TEORIA

T

Rivista di filosofia
fondata da Vittorio Sainati
XLIV/2024/2 (Terza serie XIX/2)

Topographies of Risk

Areas of application

Topografie del rischio

Terreni di applicazione

Edizioni ETS

TEORIA

Rivista di filosofia
fondata da Vittorio Sainati
XLIV/2024/2 (Terza serie XIX/2)

Iscritto al Reg. della stampa presso la Canc. del Trib. di Pisa n° 10/81 del 23.5.1981

Direzione e Redazione
Dipartimento di civiltà e forme del sapere dell'Università di Pisa, via P. Paoli 15, 56126 Pisa,
tel. (050) 2215400 - www.cfs.unipi.it

Direttore Responsabile
Adriano Fabris

Periodico semestrale

Abbonamento (cartaceo, privato): Italia e UE € 40,00; extra UE € 50,00
Abbonamento (cartaceo, istituzionale): Italia e UE € 50,00; extra UE € 60,00
PDF (online, stampabile): privato (accesso individuale) € 40,00
PDF (online, stampabile): istituzionale (accesso con riconoscimento IP) € 40,00

Bonifico bancario intestato a
Edizioni ETS - Banca C.R. Firenze, Sede centrale, Corso Italia 2, Pisa
IBAN IT 21 U 03069 14010 100000001781
BIC/SWIFT BCITITMM
causale: abbonamento «Teoria»

«Teoria» è indicizzata ISI Arts&Humanities Citation Index e SCOPUS, e ha ottenuto la classificazione "A" ANVUR per i settori 11/C1-C2-C3-C4-C5. La versione elettronica di questo numero è disponibile sul sito: www.rivistateoria.eu

L'indice dei fascicoli di «Teoria» può essere consultato all'indirizzo: www.rivistateoria.eu
Qui è possibile acquistare un singolo articolo o l'intero numero in formato PDF, e anche l'intero numero in versione cartacea.

I numeri della rivista sono monografici. Gli scritti proposti per la pubblicazione sono double blind peer reviewed. I testi devono essere conformi alle norme editoriali indicate nel sito.

Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa
info@edizioniets.com - www.edizioniets.com

Distribuzione
Messaggerie Libri SPA - Sede legale: via G. Verdi 8 - 20090 Assago (MI)

Promozione
PDE PROMOZIONE SRL - via Zago 2/2 - 40128 Bologna

ISBN 978-884677058-5
ISSN 1122-1259

Contents / Indice

Premio di Studio «Vittorio Sainati» 2023-2024

T

Premise / Premessa

The two issues of 'Teoria' 2024 are dedicated to an in-depth examination and rethinking of the category of 'risk', a fundamental notion of modernity and still today at the centre of intense debate. The final objective that has been set is that of the elaboration of a 'topography of risk' that allows a clear mapping of the various meanings and applications of this notion. The first issue of the year was mainly devoted to conceptual reconstruction, drawing on the various traditions and the most significant figures in the history of thought that have dealt with this theme. On the other hand, this second issue will focus on the exploration of the main areas of application of this concept. These include ecological reflection and climate risk, technological risk, its management in the new dimension of Artificial Intelligence, risk in health care, as well as in pedagogical education. The need to move into more applied terrain is due to the very nature of the subject, which is used in the most diverse fields, becoming a pivotal concept in the management of our daily lives. Our thinking is in fact shaped by the idea of risk, the calculation of the probability of the occurrence of an adverse event, and the attempt to control or limit the damage. On the practice of calculating and managing risk depends the success of our activities and our greater or lesser confidence in the infrastructure, systems and devices we use. This concept is so present in our society that it is almost transparent and adherent to our very lives, and thus ends up escaping criticism. In order to retrace its meaning and bring out its various uses, with this second issue we explore some of its main areas of application.

The volume therefore opens with Dean Komel's contribution that offers us a reflection on the nihilistic root of the risk society, in correlation with a series of current phenomena that undermine or redefine the sense of the world,

in particular the climate crisis and the development of AI. By critically analysing Ulrich Beck's notions of 'reflexive modernisation', Zygmunt Bauman's 'liquid modernity' and Niklas Luhmann's 'self-production of social systems', the author highlights how the 'risk society' constitutes an 'operational concept' only if it is grounded in the apparatus of ecosociology, which links eco-nomics, eco-logy and eco-technology.

Dimitri D'Andrea clarifies the difference between risk and threat, arguing that in the case of the climate crisis, it is more correct to speak of a threat, since it appears as a natural and certain consequence of our reckless behaviour towards the environment. The author analyses the reasons for the general denial that paralyses the fight against this phenomenon, identifying them in particular in the emotional and cognitive distortions that derive from the perceived excessive renunciation of goods and freedoms that such a fight implies, and the difficulty of devising realistic and effective solutions. The answer to this impasse lies mainly in a coordinated engagement of global institutions with local ones, recovering the value of the closest and most particular realities.

Marco Emilio's paper also examines ecological and climate risk. In particular, the argumentation is based on local examples that underline the conflicting nature of the transition to different energy sources. The author illustrates conflicts that include social, technological, economic and ecological aspects. Using the concepts of epistemic responsibility and vulnerability, he suggests improving the interaction between the social sciences, climate science and the philosophical exploration of risk and collective action. This improvement aims at a deeper understanding of decision-making, particularly in situations of profound uncertainty, and effective risk communication to prevent the dangers of collective inaction and a sense of powerlessness.

With Zarko Paić's essay, we move into the terrain of the technosphere, i.e. the relationship between human beings and technology. The author analyses the link between risk and chaos, entropy and contingency. These concepts, in fact, increasingly present in both contemporary science and philosophical reflections, are both the foundation of policies and practices based on risk and their threat: they are contained and controlled by the disciplines of statistics and probabilistic calculation.

Still remaining on the technological terrain, Veronica Neri focuses in particular on the risks of image-generating AI. Indeed, images created by artificial intelligence allow the development of unimaginable creative possibilities, but open up equally serious ethical issues, which must be regulated in order to limit their dangerousness. Through an examination of these

risks (bias, deep fakes, manipulation, multiple identities), the author offers a public ethics response based on strengthening information and awareness in order to provide subjects with the appropriate skills to counter the state of uncertainty caused by these systems.

Anastasia Siapka analyses the risk-based approach in the European regulation of artificial intelligence (AI Act). The author argues that in order to make this approach more operational and effective, the objective dimension must be integrated with a regulatory governance perspective. To achieve this integration, she examines AI-induced risk from the dominant approach of consequentialism, highlighting its limitations under conditions of uncertainty. He then proposes virtue ethics as an alternative approach to AI-induced risk. The essay concludes with an analysis of the implications of this approach for research, policy and practice.

With Leopoldo Sandonà's essay, we move on to analyse risk in the context of health services. While at first this concept was only considered from the medical-clinical perspective, today, it is joined by a corporate perspective. However, this complex system that holds together the clinical space and that of complex organisations seems to reveal its limitations, in particular linked to a legalistic approach that puts the ethical approach to the patient and the therapeutic relationship in second place. In response to this, the author emphasises the importance of an ethically informed approach to risk, which makes it possible to approach the individual case but relates it to a global ethical perspective.

To conclude the monographic section, Ilaria Malagrinò's contribution focuses on the relationship between risky conduct and young actors. This relationship is usually mediated by the use of new technologies, in particular social media, which modify our practices, revealing moral connotations. Precisely by analysing this technological context and the idea of risk and uncertainty that are promoted in it, the author expresses the urgency for a rethinking of our moral conduct within digital environments and a serious assumption of responsibility for future generations.

This thematic section is followed by the one in which is published the essay that won the Sainati Prize 2024, dedicated to the memory of Vittorio Sainati, professor of Theoretical Philosophy at the University of Pisa and founder of 'Teoria'. In this issue, the winning text is written by Giulia Bernard.

I due fascicoli di «Teoria» 2024 sono dedicati all'approfondimento e al ripensamento della categoria di "rischio", protagonista della modernità e tutt'oggi al centro di intensi dibattiti. L'obiettivo finale che ci si è posti è

quello dell'elaborazione di una "topografia del rischio" che permetta di delineare una chiara mappatura dei vari significati e delle varie applicazioni di tale nozione. Se il primo numero dell'anno è stato dedicato principalmente alla ricostruzione concettuale, rifacendoci alle varie tradizioni e alle figure più significative della storia del pensiero che di questo tema si sono occupate, questo secondo numero sarà invece incentrato sull'esplorazione dei principali ambiti di applicazione di tale concetto. Tra questi ricordiamo la riflessione ecologica e il rischio climatico, il rischio tecnologico, la sua gestione nella nuova dimensione dell'Intelligenza Artificiale, il rischio in sanità, così come nell'educazione pedagogica. La necessità di passare a un terreno più applicativo è dovuta alla natura stessa del tema, che trova impiego nei settori più disparati, divenendo un concetto cardine per la gestione della nostra vita quotidiana. Il nostro pensiero è infatti plasmato dall'idea di rischio, dal calcolo delle probabilità dell'accadimento di un avvenimento avverso, e dal tentativo di controllo o limitazione del danno. Dalla pratica di calcolo e di gestione del rischio dipende la riuscita delle nostre attività e la nostra maggiore o minore fiducia nelle infrastrutture, nei sistemi e nei dispositivi che utilizziamo. Tanto è pregnante, per la nostra società, tale concetto, da risultare quasi trasparente e aderente alla nostra stessa vita, finendo quindi per sfuggire alla critica. Al fine di recuperarne il significato e farne emergere i suoi svariati impieghi, con questo secondo numero esploriamo alcuni dei principali ambiti di applicazione.

Il volume si apre dunque con il contributo di Dean Komel che ci offre una riflessione sulla radice nichilistica della società del rischio, in correlazione con una serie di fenomeni attuali che mettono a repentaglio o ridisegnano il senso, in particolare la crisi climatica e lo sviluppo dell'IA. Analizzando criticamente le nozioni di "modernizzazione riflessiva" di Ulrich Beck, di "modernità liquida" di Zygmunt Bauman e di "autoproduzione di sistemi sociali" di Niklas Luhmann, l'autore mette in luce come la "società del rischio" costituisca un "concetto operativo" solo se fondato sull'apparato dell'ecosociologia, che collega l'eco-nomica, l'eco-logia e l'eco-tecnologia.

Dimitri D'Andrea chiarisce la differenza tra rischio e minaccia, sostenendo che nel caso della crisi climatica è più corretto parlare di una minaccia, in quanto essa incombe come naturale e certa conseguenza del nostro agire sconsiderato nei confronti dell'ambiente. L'autore analizza le ragioni del generale diniego che paralizza il contrasto a questo fenomeno, identificandole in particolare nelle distorsioni emotive e cognitive che derivano dalla rinuncia percepita come eccessiva in fatto di beni e libertà che implica tale contrasto e la difficoltà di elaborare soluzioni realistiche ed efficaci. La risposta

a questo stallo sembra trovarsi principalmente in un coordinato impegno delle istituzioni globali con quelle locali, recuperando il valore delle realtà più prossime e particolari.

