly, an investigation into the philosophical concept of autonomy will be carried out, with special regard for two different positions: procedural and relational, respectively. Secondly, it will be shown how recommender systems affect our way of thinking and thus our authenticity and identity.

[1] S. Bonicalzi-M. De Caro-B. Giovanola, *Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems.*in Topoi XLII, 3, 2023, p. 825.

1. *Autonomy: between procedural and relational views*

In order to understand the relationship between recommender systems and autonomy, a brief philosophical investigation about the meaning of autonomy is necessary. As stated in the introduction, in the current debate there are two main positions, procedural and relational. According to the procedural perspective, autonomy is the ability of self-governance, which consists of determining how to live according to one's beliefs, values and goals [2]. This capacity requires the fulfilment of certain requirements [3]. The first is a minimum degree of rationality; desires or beliefs should not be manifestly inconsistent. In other words, there should be a certain degree of coherence in beliefs and desires [4]; they should respect logic laws. The final ends and purposes must also be harmonised with the rest of the values, preferences, and ideas to which an individual has committed himself. It doesn't count if the beliefs are false; individuals don't lack autonomy simply due to this [5]. It also occurs that a person identifies himself with his projects, values, aims, goals, desires, and so forth. This aspect, named authenticity, is shared among different representatives of this vision of autonomy, such as Joel Feinberg or Gerald Dworkin, but with little differences. For the first, authenticity consists of the capacity to alter his beliefs for reasons of his own. For Dworkin instead, a person has to assimilate the influences that motivate him to himself; they has to recognise them as their own. Otherwise, if a person is alienated from them, he couldn't be defined as autonomous [6]. Moreover, during the process of development of the desires, an autonomous individual should not be under the influence of manipulating factors that inhibit his or her capacity to critically reflect on those desires [7]. In other terms, individuals should have a certain degree of self-control; they should not be influenced by external forces. The procedural perspective owes its name precisely to the fact that it focuses on the process of critical reflection made by the indi-

[2] J. Christman, *Introduction* in The Inner Citadel: Essays on Individual Autonomy, Oxford University Press, New York 1989, pp. 5-6.

[3] Procedural accounts usually focus on factors internal to the psychology of the agent. Nevertheless, Diana Meyers' account is a noticeable exception; she indeed argues that autonomy requires also social and interpersonal skills.

[4] J. Christman, *Autonomy and Personal History*, in Canadian Journal of Philosophy, XXI, 1, 1991, pp. 16 ss.

[5] *Ibidem*.

[6] Id. *Introduction* cit. p. 12.

[7] Id. *Autonomy and Personal History*, cit. p. 19.

ISSN 0039-4130

viduals about their beliefs and values. What matters is the individual's active "participation" in this process, not the content of those beliefs, preferences, and desires. This vision is valuable since it underlines the psychological dimension of self-government. At the same time, it has been widely criticised for various reasons by the feminists; first of all, according to this view, individuals are entities separated from the rest of the world. Individuals are supposed to be atomistic and individualistic. However, human beings are social beings.

«The conviction that persons are socially embedded and that agents' identities are formed within the context of social relationships and shaped by a complex of intersecting social determinants» [8] is indeed the fundamental conviction of all perspectives that are labelled with the umbrella term "relational autonomy".

Thus, the effects of social background on individuals' sense of themselves must be acknowledged and consequently autonomy has to be reconceptualised [9]. Secondly, with this normative primacy on individual independence, authors who support a procedural account of autonomy fail to recognise the value of relations of dependency and interconnection. Since such relations of care have been historically associated with femininity, it is argued that this concept of autonomy is masculinist [10]. However, the main criticism of the procedural conception remains the fact that these theorists underestimated the socialisation, which aspects promote human autonomy and those that undermine it [11]. Autonomy should be indeed understood as a socio-relational phenomenon. Oshana particularly insists on the fact that autonomy is compatible only with social conditions that endorse dignity rather than deny it [12]. This doesn't imply that individuals should not reflect critically on their beliefs and desires and recognize them as their own, but is not sufficient; it is necessary to take into account also the social context in which individual ideas and preferences develop. As a consequence, to achieve autonomy, individuals within a community must establish relationships with others that facilitate the pursuit of their objectives in an environment of social and psychological security [13]. In addition, two

[8] C. Mackenzie-N. Stoljar, *Introduction* in Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self, Oxford University Press, 2000, p. 4.

[9] *Ibidem*.

[10] *Ibidem.*

[11] *Ibidem.*

[12] M. Oshana, *Personal Autonomy and Society*, in Journal of Social Philosophy XXIX, 1998, p.81.

[13] *Ibidem.*

ISSN 0039-4130

other conditions must be satisfied: access to a relevant set of options and procedural independence. The latter refers to the possibility for individuals to make their own decisions without being influenced or restricted by others in autonomy-constraining ways, in accordance with the procedural view [14]. In conclusion, to be autonomous, individuals must be capable of critically evaluating their preferences and ideas and acknowledging them as their own, but this ability, however, flourishes properly only when there is a wide and meaningful range of options.

2. *Recommender systems as digital nudging.*

As stated in the introduction, recommender systems are built to suggest the best options to the users. Most of them employ three different methods: content-based recommendations, collaborative recommendations, and knowledge recommendations. The first filters options by considering past user behaviour. The criterion of the second is to present the user items other users with similar user preferences have liked in the past, while the third bases his recommendation on the users' preferences and constraints. There are also hybrid systems, which combine these various techniques [15]. Whatever techniques are used, these systems modify the users' digital architecture; they select and order contents, personalise information, and recommend alternatives. In this sense, they could be considered as a form of digital nudging. This concept, as a matter of fact, refers to «the use of user-interface design elements to guide people's behaviour in digital choice environments» [16]. This is the first definition, but in a broader sense, digital nudging refers to the usage of digital technology (not only user interfaces) to predictably change people's choices and behaviour in both digital and physical choice environments [17]. As with the traditional nudging introduced by Thaler and Sunstein, the aim is to guide individuals' choices. Despite this, digital nudging has some distinctive features, which may have an important impact on human autonomy. First of all, nudging is originally conceived to promote the best interest of the individual according

[14] *Ibidem*

[15] S.Tiribelli - D. Calvaresi, *Rethinking Health Recommender Systems for Active Aging: An Autonomy-Based Ethical Analysis* in Science and Engineering Ethics, XXX, 22, 2024, p.22.

[16] M. Weinmann - C. Schneider - J. Brocke. 2016, *Digital Nudging*, in *Business & Information Systems Engineering*, LVIII, 6, 2016, p.433.

[17] M. Ienca - E. Vayena, *Digital Nudging: Exploring the Ethical Boundaries*, in *Oxford Handbook of Digital Ethics*, Oxford University Press, 2023, p. 361.

ISSN 0039-4130

to the choice architect (e.g health), which usually are public actors. The digital environment on the contrary is largely managed by private companies, who may be motivated by self-interest. Therefore, there is a misalignment between commercial and individual objectives. In other words, the users could be guided towards a certain option not because it is the best for them but rather because this better responds to the interests of those who offer this service. The purpose of these companies is actually to increase profit, which in turn is mostly derived from advertising. Because of this, they have a competitive interest in indefinitely increasing the amount of time that its users spend on the platform, even though it could be detrimental for them. As a consequence, they nudge users towards compulsive and persistent engagement on the platform through different strategies, such as infinite scrolling and autoplay [18]. However, this massive involvement is only one of the ways to increase profit; being able to grasp individual tastes and preferences is also crucial to offering the right content. To this end, recommender system algorithms collect large amounts of diverse data and build a profile of the individual, including preferences, ideas, personality traits, lifestyles and so on. As a result, they can then offer more precise, user-tailored nudges. These nudged 'choices', being the only ones available, will in turn become the basis for new hypotheses about the users. The principle behind this circle is therefore «what has been is what will be» [19]. As a consequence, everything that is not part of individual preferences is not shown. This mechanism is known as personalisation, and it is the second distinctive mark of digital nudging. Hence, these algorithms may affect the agent's ideas and preferences. Due to this influence, individual ideas, values and preferences could not be one's own anymore, so authenticity might be undermined. In other words, what could be damaged is the psychological dimension of autonomy. In line with this hypothesis, in 2012 the Adomavicious' research group investigated precisely the conditioning operated by recommender systems on individual tastes. The researchers analysed three different conditions, in each of which the participants were asked to view an item and give their assessment [20]. In the first and second case studies, the stimulus to be evaluated was an episode of a television

[18] *Ibidem.*

[19] S. Grafanaki, *Autonomy Challenges in the Age of Big Data*, in *Fordham Intellectual Property, Media and Entertainment Law Journal*, XXVII, 2014, p. 834.

[20] G. Adomavicious et al., *Do recommender systems manipulate consumer preferences? A study of anchoring effects*, in *Information Systems Research*, XXIV, 2013, pp. 962 e ss.

ISSN 0039-4130

show, with one substantial difference: in order to verify the possible ascendancy of algorithmic suggestions, in the first, the subjects were given artificial ratings, which had not been produced by any recommender system. In the second, individual preferences were also taken into account, which were then used by a well-known algorithm to give personalised recommendations. In the third, the same method as the second was used, but this time jokes were evaluated. In all conditions, it was observed that individual preferences were significantly influenced by the recommendations received [21]. Because recommender systems might have an effect on ideas and preferences, they also indirectly act on the global identity of the individual. The mechanism of personalisation, as we just said, could strengthen our values and consequently reinforce personal identity. Nevertheless, it may restrict access to different, relevant options, which, as mentioned earlier, is one of the requirements of autonomy. In regard to this closure, Eli Pariser speaks significantly of the filter bubble, describing it as a lens that controls what we see and what we do not see [22]. Secondly, the activity of profilation may poorly reflect categories that are perceived as central by the agent. As a result, the corresponding suggestions could become irrelevant [23]. At the same time, these algorithms indeed tell people who they are, but the representation of identity provided by the algorithm is crystallised at the present moment, i.e., the moment in which we use the platforms or search the web. The reason lies in the fact that, when we click on a given piece of content, it is always our present self that does so [24], so recommender systems are unlikely to reflect our future aspirations. However, the purpose of the algorithm is not only to describe the individuals in the present but also to provide an estimate of future identities and behaviours. This estimate is consequently partial and probably inaccurate. This representation will be used in turn to nudge the users in the future, so the recommender systems might reshape the subjective experience of one's own identity. Moreover, in this context, Smith suggests that individual values could be replaced by those that can be economically exploited. In this regard, he recalls that Google/Alphabet paid for the well-known game Pokèmon Gò through

[21] *Ibidem.*

[22] E. Pariser, *The filter bubble: What The Internet Is Hiding From You*, Penguin, Londra, 2011, pp.48-49.

[23] S. Bonicalzi - M. De Caro - B. Giovanola, *Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems*, cit. p. 827.

[24] E. Pariser, *The filter bubble: What The Internet Is Hiding From You,* cit. p.66.

ISSN 0039-4130

the sale of virtual lands in real locations [25]. Thus, Starbucks paid for the game's monsters to reside near their cafés so as to gather many people and increase sales. Of course, no one had any idea of the motive behind the distribution of monsters. In this case, the desire to play video games was channelled in a distorted way towards the sale of cafés and similar [26]. In sum, recommender systems, by profiling individuals, could help users to save time and attention in the research but they could also be detrimental both for identity and autonomy.

Conclusion

In conclusion, although recommender systems facilitate users navigate the online environment, at the same time, they may restrict their freedom of choice and heavily influence their opinions and ideas. In this way, they could undermine autonomy, both in a procedural and relational sense. These algorithms determine the content to be shown to users based on their profile. On the one hand, they can reinforce individual tastes and preferences and thus their identity. On the other hand, the risk is that the representation is static, i.e. it does not adequately take future aspirations into account. The suggestions made by the algorithm may therefore not be relevant or inaccurate. Nevertheless, they might still influence users and consequently they could reshape their identity. Thus, in order to protect autonomy and freedom to form our identity, it would be necessary to reconsider the humans' relational dimension in AI ethics and in the algorithms' design.

[25] C. H. Smith, *Corporatised Identities ≠ Digital Identities: Algorithmic Filtering on Social Media and the Commercialisation of Presentations of Self*, in Ethics of Digital Well-Being, SpringerLink, 2020, pp. 58-59.

[26] *Ibidem*.

ISSN 0039-4130

Posso fidarmi? La fiducia nelle relazioni del "Dopo di noi"

Folco Cimagalli[1], *Giuseppina Signorello*[2]

1. *Fiducia e incertezza*

Come ha notato Simmel «chi sa completamente non ha bisogno di fidarsi, chi non sa affatto non può ragionevolmente fidarsi»[3]: la fiducia, formulando un'aspettativa sul comportamento dell'altro, consente di stabilire legami e di garantire le condizioni minime di prevedibilità e stabilità all'azione sociale. Pertanto, la fiducia – che è la condizione indispensabile anche solo «per alzarsi dal letto la mattina»[4] – rappresenta l'architrave di ogni sistema di relazioni sociali.

La fiducia – ricorda Simmel – interviene in situazioni di incertezza, facilitando e orientando l'azione sociale; essa opera in uno spazio intermedio tra conoscenza ed ignoranza: non vi è bisogno di fiducia quando la conoscenza dell'altro è piena e incondizionata; contemporaneamente, sarebbe del tutto priva di ogni fondamento una fiducia in situazioni di totale assenza di conoscenza.

Lungo tale filone, la fiducia, secondo Luhmann[5], rappresenta un meccanismo di riduzione della complessità sociale, che permette di gestire l'incertezza del futuro e rende possibile l'interazione sociale, fungendo da «equivalente funzionale» della certezza. Al tempo stesso, nota il sociologo tedesco, essa implica un rischio ed espone colui che

[1] Professore di sociologia generale presso il Dipartimento GEPLI, LUMSA Università.

[2] Assistente sociale specialista, PhD, Docente di Metodi e tecniche del servizio sociale presso il Dipartimento GEPLI, LUMSA Università e ricercatrice sociale presso Fondazione Roma Solidale onlus. Il presente contributo è frutto della comune riflessione dei due autori. Per una corretta attribuzione dei contenuti, Folco Cimagalli è autore del par. 1, mentre Giuseppina Signorello dei parr. 2 e 3.

[3] G. Simmel, *Sociologia*, Edizioni di Comunità, Milano 1998, p. 299.

[4] N. Luhmann, *La fiducia*, il Mulino, Bologna 2002, p. 5.

[5] *Ibidem.*

la sperimenta, e che si apre alla relazione, in una condizione di vulnerabilità.

Secondo Gambetta, la fiducia rappresenta «un particolare livello di probabilità soggettiva con cui un agente valuta che un altro agente o gruppo di agenti compia una particolare azione»[6] e poi, ancora, è «l'atteggiamento verso un'altra persona basato sulla convinzione che questa non farebbe nulla contro di noi anche se ne avesse la possibilità e ne potesse trarre un vantaggio personale»[7]; essa dunque è una scelta probabilistica del comportamento di *alter*.

Mutti – uno degli studiosi italiani più prolifici in tale ambito – definisce la fiducia come un'«aspettativa con valenza positiva per l'attore, maturata sotto condizioni di incertezza, ma in presenza di un carico emotivo e/o cognitivo tale da permettere di superare la soglia della mera speranza»[8].

Emerge in tale definizione la portata positiva di tale aspettativa e si intravedono i possibili fondamenti che la ispirano: sia di tipo emotivo (la fiducia "istintiva", di tipo empatico, fondata su basi emozionali e non necessariamente ancorata a un sostrato razionale), sia di tipo cognitivo, ossia motivata alla luce di una considerazione razionale, facendo tesoro di esperienze pregresse o considerando la reputazione del destinatario.

Proseguendo il ragionamento in termini più analitici, possiamo considerare tre tipi di fiducia: particolaristica, generalistica e politico-istituzionale. La prima si riferisce a uno specifico destinatario attraverso una relazione diretta: come propone Uslaner[9], in tale schema abbiamo la situazione in cui "A si fida di B". Nel secondo tipo, l'oggetto della fiducia è un "altro generalizzato"; si tratta, in altri termini, di un orientamento benevolo e confidente nei confronti del prossimo che non assume un riferimento specifico: secondo la formulazione di Uslaner abbiamo, semplicemente, la situazione in cui "A si fida". Nel terzo tipo, la fiducia è rivolta nei confronti di attori collettivi: in questo caso, sono oggetto dell'affidamento alcune entità impersonali che incidono nell'esperienza dell'individuo.

A questo riguardo, diverse *survey* – tanto al livello nazionale quanto a quello internazionale – rilevano che le tre forme di fiducia non

[6] D. Gambetta, *Fiducia. Un meccanismo per comprendere la complessità della società*, Einaudi, Torino 2009.

[7] *Ibidem.*

[8] A. Mutti, *Capitale sociale e sviluppo,* Il Mulino, Bologna 1998.

[9] E. M. Uslaner, *The Moral Foundations of Trust,* Cambridge University Press, Cambridge 2002.

ISSN 0039-4130

sono tra loro strettamente correlate [10] e, soprattutto, osservano come, in generale, si osservi da ormai diversi anni una crescente crisi delle relazioni fiduciarie di tipo impersonale e, parimenti, un declino della fiducia nei confronti delle istituzioni. La "società dell'incertezza", in altri termini, non sembra trovare adeguati antidoti di tipo fiduciario se non quelli, non sempre accessibili, dell'*ingroup* identitario.

2. *La fiducia nelle relazioni di aiuto*

Così definita, è evidente che la fiducia costituisce un elemento imprescindibile delle relazioni di aiuto. Non è infatti un caso che già nel preambolo del Codice deontologico degli assistenti sociali si legge come «la relazione con la persona, anche in presenza di asimmetria informativa, si fonda sulla fiducia e si esprime attraverso un comportamento professionale trasparente e cooperativo, teso a valorizzare tutte le risorse presenti e la capacità di autodeterminazione degli individui» [11].

Nella letteratura del *social work*, la centralità della fiducia nella relazione di aiuto è dunque evidente. Essa si associa, per l'appunto, alla relazione tra operatore e persona-utente e considera il legame fiduciario come un fondamento della relazione stessa. Non si tratta, è bene ricordarlo, di un pre-requisito, vale a dire di una condizione che precede l'incontro, ma di un'apertura progressiva, che richiede cura e manutenzione. La fiducia, in questo senso, non rappresenta un elemento presente/assente, che viene "acceso" da uno dei due protagonisti, ma un processo dinamico che si co-costruisce. Lievito della relazione di aiuto, essa rende possibile il raggiungimento dell'obiettivo dell'interazione, che è il cambiamento, e al tempo stesso apre spazi a un "rischio" su entrambi i fronti: la delusione delle aspettative da parte della persona che si rivolge al servizio, la constatazione di un agire esclusivamente strumentale da parte dell'operatore.

Tornando alla tipologia definitoria sopra presentata, si tratta, nella nostra ipotesi, di una fisionomia di fiducia *sui generis* che, nel suo declinarsi, è il prodotto congiunto dei tre tipi enunciati. In altri termini, la fiducia nelle relazioni di aiuto non è sovrapponibile a *una* delle tre forme sopra descritte, ma scaturisce, tanto nella situazione *micro* quan-

[10] D. Johnson - K. Grayson, *Cognitive and affective trust in service relationships*, in *Journal of Business Research*, LVIII, 4, 2005, p. 500-507.

[11] Consiglio Nazionale Ordine Assistenti Sociali (2020), *Codice deontologico dell'assistente sociale*, p. 7, in http://www.cnoas.org.

ISSN 0039-4130

to nel quadro *macro*, dalla composizione delle tre conformazioni. Essa infatti è fiducia interpersonale perché la relazione di aiuto ha una natura empatica tra due soggetti che interagiscono nella loro unicità; è, o diviene, relazione "calda" che travalica la mera dimensione istituzionale o burocratica. Al tempo stesso, tale rapporto si connota per collegare due soggetti tra loro sconosciuti; il legame, come ben ricorda l'Ordine, viene *costruito* mentre si agisce e dunque, nel suo farsi, risente non solo degli atteggiamenti del professionista e della sua capacità di creare un collegamento, ma anche della predisposizione della persona-utente a relazionarsi, a fidarsi dell'"altro generalizzato" impersonato dall'operatore. Soggetti diffidenti e con un basso livello di fiducia generalizzata esiteranno dunque a entrare in relazione con l'operatore e, di qui, tutto il processo d'aiuto potrà risultare lento o faticoso.

In terzo luogo, non da ultimo, questa forma specifica di relazione si sviluppa all'interno di una cornice istituzionale e normativa. La persona si rivolge a un servizio o è da questo interpellata. L'esperienza e la letteratura mostrano chiaramente quanto conti, nel declinarsi della relazione, la fiducia istituzionale maturata sia nei confronti dell'Ente chiamato in causa, sia nei confronti della categoria professionale in generale [12].