Anche l'intervento di Marco Emilio prende in esame il rischio ecologico e climatico. In particolare, l'argomentazione si basa su esempi locali che sottolineano la natura conflittuale della transizione verso fonti energetiche diverse. L'autore illustra conflitti che comprendono aspetti sociali, tecnologici, economici ed ecologici. Egli suggerisce, utilizzando i concetti di responsabilità epistemica e di vulnerabilità, di migliorare l'interazione tra le scienze sociali, le scienze del clima e l'esplorazione filosofica del rischio e dell'azione collettiva. Questo miglioramento mira a una comprensione più profonda del processo decisionale, in particolare in situazioni di profonda incertezza, e alla comunicazione efficace dei rischi per prevenire i pericoli dell'inazione collettiva e del senso di impotenza.

Con il saggio di Zarko Paić ci spostiamo sul terreno della tecnosfera, ossia della relazione tra essere umano e tecnologia. L'autore analizza il legame tra rischio e caos, entropia e contingenza. Questi concetti, infatti, sempre più presenti sia nelle scienze contemporanee che nelle riflessioni filosofiche, sono sia il fondamento delle politiche e delle pratiche basate sul rischio che la loro minaccia: sono contenuti e controllati dalle discipline della statistica e del calcolo probabilistico.

Rimanendo sempre sul terreno tecnologico, Veronica Neri si concentra in particolare sui rischi dell'IA generativa di immagini. Infatti, le immagini create dall'intelligenza artificiale permettono lo sviluppo di impensate possibilità creative, ma aprono altrettante serie questioni etiche, che devono essere regolamentate al fine di limitarne la dannosità. Attraverso una disamina di questi rischi (bias, deep fake, manipolazione, identità plurime), l'autrice offre una risposta in chiave di etica pubblica basata sul rafforzamento dell'informazione e della consapevolezza per fornire ai soggetti le adeguate competenze per contrastare lo stato di incertezza provocato da questi sistemi.

Anastasia Siapka analizza l'approccio basato sul rischio, presente nella regolamentazione europea in materia di intelligenza artificiale (AI Act). L'autrice sostiene che, per rendere maggiormente operativa e efficace tale approccio, si debba integrare l'aspetto oggettivo di valutazione con una prospettiva di governance normativa. Per ottenere questa integrazione, esamina il rischio indotto dall'IA a partire dall'approccio dominante del conseguenzialismo, evidenziandone i limiti in condizioni di incertezza. In seguito, propone l'etica della virtù come approccio alternativo al rischio indotto dall'IA.

Il saggio si conclude con l'analisi delle implicazioni di questo approccio per la ricerca, la politica e la pratica.

Con il saggio di Leopoldo Sandonà passiamo ad analizzare il rischio nell'ambito dei servizi sanitari. Se all'inizio questo concetto era considerato soltanto dalla prospettiva medico-clinica, oggi, a questo piano si aggiunge quello di carattere aziendale. Tuttavia, questo sistema complesso che tiene insieme lo spazio clinico con quello delle organizzazioni complesse sembra rivelare i suoi limiti, in particolare legati ad un'impostazione legalistica che pone in secondo piano l'approccio etico al paziente e la relazione terapeutica. In risposta a ciò, l'autore sottolinea l'importanza di un approccio eticamente informato al rischio, che permette di approcciarsi al singolo caso ma ricollegandolo a una prospettiva di etica globale.

A concludere la sezione monografica, il contributo di Ilaria Malagrinò si concentra sul rapporto tra condotte rischiose e soggetti giovani. Tale relazione è solitamente mediata dall'utilizzo delle nuove tecnologie, in particolare dai social media, i quali modificano le nostre pratiche, rivelandosi moralmente connotati. Proprio analizzando questo contesto tecnologico e l'idea di rischio e incertezza che in esso vengono promosse, l'autrice esprime l'urgenza per un ripensamento della nostra condotta morale all'interno degli ambienti digitali e una seria assunzione di responsabilità per le generazioni future.

A questa sezione tematica segue quella che ospita il saggio vincitore del Premio Sainati 2024, dedicato alla memoria di Vittorio Sainati, professore di Filosofia teoretica all'Università di Pisa e fondatore di «Teoria». L'articolo vincitore di quest'anno è stato scritto da Giulia Bernard.

Silvia Dadà
Università di Pisa
silvia.dada@unipi.it

T

Dean Komel

The Nihilism of Risk Society

It should be observed[1] from the outset that the term "risk society," which stems primarily from Ulrich Beck's 1986 *Risikogesellschaft. Auf dem Weg in eine andere Moderne*[2], can by no means be taken as an elaborated theoretical concept[3], even though, decades after it first appeared, it continues to be profusely employed within the sociological and humanistic disciplines, and probably even more so within today's perception of 'social reality.' At the same time, the term "risk society" – like the terms "knowledge society," "global society," "information society," "society of spectacle," etc. – is no mere slogan or buzzword. Perhaps it is best to say that "risk society" is a *preception* that dictates a particular thematised experience, alongside a dearth

[1] The text is published as part of the implementation of the research program *The Humanities and the Sense of Humanity from Historical and Contemporary Viewpoints* (P6-0341) and the research project *The Hermeneutic Problem of the Understanding of Human Existence and Coexistence in the Epoch of Nihilism* (J7-4631).

[2] U. Beck, *Risikogesellschaft. Auf dem Weg in eine andere Moderne*, Suhrkamp, Frankfurt am Main 1986; english: *Risk Society. Towards a New Modernity*, transl. by M. Ritter, Sage Publications, London 1992.

[3] In this connection, Niklas Luhmann, who has himself done a great deal of work on risk society, also notes "The descriptions of modern society that are on offer today no longer strive for theoretical elaboration. They emphasise individual phenomena which they consider particularly noteworthy and leave it at that. Even the term 'capitalist society' was not covered by the relevant economic sciences and merely suggested a socio-historical description of an epoch, a narrative, so to speak. This lack of theoretical analysis is even more obvious in the case of "risk society" or "experience society". The same applies to the 'information society'" (Niklas Luhmann, "Entscheidungen in der 'Informationsgesellschaft'", lecture held at the conference *Soft society: eine internationale Konferenz über die kommende Informationsgesellschaft*, 28.10. - 3.11.1996 in Berlin, organised by the Arbeitskreis Informationsgesellschaft der Humboldt-Universität and the Japan Society for Future Research, Tokyo. https://www.fen.ch/texte/gast_luhmann_informationsgesellschaft.htm).

 DOI: 10.4454/5s629x85

of effective socio-critical recipes and formulas.

The label "risk society" does not link the element of risk to some human action but to the *performance of society as a whole*. The first concern is this: how can what we generally understand and accept as "society" encompass risk, even mega-risk, which exceeds any other kind of "risk," from "living" to "economic" risk? How can "society," as a subject we trust in more than we trust in people, God, and the world, in the process of its subjectification, turn out to be a "risky investment"?

Today we are clearly facing a deluge of all kinds of social risks, and society itself is rushing forward boldly and blithely, risking everything because everything is at its disposal. Is not the formulation "risk society," then, the "Trojan horse" of sociological science and the entire scientific apparatus, which leaves the self-totalising subjectivity of society intact or makes it a function of its own promotion? It is easy, of course, to theoretically dismiss these sorts of concerns, but at the same time, in practical terms, this could suppress critical awareness of how our general attitude about the state of the world has become suspiciously risky and is already beyond normal human comprehensivity. This abnormality, which also has a dramatic impact on the formation of social norms and political imperatives[4], demands that we position the prevailing discourse of risk society in the context of today's *erasure of the world-horizon*[5],

[4] As a current example of this, consider the contemporary political and scientific glorification of nuclear energy as "green," which is a complete reversal, given that since Chernobyl we have seen widespread political activity aimed at *closing* nuclear power plants, etc. Consider, too, how ecology has become a political issue that generates its own economy and requires a special reflection on the functionality and functioning of European and global political institutions and fora.

Another example is the monitoring of risky war situations in various parts of the world, with the fear that they could spread and escalate into a conflict of greater proportions, perhaps even between superpowers, which, if nuclear weapons were used, could lead to the widespread destruction. Such hypocritical fears are at the same time "complemented" by constant war-goading and by concrete aid in the forms of money, arms and manpower, which leads to the conclusion that establishing world peace would be riskiest of all for today's military, economic, technological, cultural and political drive for world domination. Let us not be distracted here by all sorts of philanthropic events and sporting events such as, say, the Olympic Games, and above all not by ethical tribunals.

[5] "In a world, in which space dominates, then there is no room for any notion of time or of place other than as modifications of space – other than as amenable to the numerical, the measurable, and the quantifiable; in such a world, what lies outside the objectivity of space and number can only be subjective and so conventional – or, one might say, 'onstructed'. The difficulty, however, is that this leaves almost everything that pertains to the human as belonging to the realm of conventionality, and so as having no intrinsic foundation or limit, at the same time being completely subject to the supposed objectivity of the spatial and the numerical. Contemporary capitalism, conjoined with modern information systems, and embodied in the 'market' (itself an informational as much as economic system), becomes the all-encompassing technologi-

which we can follow live on our screens[6] and which social theories generally have little regard for, let alone take it as a key premise of their own occupation with the sociability of contemporary society. Instead, *theory* itself, with the all-encompassing aid of various media houses and housings, is becoming *theatre*. The society of the spectacle[7], where we do not seem to risk anything, but in fact – without realising it – actually risk everything, is the best company for risk society. The result of this can only be a *calculation* that defines every possibility of *communication*, including scientific or religious communication. The ubiquitous calculative-communicative *economy of risk*, to which the use of the word "risk"[8] is originally allied, is, in terms of supply and demand, in strategic league with the *ecology of risk*, which in recent decades has acquired not only catastrophic but even apocalyptic tones[9]. *Rationalisation*, *computerisation*, *capitalisation*, *globalisation* on the one side, and *the apocalyptic*, *the catastrophic*, *riskiness*, *panic* on the opposite side, offer themselves up as determining *para-characteristics* of the present age

cal 'machine' that allows spatialized human subjectivity to be worked out within the realm of the objectively quantifiable and numerical. Moreover, there can be no easy defense against the encompassing reach of technological modernity, including its instantiation in contemporary capitalism, since technological modernity refuses the very idea of boundary or limit on which such a defense must depend" (J. Malpas, *The Spatialization of the World. Technology, Modernity, and the Effacement of the Human*, "Phainomena" 27, 106/107 (2018), pp. 101-101).

6 "With the real-time transmitting and receiving power of the various signals alienating the nature of time distances, the active optics of electromagnetic waves exploits the depth of field, the very reality of our own world to the point of reducing it to nothing, or next to nothing, thereby leading to a catastrophic sense of incarceration now that humanity is literally deprived of horizon" (P. Virilio, *Open Sky*, trans. by Julie Rose, Verso, London 1997, p. 40-41; orig.: *La Vitesse de Liberation*, Éditions Galilée, Paris 1995, p. 55).

7 Guy Debord, *La société du spectacle* (1967), Les Éditions Gallimard, Paris 1992; english: *The Society of the Spectacle*, Zone Books, New York 1995. See further: R. Gruneau, J. Horne (eds.), *Mega-Events and Globalization: Capital and spectacle in a changing world order*, Routledge, Abingdon-Oxon. 2017.

8 "*Risk* (n.): 1660s, *risque*, 'hazard, danger, peril, exposure to mischance or harm', from French *risque* (16c.), from Italian *risco, riscio* (modern *rischio*), from *riscare* 'run into danger', a word of uncertain origin. The English spelling is recorded by 1728. Spanish *riesgo* and German *Risiko* are Italian loan-words. The commercial sense of 'hazard of the loss of a ship, goods, or other properties' is by 1719; hence the extension to 'chance taken in an economic enterprise'. Paired with *run* (v.) from 1660s. *Risk aversion* is recorded from 1942; *risk factor* from 1906; *risk management* from 1963; *risk-taker* from 1892" (https://www.etymonline.com/word/risk). See also the chapter "The Concept of Risk" in: N. Luhmann, *Risk: A Sociological Theory*, transl. by R. Barrett, de Gruyter, New York 1993, pp. 1-39; orig.: *Soziologie des Risikos*, de Gruyter Berlin-New York 1991, pp. 9-40.

9 J.-P. Dupuy, *How to Think about Catastrophe: Toward a Theory of Enlightened Doomsaying*, trans. by M.B. DeBevoise and M.R. Anspach, Michigan State University Press, East Lansing 2022; orig.: *Pour un catastrophisme éclairé*, Le Seuil, Paris 2002.

and of the image of humanity, i.e. (post-? hyper-? meta-? contra-?) modernity as… *bidirectional? multidirectional? unidirectional?* or *directionless!* course.

The very title of Beck's 1986 pioneering *Risikogesellschaft. Auf dem Weg in eine andere Moderne* shows that he structurally defines "risk society" as a consequence of modernity, and this structural defining is further reinforced by his correcting "modernity" into "new modernity." In this context, Anthony Giddens' 1990 study *The Consequences of Modernity*[10] is cited as a reference work. Related to Giddens' insights is Beck's further structural determination of the "new modernity" as a *reflexive modernisation*, involving both (reflexive) scientification and politicisation. The label 'reflexivity,' which has its own exclusive and defining place within the shaping of modern philosophy and the forming of modern subjectivity, is used here more in the sense of what is being (reflexively) *reacted to*, which is also true of the aspect of *risk-consciousness* that characterises the current process of modernisation in comparison to previous ones[11]. This "reflexive cut," like the "knowledge society," may be conceptually weak but it is indicative in the sense that the risk to be reflected upon, not only after the fact but also *in advance* of the fact, is not produced by any particular social subjects, but by the very self-reflexive social system of production and consumption. To "reflect in advance" thus means not only "to react" in the sense of "to look at something" and "to pay attention," but "to keep under the spotlight" – to *exercise control*. A society that is reflected as risky must take recourse to a *methodology of control* that serves as its *prevailing worldview*, or rather as a *camera and a screen* that can present anything and everything as risky – that is, the whole world *and most of all the world*. The world has been riskily diminished through this presentation. It is therefore superficial and unfitting to talk about how the world is being globalised; rather, society is globalising (itself) as a subjectivity which, "respecting the principles of the free market," establishes dominance over the entire world and, in its totalization,

[10] A. Giddens, *The Consequences of Modernity*, Polity, Cambridge 1990.

[11] "While simple modernization ultimately situates the motor of social change in categories of instrumental rationality (reflection), 'reflexive' modernization conceptualizes the motive power of social change in categories of the side-effect (reflexivity). Things at first unseen and unreflected, but externalized, add up to the structural rupture that separates industrial from 'new' modernities in the present and the future. 'Reflexive' thus also implies reflex-like and simultaneously historic modernization (which, of course, as the present enterprise evidences, can be conceptualized, that is, reflected)" (U. Beck, *The Reinvention of Politics. Rethinking Modernity in the Global Social Order*, Polity Press, Cambridge 1996, p. 45; orig.: *Die Erfindung des Politischen*, Suhrkamp Verlag, Frankfurt am Main 1993, p. 51).

exercises total control. On this basis, it establishes its own *ecosociology*, by which it no longer, and no longer only, means social theory and critical social reflection. In reflexive modernisation, ecosociology (where "eco" is to be understood strictly as a conflation of ecology, ecology and eco-technology) has become *a key medium for the self-(re)production* of society itself. As a system of self(re)production, society synthesises both the functions of subjectivity and objectivity – it is the *society of society*, as Niklas Luhmann suggested in the title of one of his most extensive works[12].

ìIn 1991, within the framework of a more broadly developed theory of social systems, Luhmann published *Soziologie des Risikos*. Significantly, Luhmann explicitly redirects the debate on risk society towards a sociology of risk, and he builds on the reflexive modernisation perspective with the autopoiesis of society as a system in process:

> Above all, this requires exact definition of the concept of risk, and analysis of the reasons why the concept and the facts it refers to have been gaining in importance in the more recent development of the societal system. We will reply to this question with the thesis that the dependence of society's future on decision making has increased, and nowadays so dominates ideas about the future that all concept of 'forms of being', which – as Nature – intrinsically limit what can happen, has been abandoned. Technology and the concomitant awareness of capability has occupied nature's territory, and both surmise and experience indicate that this can more easily prove destructive than constructive. The fear that things could go wrong is therefore growing rapidly and with it the risk apportioned to decision making. In this analysis the concepts of decision and technology (in a sense yet to be specified) play an important role. It is thus all the more necessary to point out from the start that no mental and no material (machine-like) phenomena are meant. Our analysis of society is exclusively concerned with communications. Communication, and nothing else, is the operation by which society as a system produces and reproduces itself by 'autopoiesis'. This is naturally not to deny that the environment of the societal system contains realities that an observer can describe as consciousness or as machine[13].

Through his attempt to transcend a subjectivist social theory on the basis of the concept of autopoiesis, Luhmann at the same time systematically fixe the understanding of the sociability of society as a self-reproducing subjec-

[12] N. Luhmann, *Die Gesellschaft der Gesellschaft*, 2 Bd., Suhrkamp Verlag, Frankfurt am Main 1997; english translation: *Theory of Society*, by R. Barrett, Stanford University Press, Stanford, Vol. 1: 2012, Vol. 2: 2013.

[13] N. Luhmann, *Risk: A Sociological Theory*, trans. by R. Barrett, de Gruyter, New York 1993, p. XII; orig: *Soziologie des Risikos*, de Gruyter. Berlin-New York 1991, p. 6.

tivity in the process of reflexive modernisation. By turning away from the "society of risk" to the "sociology of risk", he in no way compromises the view that the economic, technological, ecological, population, political, cultural and other processes that take place within society, or in conflict with it, rather than *society itself*, are to blame for the global instability. If we accept that the flexion of modernity dictates the procedure of reflexive modernisation, the problem arises of how and through which channels *communication* – which Luhmann counts as a systemic condition of society's self-production – should take place, since otherwise society as a system loses control, which constitutes a particular kind of risk. The channels of communication must therefore be channelled through *informational systematisation*.

The information society, which has the power to render the world uniform, communicates in a manner that abolishes the form of *tradition*, as well as the formative entities of *communitas* and *civitas*. This has a direct impact on the institution of the *public sphere* and on the functioning of so-called *civil society*, which makes varied appeals for greater social responsibility, wise decisions, solidarity, human conversion, and revolutionary changes within the social order. But all ethics, politics and social activism of various origins and orientations are powerless to do anything in this respect. We have even come to the point where the militaristic option and the strictly controlled reorganisation of the social base, directed by the highest political instances, remains as the only solution. In Slavoj Žižek's recent book *Too Late to Awaken: What Lies Ahead When There Is No Future?*, we can read:

> The situation is similar across Europe, from Germany to my own Slovenia. To cope with our ongoing, escalating crises, from threats to our environment to unfolding wars, we will need elements of what, in this book, I provocatively call 'war Communism': mobilizations that will have to violate not only the usual market rules but also the established rules of democracy (enforcing measures and limiting freedoms without democratic approval).
>
> A collection of Bertolt Brecht's (largely ignored or forgotten) short interviews and encounters was recently published under the title *Our Hope Today Is the Crisis*. Let's be courageous enough to fully endorse this insight: instead of just trying to escape, postpone, or minimize the threat posed by the four new riders of the apocalypse; instead of continuing to dwell in our melancholic apathy and frantically doing nothing, let's mobilize ourselves to attack the roots of our crisis, with all the risks that this involves. Because the greatest risk today is doing nothing and allowing history to follow its course[14].

[14] S. Žižek, *Too Late to Awaken: What Lies Ahead When There is no Future?*, Allen Lane, London 2023, p. 148.

Žižek, referring to his earlier observations on "risk society"[15], appeals to the need to do something, to take action and mobilise masses of people, also in the style of a soft or hard revolutionary upheaval, instead of snuggly giving in to the course of history, albeit with a bit of kvetching and bleating along the way. Actually, he is right. But what if the key discomfort lies precisely in the "*course of history*," which is unrelentingly sweeping us away, sparing us nothing? It is easy to identify it as "modernity," much more difficult to figure out what drives it, because it is as if there is nothing there, no subjective or substantive basis. But perhaps we must not lose sight of how it is precisely the fact that there is *nothing here* that characterises the *mode* of *modernity*, which nevertheless and with a faceless indifference takes its course and control, staging a total mobilisation. What is essentially absent in this "never ending story" after "the end of history" is the *world-horizon*. Instead of the world, there are its *mobiles*, which, constantly in circulation, are entirely fungible and even useful as a substitute for the historical world. The more we try to proclaim this circulation as modern, the less historical it is in its course, the more a *total social drive* is being processed; this is carrying out a *total mobilisation*, which, "with a little compromise," of course can be called "reflexive modernisation." This in no way abolishes the unrelenting *flexion of modernity*, but merely prepares the power to manage it that belongs to a self-managing society. *The flexion of modernity is systematically reflected in the totalization of social subjectivity*.