In tale quadro, un ambito di riflessione particolarmente significativo nel quale le tematiche fiduciarie appaiono cruciali è rappresentato dal cosiddetto "Dopo di noi". Si tratta di un'espressione utilizzata per descrivere il periodo in cui i familiari delle persone con disabilità non possono più aiutarle o prendersi cura di loro [13]. "Chi si prenderà cura dei nostri figli quando noi non saremo più in grado di occuparcene o non ci saremo più?". È per rispondere a questa domanda che nel 1984 alcune famiglie di un'associazione italiana promuovono la nascita di una Fondazione, coniando per prime il termine "Dopo di noi" (in ambito internazionale, il "Dopo di noi" è un'etichetta che sottende altri temi, quali l'autonomia, l'autodeterminazione e la vita indipendente delle persone con disabilità, temi già riconosciuti dalla convenzione ONU del 2006).

Nel nostro Paese, la Legge 112 ha cercato di rispondere a una sfida importante: «favorire il benessere, la piena inclusione sociale e l'auto-

[12] A questo riguardo, sono purtroppo ben note le complessità e le contraddizioni ben radicate nelle rappresentazioni sociali e mediatiche rispetto alla categoria degli assistenti sociali.

[13] E. Zanfroni - S. Maggiolini - M. C. Carruba - L. D'Alonzo, *Re-thinking inclusion for adult people with disability: Residential centers from makeshift solution to educational resource for the community,* in *Education Sciences & Society,* XIII, 1, 2022, pp. 244-258.

ISSN 0039-4130

nomia della persona con disabilità» (art. 1, c. 1), anche quando manca il supporto familiare o parentale. Questa legge ha segnato un cambiamento di prospettiva, introducendo nuovi attori sociali e un mix tra interventi di natura pubblica, misure di natura privata e strumenti di gestione economico-finanziaria, in un'ottica di coprogettazione e compartecipazione tra pubbliche amministrazioni, organismi del terzo settore e famiglie [14]. La finalità è aprire la sfera familiare ad una dimensione comunitaria più inclusiva.

La legge rappresenta dunque un tentativo concreto di rispondere a questa sfida importante: da un lato cerca di trovare soluzioni innovative ai bisogni delle persone con disabilità grave, dall'altro risponde formalmente all'esigenza di affrontare la questione del "Dopo di noi" quando le famiglie sono ancora in grado di prendersi cura dei propri familiari con disabilità e, dunque, quando le famiglie sono ancora in grado di progettare e condividere con i loro familiari le scelte più adatte e rispettose della loro dignità, delle loro aspirazioni e dei loro bisogni [15]. Negli ultimi anni, infatti, all'espressione "Dopo di noi" si è aggiunta quella del "Durante noi". Un cambiamento linguistico e concettuale che sostiene l'importanza di promuovere, in tempi opportuni, modelli di progettazione della qualità della vita che coinvolgano il sistema familiare, superando così approcci assistenzialistici ed emergenziali.

In tale ambito, appare interessante esplorare come la fiducia possa influire sul processo decisionale delle famiglie che si avviano verso il percorso del "Dopo di noi", per garantire che le persone con disabilità e i loro familiari si sentano sostenuti e possano scegliere se partecipare o meno a tali progetti di vita.

L'ipotesi centrale della nostra proposta è che le famiglie, nel processo del "Dopo di noi", vivano una condizione di incertezza e ansia, e che la fiducia possa essere una risorsa fondamentale per affrontare questo stato. In questo senso, la fiducia non rappresenta soltanto un meccanismo di sostegno emotivo, ma anche un elemento cruciale nel processo decisionale per le famiglie che stanno pianificando il futuro del proprio familiare con disabilità, sia in termini di riuscita del processo che di benessere delle persone coinvolte.

[14] C. Giaconi et al., *Il Dopo di Noi: nuove alleanze tra pedagogia speciale ed economia per nuovi spazi di Qualità di Vita*, in *MeTis-Mondi educativi. Temi indagini suggestioni*, X, 2, 2020, pp. 274-291.

[15] M. Verga, *Il Dopo di noi e il durante noi: brevi riflessioni a cinque anni dall'approvazione della Legge 112/2016*, in *Sociologia del diritto*, 2, 2021, pp. 149-173.

ISSN 0039-4130

3. *Conclusioni*

In un contesto così delicato per le politiche sociali, emerge dunque l'urgenza di una ricerca che approfondisca il ruolo della fiducia nella relazione tra persone con disabilità, famiglie, operatori, servizi e territorio. Senza una solida base di fiducia, il rischio è che le famiglie si trovino sole ad affrontare scelte complesse, ricorrendo a soluzioni emergenziali poco sostenibili nel lungo periodo. Comprendere i meccanismi di costruzione e "manutenzione" della fiducia è essenziale per migliorare la qualità degli interventi nel "Durante e Dopo di noi". Approfondire questi aspetti consente di orientare le politiche sociali verso modelli di coprogettazione più efficaci, che valorizzino il ruolo attivo delle famiglie e garantiscano percorsi personalizzati in grado di migliorare la qualità di vita.

Investire nella fiducia significa investire in politiche sociali più sostenibili, fondate sulla collaborazione tra famiglie, servizi e territorio. Solo così il "Dopo di noi" potrà diventare un percorso che rispetta realmente le aspirazioni e i diritti delle persone con disabilità e delle loro famiglie. La fiducia rappresenta l'elemento generativo di una rete solida e coesa, capace di rafforzare i legami sociali e favorire relazioni aperte e resilienti. Sarà proprio questa rete, fondata sulla responsabilità condivisa, a trasformare il "Dopo di noi" da una risposta tardiva a un processo partecipato e sostenibile nel tempo del "Durante e Dopo di noi".

Riferimenti bibliografici

Consiglio Nazionale Ordine Assistenti Sociali, *Codice deontologico dell'assistente sociale*, p. 7, 2020, in http://www.cnoas.org.

D. Gambetta, *Fiducia. Un meccanismo per comprendere la complessità della società*, Einaudi, Torino 2009.

C. Giaconi et al., *Il Dopo di Noi: nuove alleanze tra pedagogia speciale ed economia per nuovi spazi di Qualità di Vita*, in *MeTis-Mondi educativi. Temi indagini suggestioni*, X, 2, 2020, pp. 274-291.

D. Johnson - K. Grayson, *Cognitive and affective trust in service relationships*, in *Journal of Business Research*, 2005, LVIII, 4.

N. Luhmann, *La fiducia*, il Mulino, Bologna 2002.

A. Mutti, *Capitale sociale e sviluppo,* Il Mulino, Bologna 1998.

G. Simmel, *Sociologia*, Edizioni di Comunità, Milano 1998.

E. M. Uslaner, *The Moral Foundations of Trust,* Cambridge University Press, Cambridge 2002.

ISSN 0039-4130

M. Verga, *Il Dopo di noi e il durante noi: brevi riflessioni a cinque anni dall'approvazione della Legge 112/2016*, in *Sociologia del diritto*, 2, 2021, pp. 149-173.

E. Zanfroni - S. Maggiolini - M. C. Carruba - L. D'Alonzo, *Re-thinking inclusion for adult people with disability: Residential centers from makeshift solution to educational resource for the community,* in *Education Sciences & Society,* XIII, 1, 2022, pp. 244-258.

ISSN 0039-4130

Mi fido, quindi fai tu

La fiducia come chiave di lettura nella comunicazione dagli anni ’50 a oggi

Simone Mulargia

Nel presente intervento analizzerò la dimensione della fiducia all’interno di tre figure significative nelle scienze della comunicazione: l’opinion leader, la celebrità, l’influencer.

Opinion leader

Durante la campagna elettorale USA del 1940, Lazarsfeld e colleghi studiarono l’influenza dei media sulle scelte di voto, scoprendo che il ruolo dei contatti personali era determinante per gli elettori indecisi. Ne derivò il modello del “flusso a due fasi della comunicazione”, dove alcuni individui – gli opinion leader – filtrano e rilanciano i messaggi dei media [1].

Gli anni immediatamente precedenti al secondo conflitto mondiale erano stati caratterizzati dall’ascesa dei totalitarismi e dal dispiegamento sistematico delle potenzialità dei mezzi di comunicazione di massa. Uno sviluppo tumultuoso delle tecnologie comunicative, perfettamente in grado di intercettare il nuovo assetto della società. Mezzi di comunicazione di massa capaci di plasmare il senso sociale della società di massa.

Le testimonianze raccolte e analizzate da Lazarsfeld e colleghi descrivevano di fatto un flusso della comunicazione differente da quello appena ricordato. Gli opinion leader si ponevano come fattore di me-

[1] F. P Lazarsfeld - B. Berelson - H. Gaudet, *The people’s choice. How the voter makes up his mind in a presidential campaign*, Columbia University Press, New York Chichester, West Sussex 1948.

diazione tra i media e le persone. Figure che costruiscono un ponte comunicativo basato sulla fiducia e sul prestigio, emanati dalla conoscenza diretta e da un'interazione calata nella vita quotidiana dei soggetti.

In alcuni resoconti dei partecipanti alla ricerca, l'influenza fiduciaria non aveva neanche bisogno di lavorare sulla dimensione cognitiva o emotiva, come nel caso di votanti che furono semplicemente accompagnati a votare.

Il quadro appena descritto è destinato a rappresentare un passaggio fondamentale nella storia della riflessione sui mezzi di comunicazione. Il rapporto interpersonale si pone come fattore di mediazione rispetto al potere dei media. Una vera e propria rivincita dei contatti personali rispetto all'onnipotenza dei media. Lazarsfeld replicò il modello su un campione di 800 donne di Decatur, dimostrando che gli opinion leader influenzano non solo la politica, ma anche le scelte di consumo e culturali [2].

Malgrado alcuni spunti critici, che non sono oggetto della presente trattazione [3], la figura degli opinion leader rappresenta un punto di vista privilegiato per cogliere il rapporto problematico tra relazione, fiducia e incoraggiamento all'azione, nell'ambito di uno scenario in cui i mezzi di comunicazione (e la natura mediata della relazione fiduciaria) sembrano per un momento fare un passo indietro rispetto alla concreta materialità del legame sociale.

Un assetto destinato a incrinarsi nella figura della celebrità.

Celebrità

La nascita del divismo è tradizionalmente legata alle vicende di Florence Lawrence, attrice del cinema muto. Come per la maggior parte degli attori del periodo, il pubblico non conosce il suo nome. Nel 1910, Carl Laemmle, (Universal Pictures), ideò un'audace campagna pubbli-

[2] E. Katz-P. F. Lazarsfeld, *Personal Influence. The Part Played by People in the Flow of Mass Communications*, London and New York, Routledge 2006. (ed. or. 1955).

[3] Sintetizzati da Eliuh Katz nell'introduzione a una recente edizione di *Personal Influence*. Da un lato, autori come Gitlin hanno sostenuto che il concetto di opinion leader mascheri in realtà gli effetti diretti e potenti dei media, i quali influenzerebbero il pubblico in modo più immediato e incisivo di quanto suggerito dal modello "Two-Step Flow of Communication". Dall'altro lato, studiosi come Lang, Adorno e McLuhan hanno criticato l'enfasi posta sugli effetti a breve termine delle campagne mediatiche, come quelle politiche o pubblicitarie, sottolineando invece che il vero potere dei media risiede nei loro effetti a lungo termine, come il mantenimento dello status quo o il rallentamento del cambiamento sociale.

ISSN 0039-4130

citaria fingendo la morte dell'attrice. Lo stratagemma generò enorme attenzione mediatica e del pubblico, che per la prima volta memorizzò l'identità dell'attrice. *Era dunque nata una stella*.

La letteratura sul tema è davvero corposa e abbraccia ambiti di riflessione diversissimi tra loro. Pioneristico, Alberoni [4] analizza la condizione contraddittoria del divismo in quanto élite priva di potere istituzionale, ma dotata di un notevole carisma sociale. Appena l'anno prima, Daniel J. Boorstin, aveva tematizzato il concetto di pseudo-evento [5], espressione utilizzata per indicare eventi creati per essere riportati dai media. Boorstin offre una straordinaria definizione di celebrità come *una persona famosa per la sua notorietà*, sottolineando come la fama sia spesso scollegata da meriti o risultati concreti. Questa intuizione apre la strada a una comprensione della celebrità come costruzione artificiale, fenomeno intrinsecamente mediato che riflette i processi simbolici della modernità.

Sulla scia di quanto appena richiamato, Richard Dyer analizza le celebrità come veri e propri *testi culturali* che incarnano valori e ideologie sociali. Dyer considera le star non solo come prodotti mediatici, ma come luoghi di produzione di significati, in cui si intrecciano questioni di classe, genere, razza e sessualità [6].

Le celebrità incarnano le contraddizioni della società dello spettacolo [7]: persone eccezionali, che però non hanno nulla di speciale. Testi prodotti dal sistema dominante, eppure anche figure della contestazione o anticipatorie di conflitti sociali futuri. Estremamente personali nel loro richiamare direttamente lo spettatore e allo stesso tempo inarrivabili.

Non sfugge il fascino che tali figure esercitano sul pubblico. La gamma delle sensazioni è vasta e ci costringe a reinterpretare in chiave mediata e simbolica quella radice fiduciaria che avevamo osservato a livello personale nello scenario degli opinion leader. Andrew Tudor distingue quattro livelli di relazione tra star e audience: affinità emotiva, auto-identificazione, imitazione e proiezione [8]. Questo schema offre una visione sfaccettata del modo in cui le audience interagiscono con le celebrità, passando da forme di coinvolgimento superficiali a processi più profondi di identificazione e imitazione.

[4] F. Alberoni, *L'élite senza potere*, Vita e Pensiero, Milano 1963.

[5] D. J. Boorstin, *The image: a guide to pseudo-events in America*, Vintage Books, New York 1992 (ed. or. 1961).

[6] R. Dyer, *Stars*, British Film Institute, London 1998 (ed. or. 1979).

[7] G. E. Debord, *La Société du Spectacle*, Buchet-Castel, Paris 1967.

[8] A. Tudor, *Image and influence. Studies in the sociology of film*, Allen & Unwin, London 1974.

ISSN 0039-4130

Non stupisce, dunque, che il mondo della pubblicità riesca a tradurre questi sentimenti in un processo di ingegnerizzazione del legame fiduciario tra pubblico e celebrità, attraverso l'utilizzo massiccio delle *celebrities* come testimonial pubblicitari. È un tornare alla radice dei processi osservati nell'ambito degli opinion leader (che pure suggerivano prodotti e servizi), se siamo disposti ad accettare che il vincolo fiduciario si esprima tutto sul versante della mediazione e della rappresentazione.

Il collegamento tra celebrità e messaggio pubblicitario potrebbe richiamare uno schema manipolatorio in cui le persone sarebbero vittime di una fiducia mal riposta. Seguendo Joshua Gamson, però, la consapevolezza pubblica della costruzione mediatica della fama ci obbliga a riconsiderare criticamente l'autenticità, in un contesto in cui la fiducia nel divo è mediata dalla consapevolezza della sua costruzione artificiale [9]. Per Chris Rojek le celebrities dissolvono la distinzione tra pubblico e privato, in una performance continua che rafforza il loro potere simbolico [10].

Sul versante delle conseguenze politiche, David Marshall esplora la celebrità come una forma di soggettività pubblica che riflette le tensioni tra inclusione democratica ed eccesso consumistico [11]. Le star incarnano modelli di auto-differenziazione che rispecchiano i valori del capitalismo, ma operano anche come figure centrali nella costruzione della sfera pubblica contemporanea. In termini riassuntivi – ed evidentemente critici - Rojek evidenzia il ruolo delle celebrità nella politica e nella filantropia, usando l'espressione "celanthropy" (unione di celebrità e filantropia). In questi termini, il processo di celebrificazione caratteristico della cultura contemporanea distrae dalle mobilitazioni collettive a favore di un'attenzione individualistica [12].

Eccoci di fronte all'ennesima contraddizione. Da un lato le celebrità riescono effettivamente a catalizzare l'attenzione pubblica attraverso i media che seguono le loro attività, suggerendo il delinearsi di una nuova forma di *agency* [13]. Dall'altro le star possono di-

[9] J. Gamson, *Claims to Fame: Celebrity in Contemporary America*, University of California Press, Oakland, California 1994

[10] C. Rojek, *Celebrity*, Reaktion Books, London 2001.

[11] P. D. Marshall, *Celebrity and power. Fame in contemporary culture*. University of Minnesota Press, Minneapolis, Minnesota 1997.

[12] G. Turner, *Understanding celebrity*, Sage, London 2014 (ed. or. 2004).

[13] M. K. Goodman - J. Littler, *Celebrity ecologies: introduction*, in *Celebrity Studies*, IV, 3, 2013.

ISSN 0039-4130

storcere la distribuzione delle risorse in modi imprevedibili e poco razionali [14].

Così come i primi studi sugli opinion leader avevano delineato due sfere di influenza (le scelte politiche e quelle di consumo) le celebrità svolgono un ruolo significativo sia nell'ambito della promozione dei prodotti e dei servizi, sia in quello politico e delle cause sociali. Si tratta, quindi, di un legame che appare non episodico tra fiducia (immediata e mediata) e forme di potenziale attivismo partecipativo delle persone.

All'intersezione tra la natura mediata e immediata del rapporto fiduciario, per come abbiamo cercato di tematizzarlo in questo intervento, emerge la terza figura significativa: l'influencer.

Influencer

I primi segnali di quella che diventerà l'industria degli influencer possono essere osservati sul finire del 2008, nei tentativi di alcuni creativi di rispondere alla crisi economica cercando di riposizionarsi sul mercato del lavoro [15].

Una delle dimensioni chiave per comprendere gli influencer sembra chiamare quasi direttamente in ballo la questione della fiducia, pur all'interno delle ambiguità che abbiamo già in parte richiamato. Si tratta del concetto di autenticità.

Un influencer è credibile, e quindi meritevole di fiducia, nella misura in cui è autentico. Più nel dettaglio, ancora con Hund, si tratta di considerare la rappresentazione dell'autenticità come risorsa strategica e merce scambiabile nel mondo dei social media. Gli influencer, dunque, rappresentano plasticamente la contraddizione tra spontaneismo e strategia comunicativa: essere se stessi per trarne un vantaggio competitivo [16]. L'autenticità diventa quindi un linguaggio e un'estetica mutevole, che chi riesce a padroneggiare può trasformare in potere e influenza. È questo il meccanismo che consentì ad alcuni abili dilettanti (spesso inizialmente blogger) di inaugurare un nuovo stile comunicativo informale e spesso non accorto nella distinzione tra contenuti editoriali e sponsorizzati.

[14] C. Rojek, *Celebrity*, Reaktion Books, London 2001 e G. Fridell - M. Konings, (a cura di), *Age of icons. Exploring philanthrocapitalism in the contemporary world*, University of Toronto Press, Toronto 2013.

[15] E. Hund, *The influencer industry. The quest for authenticity on social media*, Princeton University Press, Princeton and Oxford 2023.

[16] *Ibidem*.

ISSN 0039-4130

Come figure emblematiche del panorama comunicativo contemporaneo non stupisce che gli influencer lavorino trasversalmente rispetto alla distinzione tra pubblico e privato, tra autorevolezza e inaffidabilità, tra competenza e pressappochismo.

L'autenticità, e con essa la fiducia che riponiamo nei confronti degli influencer, assume una natura precaria, oggetto di continue oscillazioni che ricordano appunto l'andamento del valore delle merci all'interno dei mercati. Gli influencer sono criticati per la loro apparente superficialità, ma influenzano profondamente la vita quotidiana delle persone, contribuendo a un significativo processo di estetizzazione della quotidianità. L'influenza viene raccolta, elaborata e commercializzata come un bene.

Il legame tra influencer e celebrità è stato ampiamente discusso nel dibattito teorico. Il concetto di *microcelebrità*, descrive la pratica attraverso cui individui comuni costruiscono una notorietà online attraverso strategie di autopromozione e gestione della propria immagine pubblica [17]. A differenza delle celebrità tradizionali, che ottengono fama attraverso media consolidati come cinema e televisione, le microcelebrità si affermano grazie alla loro capacità di attrarre e mantenere un pubblico sui social media, *coltivando* la propria audience tramite la condivisione costante della propria vita quotidiana e l'autopromozione strategica. Da questo punto di vista, si tratterebbe di ibridi tra consumatori e brand [18], e per questo percepiti come consumatori esperti, e quindi degni di maggiore fiducia rispetto ai volti pubblicitari prestabiliti [19].

La fiducia tra delega e partecipazione

In questa analisi ho cercato di problematizzare il rapporto tra fiducia, tenuta del legame sociale e partecipazione. Nel titolo dell'intervento ho provocatoriamente optato per lo scenario della delega, in cui la fidu-

[17] T. M. Senft, *Camgirls: celebrity and community in the age of social networks*, Peter Lang, Berlin 2008 e A. E. Marwick, *Status Update: Celebrity, Publicity, and Branding in the Social Media Age*, Yale University Press, New Haven, Connecticut 2013.

[18] R. K. Britt - J. L. Hayes - B. C. Britt - H. Park, *Too big to sell? A computational analysis of network and content characteristics among mega and micro beauty and fashion social media influencers*, in *Journal of Interactive Advertising*, XX, 2, 2020, p. 111.

[19] K. Freberg - K. Graham - K. McGaughey - L. A. Freberg, *Who are the social media influencers? A study of public perceptions of personality*, in *Public relations review*, XXXVII, 1, p. 90.

ISSN 0039-4130

cia è la radice per la deresponsabilizzazione dei soggetti. Nei contesti altamente mediati in cui viviamo, questo fidarsi che diventa affidarsi può contribuire a spiegare la crisi della partecipazione contemporanea. Eppure, esiste un diverso angolo di lettura di questi fenomeni che apre alla possibilità di un rinnovato afflato di partecipazione. Uno scenario in cui la provocazione del titolo "Mi fido, quindi fai tu" può tramutarsi in "mi fido, quindi facciamo insieme".

ISSN 0039-4130

La fiducia nell'esperienza giuridica contemporanea

Brevi note introduttive

Michele Ciancimino

1. *Introduzione*

Il presente contributo intende anticipare e presentare gli studi, pubblicati in questa medesima Sezione, svolti da parte di alcuni dottorandi del Dottorato in *Mediterranean Studies* dell'Università LUMSA, Dip. G.E.C. di Palermo, sul tema della fiducia in una prospettiva interdisciplinare, sulla scia delle attività dell'Osservatorio di Ateneo «FIDES» [1].

2. *Il contesto di riferimento*

Secondo la più recente Indagine Istat disponibile sul punto (dati relativi al 2023 [2]), in una scala da uno a dieci il punteggio medio della fiducia nel sistema giudiziario è di 4,9; di 4,8 la fiducia nel parlamento italiano; di 3,5 nei partiti attuali. Pur riscontrandosi una leggera crescita rispetto agli anni precedenti, tali dati non sembrano addirsi alle esigenze di un Paese democratico, formalmente attento alle istanze solidaristiche.

Queste considerazioni hanno indotto a voler riflettere su quale sia il senso della fiducia nel diritto oggi. Ciò nell'ottica di provare a indagare le possibili *cure*, non soltanto teoriche, alle relative sfide contemporanee – cure, dunque, alla carenza di fiducia nelle istituzioni statali,

[1] Richiamando l'introduzione del Prof. Gabriele Carapezza Figlia, svolta il 7.02.2025 al secondo Convegno dell'Osservatorio FIDES, ci si riferisce, in particolare, ai contributi di Giulia Anselmo, Vincenzo Mignano, Pierfrancesco Minicangeli e Francesco Reina.

[2] Cfr. ISTAT, *Rapporto BES 2023*, in *istat.it*, pp. 145 ss.

nella pubblica amministrazione e nella giustizia e, in definitiva, nell'ordinamento in quanto tale.

3. *Fiducia e diritto: spunti per una rigenerazione del legame*

Nell'ottica di inquadrare la tematica e di comprendere meglio l'approccio delle relazioni che seguiranno, è opportuna una premessa teorica. Come è noto, alla luce del gene linguistico ambivalente della parola latina *fides*, la dottrina giuridica non distingue compiutamente tra "fiducia" e "fede" (o "buona fede")[3]. E tuttavia, la fiducia sembra trasparire in filigrana in numerosi rapporti giuridici.

Secondo la risalente e notoria impostazione hobbesiana del Leviatano[4], l'ordinamento giuridico nascerebbe da un "patto sociale" stipulato per garantire un apparato che eviti il conflitto generalizzato. Lo Stato, quindi, si svilupperebbe come risposta alla sfiducia reciproca tra gli esseri umani.

Tuttavia, di là dalla attendibilità di tale assunto, nella prospettiva dell'esperienza giuridica contemporanea, la fiducia sembra lentamente rinnovare il proprio ruolo[5].

Procedendo con ordine, la fiducia è un elemento centrale nelle dinamiche umane, sociali e giuridiche, anche se tale dato non è sempre palese[6]. Essa costituisce la base per l'interazione sociale e per la stabilità delle relazioni istituzionali. La fiducia, infatti, può qualificarsi alla

[3] F. Riccobono, *Fiducia, fede, diritto*, in *I modelli*, 2009, p. 133.

[4] T. Hobbes, *Leviatano*, a cura di R. Santi, Bompiani, Milano 2001, spec. pp. 282 ss. Cfr. comunque sul tema, anche per ulteriori riferimenti, A. Lo Giudice, *Paura e terrore nella teoria del diritto di Hobbes*, in *Teoria e Storia del Diritto Privato*, N.S., 2022, p. 2 ss.; A. Di Bello, *Sovranità e rappresentanza. La dottrina dello Stato in Thomas Hobbes*, Istituto Italiano per gli Studi Filosofici, Napoli 2010, spec. pp. 69 ss. V., comunque, F. Viola, *La concordia come concetto politico. Da Aristotele a Rawls e ritorno*, in *Filos. pol.*, 2018, pp. 11 ss.

[5] Ne sono indici i numerosi scritti emersi recentemente sul tema, fra cui v. T. Greco, *La legge della fiducia. Alle radici del diritto*, Roma-Bari 2021, cui si ricollegano F. Macioce, *La legge della fiducia e la questione del privilegio*, in *Etica e politica*, 2023, pp. 1187 ss.; A. Campo, *La fiducia come metodo. Itinerari teorici nella direzione di un 'sentire' giuridico*, in *Calumet*, 2023, pp. 104 ss.; M. Luciano, *Verso un diritto alla fiducia?*, in *Teoria e Storia del Diritto Privato*, N.S., 2022, pp. 2 ss. Cfr. anche M. Milani, *La fiducia in diritto romano. Atti costitutivi, causa, oggetto*, Jovene, Napoli 2022; M. Bertolaso - L. Valera, *Verità e fiducia nell'era del transumanesimo*, in *SCIO. Rev. de Filosofia*, 13, 2018, pp. 98 ss. Nella prospettiva pubblicistica, di recente, v. il numero monografico di *Riv. trim. Scienza dell'amm.*, 4, 2024, e spec. S.S. Scoca - P. Savarese, *Guida alla lettura*, *ivi*, pp. 1 ss.

[6] Ampi riferimenti in M. Conte, *Sociologia della fiducia*, Edizioni Scientifiche Italiane, Napoli 2009, pp. 46 ss.

ISSN 0039-4130

stregua di primario *legame relazionale*[7], di un *bene relazionale*[8]. Ciò è vero sia sul piano generale che su quello giuridico, in quanto da essa dipendono non soltanto l'identità personale e sociale degli individui, ma anche le identità e le caratteristiche dello stesso ordinamento e delle persone in quanto soggetti di diritto.

Con specifico riguardo al contesto giuridico, la fiducia svolge un ruolo duplice: essa è tanto il presupposto per l'efficacia del diritto, quanto uno degli obiettivi che il diritto stesso deve perseguire[9].

In questa prospettiva, si sta rilevando la possibilità di *ribaltare* il paradigma diffuso per cui il diritto sia mero strumento di sovversione dello stato di natura, di reazione al timore reciproco. Se, come ricordava papa Francesco, dal conflitto «si esce con gli altri»[10], ecco allora che proprio nella base relazionale della fiducia va ricercata la chiave per innovare il ruolo di questa nelle dinamiche giuridico-istituzionali. Ciò sembra possibile attraverso un processo di rigenerazione dell'ordinamento afferente a due macro-direttrici di indagine: dunque, una rigenerazione, prima, sul piano culturale e, poi, su quello della configurazione e dell'interpretazione delle norme.

4. *Sulla necessità di un'impostazione culturale della problematica*

In primo luogo, dunque, la questione della fiducia nel diritto va impostata sul piano "culturale".

La cultura giuridica comprende l'insieme dei valori, delle credenze e delle prassi che guidano il funzionamento di un sistema giuridico. Essa influisce direttamente sulla costruzione della fiducia nelle istituzioni: una cultura giuridica matura può favorire la percezione di equità e legittimità delle norme, rafforzando il legame fiduciario tra i cittadini e le istituzioni.

In un contesto globalizzato, poi, la fiducia nel diritto deve confrontarsi con le sfide poste dalle diversità culturali: i sistemi giuridici moderni devono essere in grado di adattarsi a contesti culturali differenti,

[7] C. Caltagirone, *La fiducia come legame relazionale*, in *Studium*, 2020, pp. 14 ss.

[8] Su cui, ampiamente, P. Donati, *Sociologia relazionale. Come cambiare la società*, Scholé, Brescia 2021, e Id., *Teoria relazionale*, in T. Marci - S. Tomelleri (a cura di), *Dizionario di sociologia per la persona*, FrancoAngeli, Milano 2021, pp. 118 ss. V. sul punto N. Caracappa, *Il concetto di fiducia nella sociologia relazionale di Pierpaolo Donati*, in *Studium*, 2020, pp. 163 ss.

[9] Cfr. T. Greco, *La legge della fiducia*, cit., pp. 53 ss.

[10] Francesco, *Udienza ai partecipanti all'Incontro promosso dall'"International Catholic Legislators Network"*, in *vatican.va*, 24.08.2024.

ISSN 0039-4130

senza perdere la loro coerenza e legittimità. Ciò richiede un dialogo continuo tra culture giuridiche, basato sulla comprensione e sul rispetto reciproco.

In tal senso, un elemento chiave della cultura giuridica è il ruolo delle norme condivise [11]. Le norme non soltanto regolano i comportamenti, ma riflettono anche i valori e le aspettative di una comunità; così, quando le norme giuridiche rispecchiano i valori culturali, la fiducia nel sistema giuridico tende ad aumentare, poiché le persone percepiscono il diritto come un riflesso delle loro esigenze e aspirazioni.

Specularmente, il diritto non è solo un prodotto della cultura, ma agisce anche come agente di trasformazione culturale nella società come nei rapporti politici ed economici [12].

Nell'ottica di una visione personalistica del sistema giuridico [13], se il diritto è al tempo stesso uno *specchio* e uno *strumento di trasformazione* della società [14], allora esso, attraverso il rafforzamento della fiducia nelle istituzioni giuridiche, contribuisce a consolidare la coesione sociale e a promuovere una cultura del rispetto e della collaborazione.

In tale chiave di lettura, il positivismo giuridico inclusivo, prospettiva ermeneutica che tende a una lettura unitaria del fenomeno giuridico, riconosce che il diritto non può essere separato dal contesto morale e culturale in cui opera [15]. Sotto questo angolo visuale, integrare valori condivisi all'interno dell'interpretazione delle norme è essenziale, in primo luogo, per rafforzare la fiducia nelle istituzioni.

Fattore determinante per la costruzione della fiducia nel diritto si rinviene, quindi, nell'educazione giuridica. Formare giuristi che comprendano il valore della fiducia e siano capaci di promuoverla attraverso la loro esperienza professionale è essenziale per garantire un sistema equo e inclusivo.

Per altro verso, l'educazione giuridica può contribuire a sensibiliz-

[11] Sul ruolo della prassi nell'ordinamento vigente, v. G. Perlingieri, *Sulla falsa alternativa tra* ius positum *e* ius in fieri*. La «creatività vincolata» dell'attività interpretativa*, in *Ann. Sisdic*, 2023, pp. 18 ss. Sul ruolo della prassi nella creazione del diritto consuetudinario, v. in particolare P. Virgadamo, *Introduzione*, in Id., *Diritto privato consuetudinario*, Edizioni Scientifiche Italiane, Napoli 2025, in corso di pubblicazione, § 6.

[12] Cfr. P. Sbriglia, *Il fattore umano; crisi di fiducia e politica economica*, in *Riv. Corte dei conti*, 4, 2023, pp. 110 ss.

[13] P. Perlingieri, *Per un positivismo giuridico «inclusivo». Note minime su diritto e cultura*, in *Ann. Sisdic*, 2023, pp. 1 ss.

[14] F. Viola, *Diritto e cultura*, in *Rass. dir. civ.*, 2023, p. 160 ss.; P. Perlingieri, *op. ult. cit.*, p. 8 ss.

[15] In tema, v. primariamente P. Perlingieri, *Note critiche sulla funzione «restauratrice» del giurista*, in *Ann. Sisdic*, 2023, pp. 3 ss.

ISSN 0039-4130

zare i cittadini sull'importanza della fiducia nelle relazioni giuridiche. Promuovendo una migliore conoscenza delle norme e dei processi decisionali, l'educazione ha l'opportunità di ridurre le percezioni di incertezza e alienazione, favorendo un rapporto più diretto e trasparente con le istituzioni.

5. *Prospettive applicative. Cenni*

In secondo luogo, la successiva macro-direttrice di indagine riguarda, come anticipato, la prospettiva ermeneutica dei diversi àmbiti del diritto. In questo senso, l'impegno della Scuola di dottorato ha portato a delle analisi ad ampio spettro di diversi rami dell'ordinamento.

Rimandando ai relativi contributi, si consideri quanto segue.

Sul piano privatistico, gli elementi fiduciari analizzati emergono tanto in istituti classici del diritto dei contratti quanto nei rapporti di più recente emersione [16]. Su quello penalistico, si è indagato il *proprium* della fiducia dei consociati verso il sistema sanzionatorio e verso coloro che *subiscono* tale sistema, anche in relazione alla giustizia c.d. riparativa [17]. Con riguardo infine ai rapporti sovranazionali, sono state approfondite le dinamiche di fiducia fra stati in seno alle diverse forme di interazione internazionale [18].

Merita, per vero, un cenno il ruolo della fiducia nelle relazioni fra operatori commerciali e consumatori. Si consideri, ad esempio, il ruolo delle attività promozionali svolte *online* da "*influencer*". Qui la *fiducia* che gli utenti ripongono in questi soggetti è in grado di condizionare significativamente la diffusione di un prodotto o di un servizio. Bene o mal riposta che sia, invero, si tratta di una "fiducia" che altera il normale sviluppo fisiologico delle dinamiche commerciali e che, talvolta, può condurre alla realizzazione di pratiche commerciali scorrette, sanzionabili ai sensi del codice dei consumatori. L'attenzione, poi, che l'AGCOM ha posto sul marketing svolto da influencer, tramite l'ema-

[16] Cfr., almeno, A.P. Uges, *La fiducia come situazione giuridica real-obbligatoria*, ESI, Napoli 2022.

[17] V. i contributi di L. Lacchè, G. Todeschini, V. Pelligra, T. Greco, G. Mannozzi, A.R. Amato, M. Stronati, R. Cornelli, R. Bartoli e D. Pulitanò in *La fiducia. Riflessioni interdisciplinari per un dibattito contemporaneo su giustizia, diritto di punire e pena*, fascicolo monografico di *Quaderno di storia del penale e della giustizia*, 5, 2023.

[18] Cfr. L. Riccardi, *Il principio di mutua fiducia nella giurisprudenza europea: ancora una opportunità?*, in *Camminodiritto*, 4, 2017, pp. 3 ss.; S. Marinai, *Obblighi informativi e fiducia reciproca tra Stati membri nei trasferimenti Dublino*, in *Dir. imm. cittad.*, 2024, pp. 1 ss.

ISSN 0039-4130

nazione di apposite linee guida, sembra testimoniare la bontà di tali riflessioni [19].

Ancora, peculiare sembra essere il ruolo della fiducia in relazione nuove tecnologie – la *blockchain*, in particolare – relativamente alla possibilità di queste di offrire garanzie di "certezza" e stabilità ai traffici commerciali su di esse basati [20]. Di la dal fatto che si tratti di una *fiducia* ben riposta, l'interrogativo primario è qui se *l'affidarsi* a degli *smart contract*, contratti digitali che garantiscono l'adempimento, porterà ad avere più fiducia nel digitale che nelle persone umane [21].

Il progresso tecnologico, dunque, solleva questioni etiche e giuridiche proprio in relazione al rapporto tra fiducia e diritto in seno, anche, al mercato digitale [22]: la 'fiducia algoritmica' può davvero sostituire quella umana? Quali sono i rischi di una sempre maggiore dipendenza da sistemi automatizzati? Quale atteggiamento assumere dinanzi alle *transizioni* del mondo contemporaneo [23]? Tali quesiti, pur sinteticamente presentati, celano, invero, interrogativi esistenziali che, senza pre-giudizi di alcuna sorta, necessitano di essere indagati con scrupolo e rigore.

6. *Conclusioni*

Ampliando per un attimo l'angolo visuale, la fiducia è, evidentemente, alla base delle relazioni sociali, quantomeno sul piano teorico. Tuttavia, l'individualismo imperante sembrerebbe collidere con la possibilità di instaurare rapporti di piena fiducia. Ciò, evidentemente, indebolisce la coesione delle nostre comunità (siano esse locali, aziendali, statali, internazionali).

Al riguardo, ad avviso di chi scrive appare interessante un dato, evidenziato da alcune letture filosofiche della modernità: il rischio che le maggiori libertà di cui disponiamo, ove vissute male, possano indur-

[19] AGCOM, *Linee-guida*, All. A, delib. n. /24/cons, in *agcom.it*.

[20] Cfr. M. Bertolaso - G. Lo Storto (a cura di), *Etica digitale. Verità, responsabilità e fiducia nell'era delle macchine intelligenti*, LUISS University Press, Roma 2020; *La fiducia nella cittadinanza digitale*, fasc. di *Digeat*, 3, 2024, in *digeat.info*; I. Martone, *Gli smart contracts. Fenomenologia e funzione*, ESI, Napoli 2022, pp. 19 ss.

[21] Cfr. G. Frezza - P. Virgadamo, *NFT e arte. Alla ricerca di una disciplina giuridica adeguata orientata al princípio di verità*, in *LawArt*, 4, 2023, pp. 285 ss.

[22] Cfr. L. Palazzani - M. Daverio, *Per un umanesimo tecnologico. L'intelligenza artificiale tra scienza, etica e diritto*, in *Munera*, 2, 2022, p. 9 ss.

[23] G. Carapezza Figlia, *Doppia transizione europea ed* European Green Deal, in *Tecn. dir.*, 2024, 340 ss.

ISSN 0039-4130

re a una schiavitù nei confronti dell'autorealizzazione, e dunque del sé, allontanando il singolo dalla comunità [24].

Si voglia perdonare la semplificazione di un tema complesso e dibattuto. Si ritiene, sinceramente, che questi due, correlati, elementi – *i.e.*, la non sana percezione di se stessi e l'allontanamento interiore dalle comunità – incidano significativamente anche sul rapporto di ciascuna persona con il diritto.

Sul punto, invero, le analisi svolte tendono a contrastare questa prospettiva. Provando ad operare una prima riconduzione ad unità delle ricerche sinora svolte, esse evidenziano come i valori e le prassi culturali influenzino profondamente la percezione e la costruzione della fiducia nelle istituzioni giuridiche, essendo la fiducia una delle condizioni essenziali tanto per la costruzione della società quanto per la piena realizzazione della dignità umana di ciascuno [25]. Una cultura giuridica inclusiva, che integri valori condivisi e promuova il dialogo interculturale in contesti tendenzialmente pluralisti, rappresenta una condizione essenziale per rafforzare la fiducia nel diritto e garantire una più efficace coesione sociale. Attraverso l'educazione, l'adattamento normativo e l'integrazione della diversità, il diritto può continuare a svolgere un ruolo centrale nella costruzione di una società più giusta e *fiduciosa*.

Così, in controtendenza rispetto alla prospettiva filosofica prima menzionata, evidenziando i tratti positivi degli *elementi fiduciari* dei rapporti giuridici, muovendo dal carattere *relazionale* della fiducia anche in tale contesto diventa possibile accedere ad una prospettiva di sensibilizzazione culturale ampia che possa agevolare una piena rigenerazione sociale.

[24] V. G. Rizzi, *Il paradosso della modernità tra individualismo e comunità*, in *ilsole24ore.it*, 20 luglio 2023.

[25] S. Biancu, *Why Trust Matters (and Why We Should Take Care of It)*, in *Studium, Contemporary Humanism Annals*, 2023, p. 24 ss.

ISSN 0039-4130

La fiducia dei giovani studenti nel proprio futuro e nelle istituzioni

Uno sguardo ai risultati ICCS 2022

Marco Valerio

Introduzione

La fiducia nelle istituzioni dei giovani studenti è un elemento centrale per capire quale percezione abbiano le giovani generazioni della coesione sociale della società in cui vivono [1]. Il successo della collaborazione tra gli individui di una società passa anche attraverso la fiducia che essi ripongono nelle sue istituzioni [2]. Inoltre, il mondo contemporaneo spinge i giovani ad affrontare dei cambiamenti caratterizzati da un alto grado di incertezza, i quali mettono alla prova le loro aspettative sul futuro [3]. A tal proposito, l'indagine comparativa ICCS 2022 fornisce un punto di vista privilegiato per indagare a fondo la prospettiva dei giovani studenti su questi aspetti. In particolare, l'attenzione sarà posta su due domande incluse nei questionari studenti ed europeo riguardanti la fiducia nei gruppi e nelle istituzioni e le aspettative sul proprio futuro. Questa analisi permetterà di esaminare questi dati alla luce dei risultati ottenuti nel ciclo ICCS precedente e di confrontare la situazione italiana e quella internazionale.