If we make use of Baumann's formulation *liquid modernity* in this context, the question arises as to how modernity maintains its *liquidity* while at the same time falling into *delinquency*, even into *liquidations*. Baumann identifies "liquid modernity", which moves with total speed[16] and at the same time never gets anywhere, as a global functioning power or power complex, although he avoids defining it as a *power* that *empowers society itself*:

> The disintegration of the social network, the falling apart of effective agencies of collective action is often noted with a good deal of anxiety and bewailed as the unanticipated 'side effect' of the new lightness and fluidity of the increasingly mobile, slippery, shifty, evasive and fugitive power. But social disintegration is as much a condition as it is the outcome of the new technique of power, using disengagement and the art of escape as its major tools. For power to be free to flow, the world must be free of fences, barriers, fortified borders and checkpoints. Any

[15] See S. Žižek, *Living in the End Times*, Verso, New York-London 2011, pp. 408-409.

[16] P. Virilio, *Open Sky*, trans. by Julie Rose, Verso, London 1997; orig.: *La Vitesse de Liberation*, Éditions Galilée, Paris 1995.

dense and tight network of social bonds, and particularly a territorially rooted tight network, is an obstacle to be cleared out of the way. Global powers are bent on dismantling such networks for the sake of their continuous and growing fluidity, that principal source of their strength and the warrant of their invincibility. And it is the falling apart, the friability, the brittleness, the transcience, the until-further-noticeness of human bonds and networks which allow these powers to do their job in the first place[17].

"Liquid modernity," which trudges over all the borders of the world and yet gets nowhere, must be recognised as *nihilistic* in the sense that it *nihilises the world-horizon*. The nihilisation of the world *mobilises* the *subjectivation of society* such that it risks its own totalization in the function of total domination over a world in which everything – and at the same time nothing – *functions*. The world is totally at disposal, but it appears rotten and indigestible. With the ongoing nihilisation of the world, the subjectivity of society receives from *nothing* the power of over everything, with the risk of falling under everything and receding *back to nothing*. The *society-power* therefore does not tolerate unwillingness towards the world and cannot endure any will at all, let alone "rational action". It is all about the fact that it functions, and goes on, even if it gets nowhere and is without purpose and meaning, which is no problem at all; in fact, by turning everything into a problem, everything that could be a serious problem disappears. All that remains is aimless spinning with the risk of dizziness.

The nihilism of the modern circulation, or, for that matter, of the "postmodern condition"[18], was first recognised by Nietzsche, who writes in an 1887 passage:

It is clear, what I combat is *economic* optimism: as if increasing expenditure of *everybody* must necessarily involve the increasing welfare of everybody. The opposite seems to me to be the case: *expenditure of everybody amounts to a collective*

[17] Z. Bauman, *Liquid Modernity*, Polity, Cambridge 2000, p. 14.

[18] J.-F. Lyotard, *The Postmodern Condition. A Report on Knowledge*, trans. by G. Bennington and B. Massumi, University of Minnesota Press, Minneapolis 1984; orig. *La Condition postmoderne. Rapport sur le savoir*, Minuit, Paris 1979. "Everything is now organized and planned; nature has been triumphantly blotted out, along with peasants, petit-bourgeois commerce, handicraft, feudal aristocracies and imperial bureaucracies. Ours is a more homogeneously modernized condition; we no longer are encumbered with the embarrassment of non-simultaneities and non-synchronicities. Everything has reached the same hour on the great clock of development or rationalization (at least from the perspective of the 'West'). This is the sense in which we can affirm, either that modernism is characterized by a situation of incomplete modernization, or that Postmodernism is more modern than modernism itself" (F. Jameson, *Postmodernism, or, the Cultural Logic of Late Capitalism*, Duke University Press, Durham, NC 1991, p. 383).

loss: man is *diminished* – so one no longer knows *what aim* this tremendous process has served. An aim? a *new* aim? – that is what humanity needs[19].

This is pursued in the following fragment[20]:

"'*Modernity*' through the metaphor of nourishment and digestion."

The circulation of modernity, which Nietzsche defines by comparing it to eating and digestion, lacks a goal but that lack is no obstacle to "economic optimism." By combating it, Nietzsche does not especially take into focus the empowerment of society's subjectivity throughout the entire food chain. Even the sporadic malfunctions of its powerful organism and occasional indigestion in its big belly do not stop the immense appetite of society – as long as it maintains constant control, so that, as Nietzsche's diagnosis of nihilism already suggests, everything *falls below the level*. The self-totalising society's power of *subjectivation, subjectivivation* and *subjection* ("reflexive modernisation", "autopoetic system" and "total mobilisation"), in the function of establishing control over everything, degrades the world and all existing, with a man at the head[21]. This *society in power* is not Nietzsche's will to power, since Nietzsche conceives it as immanent to the world itself, not as its nihilisation. The nihilism that characterises the totalization of *society-power* is not primary destructive, but *nihilising*. The main risk of "risk society" is the *nihilisation of the world-horizon*. This is not to unburden it of its complicity in the various phenomena of destructive violence and the threat of global annihilation,

[19] F. Nietzsche, *The Will to Power*, trans. by W. Kaufmann and R.J. Hollingdale, ed., with commentary, W. Kaufmann with Facsimiles of the Original Manuscript Vintage, New York 1968, p. 464. orig.: F. Nietzsche, *Nachgelassene Fragmente*, KSA 12, ed. by G. Colli and M. Montinari, DTV-de Gruyter, München-Berlin-New York 1988, p. 463.

[20] After the German edition: *Nachgelassene Fragmente 1885-1887*, KSA 12, ed. by G. Colli and M. Montinari, DTV-de Gruyter, München-Berlin-New York 1988, p. 464 ("Die '*Modernität*' unter dem Gleichniß von Ernährung und Verdauung"); english trans. *The Will to Power*, Vintage, New York 1968 p. 47. *Unpublished Fragments, (Summer 1886 - Fall 1887)* as the 17th volume of the *Complete Works of Friedrich Nietzsche* (english translation of the full contents of the *Kritische Studienausgabe*) will not be released until next year.

[21] "The disposition of nihilism is a *liminal* one in that we cannot interpret it in either a psychological or a social perspective as an effect of a negative 'subjective experience', since as such it expresses *the crisis of the dominant self-perception of the humanness* of the human being *as subjectivity*. It manifests itself primarily in the fact that, despite our being 'chained' to all possible information means – which are omnipresent and available to everyone – we fundamentally do not know *what we are witnessing in the world and what is being witnessed as our own existential meaning*" (M. Erzetič, *Vulnerability and Testimony within the Nihilistic Experience and Existential Attestation*, in "Teoria", Rivista di filosofia fondata da Vittorio Sainati XLIII, 1, 2023 (Terza serie XVIII/1), pp. 71-72.

quite the contrary. However, the precondition for these destructive effects is the nihilisation of the world horizon, which is constantly and permanently under ultimate control, so that it can disappear into infinity.

Whether and to what extent Nietzsche himself was already able to recognise the *nihilistic machination in the totalization of society-power* is not overly important here. However, the matter was certainly confronted by those who, "with a hammer" or some other tool, were able to relate to his philosophy, which conceals within itself the riddle of the future[22]. Let us cite in particular Michel Foucault and his consideration of *biopolitics*, which defines the new power of the control society[23]. The biopolitical management of life have been critically examined by Giorgo Agamben in the context of the measures taken to curb the COVID pandemic:

> We can use the term 'biosecurity' to describe the government apparatus that consists of this new religion of health, conjoined with the state power and its state of exception – an apparatus that is probably the most efficient of its kind that Western history has ever known. Experience has in fact shown that, once a threat to health is in place, people are willing to accept limitations on their freedom that they would never theretofore have considered enduring – not even during the two world wars, nor under totalitarian dictatorships[24].

[22] See M. Heidegger, *Who is Nietzsche's Zarathustra?*, in "Review of Metaphysics" 20, 3 (1967) (pp. 411-431), p. 430; orig.: *Wer ist Nietzsches Zarathustra?*, *Vorträge und Aufsätze*, GA 7, Klostermann, Frankfurt am Main, p. 123.

[23] M. Foucault, *The Birth of Biopolitics: Lectures at the College de France, 1978-1979*, trans. by G. Burche, Palgrave Macmillan, New York 2008; orig.: *Naissance de la biopolitique*, *Cours au Collège de France (1978-1979)*, Gallimard, Seuil-Paris 2004. In connection with this, Gilles Deleuze, another Nietzsche acolyte and a contemporary of Foucault, writes: "These are the societies of control, which are in the process of replacing the disciplinary societies. 'Control' is the name Burroughs proposes as a term for the new monster, one that Foucault recognizes as our immediate future. Paul Virilio also is continually analyzing the ultra-rapid forms of free-floating control that replaced the old disciplines operating in the time frame of a closed system. There is no need here to invoke the extraordinary pharmaceutical productions, the molecular engineering, the genetic manipulations, although these are slated to enter into the new process. There is no need to ask which is the toughest or most tolerable regime, for it's within each of them that liberating and enslaving forces confront one another. For example, in the crisis of the hospital as environment of enclosure, neighborhood clinics, hospices, and day care could at first express new freedom, but they could participate as well in mechanisms of control that are equal to the harshest of confinements. There is no need to fear or hope, but only to look for new weapons" (G. Deleuze, *Postscript on the Societies of Control*, in "October" 59, 1992, pp. 3-7, p. 4). See further: G., Burchell, C. Gordon, P. Miller (eds.), *The Foucault Effect*, *Studies in Governmentality*, *with Two Lectures by and an Interview with Michel Foucault*, The University of Chicago Press, Chicago 1991.

[24] G. Agamben, *Where Are We Now? The Epidemic as Politics*, trans. by V.D. Lanham, Rowman & Littlefield, Lanham, etc. (MA) 2021, p. 9; orig.: *A che punto siamo? L'epidemia come politica*, Quodlibet, Macerata 2020.

Here it is not so important whether and to which extent we agree or disagree with Agamben's critique of the social measures taken during the COVID pandemic[25]. Of course, other arguments can be tabled[26], but we should primarily ask ourselves what is being offered here as an alternative? And is there any alternative at all? Or must we cede and acknowledge that the *cooperative power of technology and the technology of power* in the process of the totalization of social subjectivity is without an alternative? Today, this cooperative power is guaranteed in operational terms by *artificial intelligence*, which has immediately proved itself to be an extremely risky tool, if not a weapon, one which no one will forgo because of the profitability it provides, but rather will try to normatively formulate the conditions of its use, almost as if it were a *secret intelligence service*. But until when?

The Doomsday Clock, the model for which was created in 1947 by the artist Martyl Langsdorf through her famous picture, and which was subsequently adopted by scientists working on the atomic bomb within the Manhattan Project to symbolically measure how long humanity has left before total annihilation, stopped at 90 seconds before midnight in 2024. A statement from the *Bulletin of Atomic Scientists*, which maintains the clock's measurement capabilities, pointed out, inter alia:

> One of the most significant technological developments in the last year involved the dramatic advance of generative artificial intelligence. The apparent sophistication of chatbots based on large language models, such as ChatGPT, led some respected experts to express concern about existential risks arising from further rapid advancements in the field. But others argue that claims about existential risk distract from the real and immediate threats that AI poses today […]. Regardless, AI is a paradigmatic disruptive technology; recent efforts at global governance of AI should be expanded.
>
> AI has great potential to magnify disinformation and corrupt the information environment on which democracy depends. AI-enabled disinformation efforts could be a factor that prevents the world from dealing effectively with nuclear risks, pandemics, and climate change[27].