[1] W. Schulz - J. Fraillon - B. Losito - G. Agrusti - J. Ainley - V. Damiani - T. Friedman, *IEA International Civic and Citizenship Education Study 2022 Assessment Framework*, Springer, 2023. p. 4,14.

[2] *Ibidem*. p. 78

[3] V. Damiani - B. Losito - G. Agrusti - W. Schulz, *Young Citizens' Views and Engagement in a Changing Europe IEA International Civic and Citizenship Education Study 2022 European Report*, Springer, 2025. p. 63.

Il questionario studenti e il questionario europeo di ICCS 2022

La IEA è un'organizzazione internazionale che si occupa di ricerca comparativa in ambito educativo. Il suo scopo principale è studiare i sistemi educativi a livello globale attraverso indagini comparative che si concentrano sugli studenti, sulle pratiche didattiche e sui contesti scolastici[4]. Una delle sue indagini più importanti è proprio ICCS. Questa indagine comparativa internazionale ha tra le sue finalità principali quella di valutare le conoscenze civiche degli studenti, nonché i loro atteggiamenti civici e il loro livello di partecipazione civica[5]. ICCS, infatti, si prefigge l'obiettivo di analizzare il modo in cui i giovani vengono preparati a esercitare il loro ruolo di cittadini attivi in contesti democratici in continua evoluzione. L'indagine ha recentemente concluso la sua terza edizione, quella del 2022 – le indagini precedenti si sono svolte nel 2016 e nel 2009. Nell'edizione appena conclusa i paesi partecipanti sono stati 24, con una preponderante partecipazione dei paesi europei.

I principali soggetti dell'indagine sono gli studenti all'ottavo grado di istruzione, i docenti e i dirigenti scolastici, ai quali sono stati somministrati, rispettivamente, il questionario studenti, il questionario docenti e il questionario per la scuola. Il questionario studenti è composto da 218 item di cui 34 opzionali a seconda del paese. Oltre agli atteggiamenti civici e alla partecipazione civica degli studenti, il questionario rileva le loro opinioni su molti temi civici e sociali, tra cui la fiducia in gruppi e istituzioni. Una parte viene dedicata anche ad indagare il background degli studenti, il loro ambiente familiare e scolastico [6].

Inoltre, agli studenti europei è stato somministrato un ulteriore questionario – il questionario europeo – che ha lo scopo di indagare tematiche ritenute particolarmente rilevanti per questa regione, ampliando le questioni già trattate nel questionario internazionale. Tra le varie tematiche affrontate all'interno delle 12 domande – per un totale di 92 item – il questionario esplora le aspettative degli studenti per il futuro personale e dell'Europa[7].

A seguire, verranno presentati nel dettaglio i risultati relativi alle risposte date dagli studenti a due quesiti: uno del questionario studenti

[4] https://www.iea.nl/

[5] W. Schulz et al., IEA *International Civic cit*.

[6] W. Schulz - T. Friedman - J. Fraillon, *ICCS 2022 Technical Report*, Springer, 2024.

[7] V. Damiani - B. Losito - G. Agrusti - W. Schulz, *Young Citizens' Views and Engagement in a Changing Europe IEA International Civic and Citizenship Education Study 2022 European Report*, Springer, 2025.

ISSN 0039-4130

relativo alla fiducia degli studenti nei gruppi e nelle istituzioni; e uno del questionario europeo relativo alle aspettative degli studenti sul proprio futuro.

Fiducia degli studenti nei gruppi e nelle istituzioni

Ai partecipanti è stato chiesto di esprimere il loro grado di fiducia («completamente», «molto», «abbastanza», «per niente») verso: «il governo nazionale»; «il parlamento/consiglio nazionale» (a seconda del paese); «i tribunali»; e infine, «i media tradizionali (televisioni, giornali, radio)». Tale quesito era presente anche nel questionario studenti di ICCS 2016.

In media gli studenti hanno mostrato il livello di fiducia più alto nei confronti dei tribunali, con il 69% che si fida completamente o molto di questa istituzione, seguito dal governo nazionale (58%), il parlamento/consiglio nazionale (53%), e i media tradizionali (52%) (Tabella 1) [8]. Si può notare come, rispetto ad ICCS 2016, in media la percentuale di studenti che si fidano completamente o molto delle quattro istituzioni in questione sia diminuita in ICCS 2022. Infatti, in media la percentuale di studenti che si fidava completamente o molto dei tribunali era del 72%, del governo nazionale era del 63%, dei media tradizionali era del 57% e del parlamento/consiglio nazionale era del 58% (Tabella 1) [9].

Tabella 1. Percentuale di studenti dei paesi ICCS che si fidano completamente o molto delle seguenti istituzioni (media ICCS 2022 e 2016): [10]

	Tribunali	Governo nazionale	Media tradizionali	Parlamento/consiglio nazionale
ICCS 2022	69 (0.3)	58 (0.3)	52 (0.3)	53 (0.3)
ICCS 2016	72 (0.3)	63 (0.3)	57 (0.3)	58 (0.3)
Differenza tra ICCS 2022 e ICCS 2016	-3	-5	-5	-5

[8] W. Schulz - J. Fraillon - B. Losito - G. Agrusti - V. Damiani - T. Friedman, *Education for Citizenship in Times of Global Challenge IEA International Civic and Citizenship Education Study 2022 International Report*, Springer, 2023.

[9] *Ibidem.*

[10] Gli errori standard figurano tra parentesi ( ) in tutte le tabelle.

ISSN 0039-4130

In Italia, la fiducia degli studenti nelle istituzioni varia a seconda dell'ente considerato. Il 61% degli studenti italiani dichiara di fidarsi completamente o molto dei media tradizionali, mentre la fiducia nel parlamento si attesta al 52%. Per quanto riguarda i tribunali, il 64% degli studenti afferma di nutrire una completa o molta fiducia in questa istituzione. Infine, la fiducia nel governo nazionale si attesta al 53%, risultando perfettamente in linea con la media generale di ICCS 2022 (Tabella 2) [11]. Rispetto a ICCS 2016, la fiducia degli studenti italiani è calata nei confronti di tutte le istituzioni. Il calo più drastico si è verificato nella percentuale di studenti che si fidano completamente o molto dei media tradizionali, di ben 15 punti percentuali. Seguito dalla fiducia nel parlamento, scesa di 13 punti percentuali. Anche la fiducia nei tribunali è calata di 9 punti percentuale. E infine, la fiducia nel governo che perde 4 punti percentuali (Tabella 2) [12].

Tabella 2. Percentuale di studenti italiani che si fidano completamente o molto delle seguenti istituzioni (media ICCS 2022 e 2016):

	Tribunali	Governo nazionale	Media tradizionali	Parlamento
ICCS 2022	64 (0.9)	53 (1.1)	61 (1.0)	52 (1.1)
ICCS 2016	72 (1.1)	57 (1.0)	75 (0.7)	65 (0.9)
Differenza tra ICCS 2022 e ICCS 2016	-9	-4	-15	-13

Rispetto ai dati complessivi di ICCS 2022, la fiducia degli studenti italiani nei media tradizionali (61%) è significativamente superiore alla media ICCS (52%). Tuttavia, nel caso dei tribunali, gli studenti italiani mostrano un livello di fiducia inferiore (64%) rispetto alla media ICCS (69%). Ciò si verifica anche con la fiducia nel parlamento, la quale risulta più bassa in Italia (52%) rispetto alla media ICCS (53%). Infine, la fiducia nel governo nazionale tra gli studenti italiani (53%) è inferiore a confronto con la media ICCS 2022 (58%).

[11] W. Schulz et al., *Education for Citizenship cit.*
[12] *Ibidem.*

ISSN 0039-4130

Aspettative degli studenti sul proprio futuro

All'interno del questionario europeo veniva chiesto agli studenti di riportare le proprie aspettative sul proprio futuro. In particolare, veniva chiesto di indicare il grado di probabilità («molto improbabile», «improbabile», «probabile», «molto probabile») con cui, secondo gli studenti, alcuni scenari relativi al proprio futuro avrebbero potuto realizzarsi: «troverò un lavoro stabile», «la mia situazione finanziaria sarà migliore rispetto a quella dei miei genitori», «troverò un lavoro che mi piace», «avrò l'opportunità di viaggiare all'estero per piacere» e «guadagnerò abbastanza da mettere su famiglia» [13].

Gli studenti europei hanno mostrato globalmente ottimismo rispetto al proprio futuro. Il 94% degli studenti pensa sia probabile o molto probabile che troverà un lavoro stabile e il 92% afferma che guadagnerà abbastanza da mettere su famiglia. L'89% crede che sia probabile o molto probabile che avrà l'opportunità di viaggiare all'estero per piacere, mentre il dato più basso, benché decisamente positivo, è relativo agli studenti che credono sia probabile o molto probabile che la loro situazione finanziaria sarà migliore rispetto a quella dei propri genitori (82%) (Tabella 3) [14].

Lo stesso set di item è stato somministrato agli studenti europei del ciclo precedente ICCS 2016. In tale occasione il 95% ha dichiarato che probabilmente o molto probabilmente troverà un lavoro stabile e il 78% che la sua situazione famigliare sarà migliore di quella dei propri genitori. L'88% crede che avrà l'opportunità di viaggiare all'estero per piacere mentre quasi la totalità dei rispondenti – il 96% – ha affermato che è probabile o molto probabile che guadagnerà abbastanza da mettere su famiglia (Tabella 3) [15]. Confrontando le risposte ottenute nelle due edizioni di ICCS, i risultati mostrano una certa stabilità nella fiducia degli studenti europei riguardo al loro futuro lavorativo, con delle leggere variazioni.

[13] V. Damiani et al., *Young Citizens' Views and Engagement cit.*

[14] *Ibidem.*

[15] B. Losito - G. Agrusti - V. Damiani - W. Schulz, *Young People's Perceptions of Europe in a Time of Change IEA International Civic and Citizenship Education Study 2016 European Report*, Springer, 2018.

ISSN 0039-4130

Tabella 3. Percentuali di studenti europei che ritengono che i seguenti eventi si realizzeranno nella propria vita futura probabilmente o molto probabilmente (media ICCS 2016 e 2022):

	Troverò un lavoro stabile	La mia situazione finanziaria sarà migliore rispetto a quella dei miei genitori	Troverò un lavoro che mi piace	Avrò l'opportunità di viaggiare all'estero per piacere	Guadagnerò abbastanza da mettere su famiglia
ICCS 2022	94 (0.1)	82 (0.2)	89 (0.2)	88 (0.2)	92 (0.2)
ICCS 2016	95 (0.1)	78 (0.2)	91 (0.2)	89 (0.2)	96 (0.1)
Differenza tra ICCS 2022 e ICCS 2016	-1	+4	-2	-1	-4

Allo stesso modo, i dati italiani si trovano in linea con quelli europei, mostrando in generale una fiducia elevata nel loro futuro lavorativo. Il 94% ritiene probabile o molto probabile di guadagnare abbastanza per formare una famiglia, mentre addirittura la quasi totalità – il 97% – pensa di riuscire a ottenere un lavoro stabile. Inoltre, il 91% crede di avere buone possibilità di viaggiare all'estero per piacere. Il dato più basso, seppur positivo, riguarda la prospettiva di migliorare la propria situazione finanziaria rispetto ai genitori, con l'87% degli studenti che considera questo scenario probabile o molto probabile (Tabella 4) [16].

A differenza degli altri paesi europei, tra ICCS 2016 e ICCS 2022 gli studenti italiani hanno mostrato un aumento di fiducia nel proprio futuro lavorativo in quasi la totalità degli item. Infatti, il 92% degli studenti italiani nel 2016 credeva che fosse probabile o molto probabile trovare un lavoro stabile. Nello stesso ciclo, l'81% degli studenti affermava che fosse probabile o molto probabile che la propria situazione familiare sarebbe stata migliore rispetto a quella dei propri genitori. La differenza più significativa riguarda la percentuale di studenti che ritiene di avere la possibilità di viaggiare all'estero per piacere, pari al 79% nel 2016. Infine, l'unico dato che è restato abbastanza in linea con l'edizione successiva è quello relativo agli studenti che ritengono

[16] L. Palmerio - S. Greco, *ICCS 2022: Rapporto nazionale sui Risultati del Questionario Europeo per gli Studenti, Indagini INVALSI*, 2024. https://www.iea.nl/sites/default/files/2024-04/Italy.pdf

ISSN 0039-4130

probabile o molto probabile che in futuro guadagneranno abbastanza da mettere su famiglia (95% – un punto percentuale in meno di ICCS 2022) (Tabella 4) [17].

Tabella 4. Percentuali di studenti italiani che ritengono che i seguenti eventi si realizzeranno nella propria vita futura probabilmente o molto probabilmente (media ICCS 2016 e 2022):

	Troverò un lavoro stabile	La mia situazione finanziaria sarà migliore rispetto a quella dei miei genitori	Troverò un lavoro che mi piace	Avrò l'opportunità di viaggiare all'estero per piacere	Guadagnerò abbastanza da mettere su famiglia
ICCS 2022	97 (0.4)	87 (0.9)	92 (0.7)	91 (0.6)	94 (0.5)
ICCS 2016	92 (0.5)	81 (0.8)	89 (0.5)	79 (0.8)	95 (0.6)
Differenza tra ICCS 2022 e ICCS 2016	+5	+6	+3	+12	-1

Gli studenti italiani mostrano un livello di fiducia nel proprio futuro generalmente in linea con quello degli studenti europei, sebbene complessivamente più elevato. In Italia, il 97% degli studenti ritiene probabile o molto probabile trovare un lavoro stabile, rispetto al 94% registrato a livello europeo. Anche la fiducia nel guadagnare abbastanza per formare una famiglia e la possibilità di viaggiare all'estero per piacere sono maggiori in Italia, rispettivamente il 94% e il 91% per gli italiani, e il 92% e 89% per gli europei. Infine, il dato più basso riguarda la prospettiva di migliorare la propria situazione finanziaria rispetto ai genitori, che, pur rimanendo positivo, risulta più elevato tra gli italiani (87%) rispetto agli europei (82%) (Tabella 3, Tabella 4).

Discussione

I cambiamenti emersi tra le risposte degli studenti nelle edizioni ICCS 2016 e ICCS 2022 devono essere interpretati tenendo conto di fattori

[17] B. Losito et al., *Young People's Perceptions cit.*.

di contesto, come ad esempio la pandemia di Covid-19, che potrebbero aver avuto un impatto significativo sugli studenti. Inoltre, i dati analizzati in ICCS non rappresentano esclusivamente la realtà scolastica, ma offrono una visione più ampia sulla prospettiva degli studenti nei confronti della società. Infatti, la fiducia degli studenti nelle istituzioni può essere fortemente influenzata da fattori di contesto quali la percezione del livello di corruzione del paese, l'uso dei social media, l'età, ecc. [18]. Pertanto, promuovere la fiducia nelle istituzioni e nei gruppi sociali non può essere attribuita esclusivamente alla scuola, ma deve riguardare l'intera comunità. Lo stesso vale per le aspettative dei giovani sul proprio futuro, le quali sono fortemente modellate dall'insieme delle condizioni sociali, economiche e culturali in cui essi vivono. Dunque, incentivare uno sguardo positivo sul futuro da parte dei giovani non è un compito attribuibile esclusivamente all'ambito scolastico. Ciononostante, va sottolineato che esistono evidenze a supporto del fatto che la scuola possa promuovere la socializzazione politica degli studenti [19]. L'indagine ICCS ha mostrato come, nei paesi dove l'indice di corruzione è basso, un livello di conoscenza civica più elevato viene associato a una maggiore fiducia nelle istituzioni [20]. O ancora, alcuni studi suggeriscono che la percezione degli studenti di un clima di classe positivo possa potenziare la fiducia riposta nelle istituzioni [21]. Questo significa che lavorando attivamente a livello scolastico su alcuni aspetti – ad esempio la conoscenza civica – si possa migliorare il livello di fiducia nelle istituzioni degli studenti.

Conclusioni

L'analisi dei dati ICCS 2022 mostra un calo della fiducia degli studenti nelle istituzioni rispetto al ciclo precedente, mentre le aspettative per il futuro rimangono generalmente positive. In Italia, si riscontra una maggiore fiducia nei media tradizionali rispetto alla media ICCS, ma la fiducia nei tribunali, nel parlamento e nel governo è inferiore. Per quanto riguarda le aspettative degli studenti sul proprio futuro, gli stu-

[18] W. Schulz - J. Fraillon - B. Losito - G. Agrusti - J. Ainley - V. Damiani - T. Friedman, *IEA International Civic and Citizenship Education Study 2022 Assessment Framework*, Springer, 2023. p. 45, 81.

[19] *Ibidem*. p. 3.

[20] *Ibidem*. p. 45.

[21] C. Barber - S. O. Sweetwood - M. King, *Creating classroom-level measures of citizenship education climate*, in *Learning Environments Research,* 18, 2015, pp. 197–216.

ISSN 0039-4130

denti italiani si collocano sopra la media europea nella totalità degli aspetti indagati. Sebbene sia importante capire come i risultati relativi alla fiducia nelle istituzioni degli studenti e le loro aspettative sul proprio futuro siano fortemente influenzati da variabili che esulano dalla realtà scolastica, la scuola ricopre un ruolo fondamentale nel promuovere in questi ultimi la fiducia nelle istituzioni.

ISSN 0039-4130

La fiducia come cura della patologia sociale nel sistema penale

Francesco Luigi Reina

Nonostante l'apparente ossimoro, la fiducia è un elemento centrale di quel tessuto normativo caratterizzato ontologicamente da sfiducia qual è il diritto penale e processual-penalistico. Da un lato, la fiducia, e in particolare quella cd. istituzionale, è un elemento essenziale del sistema penalistico e immanente ad esso, poiché senza la fiducia che i consociati nutrono nell'effettività di questo sistema, lo stesso collasserebbe: se il cittadino non riponesse affidamento nella piena e certa applicazione della legge, l'apparato penalistico e quello giudiziario rimarrebbero privi di legittimazione sociale, poiché le norme penali e le pene in esse comminate non avrebbero più alcuna funzione deterrente; specularmente, le vittime di reati potrebbero persino rinunciare a denunciare i torti subiti, in ragione della maturata sfiducia verso il sistema stesso.
Contestualmente, però, anche la sfiducia è una componente propria del sistema, ed ancor più pervasiva della prima. Le norme penalistiche, infatti, trovano applicazione solo dopo che sia stato commesso un reato, e la realizzazione di quest'ultimo crea a sua volta quella che viene definita una "frattura sociale": la commissione di un fatto di rilievo penale comporta una rottura del cd. "patto sociale" che lega i consociati, ledendo così il rapporto di fiducia che intercorre tra la società e il consociato colpevole.

In questo assetto, il ricorso alla giustizia punitiva, specie alla privazione della libertà personale, rappresenta l'unico strumento per sanare questa frattura e, dunque, ripristinare il patto sociale e la fiducia reciproca[1]. Il tutto inevitabilmente connotato da sentimenti retributivi, e spesso

[1] «Il rispetto del patto sociale ha come motore primo "*the terror of some punishment*". Qui la fiducia reciproca è sempre accompagnata dal *metus*, con la spada a fare da visibile deterrente», L. Lacchè, *Avere fiducia nella fiducia: quale visione del penale?*, in *Quaderno di storia del penale e della giustizia*, V, 2023, eum, Macerata, p. 10.

contaminato da pulsioni vendicative. Pertanto, non è affatto azzardato sostenere che il tessuto normativo penalistico sia composto ontologicamente da sfiducia verso il reo.

La compresenza simbiotica di questi due elementi, antitetici, fonda un sistema in cui la violenza sfiduciaria, punitiva, che viene esercitata nei confronti del consociato colpevole, è funzionale a generare la fiducia nella collettività. Si crea così un circolo vizioso nel quale il diritto penale assume un «carattere fisiologicamente patologico»[2].

Sebbene questa appena descritta costituisca tutt'oggi l'ossatura dell'apparato penalistico, le recenti evoluzioni normative e dottrinarie stanno illuminando il percorso verso un sistema penale che potrebbe essere definito come "fiduciario". Una struttura il cui assetto non è più unicamente costituito dalla repressione, dalla sfiducia e dalla violenza punitiva, ma viene implementato con una forte componente di fiducia e, per certi versi, di empatia. Un sistema in cui, a differenza di prima, è la fiducia nei confronti del reo che ingenera la fiducia nella collettività.

La conseguenza di questa evoluzione è che la fiducia, che prima rilevava unicamente in senso negativo, come valore che è stato violato, ora assurge a valore positivo che permea la vicenda giudiziaria e ne determina gli esiti. La rilevanza negativa della fiducia si appalesa infatti all'interno della struttura di alcuni reati (es, reati in contratto; truffe[3]), nei rapporti con gli altri consociati in ragione della frattura del patto sociale, nonché nel rapporto tra il reo e lo Stato: quest'ultimo esercita la sua violenza punitiva nella speranza di una rieducazione del colpevole, e che questi, a seguito dell'espiazione della pena, comprenda i valori costituzionali che il suo comportamento ha offeso[4].