The following questions arise: to what extent does the Doomsday Clock

[25] See S. Sabeva, *Life with the virus. A Phenomenology of Infectious Sociality*, in "Phainomena" 30, 116/117 (2021), pp. 41-60.

[26] J.-L. Nancy, *Eccezione virale*, in "Antinomie. Scritture e immagini", 27/02/2020, https://antinomie.it/index.php/2020/02/27/eccezione-virale/

[27] A *moment of historic danger: It is still 90 seconds to midnight. 2024 Doomsday Clock Statement*, in "Science and Security Board, Bulletin of the Atomic Scientists", https://thebulletin.org/doomsday-clock/current-time/

measure time? and which time? and, in particular, does it measure historical time? and in what way does the effort to raise awareness of the urgent global condition of humanity, which can in fact be healed, intervene in social space – to the degree, of course, that we take into account that the subjectivity of society extends its power over everything, without which AI would not have the chance to elevate. The rupture that looms as the locus of humanity's failure is foresaged in the very process of the AI development; according to the *theory of the technological singularity*, sooner or later there will surely come a moment when AI will outdo human intellect and take over all control functions.

Which we could understand as: no longer will the totalization of subjectivity of the society involve any risk, since the final risk concerning man's position in the world will have been annulated. Well, let us allow ourselves a little remark: given the fact that artificial intelligence can infinitely outdo the natural human intellect,[28] perhaps it would be advisable to *slow down time*, to *give time time*, *to find a place for time*, *to free up time*, just as, long, long ago, for the sake of philosophy, for the sake of wondering at what is and what has its time, it was freed up in the *schole*.

Abstract

In recent decades, discussion about "risk society" has been allotted considerable attention both in studies and among the critically minded public. This article examines risk society in terms of a totalizing society-power of a social subjectivity that erases the world horizon and proves to be nihilistic in its empowerment. Society, as self-establishing, not only follows the course of modernity, which in its unrelenting drifting gets nowhere but also, within its own totalization, risks anticipation of the same. In this connection, the notions of "reflexive modernisation" in Ulrich Beck, of "liquid modernity" in Zygmunt Bauman, and of the "self-production of social systems" in Niklas Luhmann

[28] "A key capability in the 2030s will be to connect the upper ranges of our neocortices to the cloud, which will directly extend our thinking. In this way, rather than AI being a competitor, it will become an extension of ourselves. By the time this happens, the nonbiological portions of our minds will provide thousands of times more cognitive capacity than the biological parts.

As this progresses exponentially, we will extend our minds many millions-fold by 2045. It is this incomprehensible speed and magnitude of transformation that will enable us to borrow the singularity metaphor from physics to describe our future" (Ray Kurzweil, *The Singularity Is Nearer: When We Merge with AI*, Viking, New York 2024, p. 18).

have been particularly adopted for critical discussion. "Risk society" as an "operational concept" presupposes the apparatus of ecosociology, which links eco-nomics, eco-logy and eco-technology and has taken over the executive function of control*, so that everything has the power to function.*

Keywords: risk society; control society; modernity; nihilism; totalisation.

Dean Komel
University of Ljubljana
dean.komel@guest.arnes.si

T

Dimitri D'Andrea

Il cambiamento climatico come minaccia comune. Aspetti cognitivi ed emotivi di una mancata percezione

> Una società che ha per obiettivo la crescita è come un individuo che ha per modello l'obesità.
>
> Luigi Pintor

1. *Una minaccia potenzialmente comune*

Il cambiamento climatico è ormai un fenomeno accertato, misurato e misurabile la cui origine antropica – le attività umane principalmente attraverso le emissioni di gas serra – viene contestata soltanto da un sempre più esiguo gruppo di negazionisti[1]. Si tratta di un fenomeno complesso, cumulativo e fortemente inerziale che consiste nell'incremento della temperatura media globale sulla superficie terrestre[2] e nelle alterazioni del clima ad esso più o meno direttamente connesse. L'impatto sulla vita – umana e non umana – è consistente e rilevante: dal moltiplicarsi degli eventi atmosferici estremi all'aumento della siccità, dalla riduzione della criosfera all'innalzamento del livello dei mari, da nuovi rischi per la salute umana fino alla perdita di biodiversità, dalla acidificazione degli oceani alla riduzione della fauna ittica.

Il cambiamento climatico è un fenomeno globale, ma non definisce attualmente una condizione comune: interessa tutti, ma non tutti allo stesso modo. Non soltanto gli effetti dannosi per gli individui e le società sono ine-

[1] Cfr. IPCC, *Climate Change 2023 Synthesis Report Summary for Policymakers*, p. 4. Disponibile all'indirizzo https://www.ipcc.ch/report/ar6/syr/.

[2] *Ibidem*.

 DOI: 10.4454/rjbh9v14

gualmente distribuiti a livello globale, ma ci sono anche regioni o zone del pianeta che stanno sperimentando (anche) effetti positivi.

Tuttavia, più aumenta la temperatura media terrestre e più le conseguenze divengono universalmente e uniformemente negative. Più si potenzia, cioè, il carattere devastante del riscaldamento globale sul clima e sugli equilibri della biosfera, fino a minacciare, in una prospettiva temporale più lunga (abbondantemente oltre la fine di questo secolo), se non la sopravvivenza della specie umana, sicuramente la civiltà e la condizione umana come l'abbiamo conosciuta[3]. E questo senza che si possa stabilire con certezza il punto di non ritorno: vale a dire la soglia temporale oltre la quale ogni intervento di mitigazione risulterà inutile.

Nel caso del cambiamento climatico, la conoscenza delle cause e degli esiti ultimi si accompagna all'incertezza relativa alla progressione del fenomeno, non soltanto per la rilevanza delle diverse azioni individuali e collettive che verranno o non verranno intraprese per contrastare l'innalzamento della temperatura terrestre, ma anche per il possibile effetto domino che il surriscaldamento globale potrebbe avere su altri fenomeni climaticamente rilevanti come la circolazione degli oceani, o l'estensione del permafrost.

Infine, per il carattere fortemente inerziale del fenomeno, da una parte, la temperatura media terrestre continuerà a crescere per decenni con i relativi effetti sulla biosfera anche se dovessimo adottare oggi politiche di contrasto radicale alle emissioni climalteranti; dall'altra, gli effetti politici delle eventuali scelte di oggi in direzione della neutralità climatica si vedranno soltanto fra decenni[4]. Insomma, il cambiamento climatico ci restituisce uno scarto costitutivo fra chi prende le decisioni e chi ne sperimenta le conseguenze.

Il cambiamento climatico è quindi un fenomeno che – al di là degli effetti sulle condizioni di vita delle generazioni presenti – produrrà un danno *certo* e *radicale* per l'umanità nel suo insieme, sebbene *indeterminato* quanto alle forme e ai tempi. Insomma, conosciamo con certezza l'esito finale del

[3] Sulle previsioni a lungo termine e sull'alternativa fra estinzione della specie umana e distruzione della civiltà cfr. A. Kroll, A. Schlosser, *Will climate change drive humans extinct or destroy civilization?*, disponibile all'indirizzo: https://climate.mit.edu/ask-mit/will-climate-change-drive-humans-extinct-or-destroy-civilization. Per una discussione filosofica di questa alternativa cfr. F. Cerutti, *Global Challenges to Leviathan*, Lexington Books, Plymouth (UK) 2007, pp. 15-17. Per la centralità filosofica dell'estinzione del genere umano cfr., invece, H. Jonas, *Das Prinzip Verantwortung*, Insel Verlag, Frankfurt am Main 1979; trad. it. *Il principio responsabilità*, Einaudi, Torino 2009.

[4] Cfr. IPCC, *Climate Change 2023*, cit., p. 12.

processo di riscaldamento antropogenico del pianeta, anche se non siamo in grado di descrivere esattamente il quando e il come. La scomparsa della civiltà per il venir meno delle condizioni di abitabilità del pianeta come si sono date nella storia umana non è una semplice possibilità o probabilità: è l'esito certo di un processo già in atto. Non un evento che si potrà produrre a partire da qualcosa che potrebbe accadere, ma l'esito *inevitabile* di qualcosa che sta già accadendo, che possiede già le proprie cause efficienti.

In questo senso propongo di distinguere fra rischio e minaccia. Mentre il termine *rischio* si riferisce alla possibilità (la cui entità può anche essere sconosciuta) o alla probabilità di un futuro evento dannoso, il concetto di *minaccia* individua un fenomeno che produrrà inevitabilmente un certo effetto, a meno che non intervengano circostanze o decisioni in grado di interrompere l'attuale corso degli eventi. Perché si produca un inverno nucleare è necessario che almeno due attori politici decidano di impiegare su vasta scala le armi nucleari. Questo implica che si verifichi qualcosa di nuovo, una decisione attualmente soltanto possibile. Nella minaccia, invece, siamo in presenza di un male futuro che è semplicemente un esito necessario: qualcosa che si realizzerà inevitabilmente se tutto continua a funzionare come adesso. La novità che deve intervenire è nel caso del rischio la causa del danno, nel caso della minaccia la condizione per scongiurarlo. Nella minaccia il danno futuro è la conseguenza di uno sviluppo lineare e coerente del presente, di qualcosa già in essere.

Nel caso del cambiamento climatico siamo, dunque, di fronte alla minaccia di un danno che al tempo stesso accomuna tutta l'umanità futura e possiede un'entità tale da non poter essere ritenuto accettabile da nessuno. Nessuno può ritenere la distruzione delle condizioni fisico-naturali della civiltà umana, la distruzione della abitabilità del pianeta qualcosa verso cui essere indifferente (magari perché riguarderebbe soltanto una parte dei futuri abitanti del pianeta) o un costo accettabile rispetto alla garanzia di beni presenti o al conseguimento di altri beni futuri.

Per le sue caratteristiche intrinseche, tuttavia, questa minaccia non configura un interesse egoistico-prudenziale delle generazioni presenti, ma "soltanto" un'obbligazione morale alla garanzia per le generazioni future delle condizioni di abitabilità del pianeta. Insomma, fra i fondamenti motivazionali possibili per il contrasto al cambiamento climatico non possiamo annoverare la paura, l'interesse delle generazioni presenti a proteggersi da danni futuri che le riguardino direttamente. Sia perché questa protezione ha dei costi – e possono esserci valutazioni diverse sulla loro congruità rispetto al beneficio della protezione dagli effetti che il cambiamento climatico potrà

produrre nei prossimi trenta anni –, sia perché questi costi saranno inutili se non ci sarà una identica disponibilità ad agire di tutti o quasi tutti gli attori su scala globale, sia perché i vantaggi delle azioni di mitigazione del cambiamento climatico non ricadranno su coloro che le hanno adottate.

L'intreccio fra questa diversità di interessi su scala globale e la ripartizione temporalmente asimmetrica dei costi e dei benefici mette fuori gioco l'idea che il contrasto al cambiamento climatico possa trovare il proprio fondamento nelle motivazioni che il realismo politico mette tradizionalmente a fondamento dell'agire individuale e collettivo: perseguimento (razionale) del proprio interesse particolare, paura e ricerca della sicurezza, diffidenza e sospetto nei confronti degli altri attori politici.

Confrontarsi con la difficoltà di intraprendere politiche efficaci di contrasto al cambiamento climatico significa dunque ricostruire analiticamente le ragioni della difficoltà di un orientamento etico dell'azione individuale e politica. Non una novità in termini assoluti, ma qualcosa di cui rendere ragione nella specifica modalità che riguarda il cambiamento climatico se vogliamo avere una *chance* di far fronte a questa difficoltà. Si tratta in sostanza di interrogarsi sui concreti meccanismi specifici che conducono a non riconoscere i nostri doveri nei confronti delle generazioni future.

2. *Debolezza cognitiva, insufficienza emotiva, deficit immaginativo*

Un primo fattore su cui si è puntato per comprendere l'assenza di una percezione adeguata della nostra responsabilità nei confronti delle generazioni future è stato non tanto l'idea della infondatezza teorica di un obbligo verso di loro, quanto piuttosto la presa d'atto del carattere inedito con cui si configura il legame fra azione individuale presente e risultato complessivo futuro. Dale Jamieson[5] e Peter Singer[6] hanno ad esempio evidenziato come nel caso del cambiamento climatico il legame fra azione individuale ed effetto dannoso futuro sfidi la tradizionale idea di responsabilità: costruita sulla esistenza di un nesso diretto, lineare e temporalmente e spazialmente ravvicinato fra azione ed effetto dannoso. Nel caso del cambiamento clima-

[5] Cfr. D. Jamieson, *The Moral and Political Challenges of Climate Change*, in S. Moser, L. Dilling (eds.), *Creating a Climate for Change: Communicating Climate Change and Facilitating Social Change*, Cambridge University Press, New York 2007, pp. 475-482.

[6] Cfr. P. Singer, *One World. The Ethics of Globalization*, Yale University Press, New Haven (CT) 2002; trad. it. *One World. L'etica della globalizzazione*, Einaudi, Torino 2003, in particolare pp. 23-24.

tico non soltanto questo nesso viene meno: il danno per le generazioni future non è l'esito di una singola azione riconoscibile, ma dell'effetto cumulativo di milioni di azioni singole in tempi e luoghi remoti.

Si tratta di un ragionamento che coglie sicuramente il modo nuovo in cui si pone la responsabilità nel caso del cambiamento climatico, anche se molti hanno sostenuto con buoni argomenti di svincolare la necessità del prendersi cura dalla responsabilità causale per la condizione delle generazioni future. In sostanza: dobbiamo assumerci la "responsabilità per" anche se non siamo "responsabili di".

Un secondo fattore a cui si è fatto ricorso per spiegare l'indifferenza nei confronti del destino che stiamo preparando per le generazioni future è, invece, di tipo emotivo. La difficoltà di agire nel presente per impedire effetti catastrofici futuri sui nostri discendenti non dipende dal carattere inedito del legame fra le nostre azioni e gli effetti che essi producono, ma dalla debolezza del nostro legame emotivo con loro. Quanto più l'altro si fa lontano nello spazio e nel tempo, tanto più diviene estraneo e tanto più diminuisce la spinta emotiva a farsi carico del suo bene. L'amore del prossimo è anche amore di chi ci è vicino nello spazio e nel tempo. In questo senso, qualcuno che ancora non esiste e che non conosceremo mai sembra essere il pretendente ideale alla nostra indifferenza affettiva, sembra possedere il profilo ideale per risultarci emotivamente estraneo.

Questo ragionamento non tiene, tuttavia, conto del carattere culturalmente condizionato dei nostri affetti morali e del fenomeno storico indubitabile dell'estensione delle solidarietà e degli affetti positivi a cerchie di esseri umani via via sempre più ampie. La vicenda morale dell'umanità è anche quella della estensione della cerchia degli individui emotivamente percepiti come rilevanti e destinatari di sentimenti e passioni di amore al di là dell'ambito familiare e della comunità di vicinato a cui erano limitati in origine i sentimenti empatici e gli affetti e le passioni che spingono alla solidarietà e alla cura. Non siamo, quindi, di fronte ad un argomento decisivo e definitivo, ma semmai alla posizione di un problema che apre alla ricognizione delle condizioni di possibilità di un ampliamento ulteriore della cerchia degli individui emotivamente rilevanti anche alle generazioni future.

Infine, un meccanismo per rendere ragione della grande cecità è quello del deficit immaginativo che caratterizza il rapporto dell'uomo con gli effetti della tecnica di cui è oggi in possesso. È quello che Günther Anders ha chiamato il dislivello prometeico[7]: l'incapacità della nostra immaginazione

[7] G. Anders definisce il dislivello prometeico come «il fatto che non siamo all'altezza del

di fornire una rappresentazione adeguata degli enormi effetti che le nostre capacità tecniche ci mettono in grado di produrre[8]. Produciamo effetti nei confronti dei quali risultiamo indifferenti perché la nostra capacità immaginativa non ci fornisce una adeguata percezione anticipata delle conseguenze delle nostre azioni. Non riusciamo, cioè, a farci un'immagine definita e perciò efficace nelle nostre decisioni, di ciò che siamo diventati capaci di fare. Con questo dispositivo Anders ha fornito una spiegazione dell'indifferenza non soltanto dell'umanità nei confronti del rischio nucleare, ma anche di Eichmann nei confronti degli ebrei che attivamente pianificava di sterminare[9]. Si tratta, tuttavia, di un dispositivo che contribuisce sicuramente a spiegare anche l'indifferenza nei confronti degli effetti delle nostre azioni sul destino delle generazioni future.

In queste prospettive la neutralizzazione della nostra percezione di una obbligazione morale nei confronti delle generazioni future è, dunque, attribuita di volta in volta ad una debolezza cognitiva, ad una insufficienza emotiva e ad un deficit immaginativo. Nel primo caso facciamo fatica a riconoscerci autori del destino delle generazioni future, nel secondo non abbiamo una sufficiente spinta emotiva a farcene carico, nel terzo l'inadeguata percezione della minaccia media un esonero dalla responsabilità, un deficit immaginativo produce la mancata percezione di un dovere.

3. *Immagini del mondo*

Nella inadeguata percezione degli effetti delle nostre azioni e degli obblighi che ne discendono interviene, poi, anche un livello cognitivo più profondo, quello delle immagini del mondo (*Weltbilder*)[10]. A rendere difficile

"Prometeo che è in noi"» (G. Anders, *Die Antiquirtheit des Menschen. I*, C.H. Beck, München 1956; trad. it. *L'uomo è antiquato. I*, Bollati Boringhieri, Torino 2007, p. 253), ovvero «l'*asincronizzazione* ogni giorno crescente *tra l'uomo e il mondo dei suoi prodotti*» (*ivi*, p. 24).

[8] Icastica la formulazione del dislivello prometeico richiamata in *L'uomo è antiquato. II*: «lo scarto tra il massimo che possiamo produrre e il massimo (vergognosamente piccolo) che possiamo immaginare» (G. Anders, *Die Antiquirtheit des Menschen. II*, C.H. Beck, München 1980; trad. it. *L'uomo è antiquato. La terza rivoluzione industriale*, Bollati Boringhieri, Torino 1992, p. 12.

[9] Cfr. G. Anders, *Wir Eichmannsöhne*, C.H. Beck, München 1964; trad. it. *Noi figli di Eichmann*, Giuntina, Firenze 1995.

[10] Per il concetto di *Weltbild* nell'accezione qui utilizzata si veda M. Weber, *Einleitung zur Wirtschaftsethik der Weltreligionen* (1915-16; 1920-21), in Id., *Gesammelte Aufsätze zur Religionssoziologie*, Mohr Siebeck, Tübingen 1920-1921, vol. I (Max Weber-Gesamtausgabe I/19, H. Schmidt-Glintzer (Hrsg.), Mohr Siebeck, Tübingen 1989); trad. it. *Introduzione* a

la percezione della nostra responsabilità per il cambiamento climatico e le sue conseguenze sulle generazioni future c'è anche e soprattutto un sistema complessivo di credenze che riguarda il mondo nella sua totalità e in particolare il rapporto uomo-natura e natura-società. Più che la mancanza di consapevolezza di specifici nessi causali, l'indifferenza emotiva o una carenza immaginativa, sul banco degli accusati si trova in questo caso l'immagine del mondo di una modernità che, nella diversità delle interpretazioni, risulta tuttavia caratterizzata da un pensiero dicotomico (soggetto-oggetto, natura-società, spirito-materia), dall'idea di un illimitato dominio sulla natura, da un pensiero gerarchico e antropocentrico.

Bruno Latour è stato fra coloro che con più forza e coerenza hanno indicato nella concezione moderna della natura la ragione di fondo della nostra incapacità di cambiare il nostro modo di vivere e di affrontare la sfida del cambiamento climatico[11]. È una certa immagine (moderna) del mondo, un certo modo di pensare la realtà che rende inconcepibile – e, quindi, impossibile – un nuovo regime climatico. Affinché le cose possano iniziare a cambiare, occorre che cambi il nostro modo di concepirle. Da qui l'insistenza latouriana sulla necessità di superare le dicotomie moderne e l'idea di una natura passiva e inerte, di pensare la realtà come insieme di reti di dipendenza, di immaginare un'alleanza fra umani e non umani per convivere all'interno dell'unico pianeta disponibile.

La consapevolezza che la radice ultima delle nostre difficoltà con il cambiamento climatico sia da rinvenire nel *Weltbild* dei moderni è alla base dei molteplici tentativi filosofici di proporre immagini dell'uomo e del mondo alternative a quelle della modernità che puntano di volta in volta sull'irriducibilità della soggettività umana al modello utilitaristico[12], sulla necessità di pensare un soggetto vulnerabile e relazionale[13], sul su-

L'etica economica delle religioni universali, in Id., *Sociologia della religione*, a cura di P. Rossi, 4 voll., vol. II, Comunità, Torino 2002, pp. 19-21; H. Blumenberg, *Weltbilder und Weltmodelle*, in «Nachrichten der Giessener Hochschulgesellschaft», XXX, Schmitz, Gießen 1961, pp. 67-75; trad. it. *Immagini del mondo e modelli di mondo*, in «Discipline filosofiche», 1, XI (2001), pp. 13-23; D. D'Andrea, *Soggettività e immagini del mondo in Max Weber*, in «Iride», 65, XXV (2012), pp. 5-24.