Focalizzando l'analisi su tale ultimo rapporto verticale, tra Stato e reo, è proprio in questa direttrice che la fiducia inizia a fare capolino come elemento evanescente ma permeante di alcuni istituti. Tra questi, la sospensione condizionale della pena: l'istituto ha fatto il suo ingresso nell'ordinamento italiano già nel 1904 con la Legge Ronchetti, ma è solo con la formulazione legislativa che gli è stata data con la L. 220

[2] T. Greco, *È possibile un diritto penale fiduciario?*, in *Quaderno di storia del penale e della giustizia*, V, 2023, eum, Macerata, p. 63.

[3] La persona offesa da queste categorie di reati, infatti, si determina a compiere un atto proprio perché si fida del colpevole, il quale invece, a sua volta, utilizza artifici o raggiri per ingannarlo.

[4] La tendenza alla rieducazione che la Costituzione impone alle pene, ai sensi dell'art 27, co. 3, non si concretizza però nella previsione di strumenti e istituti volti in modo precipuo a sanare il conflitto relazionale che quel determinato reato cagiona, ma si limita a favorire il reinserimento sociale del singolo condannato.

ISSN 0039-4130

del 1974 che ha acquisito la fiducia verso il reo come elemento costitutivo implicito ma determinante. Dopo aver irrogato una pena, se questa rientra in certi limiti edittali, il giudice stesso può ordinare che l'esecuzione della stessa venga sospesa: il requisito codicistico è che il colpevole si astenga dal commettere ulteriori reati negli anni successivi. Pertanto, il giudice deve formulare una prognosi favorevole sul futuro comportamento del colpevole, e l'elemento centrale che orienta questa valutazione è proprio la fiducia che il giudice ripone verso il reo, e verso il fatto che questo si asterrà da una futura recidivanza. La circostanza per cui si abbandona una visione strettamente positivistica è sintomatica del fatto che questo giudizio si configuri come una valutazione fiduciaria: la norma infatti non positivizza dei requisiti per la valutazione, ma lascia ampi margini di discrezionalità al giudice.

Ancora, con la L. 67 del 2014 ha fatto ingresso anche per gli imputati adulti la cd. "messa alla prova", un istituto innovativo e anch'esso fondato su una fiducia diafana: con la differenza, però, che è lo stesso imputato a poter chiedere che gli venga data fiducia. Prima ancora della condanna, l'imputato stesso può chiedere al giudice di sospendere il procedimento così da poter essere messo alla prova, eliminando le conseguenze dannose o pericolose derivanti dal reato e svolgendo lavori di pubblica utilità: il giudice, ai fini dell'ammissione all'istituto, compie una valutazione fiduciaria, dovendosi fidare del ravvedimento e della resipiscenza dell'imputato, nonché del fatto che questo non commetterà ulteriori reati.

Una volta analizzate sinteticamente le dinamiche fiduciarie nel rapporto verticale tra Stato e reo, occorre però evidenziare che la progressione verso un diritto penale maggiormente fiduciario deve necessariamente coinvolgere soprattutto i rapporti orizzontali: quelli, cioè, tra la vittima del reato, il reo, e la comunità di riferimento. Come è stato argutamente evidenziato, «bisogna lavorare per le riconciliazioni e per soluzioni non punitive, ma nella comunità», poiché «la fiducia [...] non può essere inserita in contesti di strategie punitive»[5]: la punizione stessa, infatti, rende superflua qualsiasi forma di fiducia.

Già nella Direttiva 2012/29 in materia di vittime di reato, le istituzioni dell'UE in diversi Considerando parlano esplicitamente della necessità di fiducia nei rapporti verticali, ma sottolineano soprattutto il fatto che «il reato non è solo un torto alla società, ma anche e soprattutto una violazione dei diritti individuali delle vittime»[6]. Pertanto si ren-

[5] E. Resta, *Fiducia nella giustizia*, in *Minori Giustizia*, I, 1996, p. 71.

[6] Direttiva 2012/29/UE del Parlamento Europeo e del Consiglio, che istituisce norme minime in materia di diritti, assistenza e protezione delle vittime di reato, sub Considerando (9).

ISSN 0039-4130

de necessaria una lettura del fenomeno criminoso in chiave relazionale.

Il nostro ordinamento ha fatto negli ultimi anni un «ambizioso investimento culturale»[7] in tal senso, e con il D.lgs. 150/2022 (cd. riforma Cartabia), ha introdotto gli istituti di giustizia riparativa. Una giustizia che si interseca con quella punitiva, ma che permette di «rispondere al male non con altro male [...] ma con una cospicua dose di fiducia reciproca tra i protagonisti del percorso»[8]: una giustizia, soprattutto, volta a ricomporre la pace sociale e a sanare il "conflitto relazionale" tra il reo, la vittima e la comunità. Il fine ultimo è quello di rendere complementari la giustizia sanzionatoria con una «giustizia consensuale», la cui «realizzazione dipende dalla volontà dei protagonisti del fatto illecito, autore e vittima»[9], nell'ottica di «ridare significato ai legami fiduciari fra le persone»[10] che il reato ha interrotto.

La fiducia costituisce un elemento centrale nel percorso riparativo: l'esito positivo è subordinato al riavvicinamento tra il reo e i soggetti lesi dal reato (la vittima, i familiari di quest'ultima e dell'autore, gli enti e associazioni che rappresentano gli interessi offesi); il riavvicinamento, a sua volta, si basa sulla fiducia nel pentimento e nella volontà riconciliativa e riparativa dell'autore del reato. Verticalmente, lo Stato, comunque presente attraverso la figura del mediatore, «assume un ruolo di controllo ma non di supremazia nei confronti degli altri soggetti»[11], e, in conseguenza di questo percorso fiduciario, attenua il rigore sanzionatorio della condanna, o non persegue il reato tramite la remissione tacita della querela, sulla base della fiducia che lo stesso ripone nella positiva presa di coscienza dell'autore del proprio agito dannoso.

La strada per un sistema pienamente fiduciario è ancora lunga e, forse, non del tutto praticabile a causa di diverse zavorre culturali; ciononostante, le innovazioni legislative hanno comunque iniziato a far mutare il suo volto. Oggi possiamo dire di essere innanzi a un sistema penale sfiduciario, ma fiducioso: sfiduciario, perché il ricorso alla forza alla violenza punitiva e sfiduciaria rimane per ovvie ragioni lo strumento prediletto di attuazione del diritto penale; ma fiducioso, perché lo Stato apre le porte alla fiducia e alla riconciliazione sociale come strumento di giustizia.

[7] G. L. Gatta, *La giustizia riparativa: una sfida del nostro tempo*, in *Sistema Penale*, X, 2024, p. 1.

[8] T. Greco, *È possibile un diritto penale fiduciario?*, cit., p. 72.

[9] F. Cingari, *La giustizia riparativa nella riforma Cartabia*, in *Sistema Penale*, 2023, p. 9.

[10] Relazione dei Lavori del Tavolo 13 degli Stati Generali dell'Esecuzione Penale, All. 3, p. 3.

[11] F. Cingari, *La giustizia riparativa*, cit., p. 6.

ISSN 0039-4130

La fiducia nel diritto internazionale e dell'Unione Europea: inquadramento e strumenti

Vincenzo Mignano

1. *Introduzione*

La definizione del concetto di fiducia nell'ambito del diritto internazionale e del diritto dell'Unione Europea (UE) costituisce un tema complesso e di non facile sintetizzazione. Esso, di fatto, implica un rapporto dalle sfumature più generali che attiene alla relazione che sussiste tra la fiducia stessa e il diritto latamente considerato, nonché agli effetti che da tale relazione discendono.

Muovendo da tale presupposto, la dottrina ha elaborato diverse definizioni di fiducia, ponendo l'attenzione sulla funzione che la stessa svolge. Tra queste, il concetto in esame ha assunto i connotati sia di «fatto-valore» [1] in assenza del quale non può essere costituito un ordine sociale, sia di «bene» [2] sostenuto da una logica di reciprocità, nonché prodotto a vari livelli sociali interconnessi. Nella prima prospettiva, che appare qui condivisibile, la fiducia viene concepita quale *prius* ontologico del diritto, poiché legato alla sua natura relazionale; un fatto, prima ancora che un valore, attraverso cui i consociati instaurano una relazione fiduciaria nell'ottica di stabilire un ordinamento giuridico [3], veicolare l'interpretazione del diritto e garantire la tutela dei diritti fondamentali.

Tale interpretazione, seppur con alcuni chiarimenti che verranno analizzati in seguito, trova conferma nel contesto del diritto internazio-

[1] M. Luciano, *Verso un diritto alla fiducia?*, in *Teoria e Storia del Diritto Privato*, XV, 2022, p. 5.

[2] L. Roniger, *La fiducia nelle società moderne. Un approccio comparativo*, Rubbettino, Soveria Mannelli 1992, pp. 18-21.

[3] T. Greco, *La legge della fiducia. Alle radici del diritto*, Laterza, Bari-Roma 2021, p. 99.

nale e del diritto dell'UE. Alla luce di ciò, il presente contributo mira ad esaminare il concetto di fiducia sotto un duplice profilo di indagine: il primo attiene all'inquadramento della natura che definisce la fiducia negli ordinamenti considerati; il secondo, strettamente connesso al precedente, concerne alcuni strumenti normativi attraverso cui tale concetto trova attuazione.

L'analisi delle questioni sopra elencate consentirà di addivenire alla conclusione secondo cui la fiducia negli ordinamenti giuridici oggetto di esame appare di difficile definizione, soprattutto in relazione alla concezione sia "fiduciaria" che "non fiduciaria" che alcuni degli strumenti che verranno considerati hanno assunto negli anni.

2. *La fiducia nel diritto internazionale e nel diritto dell'UE: la natura*

Inquadrare la natura della fiducia nel diritto internazionale e nel diritto dell'UE, nel tentativo di fornirne una definizione compiuta e univoca, rappresenta un processo complesso. Quanto appena affermato trova una giustificazione nel dinamismo che ha caratterizzato l'evoluzione di tale concetto, spingendo la dottrina giuridica - e non solo - ad interrogarsi di volta in volta sul ruolo che la fiducia ha esercitato nel settore o nella materia in cui essa ha operato. In relazione al periodo storico di riferimento, inoltre, l'interesse per un'indagine di dettaglio sul rapporto fiducia-diritto si è rinnovato, come nel caso del conflitto russo-ucraino.

Entrando nel merito della questione concernente la natura della fiducia negli ordinamenti giuridici in esame, si può affermare in via preliminare come questa sia stata qualificata, seppur con concezioni diverse, quale elemento centrale della teoria delle relazioni internazionali, svolgendo un ruolo significativo nella diplomazia, nella cooperazione tra Stati, nonché nella creazione di grandi istituzioni multilaterali [4].

Posta in altri termini, nell'ambito del diritto internazionale la fiducia diviene il primo luogo di riconoscimento dell'altro, di un'altra entità statale con cui negoziare e concordare obblighi internazionali od operare all'interno di organizzazioni internazionali [5], al fine di persegui-

[4] B. C. Rathbun, *Trust in International Relations*, in E. M. Uslaner (edited by), *The Oxford Handbook of Social and Political Trust*, Oxford University Press, Oxford 2017, pp. 687-706.

[5] Con riferimento, ad esempio, all'Organizzazione delle Nazioni Unite (ONU), si veda B. Torgler, *Trust in international organizations: An empirical investigation focusing on the United Nations*, in *The Review of International Organizations*, III, 1, 2008, pp. 65-93.

ISSN 0039-4130

re obiettivi comuni e di ottenere risultati vantaggiosi per tutti. Un tale processo richiede un vero e proprio affidamento, non solo propedeutico alla conclusione dell'accordo, ma che guardi altresì al futuro, perché presuppone che vi sia concordanza anche nell'esecuzione dell'accordo stesso.

Non manca, tuttavia, chi ritiene che la fiducia nelle relazioni internazionali sia equiparata alla disponibilità ad assumersi rischi sul comportamento degli altri, nella convinzione che le motivazioni e le intenzioni degli Stati investiti di reciproca fiducia protenderanno verso la realizzazione di obiettivi comuni, con conseguente abbandono delle logiche e degli interessi di matrice prettamente nazionale [6].

Con specifico riguardo al diritto dell'UE, la concezione della fiducia sostenuta dalla logica della reciprocità risulta accresciuta e assume una rilevanza maggiore per due ragioni. In primo luogo, a differenza di quanto accade nelle altre organizzazioni internazionali connotate da un forte carattere intergovernativo nel quale gli Stati mantengono tutte le proprie prerogative sovrane, nell'UE si realizza una reale cessione di sovranità in relazione a determinate materie. L'art. 5 del Trattato sull'Unione Europea (TUE), infatti, disciplina il principio di attribuzione, alla luce del quale «l'Unione agisce esclusivamente nei limiti delle competenze che le sono attribuite dagli Stati membri nei trattati per realizzare gli obiettivi da questi stabiliti». Ciò implica che, contrariamente a quanto si verifica nel quadro internazionale sopra descritto, il concetto di fiducia nel diritto dell'UE si rafforza, assumendo una triplice dimensione. In particolare, oltre a riferirsi al piano della fiducia tra Stati membri, che hanno ceduto la propria sovranità all'UE, esso rileva altresì nel rapporto tra Stati membri-UE, così come in quello tra cittadini dell'UE e le sue istituzioni.

La seconda ragione in virtù della quale la fiducia reciproca appare maggiormente pervasiva nel contesto dell'UE è legata all'evoluzione che tale concetto ha avuto per effetto della giurisprudenza della Corte di Giustizia dell'Unione Europea (CGUE). Nel dettaglio, dopo una prima fase in cui la CGUE ha qualificato la fiducia reciproca quale «principio più generale tra le autorità degli Stati membri» [7] senza, di fatto, identificarne presupposti, contenuto ed effetti, si è assistito ad un cambio di tendenza definitorio attraverso il parere sull'accordo di

[6] Sul punto, A. F. Hoffman, *A Conceptualization of Trust in International Relations*, in *European Journal of International Relations*, VIII, 3, 2002, pp. 375-401.

[7] Corte di Giustizia dell'UE, causa 25/88, *Procedimento penale contro Esther Renée Wurmser, vedova Bouchara, e società Norlaine*, sentenza dell'11 maggio 1989, par. 18.

ISSN 0039-4130

adesione dell'UE alla Convenzione europea per la salvaguardia dei Diritti dell'Uomo e delle Libertà fondamentali (CEDU) reso il 18 dicembre 2014. In tale occasione, la CGUE ha espressamente inquadrato il principio della fiducia reciproca quale obbligo che impone agli Stati membri di ritenere, tranne in circostanze eccezionali, che tutti gli altri Paesi UE rispettano il diritto dell'UE e, più in particolare, i diritti fondamentali riconosciuti da quest'ultimo [8].

Come si evince da quanto precede, quindi, sin dalle sue prime pronunce in merito, la CGUE ha riconosciuto alla fiducia reciproca la natura di principio capace di veicolare l'interpretazione delle disposizioni dei Trattati dell'UE, così come del sistema di obblighi che da queste derivano. Quanto detto è emerso soprattutto con riferimento alla giurisprudenza relativa alle libertà di circolazione, nella quale la fiducia reciproca è stata inquadrata sia come presupposto logico del principio del mutuo riconoscimento [9], sia quale criterio che limita la capacità di uno Stato membro di ostacolare la libera circolazione di beni e persone [10].

In virtù delle ragioni sin qui esposte, non sorprende che parte della dottrina abbia proposto una rilettura dei possibili significati e delle funzioni che la nozione di fiducia reciproca assume nel processo decisionale dell'UE [11].

3. *La fiducia nel diritto internazionale e nel diritto dell'UE: gli strumenti*

L'inquadramento della natura che la fiducia assume nel quadro giuridico internazionale e dell'UE, per come sopra ricostruito, costituisce il presupposto fondamentale per analizzare il secondo profilo oggetto della presente indagine, concernente gli strumenti tramite i quali la fiducia reciproca trova espressione. Si precisa preliminarmente che l'esame che segue non pretende di vagliare in modo esaustivo tutte le fattispecie e le ipotesi di riferimento, quanto piuttosto di fornire un elenco esemplificativo nel tentativo di sviluppare alcune riflessioni sulle modalità in cui la fiducia si articola in relazione agli strumenti considerati.

[8] Corte di Giustizia dell'UE, parere 2/13, *Adesione dell'Unione alla CEDU*, parere del 18 dicembre 2014, par. 191.

[9] Corte di Giustizia dell'UE, causa 25/88, *cit.*, par. 20.

[10] Corte di Giustizia dell'UE, causa 46/76, *W. J. G. Bauhuis contro Stato olandese*, sentenza del 25 gennaio 1977, parr. 37-40.

[11] Sul punto, M. Weller, *Mutual trust: in search of the future of European Union private international law*, in *Journal of Private International Law*, XI, 1, 2015, pp. 64-102.

ISSN 0039-4130

Con specifico riguardo al diritto internazionale, meritano particolare menzione la consuetudine e i Trattati internazionali. La prima suscita, di fatto, maggior interesse in rapporto alla nozione di fiducia, poiché poggia sull'idea che esista una condotta su cui tutti concordano e che da tutti gli Stati è ritenuta vincolante. Pertanto, la consuetudine viene a costituirsi attraverso l'adozione costante e uniforme degli Stati di un determinato comportamento, accompagnata dalla convinzione che lo stesso sia giuridicamente obbligatorio [12]. E nel contesto della presente disamina, ciò che rileva della consuetudine è la sua natura di fonte del diritto internazionale non scritta, a conferma dell'ampia portata attraverso cui la fiducia che intercorre tra gli Stati si manifesta.

Il secondo strumento nel quale la fiducia trova espressione nell'ambito del diritto internazionale è rappresentato dalla categoria dei Trattati internazionali. Come già anticipato in precedenza, per il loro tramite gli Stati veicolano la propria volontà di cooperare per il raggiungimento di obiettivi comuni, secondo una logica che guarda all'affidamento reciproco sia in fase di redazione che in fase di esecuzione dell'accordo [13].

Tale lettura positiva dei Trattati internazionali quali strumenti di fiducia reciproca tra Stati è stata, tuttavia, ribaltata in occasione degli *shock* finanziari – Grande Recessione e crisi dei debiti sovrani – che hanno scosso l'UE tra il 2008 e il 2012. Nel dettaglio, alla logica di reciprocità sottesa alla fiducia è subentrata una visione "non fiduciaria" incentrata sulla convinzione degli Stati membri che dispongono di un'economia più solida che i Paesi UE in gravi difficoltà finanziarie non potessero adottare politiche di bilancio virtuose senza la previsione di appositi vincoli giuridici. L'esempio tipico è rappresentato dal quadro giuridico istituito dal Meccanismo Europeo di Stabilità (MES), un'organizzazione internazionale introdotta da un apposito Trattato e prevista nel tentativo di fornire assistenza finanziaria agli Stati membri maggiormente colpiti dalle crisi. La logica "non fiduciaria" sopra richiamata si è concretizzata, nel dettaglio, nella subordinazione dell'erogazione delle risorse finanziarie ad una stringente condizionalità economica che si è tradotta in una vera e propria intrusione nella sovranità degli Stati beneficiari, come la Grecia [14].

[12] B. Conforti-M. Iovine, *Diritto internazionale*, Editoriale Scientifica, Napoli 2023, pp. 41-53.

[13] *Ibid.*, p. 74.

[14] P. De Sena-M. Starita, *Fra stato di necessità ed (illecito) intervento economico: il terzo "Bail Out" della Grecia*, in *Quaderni di SIDIBlog*, II, 2015, pp. 119-133.

ISSN 0039-4130

Con riferimento al quadro giuridico dell'UE, oltre ai rapporti con il principio del mutuo riconoscimento sopra richiamati, la fiducia reciproca risulta correlata al principio di leale cooperazione, sancito all'art. 4 TUE. Basti pensare che tale principio sancisce - tra le altre cose - l'obbligo, per gli Stati membri e l'UE, di rispettarsi e assistersi reciprocamente nell'adempimento dei compiti derivanti dai Trattati, ponendo su un piano di uguaglianza i Paesi UE. In aggiunta, i rapporti tra l'UE e i suoi partner sono in qualche modo regolati dalla logica della fiducia. Ciò avviene allorquando l'UE utilizza con i suoi partner la politica di condizionalità, che subordina la concessione di benefici al raggiungimento di determinati obiettivi, ossia alla realizzazione di specifiche condotte. In questa prospettiva, la determinazione dei contenuti dell'accordo è un momento fiduciario, non di sfiducia.