[11] Cfr. B. Latour, *Nous n'avons jamais été modernes*, La Dècouverte, Paris 1991; trad. it. *Non siamo mai stati moderni*, Elèuthera, Milano 1995 e Id., *Face à Gaïa: Huit conférences sur le Nouveau Régime Climatique*, La Dècouverte, Paris 2015; trad. it. *La sfida di Gaia*, Meltemi, Milano 2020, in particolare pp. 73-116 e 259-305.

[12] Cfr. A. Caillé, *Critique de la raison utilitaire*, La Dècouverte, Paris 1989; trad. it. *Critica della ragione utilitaria*, Bollati Boringhieri, Torino 1991.

[13] Cfr. E. Pulcini, *La cura del mondo*, Bollati Boringhieri, Torino 2013.

peramento dell'antropocentrismo e della dicotomia umano-non umano[14]. Tutte immagini del mondo comunque accomunate dalla scelta di pensare la relazione, la dipendenza, il limite dove la modernità aveva collocato la centralità dell'umano, il dominio della natura, l'irrilevanza del non umano, la gerarchizzazione degli enti.

Si tratta sicuramente di immagini del mondo che, al di là delle profonde differenze che in qualche caso le separano, risultano più adeguate alla realtà della crisi climatica e restituiscono un modo di intendere il mondo capace, in teoria, di indurre comportamenti individuali e collettivi più in linea con le esigenze di una presa in carico del destino delle generazioni future. E, tuttavia, altrettanto evidente è la loro scarsa capacità di diffusione, la resistenza che l'adozione di una diversa immagine del mondo incontra nonostante la sua maggiore adeguatezza allo stato del mondo almeno dal punto di vista del cambiamento climatico. Quali sono le ragioni di questa difficoltà? Che cosa ostacola il superamento dell'immagine moderna del mondo, nonostante la sua inadeguatezza alla comprensione e al contrasto della crisi climatica?

Per rispondere a questa domanda e più in generale per affrontare la questione dell'indifferenza etica nei confronti della crisi climatica e dei suoi effetti sulle generazioni future può essere utile confrontare le attuali difficoltà nella costruzione di un regime globale di controllo delle emissioni di gas serra con la soluzione positiva che, invece, è stata data ad un'altra minaccia ambientale potenzialmente globale: il deperimento della fascia di ozono.

Con l'applicazione delle misure previste dal Protocollo di Montreal del 1987 e le successive integrazioni, il deperimento della fascia di ozono può essere ormai considerato sotto controllo. Con il bando dei clorofluorocarburi (CFC) e la loro sostituzione con gli idrofluorocarburi (HFC) e gli idroclorofluorocarburi (HCFC), le cause del deperimento della fascia di ozono sono, infatti, state rimosse, anche se – per l'inerzia del fenomeno – una ricostituzione completa della fascia di ozono stratosferico non è prevista prima di cinquanta anni. Ciò che ha reso possibile questa *success story* è stata la *credibilità* che una strategia di contrasto del fenomeno possedeva per alcune ragioni di fondo: in primo luogo, la dipendenza del fenomeno dall'immissione in atmosfera di una singola sostanza (i CFC) e il numero relativamente limitato di prodotti industriali nei quali tale sostanza era utilizzata; in secondo luogo, la disponibilità di una tecnologia sostitutiva a parità di prestazioni e

[14] Cfr. D. Haraway, *Staying with Trouble. Making Kin in the Chthulucene*, University of Chicago Press, Chicago 2016; trad. it. *Chthulucene. Sopravvivere su un pianeta infetto*, Nero, Roma 2019.

di costi. *Semplicità* delle cause, fattibilità tecnica della sostituzione, *limitatezza* dei costi, compatibilità con il normale funzionamento dell'economia e con lo stile di vita acquisito hanno contribuito a determinare la credibilità di una soluzione del problema che è stata condizione di possibilità della soluzione stessa.

Ma queste sono proprio le caratteristiche che mancano ad una strategia efficace e coerente di contrasto al cambiamento climatico.

4. *La serra del comfort e la cura del mondo*

Quasi quarant'anni fa Ulrich Beck ha definito i rischi ambientali nel loro insieme come una sorta di passeggeri clandestini del consumo di tutti i giorni[15], enfatizzando il legame controintenzionale fra strategie e atti di consumo e distruzione degli equilibri della biosfera. Il capitalismo è la forma di vita che ha trasformato il mondo in un Palazzo di cristallo e in una serra di comfort[16] in cui agli esseri umani vengono offerti beni di consumo in abbondanza in una misura mai vista prima nella storia dell'umanità. L'effetto serra è il prezzo della serra del comfort: l'esito complessivo dell'economia capitalistica globale dominata dalla logica di un incremento illimitato dei consumi e di una accelerazione senza fine del ciclo produzione-consumo[17]. La pandemia di CoViD-19 ha mostrato, in modo eclatante, gli effetti disastrosi di una drastica riduzione dei consumi sull'economia. Durante il primo lockdown, molti hanno sottolineato gli effetti positivi sulla qualità dell'aria e dell'acqua e, in generale, sugli indicatori dell'impatto umano sui processi naturali. Il disastro economico che ne è derivato testimonia però l'incompatibilità tra economia capitalistica e riduzione dei consumi, tra capitalismo globale e qualunque consapevole assunzione della finitezza delle risorse. Il mantra della crescita, così come viene recitato religiosamente e unanimemente dalla politica su scala globale, è la testimonianza più evidente dell'impossibilità di separare capitalismo e crescita dei consumi. Ma questo

[15] Cfr. U. Beck, *Risikogesellschaft. Auf dem Weg in eine andere Moderne*, Suhrkamp, Frankfurt am Main 1986, trad. it. *La società del rischio*, Carocci, Roma 2000, p. 53.

[16] Per l'origine dell'immagine del Palazzo di cristallo come simbolo della civiltà occidentale cfr. F. Dostoevskij, *Zapiski iz podpol'ja*, 1864; trad. it. *Memorie del sottosuolo*, Einaudi, Torino 2005; per la ripresa dell'immagine cfr. P. Sloterdijk, *Weltinnenraum des Kapitals*, Suhrkamp, Frankfurt am Main 2005; trad. it. *Il mondo dentro il capitale*, Meltemi, Roma 2006, pp. 38-43.

[17] Cfr. H. Rosa, *Beschleunigung und Entfremdung*, Suhrkamp, Frankfurt am Main 2013; trad. it. *Accelerazione e alienazione*, Einaudi, Torino 2015.

confligge con l'impossibilità della crescita materiale infinita in un ambiente finito, con l'assenza delle risorse fisiche e biochimiche per sostenere una forma di vita di questo tipo: «ce ne vorrebbero molti, di pianeti; ma purtroppo ce n'è uno solo»[18].

Le brutte notizie cominciano con la constatazione che l'incremento della disponibilità di beni, l'espansione dei consumi, l'accesso a nuove esperienze di piacere che il capitalismo su scala globale offre – o anche soltanto promette – è ciò che meglio risponde, in termini di relazioni materiali – economiche e sociali in senso lato –, alla moderna idea di felicità, o, in modo meno enfatico, di benessere. E non si tratta solo di consumo. Il capitalismo è anche la civiltà della libertà negativa, dell'emancipazione dai legami personali. Il denaro è il più impersonale dei poteri che dominano la nostra vita, ma anche un potente fattore di emancipazione dal peso dei legami personali e comunitari. Questa promessa di benessere e libertà spesso non viene mantenuta, ma è comunque realizzata per molti e desiderata da tutti. Libertà e denaro sono universali indifferenti capaci di servire fini estremamente diversi e perfettamente adeguati per preservare o espandere quella singolarità assoluta del sé che è al centro della nostra immagine del mondo.

Infine, come affermava Max Weber, il capitalismo è un cosmo[19]: una totalità non scomponibile, un sistema unitario e completo di regole che non si può adottare solo parzialmente. Un cosmo è un insieme di parti che non può essere scomposto o disaggregato. In ciò risiede il fondamento della schiacciante coercizione che è in grado di esercitare e per questo Weber lo descriveva come una *gabbia d'acciaio*[20].

Le leggi del mercato non si possono sospendere o cambiare, se si vuole che il mercato continui a garantire le sue prestazioni: non sono «una carrozza che si possa far fermare a piacere per salirvi o scenderne»[21]. L'ordine economico capitalista è governato da una logica che può essere contenuta, ma non modificata. E questo è tanto più vero in una condizione globale in cui gli attori sono ormai disseminati per tutto il pianeta. Forme di regolazio-

[18] B. Latour, *Où atterrir?*, La Découverte, Paris 2017; trad. it. *Tracciare la rotta*, Raffaello Cortina, Milano 2018, p. 13.

[19] Cfr. M. Weber, *Die protestantische Ethik und der Geist des Kapitalismus* (1904-05; 1920-21), in Id., *Gesammelte Aufsätze zur Religionssoziologie*, Mohr Siebeck, Tübingen 1920-1921; trad. it. *L'etica protestante e lo spirito del capitalismo*, in Id., *Sociologia della religione*, a cura di P. Rossi, 4 voll., vol. I, Comunità, Torino 2002, pp. 184-185.

[20] *Ibidem*.

[21] M. Weber, *Politik als Beruf* (1919), Max Weber-Gesamtausgabe I/17, W.J. Mommsen, W. Schluchter (Hrsg.), Mohr Siebeck, Tübingen 1992; trad. it. *Politica come professione*, in Id., *Scienza come professione. Politica come professione*, Einaudi, Torino 2004, p. 107.

ne del funzionamento del capitalismo come quelle conosciute negli Stati-nazione del XX secolo o cambiamenti radicali nell'organizzazione complessiva della vita sociale e del modo di produrre sono inimmaginabili sia su scala nazionale che su scala globale. Con le parole di Fredric Jameson: «è più facile immaginare la fine del mondo che la fine del capitalismo»[22].

Al di là della questione specifica del capitalismo, Peter Sloterdijk ha persuasivamente sottolineato come l'incremento della complessità dei fenomeni attraverso l'ampliamento delle interdipendenze genera inerzia e rende difficilmente immaginabili trasformazioni radicali o legate a propositi unilaterali[23]. Complessità e interdipendenza sono riduttori di libertà, fattori che inibiscono l'autonomia nella misura in cui annullano lo spazio vuoto di un'azione che non produce immediatamente effetti su – e reazioni da – altri. Un mondo saturo di relazioni è un mondo dominato dalla logica dell'adattamento e della evoluzione senza progetto. Più i fenomeni sono complessi più diviene difficile immaginare una loro trasformazione in conformità ad un disegno o ad un principio.