4. *Il rapporto tra fiducia e sanzione*

A margine dell'analisi, merita un breve cenno la relazione che sussiste nella realtà giuridica internazionale e dell'UE tra fiducia e sanzione. Sebbene quest'ultima possa generalmente alludere ad un'assenza di fiducia o alla conferma che quella fiducia verrà tradita, a ben vedere il sistema della sanzione poggia sulla convinzione che quella sanzione verrà applicata; convinzione, questa, che rappresenta un vero e proprio atto di fiducia.

Si pensi, a titolo di esempio, al potere discrezionale attribuito alla Commissione europea di proporre ricorso per infrazione dinanzi alla CGUE contro uno Stato membro che non abbia rispettato gli obblighi che discendono dai Trattati. Se, da un lato, tale discrezionalità potrebbe essere giustificata dall'aspettativa degli Stati membri che l'istituzione sopra menzionata non agirà secondo il mero arbitrio, dall'altro va precisato che gli Stati medesimi hanno scelto di preservarsi dall'azione certa che sarebbe derivata nel caso in cui la Commissione europea avesse avuto l'obbligo e non la facoltà di ricorrere dinanzi alla CGUE in caso di violazione.

Peraltro questo concetto è ancora più sentito nel diritto internazionale, posto che in generale per sottoporsi alla giurisdizione in tale ordinamento è necessario un atto di volontà degli Stati, come accade nel caso delle decisioni della Corte Internazionale di Giustizia (CIG). E ciò si giustifica solo attraverso la fiducia nell'esistenza di un comune ordine internazionale, che va tutelato e preservato. Proprio in tale pro-

ISSN 0039-4130

spettiva, il conflitto russo-ucraino ha offerto l'opportunità di riflettere sull'esistenza di una rinnovata fiducia nei confronti dei Tribunali internazionali [15].

5. *Conclusioni*

Nonostante la fiducia per come sopra descritta conduca ad esiti positivi nel diritto internazionale e nel diritto dell'UE, alcuni elementi di criticità permangono. In primo luogo, la fiducia cui si fa primariamente riferimento coinvolge soggetti non assimilabili alle persone fisiche, ossia gli Stati. In quanto tali, essi sono portatori di interessi che, molto spesso, incrinano la piena cooperazione, limitando in tal senso quella fiducia posta alla base dell'ordinamento internazionale e dell'UE e determinandone la crisi.

Sotto altro aspetto, strumenti tipicamente espressione di fiducia – come i Trattati internazionali – finiscono per divenire l'emblema di una totale assenza di fiducia. Basti pensare ai piani di aggiustamento posti in essere durante le crisi finanziarie dello scorso decennio.

In conclusione, la fiducia reciproca costituisce un valore fattuale imprescindibile nella realtà giuridica, poiché, quale presupposto del diritto, contribuisce a plasmare la società e a trasformarla. E questa esigenza si riscontra in maniera preminente nel diritto internazionale e nel diritto dell'UE, vista la maggiore distanza che separa i cittadini degli Stati dalle Istituzioni europee e internazionali.

[15] L. Acconciamessa, *Il conflitto armato in ucraina come catalizzatore di una rinnovata fiducia nei tribunali internazionali? Riflessioni a partire dalle misure provvisorie della Corte internazionale di giustizia*, in *Quaderni di SIDIBlog*, IX, 2022, pp. 19-48.

ISSN 0039-4130

Gli *influencer* tra credibilità, identificazione e trasparenza: linee di ricerca sulla negoziazione della fiducia

Mael Bombaci

Oggi più che mai, la società è caratterizzata da incertezza e imprevedibilità, rendendo impossibile conoscere ogni dettaglio della realtà o anticipare tutte le conseguenze delle proprie azioni. In questo scenario, la fiducia diventa un meccanismo particolarmente utile per ridurre la complessità [1]. Delegando parte della propria responsabilità a terzi – che siano persone, istituzioni o sistemi – l'individuo riesce a ridurre lo sforzo richiesto, sia a livello cognitivo che comportamentale. L'espressione "Mi fido, quindi fai tu" sembra sintetizzare in maniera particolarmente efficace il meccanismo fiduciario attraverso cui un individuo delega, almeno in parte, a un altro il compito di filtrare e interpretare la realtà al proprio posto [2]. Resta tuttavia fondamentale interrogarsi sui limiti di questo meccanismo e, in particolare, sulle sue implicazioni nei processi di partecipazione collettiva. Per esplorare tale questione, di seguito saranno presentati i principali risultati di tre ricerche, le quali analizzano il rapporto tra *influencer* e *audiences*, approfondendo tre dimensioni strettamente legate al tema della fiducia: credibilità, identificazione e autenticità. Il filo conduttore di tutte le ricerche è, quindi, la figura dell'*influencer*, definita da Abidin [3] come individui che creano contenuti sui *social media* legati alla loro vita quotidiana e accumulano un

[1] N. Luhmann, *Trust and power: Two works by Niklas Luhmann, John Wiley & Sons*, 1979.

[2] N. Luhmann, *Trust and power*, Polity Press, 2002.

[3] C. Abidin, *Internet celebrity: Understanding fame online*, Bristol University Press, 2018.

ampio seguito, influenzando le opinioni e le decisioni dei loro *follower*. Tuttavia, la figura dell'*influencer* è in continua evoluzione e non si limita alla semplice promozione commerciale. Attraverso le tre ricerche, si esploreranno sfumature differenti di queste figure, mettendo in luce aspetti più complessi e articolati.

1. *Credibilità: quando fidarsi si confina a una dimensione individuale*

Sempre più *influencer* stanno ampliando il loro ambito di azione, affrontando tematiche di interesse pubblico, sfumando così il confine tra attivismo digitale e *influence culture*[4]. Un caso esemplare è quello dei *green influencers*, figure che promuovono contenuti legati alla sostenibilità e sensibilizzano il pubblico sulla crisi climatica[5]. La ricerca, condotta attraverso l'analisi del contenuto di un *dataset* di video su *TikTok*, si proponeva di esplorare, tra le altre cose, l'interazione degli utenti con le *affordances* offerte dalla piattaforma. Le *affordances*, intese come le possibilità messe a disposizione dalle piattaforme, sono negoziate dagli individui all'interno delle pratiche sociali. Sebbene non determinino direttamente il comportamento umano, esse offrono una serie di opzioni che gli utenti sfruttano in base alla loro autonomia[6]. Un aspetto interessante emerso è che le interazioni con questi contenuti avvengono principalmente in una dimensione individuale piuttosto che collettiva. In particolare, il numero di salvataggi dei post risulta nettamente superiore a quello delle condivisioni, suggerendo che gli utenti riconoscono la rilevanza del tema trattato, ma non sentano la necessità di diffonderlo ulteriormente con la propria rete. Qui emerge una tensione tra la fiducia riposta negli *influencer* e la capacità di trasformare questa fiducia in azione collettiva. Da un lato, il pubblico affida agli *influencer* la selezione e semplificazione delle informazioni, riducendo la complessità del proprio processo decisionale. Dall'altro, questo affidamento potrebbe limitare la partecipazione attiva, confinando l'interesse per determinate tematiche a una sfera individuale e

[4] M. Colucci - M. Pedroni, *Got to be real: An investigation into the co-fabrication of authenticity by fashion companies and digital influencers*, Journal of Consumer Culture, 2022. https://doi.org/10.1177/14695405211033665

[5] M. Pittman - A. Abell, *More Trust in Fewer Followers: Diverging Effects of Popularity Metrics and Green Orientation Social Media Influencers*, Journal of Interactive Marketing, LVI, 1, 2021, pp. 70-82. https://doi.org/10.1016/j.intmar.2021.05.002

[6] I. Hutchby, *Technologies, texts and affordances*, Sociology, XXXII, 2, 2001, pp. 441-456. https://doi.org/10.1177/S0038038501000229

ISSN 0039-4130

privata. In un mondo caratterizzato da un eccesso di informazioni e possibilità, la fiducia consente di filtrare ciò che è rilevante, riducendo così il carico cognitivo dell'individuo, ma potenzialmente disincentivando l'impegno collettivo.

2. *Identificazione: il desiderio come motore della fiducia*

Un secondo aspetto legato al ruolo dell'*influencer* riguarda il processo di identificazione, che sembra fungere da leva nel consolidamento della fiducia. La ricerca, condotta in collaborazione con l'azienda Sony Interactive Entertainment, si propone di comprendere il ruolo delle narrazioni degli *influencer* nella costruzione della *coolness* rispetto ai videogiochi. Il concetto di *coolness* include valutazioni multidimensionali degli utenti su quattro motivazioni principali: utilità, attrattività, appeal subculturale e originalità[7]. La ricerca nasce dalla consapevolezza che le narrazioni non sono limitate al videogioco stesso, ma si estendono anche attorno ad esso, riflettendo le rappresentazioni legate alla sua fruizione e condivisione. Per rispondere a tali interrogativi, si è adottato un disegno di ricerca basato sull'approccio metodologico della *Grounded Theory*[8], che consente di costruire un modello teorico a partire dai dati raccolti sul campo. In particolare, la metodologia scelta per raccogliere informazioni è quella del *focus group*, che, simulando una discussione tra pari, ricrea una dinamica simile al processo ordinario di formazione delle opinioni. I *focus group*, condotti in tre diverse regioni d'Italia, hanno mostrato che spesso la visione di un video di *gameplay* pubblicato da un *influencer* genera nei giocatori un senso di immedesimazione che va oltre la semplice attrattiva del prodotto. Si innesca un meccanismo di costruzione del valore sociale, in cui *l'influencer* non si limita a suggerire un acquisto, ma contribuisce a legittimare o meno determinate pratiche culturali, costruendo mete più o meno desiderabili. In questo modo, l'identificazione con *l'influencer* può incidere significativamente sulla costruzione del desiderio e sulla percezione sociale di determinate pratiche culturali. In un contesto come quello dei vide-

[7] S. S. Sundar - D. J. Tamul - M. Wu, *Capturing "cool": Measures for assessing coolness of technological products*, International Journal of Human-Computer Studies, n.d. Retrieved from https://www.elsevier.com/locate/ijhcs.

[8] B. G. Glaser - A. L. Strauss, *The discovery of grounded theory: Strategies for qualitative research*, Aldine Publishing, 1967.

ISSN 0039-4130

ogiochi, caratterizzato da alta competitività, la fiducia riveste un ruolo centrale, e l'identificazione con *l'influencer* diventa un elemento chiave per la diffusione di norme e valori. Il meccanismo fiduciario, in questo caso, non si limita a una semplice delega informativa, ma si estende alla costruzione di significati condivisi, che influenzano le opinioni e i comportamenti degli individui, contribuendo a plasmare la percezione collettiva di pratiche culturali e tendenze sociali.

3. *Autenticità: tra fiducia e spettatorialità*

La terza ricerca si concentra su un fenomeno che mette in discussione il potenziale partecipativo del pubblico: i *reaction video (RV)*. Questo genere di contenuto, in forte crescita negli ultimi anni, solleva interrogativi cruciali sul rapporto tra autenticità, fiducia e partecipazione. I *RV* si basano sulla condivisione pubblica di reazioni, spesso enfatizzate, a contenuti di forte impatto emotivo. Studi recenti [9] suggeriscono che, indipendentemente dall'autenticità delle reazioni mostrate, questi video vengono utilizzati sia per prendere le distanze da un contenuto, sia per posizionarsi rispetto ad esso. Un'analisi dei *RV* dal punto di vista sociale, culturale e comunicativo permette di comprenderne il significato e le modalità di fruizione e interpretazione da parte delle *audiences*, evidenziandone la natura meta-spettatoriale. Gli utenti non si limitano a guardare un contenuto, ma osservano qualcuno che lo guarda, in un processo di rimediazione [10] che ridefinisce il concetto stesso di partecipazione. Tuttavia, sebbene la cultura convergente [11]suggerisca che questo tipo di interazione possa incentivare nuove forme di *engagement*, i dati emersi dai *focus group*, rivelano un aspetto critico: l'interazione avviene più frequentemente in forma individuale che collet-

[9] H. Warren-Crow, *Screaming like a girl: viral video and the work of reaction, Feminist Media Studies*, XVI, 6, 2016, pp. 1113-1117.

S. Antonioni & M. Farci, *Post-Millennial Spectatorship and Horror Films: The Case of It*, Comunicazioni sociali, 2, 2018, pp. 180-192.

B. McDaniel, *Popular music reaction videos: Reactivity, creator labor, and the performance of listening online,* New Media & Society, XXIII, 6, 2020, pp. 1624-1641.

[10] R Grusin,. (2004). *Remediation in the Late Age of Early Cinem*a. In A. Gaudreault - C. Russell - P. Véronneau (Eds.) *Early Cinema: Technology and Apparatus* (pp- 343-360). Payot Lausanne. Trad. it. di A. Maiello, *La rimediazione nella tarda epoca del cinema delle origini*, in A. Maiello (a cura di), *Richard Grusin. Radical Mediation. Cinema, estetica e tecnologie digitali*, Luigi Pellegrini Editore, 2017.

[11] H. Jenkins & M. Deuze, *Convergence culture,* Convergence, XIV, 1, 2008, pp. 5-12.

ISSN 0039-4130

tiva. In questo contesto, il rapporto di fiducia tra *creator* e pubblico si configura come una relazione uno-a-uno, più che come un'esperienza comunitaria. La fiducia riposta nell'*influencer* porta gli utenti a identificarsi con le sue emozioni e reazioni, ma questo processo non sembra tradursi sempre in una partecipazione attiva. Il rischio è che l'autenticità percepita dell'*influencer* rafforzi una modalità di partecipazione passiva, in cui il pubblico si limita a consumare le emozioni altrui senza tradursi in un coinvolgimento diretto.

4. *Conclusioni: la fiducia come delega o come leva di partecipazione?*

I risultati finora presentati permettono di riflettere sulla fiducia come elemento centrale nella relazione tra *influencer* e *audiences*, evidenziando come concetti quali credibilità, identificazione e autenticità siano costantemente negoziati dalle *audiences* stesse. Sebbene gli *influencer* possano indirizzare la percezione di questi aspetti, è il pubblico che, attraverso le proprie interazioni, attribuisce significato e valore, decidendo se accettare o rifiutare la fiducia accordata anche in base alle dimensioni citate. Se si considera il meccanismo esplicitato dall'espressione "Mi fido, quindi fai tu" in relazione agli *influencer*, diventa necessario riflettere su come la fiducia possa, da un lato, favorire nuove forme di partecipazione e, dall'altro, contribuire a dinamiche di passività, influenzando il modo in cui il pubblico si rapporta ai contenuti e alle tematiche proposte. Questo processo solleva interrogativi rilevanti sulle implicazioni della fiducia nella partecipazione collettiva, che meriterebbero di essere approfonditi ulteriormente, mettendo in luce la possibilità di passare da una sfera individuale a una dimensione collettiva, in cui la fiducia può trasformarsi in un elemento dinamico capace di generare nuove forme di impegno o, al contrario, consolidare il distacco.

ISSN 0039-4130

Il tempo della permanenza

Riflessioni pedagogiche sulla fiducia come "sustanza di cose sperate"

Giovanna Arigliani

«Fede è sustanza di cose sperate
e argomento de le non parventi,
e questa pare a me sua quiditate».
(Dante, Par. XXIV, 64-66)

Vorrei cominciare dal porre alcune domande; cos'è fiducia e cosa porta con sé questa piccola ma grande parola? Un termine che va inteso come un complesso contenitore composto da più elementi e che tenterò di sollevare almeno in parte in questa sede.

Percorrendo alcune possibili tracce, possiamo collocare la fiducia all'interno di una sorta di cuore segreto che anima molte dimensioni della vita collettiva [1], un concetto e in un certo senso un problema che si esprime nei termini dell'imprescindibilità e che esige oggi forse più di ieri, di una riflessività profonda anche e in forza di una società contemporanea che sembra sollecitare l'impulso di una diffidenza primaria che sospinge all'arroccamento nell'individualismo, nell'isolamento e nella solitudine in cui il singolo soggetto diviene il «miglior partner di se stesso e la più affidabile delle sue compagnie» [2], accrescendo così una problematica sfiducia verso l'altro il cui sovente "avvertimento" lascia passare il messaggio che "fidarsi è bene, ma non fidarsi è meglio".

Questo, un linguaggio sempre più collettivo e che sembra permeare al punto da essere incorporato in una forse semplicistica tensione volta al *disimpegno* che ci allontana dalla *civitas* [3] a cui invece dovremmo poter tendere.

[1] E. Resta, *Le regole della fiducia*, Laterza, Bari 2009.
[2] C. Widmann, *f come fiducia*, Cittadella, Assisi 2012, pp. 10-13.
[3] Sul concetto di *civitas* ricordiamo come essere "cive", non significhi la sola partecipazione al processo di civilizzazione e al progresso della civiltà, ma designa un'appartenenza alla *civitas* nei termini di un'agentività della collettività cittadina e quindi di pieno sodalizio umano. (Ibid.).

Si parla infatti di "crisi della fiducia" e da qui giunge necessario tornare a parlare dell'insita *natura* fondativa di cui la fiducia è portatrice, così come gli effetti negli ambiti della relazionalità [4] prodotti dal suo contraltare quando parliamo del sentimento di sfiducia.

Comprendiamo quindi come il termine "fiducia" sia molto più di una parola, una chiave sensibile, oltre che concettuale e che coinvolge tutti nella percezione delle difficoltà del presente [5].

Il punto è che i sentimenti e le spinte che la fiducia mette in gioco non sono mai univoci e monologanti, essa si muove in una continua sospensione tra la dimensione spontanea, irrazionale e talora persino fideistica.

Lo sguardo offerto dalla pedagogia sul tema della fiducia muove le sue riflessioni sulla necessità di uscire dalla dimensione dell'aspettativa del *dare-avere*, difatti, riconoscere l'*alterità* significa poter accogliere il proprio sé attraverso l'incontro e la condivisione con l'*altro* che è traccia e significato di ciò che siamo [6].

Parlare di fiducia non può includere la dimensione dell'aspettativa nei termini di un "investimento fiduciario" [7] con l'altro, ma una relazionalità che si fonda su un tipo di riconoscimento e di accoglienza dell'alterità che avviene nella dimensione della permanenza temporale fino a che questa possa giungere ed assumere i connotati della familiarità: avere e dare fiducia implica in fondo un atto di fede, significa affidarsi all'altro in un incessante nutrimento a fondo perduto.

Ma tornando alle domande, la fiducia può essere intesa come un'espressione di rischiosa cecità? Secondo Max Weber la cecità rimuove il rischio se la si comprende come scelta sentita dall'attore, sebbene que-

[4] A tal proposito ricordiamo l'attenzione posta da E. Erikson in *Childood and Society* (1950) sulle capacità psicosociali che lo studioso suddivide in otto stadi, la cui soluzione positiva per ciascuna *crisi* porta all'acquisizione di una virtù; la fiducia rientra tra queste e costituisce una delle capacità essenziali per affrontare le sfide della vita, oltre ad essere condizione necessaria che consente di instaurare relazioni significative: una capacità che incide sullo sviluppo personale e sulla costruzione del sé.

[5] C. Widmann, *f come fiducia*, cit., pp. 10-13.

[6] D'altra parte Aristotele sull'incompletezza dell'umanità sin dalla nascita, mise ben in evidenza la necessità dell'alterità e della relazione. È come se la natura non avesse funzionato a dovere e non avesse terminato la sua opera gettandoci nell'arena della vita deboli e disarmati più di qualsiasi altro mammifero: «nasciamo senza saper camminare, e ci impiegheremo mediamente un anno a imparare […], per cui compensiamo la nostra deficienza naturale tramite la cultura: la famiglia, l'aiuto reciproco, l'educazione. Grazie alla nostra predisposizione al rapporto umano, ci adoperiamo a completare quell'opera che la natura ha lasciato incompiuta, a conquistare quella fiducia che la natura non ci ha dato» (C. Pépin, *La fiducia in se stessi. Una filosofia*, La nave di Teseo, Milano 2018, p. 18) e di cui insitamente abbiamo bisogno.

[7] Per usare un linguaggio provocatoriamente economista.

ISSN 0039-4130

sta sia sempre pronta a incorporare il rischio della delusione [8], ed ecco quindi che la fiducia nella sua dimensione di imprevedibilità richiama a sé la necessità della fede verso l'altro nonostante questa non garantisca un risultato certo.

Sappiamo inoltre che una «società senza fiducia è una società senza ossatura» [9] i cui termini pari e contrari risiedono nella diffidenza che può degradare fino al sospetto, al pregiudizio e alla paranoia [10] in cui il soggetto si percepisce al centro di «un'attenzione astiosa, di macchinazioni sotterranee, di una mostruosa e assurda cattiveria» [11] in un mondo persecutorio. La diffidenza rappresenta quindi una vera e propria tomba della fiducia pari ad un contraltare infestante che indebolisce e sgretola il mondo sociale.

La politica della diffidenza ' incompatibile con la pulsione sociale ' produce separazione e alimenta la distanza [12], ma in virtù del fatto che come noto l'uomo è insufficiente a se stesso ed ha bisogno «dell'altro e di molti altri» [13], il sentiero dell'alienità entro le categorie della circospezione e del sospetto al mondo delle relazioni è insufficiente a se stessa ed è insufficiente alla nostra stessa evoluzione.

Se è vero che la diffidenza originaria ha consentito la sopravvivenza all'umanità, oggi risulta una modalità certamente inefficace e inattuale. Oggigiorno, diversamente, la fiducia è un'esigenza evolutiva cognitiva che necessità di maturità affettiva indispensabile per il superamento di arroccamenti narcisistici, angosce paranoidi e disfattismi. La fiducia è linfa vitale per la costellazione psichica della personalità di ciascuno e per la società; impone all'Io atteggiamenti coerenti e si articola attorno a modalità precoci [14] come il sorriso, la gratificazione, la gratitudine,

[8] M. Weber, *Economia e società. La città*, Donzelli Roma 2016.

[9] M. Marzano, *Avere fiducia*, Mondadori, Milano 2014, p. 219.

[10] Nella psicologia anglosassone «*paranoid* è sinonimo di *persecutory* e designa un assetto psichico che filtra la realtà attraverso la griglia di un'esasperata diffidenza» (C. Widmann, *f come fiducia*, cit., p. 29).

[11] Ibid., p. 28.

[12] Continuando sulla scia delle riflessioni Ericksoniane nell'ambito di uno sviluppo psicosociale sano, lo psicanalista tedesco sostiene che la fiducia sviluppata durante l'infanzia costituisce una base necessaria per una crescita psicosociale sana. Nel primo stadio detto "fiducia vs. sfiducia", il bambino impara a comprendere se il mondo è un luogo sicuro o minaccioso. Una fase critica di assoluta incidenza per il benessere emotivo del bambino nell'adultità. (Erikson, E.H., 1950. Childhood and Society).

[13] Ibid., p. 44.

[14] E ancora, sugli effetti che l'ambiente ha sulla costruzione dello stile di attaccamento e di come questo influisca nell'ambito delle relazioni, sappiamo che all'interno della relazione genitore-figlio se la figura genitoriale rappresenta una figura stabile e favorisce un clima affettuoso e accogliente si improntera nel neonato «un'elementare fiducia di base"

ISSN 0039-4130

condizioni che consentono alla persona di riconfigurare la "pericolosità dell'altro" muovendosi invece all'interno dello spazio della curiosità e dell'attrazione. Comprendiamo quindi che la diffidenza originaria debba essere sostituita da una fiducia originaria fondata dal *logos* e quindi in una intrinseca fede verso l'altro in cui la fede si riappropria della *sustanza di cose sperate* [15].

E ancora, prendendo in prestito una metafora utilizzata dal sociologo Sergio Sorgi, l'ambito della fiducia può essere evocato nei termini di «un pavimento di vetro che ci consente di camminare [*ma*] senza cadere» [16].

Come già detto nelle premesse di questo contributo, si tratta di una "parola contenitrice" che diviene sistema entro cui l'insicurezza esterna viene sostituita da una relativa sicurezza interna in grado di aiutarci a tollerare le incertezze ambientali e relazionali.

Giungendo ora sul versante relazionale affettivo, nella condivisione della nostra esistenza con persone che amiamo e nella convinzione che insieme si sia più forti e protetti e che si possa vivere per sempre felici e contenti, siamo di fronte ad una e più scelte (come quella di sposarsi, di laurearsi, di adottare o concepire un bambino, divorziare, studiare, metter su famiglia, accudire un animale domestico [17]) che includono una cornice decisionale di fondo che necessita di riporre fiducia verso sé e verso l'altro.

Talora si potrebbe anche dire che la fiducia possa muoversi all'interno di un'analisi più o meno consapevole di "costi e benefici" ma in tal caso torneremmo nel commettere un errore poiché non vi può essere una "realistica" possibilità di quantificare "il grado di fiducia" il che rende irrealistico un qualsivoglia confronto tra "tipi" di fiducie

su cui edificherà ogni futura relazione fiduciaria» (C. Widmann, *f come fiducia cit*., p. 45). Al contrario, se l'infanzia sarà caratterizzata da cure inaffidabili, negligenti o incoerenti, il bambino potrà sviluppare un senso di sfiducia verso gli altri, temendo il rifiuto o il tradimento. Questa insicurezza può manifestarsi nell'adolescenza e nell'età adulta sotto forma di difficoltà a creare legami intimi; paura dell'abbandono; atteggiamenti difensivi; di paura verso l'alterità o tendenze all'isolamento incidendo sullo sviluppo dell'autostima, della sicurezza emotiva e sulle connessioni empatiche. (Erikson, E.H., 1950. Childhood and Society).

[15] Ricordiamo quando Dante in *La divina Commedia* nel canto XXIV del Paradiso, riprendendo la definizione della fede dalla Lettera agli Ebrei (11,1) della Bibbia: «fede è sustanza di cose sperate e argomento de le non parventi, e questa pare a me sua quiditate» (Dante, Par. XXIV, 64-66), spiega la fede nei termini di una "certezza interiore" in grado di dare sostanza a ciò che speriamo offrendo così una dimostrazione delle realtà spirituali invisibili: una definizione che ripone nella fiducia una delle massime espressioni di fede.

[16] S. Sorgi-F. Bertè, *Fiducia sostantivo plurale. Meritarla, curarla e conservarla*, Egea, Milano 2022, p. 5, corsivo aggiunto.

[17] Cfr. ibid.

ISSN 0039-4130

utili a «poterci orientare scientificamente verso l'alternativa più conveniente» [18].

A proposito della pretesa della quantificazione, tra le ossessioni caratterizzanti della nostra contemporaneità, non fa eccezione di incidenza anche nell'ambito delle relazioni, comprese quelle affettive, difatti, sovente parliamo di tempo come termine di qualità e misura di una relazione.

Tuttavia, nell'ambito delle relazioni non si dovrebbe tanto far leva sulla permanenza del passato ma di una permanenza presente che è precondizione di un varco futuribile in grado di svincolarsi da quel tempo inteso come "misura artificiale" e quindi inautentica [19] e che invece trova la sua genuina dimensione nella cura che diventa permanenza e che si declina più che nel passato, nel presente, in quell'"oggi" in cui *scelgo* di permanere.

I bambini in tal senso sono permanenza, essi vivono nel presente, non si preoccupano del futuro e non si interrogano sul passato: incarnano la più profonda essenza della permanenza del tempo.

Poiché l'essere futuro è propriamente il tempo, la questione della quantità del tempo, della durata e del «quando?» [20], non può che rimanere inadeguata, e se è vero che l'orologio ci indica l'ora è anche vero che nessun orologio indica mai il futuro né ha mai indicato il passato, *ergo*, ogni misurazione del tempo significa ricondurre il tempo alla quantità.

Declinando queste coordinate nell'ambito delle relazioni e in particolare a quelle sentimentali, sovente vi è la tendenza di fare del tempo il termine di validità in cui misurare una relazione più o meno solida. Certamente se una coppia sceglie ogni giorno la permanenza della cura, in tal caso siamo innanzi al tempo della permanenza, ma se questa si fa forza solo in virtù di una memoria passata senza presente, allora siamo innanzi ad una relazione connaturata in via esclusiva dal tempo misurato e quindi inautentico. Domandare della quantità del tempo significa lasciarsi completamente assorbire dal prendersi cura di un «che cosa» [21] presente, che è ben diverso dal prendersi cura in un esserci a tempo pieno che diversamente segue le risonanze di un tempo come cura

[18] Ibid., p. 7.

[19] L'orologio che tenta di determinare il tempo quando questo è caratterizzato dall'indeterminazione.

[20] M. Heidegger, *Il concetto di tempo*, Adelphi, Milano 1998, p. 41.

[21] Ibid., p. 42.

ISSN 0039-4130

e quindi del tempo [22] della permanenza, o per dirlo attraverso le parole di Martin Heidegger «l'esserci non è il tempo, ma la temporalità» [23] che risponde al «come» aver cura.

Per giungere alle conclusioni di questo breve contributo, la cui finalità è volta ad una rinnovata riflessione sui significati intrinseci ed estrinseci sul tema della fiducia, vorrei aprire un'ultima suggestione in cui a ragion veduta Aristotele in *Etica Nicomachea,* a proposito di una necessaria etica delle relazioni e proprio in virtù di un'inevitabile vulnerabilità e – come detto – di quella porzione di imprevedibilità insita nella fiducia, ci parla di *alterità* nei termini di una possibilità di rifugio [24] in cui la familiarità e l'amicizia rappresentano dei valori necessari per la crescita di ciascuno e per una «società della concordia» [25] sostenute dai principi insiti nella fiducia, della benevolenza, di impegno e di reciproca responsabilità, che costituiscono la condizione necessaria di una promessa e di una scelta che si compie nel dare «se stessi a un altro e di ricevere l'altro come un altro se stesso» [26], ed è in virtù di questa reciproca appartenenza di valori che la violazione della fiducia corrisponde alla violazione di se stessi.

[22] Un'altra pista interessante sulla interrelazione tra tempo e relazioni è rintracciabile all'interno della cultura maori che intende il tempo come un permeare contiguo e interconnesso tra passato, presente e futuro, un'essenza in cui vi è una profonda connessione tra la terra, gli antenati, la persona presente e la società, una prospettiva che contiene in sé la fiducia nel futuro proprio perché esce dalla tirannide di un tempo inautentico e che contrariamente rafforza le radici culturali curando e tutelando ciò che ancora non è nato. (S. Sorgi-F. Bertè, *Fiducia sostantivo plurale*, cit.).

[23] M. Heidegger, *Il concetto di tempo cit.*, p. 50.

[24] Difatti sappiamo che potersi affidare all'altro a maggior ragione quando si trascorre una fase di vita particolarmente critica può rassicurare e rafforzare il senso di autoefficacia.

[25] Aristotele, *Etica Nicomachea*, 1133b29, in Id, *Etica Nicomachea*, a cura di C. Natali, Laterza, Roma-Bari 2005.

[26] A. Sgobba, *La società della fiducia. Da Platone a whatsApp*, Il Saggiatore, Milano 2020, p. 146.

ISSN 0039-4130

La fiducia nel diritto civile

Giulia Anselmo e Pierfrancesco Minicangeli

1. Nel diritto civile la fiducia, sin dall'età antica, ha rappresentato e tuttora rappresenta uno degli elementi più intricati e poliedrici. La fiducia costituisce un fondamento implicito che regola le relazioni giuridiche, e ne costituisce presupposto per la coesione e il funzionamento ordinato della società.

Infatti, nel diritto privato vi è una forte prevalenza non dell'aspetto sanzionatorio – il quale, tradizionalmente, connota maggiormente il diritto penale –, ma di quello regolatorio dei rapporti tra consociati: in questo contesto, dunque, la fiducia opera come una sorta di "collante invisibile" che permea le interazioni interpersonali, consentendo ai soggetti di confidare nell'adempimento dei reciproci obblighi e nella prevedibilità delle conseguenze normative scaturenti dall'attuazione di determinati comportamenti.

La persistente presenza di una dimensione fiduciaria nell'evoluzione delle regole di diritto privato si può cogliere, altresì, volgendo lo sguardo agli istituti giuridici più antichi: basti pensare, nel diritto romano, alla *fiducia cum amico* – istituto dal quale poi nascerà l'odierno negozio fiduciario (nella tradizione di *common law* meglio noto, seppur con alcune differenze, quale "*trust*") – e alla *fiducia cum creditorem*, atti solenni effettuati *causa fiduciae* con i quali si alienava una determinata *res fiduciae data* da parte di un soggetto, detto fiduciante, ad un altro soggetto, detto fiduciario, in capo al quale sorgeva un obbligo, il quale, il più delle volte, consisteva nel ritrasferimento della cosa [1].

In ogni caso, la fiducia ha permeato e plasmato nel corso dei secoli numerosi istituti, i quali tutt'oggi rivestono un ruolo centrale nel nostro sistema ordinamentale. Ne sono esempi significativi, seppur non esaustivi, il criterio di buona fede, il principio dell'affidamento e il principio di equità.

[1] Tra i più recenti contributi in tema di *fiducia* nel diritto romano, v. M. Milani, *La fiducia in diritto romano. Atti costitutivi, causa, oggetto*, Jovene, Napoli 2022.

La buona fede (artt. 1175 e 1375 c.c.[2]) – tanto nella sua accezione c.d. "oggettiva" quanto in quella c.d. "soggettiva" – rappresenta una delle più emblematiche espressioni della fiducia nel diritto civile, imponendo ai consociati un fondamentale dovere di lealtà e correttezza nell'esecuzione dei rapporti giuridici, in particolar modo contrattuali. La buona fede c.d. "oggettiva", segnatamente, si configura quale criterio di comportamento che impone alle parti di un rapporto giuridico di agire con onestà ed intelligenza, prevenendo, in tal modo, abusi e ingiustificate alterazioni dell'equilibrio contrattuale[3].

Parallelamente, il principio dell'affidamento assume una funzione altrettanto essenziale, traducendosi in uno strumento di tutela della ragionevole aspettativa che i soggetti ripongono nella stabilità nonché nella correttezza delle situazioni giuridiche. Tale principio trova applicazione in molteplici settori del diritto civile, a partire dallo stesso ambito contrattuale – ove tutela il contraente che legittimamente si sia affidato a dichiarazioni o comportamenti altrui (ad es., art. 1337 c.c.[4]) – fino ad arrivare alla sfera della responsabilità civile, in cui rileva nella valutazione della colpa e della diligenza richiesta in determinati contesti tanto da concretizzarsi in una fonte di rapporti obbligatori (art. 1173 c.c.), come accade nel caso del c.d. "contatto sociale qualificato". Ma non solo: infatti, la fiducia connota in maniera significativa anche l'intera materia del diritto di famiglia e delle persone. A riguardo non si può non sottolineare come l'intera disciplina del diritto di famiglia sia pervasa dalla fiducia, là dove tutti gli istituti previsti dal nostro sistema ordinamentale sono posti a salvaguardia di quell'affidamento reciproco intercorrente tra i componenti della famiglia, in ordine all'adem-

[2] In tema, *inter alia*, v. G Perlingieri, *Profili applicativi della ragionevolezza nel diritto civile*, ESI, Napoli, 2015, pp. 115 ss.; G. Perlingieri-A. Fachechi, *Ragionevolezza e proporzionalità nel diritto contemporaneo*, ESI, Napoli, 2017, *passim*; G. Frezza, *L'usucapione decennale e i rapporti fra la trascrizione e la buona fede. L'ineludibile necessità di un approccio casistico*, in *Rass. dir. civ.*, 2021, 2, pp. 532 ss.; M. Foti, *Buona fede precontrattuale, conformità normativa e «interferenze» tra responsabilità e contratto*, *ivi*, 2020, 3, pp. 761 ss.

In particolar modo, l'art. 1175 c.c. sancisce che, in seno ad un rapporto obbligatorio, il debitore e il creditore devono comportarsi secondo le regole della correttezza, mentre l'art. 1337 c.c. stabilisce che le parti, nello svolgimento delle trattative e nella formazione del contratto, devono comportarsi secondo buona fede.

[3] Cfr. F. Piraino, *Buona fede, ragionevolezza e «efficacia immediata» dei principi*, ESI, Napoli, 2017; M. Pennasilico, *Buona fede e ragionevolezza nell'interpretazione dei contratti*, in G. Perlingieri e M. D'Ambrosio (a cura di), *Fonti, metodo e interpretazione*, cit., pp. 851 ss., pubblicato anche in G. Perlingieri e A. Fachechi (a cura di), *Ragionevolezza e proporzionalità nel diritto contemporaneo*, II, Napoli, ESI, 2017, pp. 851 ss.

[4] Cfr. G. Perlingieri, *Regole e comportamenti nella formazione del contratto. Una rilettura dell'art. 1337 codice civile*, ESI, Napoli, 2003.

ISSN 0039-4130

pimento di quei doveri fondamentali di collaborazione, assistenza sia materiale nonché morale, di cura (artt. 143 e 144 c.c.) e – con specifico riferimento ai figli – di educazione, istruzione e mantenimento (art. 147 c.c.). Parimenti, sono incentrati sul valore della fiducia altresì tutti quegli istituti di diritto privato i quali hanno lo scopo di tutelare le persone la cui capacità di agire è limitata, a causa di infermità fisiche e/o psichiche più o meno gravi: così, si rinviene nel nostro sistema di diritto privato la figura del tutore e del curatore, rispettivamente preposti alla tutela degli interessi delle persone interdette e inabilitate (artt. 414 ss. c.c.), o, ancora, la più recente figura dell'amministratore di sostegno (artt. 404 ss. c.c.). In questi casi, difatti, la fiducia emerge chiaramente nella stessa *ratio* delle norme, in quanto, al fine di assistere il soggetto la cui capacità di agire è limitata, il medesimo è affiancato da un altro soggetto a cui si affida [5].

Allo stesso modo l'equità riveste un ruolo centrale nel riequilibrare situazioni giuridicamente rilevanti parzialmente ingiuste, attenuando lo stesso formalismo giuridico e introducendo criteri di giustizia sostanziale del caso concreto. Non può ritenersi, infatti, che l'equità si limiti a un mero criterio integrativo, in quanto essa si manifesta quale strumento di interpretazione ed applicazione del diritto positivo, come chiaramente emerge dall'art. 1374 c.c., che impone di integrare il contratto anche secondo equità [6].

In ultimo, non può non menzionarsi – seppur brevemente – come la fiducia sia il presupposto per il corretto funzionamento di schemi negoziali tipici di quotidiana applicazione nella prassi commerciale ma non solo, quali, ad esempio, il mandato (artt. 1703 ss. cc.), il deposito (art. 1766 ss. c.c.) e lo stesso negozio fiduciario.

2. Fermo quanto in via generale appena illustrato, la dimensione dinamica della fiducia nell'esperienza civilistica si può cogliere, in particolar modo, a due particolari figure: l'appena citato negozio fiduciario e la nuova figura della *blockchain*.

Come anticipato, infatti, sin dall'antichità, si è cercato di valorizzare questa dimensione relazionale attraverso strumenti capaci di formalizzarne i contenuti e le modalità operative. Tra questi, il negozio fiducia-

[5] Si vedano, sull'istituto dell'amministrazione di sostegno, le recenti riflessioni di G. Carapezza Figlia, *Amministrazione di sostegno: una rilettura assiologica*, in *Rass. dir. civ.*, 3, 2024, pp. 817 ss.

[6] In tema, v. P. Perlingieri, *Equità e ordinamento giuridico*, in *Rass. dir. civ.*, 4, 2004, pp. 1149 ss.

ISSN 0039-4130

rio emerge quale archetipo giuridico, pur in assenza di una disciplina codicistica diretta – ma per tramite dell'art. 1322 c.c., il quale consente alle parti di concludere negozi atipici purché leciti e meritevoli di tutela –, grazie a un'elaborazione dottrinale e giurisprudenziale che ne ha definito i contorni. Esso si articola attraverso un doppio rapporto obbligatorio: da un lato, un soggetto. detto "fiduciario", riceve beni o diritti da un altro soggetto, detto "fiduciante"; dall'altro, il fiduciario è vincolato ad un altro obbligo, spesso non esplicitamente regolato ma desumibile dall'accordo. Proprio questo secondo patto, tacito od espresso che sia, viene per l'appunto denominato *pactum fiduciae*, il quale si connota, in tal modo, della capacità di influenzare l'esito del primo rapporto – quello con il quale si è trasferito il bene o il diritto – tra le parti. In altri termini, si manifesta come la fiducia sia in grado di travolgere un intero rapporto giuridico tra due soggetti [7].

3. Sebbene il negozio fiduciario rappresenti un modello intrinsecamente flessibile, esso riflette un paradigma di fiducia personale che, negli anni, ha progressivamente perso centralità. Soprattutto in epoca più recente, la crescente complessità delle relazioni economiche, unitamente all'incremento del numero di parti coinvolte nelle transazioni, ha richiesto un trasferimento della funzione fiduciaria da relazioni interpersonali ad entità istituzionali: così, banche, notai e autorità regolatrici hanno assunto il ruolo di garanti, fornendo un quadro normativo e organizzativo volto a garantire sicurezza e prevedibilità.

La crisi finanziaria del 2008, a riguardo, ha segnato un punto di svolta. Lo sgretolamento di colossi bancari ritenuti inscalfibili e la divulgazione di comportamenti opachi da parte di attori istituzionali hanno eroso la fiducia nei sistemi centralizzati. La mancanza di trasparenza e l'incapacità di prevenire abusi hanno reso evidente la necessità di un nuovo paradigma, che si svincoli dalla dipendenza da intermediari umani [8].

[7] Sul negozio fiduciario, tra i contributi più recenti, *inter alia*, V. Barba, *Affidamento fiduciario testamentario*, in *Rass. dir. civ.*, 1, 2020, pp. 1 ss.; V. Occorsio, *Titolarità e gestione nei rapporti fiduciari*, ESI, Napoli, 2020; A. P. Uges, *La fiducia come situazione giuridica real-obbligatoria*, ESI, Napoli 2022.

[8] La definizione normativa italiana di *blockchain* è data dalla l. 11 febbraio 2019, n. 12 all'art. 8-*ter* (Tecnologie basate su registri distribuiti e *smart contract*) legge pubblicata in Gazzetta Ufficiale n. 36 del 12 febbraio 2019 di conversione del decreto legge 14 dicembre 2018 n. 135, recante disposizioni urgenti in materia di sostegno e semplificazione per le imprese e per la pubblica amministrazione: «si definiscono "tecnologie basate su registri distribuiti" le tecnologie e i protocolli informatici che usano un registro condiviso, distribuito, replicabile, accessibile simultaneamente, architetturalmente decentralizzato su basi crittografiche, tali da

ISSN 0039-4130

In tale contesto si colloca l'avvento della *blockchain*, una tecnologia che propone di trasferire la funzione fiduciaria dal rapporto personale o istituzionale a un sistema decentralizzato e immutabile. Essa opera attraverso un registro distribuito, garantendo che ogni transazione sia verificata e registrata in modo crittograficamente sicuro da una rete di nodi indipendenti. La sua forza risiede nell'eliminazione degli intermediari tradizionali e nella possibilità di garantire trasparenza e integrità attraverso il funzionamento stesso della rete [9].

Dal punto di vista normativo, la *blockchain* pone interrogativi di estrema rilevanza, in particolar modo in ambito contrattuale, e di ciò gli *smart contract* ne rappresentano un esempio paradigmatico: programmati per eseguire automaticamente le clausole contrattuali al verificarsi di condizioni predefinite, essi incarnano una forma avanzata di fiducia meccanica. Pur rispettando il principio di autonomia contrattuale, gli *smart contract* sollevano questioni in merito alla loro compatibilità con i requisiti di causa e con il controllo giurisdizionale [10].

La "fiducia meccanica" introdotta dalla *blockchain*, benché prometta maggiore trasparenza ed efficienza, evidenzia limiti intrinseci. Essa riduce la fiducia a un dato puramente tecnico, eliminando la componente relazionale tra soggetti privati che da sempre caratterizza il diritto civile. La rigidità degli *smart contract*, che eseguono obbligazioni senza possibilità di interpretazione o adattamento, rischia di compromettere l'equilibrio tra certezza e giustizia, depersonalizzando in maniera eccessiva i rapporti giuridici e, di conseguenza, riducendo il fondamentale ruolo della fiducia.

4. Così, il diritto civile, storicamente orientato a bilanciare esigenze normative e valori umani, si trova dinanzi a una sfida inedita: integra-

consentire la registrazione, la convalida, l'aggiornamento e l'archiviazione di dati sia in chiaro che ulteriormente protetti da crittografia verificabili da ciascun partecipante, non alterabili e non modificabili».

[9] L'adozione della tecnologia *blockchain* parrebbe proporsi come un passo fondamentale per ricostruire la fiducia nel sistema di gestione dei diritti d'autore. Specificamente in merito al mercato dell'arte, si vedano G. Frezza, Blockchain*, autenticazioni e arte contemporanea*, in *Dir. fam. pers*., 2020, p. 489 ss.; G. Frezza e P. Virgadamo, *NFT e arte. Alla ricerca di una disciplina giuridica adeguata orientata al princípio di verità*, in *LawArt*, 4, 2023, pp. 285 ss.

[10] Sugli *smart contract*, *ex multis*, v. E. Battelli, *Le nuove frontiere dell'automatizzazione contrattuale tra codici algoritmici e big data: gli smart contracts in àmbito assicurativo, bancario e finanziario*, in *Giust. civ.*, 2020, pp. 661 ss.; I. Martone, *Gli* smart contracts*. Fenomenologia e funzione*, ESI, Napoli, 2022; G. Settanni, *Analisi e qualificazione giuridica dell'oracolo dei "contratti intelligenti"*, in *Dir. informazione e informatica,* 2024, pp. 240 ss.; M. Vatiero, *The "dark side" of smart contracts: A contract theory perspective*, in *European Management Rev.*, 2024, pp. 516 ss..

ISSN 0039-4130

re l'innovazione tecnologica senza perdere di vista la funzione sociale del diritto. La *blockchain*, se adottata acriticamente, rischia, infatti, di ridurre il diritto a un algoritmo, svuotandolo della sua capacità di rispondere alle complessità delle relazioni umane. La rivoluzione introdotta dalla fiducia meccanica rappresenta una sfida epocale che il diritto civile non può ignorare. Lungi dall'essere un semplice strumento tecnico, la *blockchain* interroga le fondamenta stesse della normatività, sollecitando una riflessione interdisciplinare che integri diritto, etica e tecnologia.

Solo attraverso un approccio critico e consapevole, consapevoli del ruolo che la fiducia ha ricoperto nella complessiva evoluzione del diritto civile nel nostro sistema ordinamentale, sarà possibile garantire un futuro in cui la fiducia stessa non sia confinata a rigidi protocolli tecnologici, ma continui a rispecchiare la complessità e la ricchezza delle relazioni umane.

ISSN 0039-4130

Like Hermes «the ox-thief» or a child «with jam on his hands»: Notes on Trust from Piero Bigongiari's Metapoetic Reflections

Lucia Battistel

This contribution aims to initiate a discourse on the theme of trust, drawing upon metapoetic reflections by Piero Bigongiari (1914-1997). The focus will be on the image of Hermes «the ox-thief» and that of the child «with jam on his hands», which Bigongiari uses to describe his own activity as a poet. Both, albeit in different ways, evoke the figure of the trickster and offer an interesting key to renegotiating the relationship of trust within the "literary relationship".

Trust is, undoubtedly, a theme with relevance to every discipline. Yet, its ubiquity entails the danger of making it the victim of sugarcoated narratives, which debase its importance by interpreting it as blind and naive reliance on the other, and thus weakening the revelance of rhetorical and persuasive discourse, which is crucial for its very functioning. This assertion was reiterated by, among others, Salvatore Di Piazza, in an article on the New Testament [1]. As Di Piazza points out, the Greek notion of *pistis*, which encompasses both the concepts of "trust" and "faith", in the New Testament maintains a close connection with the semantic dimension of persuading and thus of rhetoric. Therefore, the act of having faith and trust does not entail an illogical surrender

[1] S. Di Piazza, *Fede/fiducia, razionalità, conversione*, in F. Piazza – A. Pugliese (edited by), *Il prisma delle passioni. Prospettive per un'antropologia delle emozioni,* Palermo University Press, Palermo 2021, pp. 29-51.

to an other/Other, but rather presupposes a structured rhetorical-persuasive process. Jesus' miracles themselves are, in this sense, *semeia* and *argumenta*, that is to say, proofs that aim to persuade and actually convince early Christians to take on another's truth and make it their own. Trust is, in fact, closely intertwined with the problem of language. And consequently, it occupies a significant place, understood as the domain in which this is articulated and takes shape, in literary works as well, since literature is, first and foremost, a thought-of construction of language. In addition to playing a pivotal role in its manifestation as a narrative apparatus, and thus constituting a pertinent object of critical scrutiny, given the capacity of all literary elements to be the subject of thematic analysis, trust is configured, even more pertinently, as a necessary condition for the very activation of the text. It is intriguing to note that literature itself, and what might be termed the "literary relationship" – the exchange of meaning between a subject who writes and a subject, individual or collective, who reads – is founded on a mechanism of trust. As is well known, Coleridge, speaking of «poetic faith», called the «suspension of disbelief» the necessary condition for the reader to accept, at least momentarily, the narrative truth proposed by a work. In this regard, a foundational contribution was made by Booth's seminal work *The Rhetoric of Fiction* (1961)[2], where the concept of the «unreliable narrator» was introduced to describe the distance between the narrator and the implied author. For the activation of the text-machine, therefore, the reader must momentarily recognise the implied author as his own *auctoritas*, that is, as a bearer of a truth other than the factual one.

Now, a considerable amount of attention and study has been dedicated to investigating the extent to which readers are predisposed to place trust in the authority of the author. As Dietrich and Sundell previously observed in their 1967 publication *The Art of Fiction*[3], readers have a tendency to accept the narrator's account as true until evidence suggests otherwise. This phenomenon serves to elucidate and corroborate the efficacy of narrative strategies that employ unreliability.[4] However, this tendency to trust in the "literary relation" is not the

[2] W.C. Booth, *The Rhetoric of Fiction*, University of Chicago Press, Chicago 1961.

[3] R.H. Dietrich – R.F. Sundell, *The Art of Fiction*, Holt, Rinehart and Winston, New York 1967.

[4] For a recent and detailed overview of the topic, cfr. L. Fetzer, *Unreliable Narrators; Origins and Impacts*, Honors Program Theses, 922, 2024, available at https://scholarworks.uni.edu/hpt/922 (last consultation: 02/21/25).

ISSN 0039-4130

result of naivety, but corresponds to the very nature of the literary text as a performing linguistic machine constructed according to criteria of appreciability: in the case of poetry in particular, it is not only the theme that counts – to reduce the study of poetry to mere thematology would be to distort it – but also, and above all, the linguistic, rhetorical and expressive strategies through which its thematic content is articulated and arranged in an original, incisive and memorable way. In other words, the reader trusts not because human beings are naturally predisposed to trust – we leave this to be confirmed or disproved by scholars from other disciplines – but because in the poetic text that works, the mechanism of trust occurs naturally: the text itself possesses a high persuasive potential, given primarily by its expendability, its ability to engage, interest and imprint itself in the reader's mind. In this sense, and in the light of the definition of trust linked to the New Testament's *pistis* mentioned above, trust is the condition on which the "literary relationship" is based, but this does not mean that it cannot be manipulated or undermined, nor that it cannot be subject to negotiation. There have been many studies, especially in the English-speaking world, on authenticity in poetry and the strategies of mystification employed by authors. Scholars such as Hetherington have pointed out that even literature of a more confessional tone actually responds to the strategies of so-called «hoax poetry» [5], the poetry, we might say, of the misleaders.

Going back to the mid-twentieth century in Italy, Piero Bigongiari, a highly complex poet whose writing, dense with intertextual references and semantic stratifications, is an important example of the poetic game between misleading and clues. His poems require a considerable act of interpretive trust: in approaching the text, the reader must come to terms with the knowledge that each word can open up mutually contradictory perspectives of meaning, which, as such, will not fail to lead him astray. On the stylistic level, Bigongiari expresses this contradiction in the poetic text through a systematic use of material antitheses, inspired by Bachelard's studies [6], and an unusual and unexpected colourism,

[5] P. Hetherington, *Poetic Self-Inventions: Hoaxing, Misrepresentation and Creative License in Poetry*, in *New Writing*, X, 1, 2013, pp. 18-32.

[6] A significant portion of Bachelard's works can be found in Bigongiari's private library, reflecting his close engagement with the author; the collection is now preserved in his personal library and archive in Pistoia. Bigongiari also discussed about Bachelard's impact in literature in: P. Bigongiari, *Il diritto di sognare*, in *L'Approdo letterario*, 52, 1970, pp. 133-135.

ISSN 0039-4130

already in the tradition of the French Symbolists [7] and Campana [8], as well as through the juxtaposition of verbal and nominal forms, which suggest opposite and opposing directions – active and passive, present and past – and also contribute to dislocating the discourse on different spatio-temporal levels. This is the case of collections such as *La legge e la leggenda* (1992) [9], a book of poetry that reworks the biblical episode of Moses to question his condition as a son. Already in the title there is a tension between what is given and transmitted as received truth, i.e. the *legge*, "what is read", and, at the same time, what is to be made of this received truth, to be understood in a dimension of becoming, i.e. the *leggenda*, understood in its Latin gerundive value as "what is to be read", thus implying a strongly relational dimension: the text needs, in this sense, a reader who completes it, who questions it, without ever exhausting its intrinsic meaning.

An image that recurs several times in Bigongiari's metapoetic reflections is, then, that of Hermes stealing oxen by erasing the traces of the theft, and the more mundane image of a child caught stealing blackberry jam. We find them in a 1971 interview [10] in which the poet anticipates something of his forthcoming book of poems, *Antimateria* (1972) [11] , and they also recur in some of the pages that, according to Noferi, Bigongiari originally conceived for the preface to the original publishing project, which was intended to bring together *Col dito in terra* (1986) [12] and *Nel delta del poema* (1989) [13] in a single volume – but

[7] It may be useful to refer to P. Bigongiari, *Poesia francese del Novecento*, Firenze, Vallecchi, 1968.

[8] For Bigongiari, Campana's *Canti orfici* are among the twenty 20th-century books worth preserving, as stated in P. Bigongiari, *La materia plastica di Dino Campana*, in *L'Approdo letterario*, 4, 1959, pp. 69-78.

[9] P. Bigongiari, *La legge e la leggenda*, Mondadori, Milano 1992.

[10] «Questo che vi offro è un libro di forze "antielementari": un libro in cui le cose non stanno. Ma come un canzoniere famoso è diviso a metà da un avvenimento che non avviene (sul piano della poesia), e cioè la morte di Laura, può darsi che la traccia della contraddizione non sia che la dizione della stessa antitraccia che unisce due metà di un tutto. Un uomo, può darsi, cammina per la prima volta tra la morte e la vita facendo perdere le tracce della contraddizione nella dizione del suo cammino. Può darsi, o può darsi che con le mani macchiate di more un fanciullo-uomo segni più indelebilmente la traccia che vuol cancellare. Allora, se lo cogliete con le mani nel sacco, non lo punite, lasciatelo andare. Egli comunque non cammina all'indietro, né ha rubato i buoi di Gerione, e non ha nessuna caverna a meta riparatrice di un furto 'eroico'» (P. Bigongiari, *Il non luogo della poesia*, in Id., *Nel mutismo dell'universo. Interviste sulla poesia 1965-1997*, edited by A. Dolfi, Bulzoni, Roma 2001, p. 18).

[11] P. Bigongiari, *Antimateria*, Mondadori, Milano 1972.

[12] P. Bigongiari, *Col dito in terra*, Mondadori, Milano 1986.

[13] P. Bigongiari, *Nel delta del poema*, Mondadori, Milano 1989.

ISSN 0039-4130

which were instead published separately for editorial reasons [14].

Both images used to describe the poet's action refer to an idea of cunning and concealment, as well as implicit guilt and responsibility for the crime, and define the poet as a hoaxer or trickster, a polymorphous figure who plays with the reader, deceiving and misleading him, constantly questioning his trust in the text. What emerges from these excerpts, however, is a dual conception of the hoax, concocted by Hermes and the child in two very different ways: the poet-Hermes, a master of misdirection, is able to successfully cover his own tracks, while the poet-child, in a clumsy attempt to conceal his own theft, ends up making his own guilt even more obvious. The latter image, moreover, imbues the poetic text with a strong expiatory and apologetic dimension, suggesting that the poetic act itself is marked by a guilt that cannot be concealed and for which the reader is in some way being apologised to [15]. The blackberries and the dark stain they leave would thus represent the persistence of the poetic word, which manifests its own expressive autonomy and, paradoxically, backfires on the author himself: the textual machine constructed by the poet ends up, in a sense, taking on a life of its own and betraying its own author. Attempting a hermeneutic synthesis of these two images, one could say that the Bigongiarian poet is a child playing at being Hermes, but with little success. We cannot fail to read here a veiled reference to the so-called hermetic poetics to which Bigongiari is said to have belonged [16], that is to say, to that taste for the closed and initiatory word, written by the chosen few for the chosen few, which escaped the grasp of others, just

[14] «Che conta è dunque questo riversare continuo in altro, questo proporre non tanto un proposito quanto un'interrogazione, che trova la risposta nel costituirsi stesso della domanda: è una domanda allargata fino al punto di non chiedere risposta che nel suo ritorno alle origini […]. La poesia pare ritrovare le sue orme, e nel cancellarle, i passi degli eroi che si avanzano, i loro scoramenti, la loro eventuale felicità: che è la felicità dell'evento visto dall'altra parte, riportato alla sua fonte […]. Forse è una *ruse* come quella di Hermes, che, dopo aver rubato i buoi ad Apollo […], per non farli ritrovare li fa camminare all'indietro, accompagnandoli coi piedi imbottiti di frasche per non lasciar traccia delle proprie orme. Il poeta forse non è da meno dell'astuto infante, destinato ad essere il messaggero degli Dei […]. Anche il furto del sacro può essere risarcito nelle briciole di una sacralità sconosciuta» (A. Noferi, *Piero Bigongiari. L'interrogazione infinita. Una lettura di* Dove finiscono le tracce, Bulzoni, Roma 2003, p. 18).

[15] On the confessional form in Bigongiari's latest works, see A. Noferi, *Teoria, poesia e "confessione" nell'ultimo Bigongiari*, in P. Bigongiari, *La poesia pensa. Poesie e pensieri inediti. Leopardi e la lezione del testo*, a cura di E. Biagini - P. F. Iacuzzi - A. Noferi, Firenze, Olschki, 1999, pp. 241-259.

[16] Due to space constraints, we cannot explore this and other issues in depth here. However, for a detailed analysis of the debate on Hermeticism, we recommend referring to A. Dolfi (edited by), *L'ermetismo e Firenze*, I-II, Firenze University Press, Firenze 2016.

ISSN 0039-4130

as Hermes escaped his own condemnation by concealing the signs of the passage of the oxen. This allusion to Hermes but, at the same time, the reference to the poet's clumsiness in attempting to behave like him, could perhaps be interpreted as an acknowledgement of Bigongiari's closeness to that tendency of Hermeticism, traditionally understood as writing that wants to conceal itself, but at the same time of the poet's difficulty – which is not a difficulty, but rather a willingness – to adhere to the game of concealment. And it is perhaps in this sense that we should read Bigongiari's poetic self-definitions when he says that he writes a «poetry of discourse» [17], meaning a poetry that reveals itself. In fact, his writing seems too functionalised for psychoanalytic self-discovery to be truly hermetic – it seems that the practice of writing is used by the poet precisely as a tool, albeit an aestheticized one, of self-analysis, and thus ends up being self-revealing: his biographical junctures do not remain hidden, but emerge conspicuously, precisely in their obsessive return, disguised only in appearance [18]. The author thus ends up deceiving and revealing himself, allowing himself to be betrayed by his own recurring fixations, such as the figures of Moses and Noah and, more generally, the Old Testament references, or, again, on a less thematic and more stylistic level, the obsessive recurrence of etymological figures, a tendency that suggests an implicit insistence on the familiar and seems to express an attempt to rework unresolved biographical traumas, especially those related to the paternal dimension. This aspect, moreover, opens the way to a possible reading of Bigongiarian poetry in the light of a systematic and well-structured sifting of the literature produced by trauma studies, a research path that we believe deserves further study and to which we promise to devote ourselves elsewhere.

But the question arises: if the poet is, or at least tries to be, a trickster, what is the role of trust in the literary text? Certainly, the figure of the trickster Hermes provokes mistrust and disorientation, suggesting a constant suspicion of the text. With his deceptions, the god betrays the trust of those who rely on the game of interpretation, those who follow the "voices in the labyrinth" only to find themselves lost. At the same time, however, if we adhere to the redefinition of trust with

[17] P. Bigongiari, *Domande sulla poesia, domande della poesia*, in A. Dolfi (edited by), *Nel mutismo dell'universo,* p. 105.

[18] Bigongiari was undoubtedly a critic with a psychoanalytic orientation. He recalled on several occasions thedelay of Italian culture in recognizing the fertility of the psychoanalysis-literature dialogue, cfr. P. Bigongiari, *Il lieve soffio dell'ala*, in *Nel mutismo dell'universo*, pp. 257-259.

ISSN 0039-4130

which we opened this brief reflection, and thus understand it not as an active involvement and participation in another's truth, which by extension also becomes our own, Bigongian texts also imply a mechanism of trust, but a trust that must continue to be negotiated between author and reader, and that will have as its first auctoritas and 'guarantor' of truth the text itself.

If the poet is a naive child who betrays himself, the reader is invited to find meaning through the traces scattered throughout the text, the eloquent blackberry stains. The poem is thus configured as a space in which trust is negotiated on the basis of a complicity that constantly questions itself; it is not an acquired, static and passive given, but an understanding that is renewed and that must have as its first site of analysis the text itself. The reader thus becomes an active participant in the construction of meaning; trust, on his part, does not amount to a passive adherence to a given meaning taken as true, as a law, but to an acceptance of the interpretive play of the text – and that, in its psychoanalytic poignancy, we might also call analysis. And it is precisely in this constant instability that trust is transformed into a true act of affection for the text: a hermeneutic fidelity that is not discouraged by the apparent difficulties of some of the red herrings, but rather goes through them, knowing that somehow one will see those proofs of reasonableness, those signs of meaning, those blackberry spots that will denounce their author.

ISSN 0039-4130