L'assunzione delle nostre responsabilità nei confronti degli effetti del cambiamento climatico sulle generazioni future ha, così, non soltanto dei costi incomparabili con quelli del contrasto al deperimento della fascia di ozono stratosferico, ma anche una complessità immaginativamente ingovernabile. Non soltanto implica la rinuncia a qualcosa che soddisfa i nostri desideri presenti, ma costringe a confrontarsi con la prospettiva di un cambiamento di organizzazione complessivo della vita sociale difficilmente immaginabile.

5. *Indesiderabilità, inimmaginabilità e diniego*

Contrastare il cambiamento climatico impone un ripensamento radicale della nostra forma di vita e questo non è percepito *in modo generalizzato* come desiderabile. È l'abbondanza – o la promessa di abbondanza – di beni materiali e di libertà che ci "incatena" all'esistente e rende così difficile trascendere il capitalismo.

Le difficoltà emotive e cognitive alla assunzione della nostra obbligazione morale nei confronti delle generazioni future non sono legate alle caratteristiche intrinseche della minaccia, ma alla circostanza che la presa d'atto del

[22] Cfr. F. Jameson, *Future City*, in «New Left Review», 21, May-June 2003.
[23] Cfr. P. Sloterdijk, *Il mondo dentro il capitale*, cit., pp. 39-41.

danno che infliggiamo alle generazioni future confligge con il soddisfacimento dei desideri – di beni e libertà – di quelle presenti. Detto altrimenti: non è un problema strettamente emotivo o cognitivo, ma rimanda all'esistenza di un conflitto fra consapevolezza e responsabilità, da una parte, e interessi e desideri, dall'altra. Le difficoltà emotive e cognitive al contrasto al cambiamento climatico possono essere comprese in gran parte come l'esito delle capacità distorsive dei desideri presenti o anche soltanto delle aspettative di un loro possibile soddisfacimento. La debolezza cognitiva, l'insufficienza emotiva e il deficit immaginativo che scontiamo nei confronti degli effetti del cambiamento climatico e delle conseguenze che ne derivano sulle generazioni future dipendono dalla condizione materiale che ci costringerebbero a rimettere in discussione.

Discorso in parte analogo può essere fatto per le immagini del mondo, ovvero per quegli assunti cognitivi che orientano il nostro rapporto con il mondo. La credibilità di un *Weltbild* non è mai un fenomeno esclusivamente teorico. La difficoltà di modificare un'immagine del mondo incentrata sul dominio illimitato dell'uomo sulla natura non è fondata sulla sua solidità teorica o sull'assenza di evidenze che stridono con questa immagine, siano esse quelle che emergono nella letteratura scientifica sul cambiamento climatico o quelle legate all'esperienza biografico-soggettiva del deterioramento della qualità dell'ambiente. Gli 'argomenti' migliori a sostegno di un'immagine dell'umanità come sovrano assoluto di una natura inerte e indifferente sono di tipo, per così dire, performativo. Consistono, cioè, in ciò che questa immagine ci consente di fare. La solidità di questa immagine del mondo non si basa sull'evidenza dei suoi 'articoli di fede', ma sul fatto che è l'immagine del mondo implicita in una forma di vita che ha garantito livelli di benessere e libertà senza precedenti. Ed è contro questo 'argomento' che devono misurarsi coloro che perseguono la creazione di un nuovo regime climatico.

Se tra le condizioni che hanno reso possibile la modernità e il capitalismo dobbiamo includere anche la distruzione dell'antica idea di cosmo e la concezione del mondo come illimitato campo dell'autoaffermazione umana[24], l'intreccio fra sviluppo economico e libertà costituisce oggi il miglior argomento per quell'immagine del mondo che per molti altri versi potremmo definire fuori tempo massimo. Una certa immagine del mondo è stata essenziale per la nascita del capitalismo e ora il capitalismo è diventato il principale sostegno di un certo *Weltbild*.

[24] Cfr. H. Blumenberg, *Die Legitimität der Neuzeit*, Suhrkamp, Frankfurt am Main 1974; trad. it. *La legittimità dell'età moderna*, Marietti, Genova 1992, pp. 143-146.

Ma i nostri problemi con i doveri nei confronti delle generazioni future non si limitano alla indesiderabilità della fuoriuscita da una forma di vita incentrata su sempre più consumo e sempre più libertà. Negli ultimi decenni si sono sicuramente sviluppate teorie e pratiche che contestano che la dinamica del 'sempre di più', 'sempre più veloce' sia in grado di assicurarci felicità o benessere. Sono nate riflessioni ed esperienze che hanno indicato in altre forme di abitare il pianeta non soltanto un modo per far fronte ai nostri obblighi nei confronti di chi verrà dopo di noi, ma anche per una esistenza più felice.

In questo caso, a neutralizzare la spinta al cambiamento non è la sua indesiderabilità, ma la sua inimmaginabilità: l'impossibilità di immaginare una forma e un percorso realistico di cambiamento generalizzato, che vada oltre le singole esperienze e pratiche individuali o di piccoli gruppi. La difficoltà nell'impostare una risposta efficace al cambiamento climatico è legata in questo caso alla difficoltà di immaginare un altro ordine economico in cui l'economia costituisca una funzione della società[25] – e non viceversa –, in cui gli umani siano consapevoli degli effetti causati dalle controazioni dei non umani e in cui la riduzione dei consumi non debba essere interpretata attraverso la lente della povertà.

Questo deficit immaginativo – l'incapacità di immaginare un diverso sistema economico o anche solo un nuovo equilibrio tra mercato e società, e tra umanità e ambiente – costituisce il modo migliore per perpetuare l'esistente in quanto induce la rimozione della percezione del problema. L'inimmaginabilità delle soluzioni induce il diniego: «alle persone, alle organizzazioni o ad intere società sono fornite informazioni troppo inquietanti, minacciose o anomale perché siano interamente assorbite o apertamente riconosciute. Pertanto, tale informazione è rimossa, negata, messa da parte o reinterpretata. Oppure essa viene sufficientemente "registrata", ma le sue implicazioni – cognitive, emotive o morali – sono evitate, neutralizzate o razionalizzate»[26]. Il caso più evidente e macroscopico di diniego è costituito dal rischio di un conflitto nucleare su (relativamente) vasta scala: nessuno ha argomenti per essere indifferente a questa eventualità, ma la nostra vita continua scorrere come se il problema non esistesse, perché nessuno riesce ad immaginare una soluzione realistica al problema.

La lentezza, la debolezza e l'inefficacia della lotta contro il cambiamento

25 Cfr. K. Polanyi, *The Great Transformation. The Political and Economic Origins of Our Time*, Rinehart, New York-Toronto 1944; trad. it. *La grande trasformazione*, Einaudi, Torino 2000.

26 S. Cohen, *States of Denial. Knowing about Atrocities and Suffering*, Polity Press, Cambridge 2001; trad. it. *Stati di negazione. La rimozione del dolore nella società contemporanea*, Carocci, Roma 2002, p. 23.

climatico dipendono, quindi, non solo dalla non negoziabilità per molti di un certo stile di vita, ma anche dalla rassegnazione frustrata di chi auspicherebbe un nuovo regime climatico ma non riesce a percepirne la possibilità.

6. *Condizioni di possibilità e ruolo della politica*

Negli ultimi decenni, nelle nostre società si sono diffuse pratiche individuali e collettive che vanno nella direzione di costruire un'economia e una forma di vita capaci di prendersi cura della casa comune: l'autogestione dei beni comuni, i gruppi di acquisto solidale, la *sharing economy*, le pratiche contro la cultura dello spreco, il *downshifting* e le 'economie diverse'[27]. Tuttavia, queste esperienze non hanno contaminato la politica: non hanno ispirato le *policies*, né modificato l'agenda politica, almeno a livello di Stati nazionali. Nella lotta al cambiamento climatico, la grande assente è stata la politica. Le istituzioni politiche si sono rivelate inospitali, incapaci di valorizzare, sostenere o promuovere le pratiche ecologiche sperimentate nella società. Nella sua generalità, la rappresentanza parlamentare è stata un fattore decisivo nell'emarginare le istanze di un altro modo di vivere e produrre: troppo generale per aderire ai contesti e incapace di sfuggire alla logica del breve periodo e alla centralità di altre questioni. E, tuttavia, la politica – regolazione *erga omnes* – resta uno snodo critico, una dimensione imprescindibile per la costruzione di un nuovo regime climatico. La scelta etica individuale e lo spostamento del soggetto verso l'assunzione di responsabilità restano passaggi fondamentali, ma l'etica da sola non basta a produrre il cambiamento necessario[28]. Per contrastare efficacemente il cambiamento climatico è necessario immaginare e attuare modi diversi di vivere, produrre, muoversi e consumare. In altre parole, le scelte politiche sono necessarie per produrre nuove forme di regolamentazione (*policies*). Tuttavia, la possibilità di nuove *policies* dipende anche dalla costruzione di nuove istituzioni politiche (*politics*). Per cambiare le politiche è anche necessario (prima) cambiare la politica, cioè la forma e il funzionamento delle istituzioni politiche. Un diverso assetto istituzionale è essenziale per aprire gli spazi di manovra e gli orizzonti di possibilità indispensabili alla mobili-

[27] Cfr. M. Deriu, *Verso un'intelligenza compositiva. Il comune multiplo politico delle economie solidali*, in L. Bertell, M. Deriu, A. De Vita, G. Gosetti, *Davide e Golia. La primavera delle economie diverse*, Jaca Book, Milano 2013, pp. 34-63.

[28] A. Ghosh, *La grande cecità*, cit., parte III.

tazione individuale così come alla conversione interiore nella direzione di una forma di vita responsabile verso il pianeta e le generazioni future. Se «il nuovo regime climatico non ha una istituzione condivisa»[29], si tratta, allora, di trovare istituzioni politiche capaci di farlo esistere.

La mia idea è che oggi le istituzioni dello Stato-nazione sono diventate inadeguate per il cambiamento necessario e che la direzione della loro trasformazione dovrebbe andare non soltanto verso il 'più grande', ma anche verso il 'più piccolo'. In altre parole, la possibilità di cambiamento e lo spazio di manovra devono essere intravisti nell'orizzonte della prossimità piuttosto che in quello della globalità. La sostenibilità globale può essere prodotta solo se ogni luogo è sostenibile, ma ogni luogo ha dimensioni, qualità e forme di vita diverse: «non si può pensare a ricette uniformi, perché ci sono problemi e limiti specifici di ogni Paese e regione»[30]. Per immaginare e produrre una riconversione ecologica è necessario ridurre la complessità delle interdipendenze e realizzare strategie di regolazione a livello locale, che passano essenzialmente attraverso la ridefinizione degli spazi politici e la trasformazione della loro natura. In questa prospettiva, si tratta di ripensare non solo le dimensioni delle istituzioni politiche fondamentali, ma anche di ridefinirne la logica immaginando spazi politici funzionalmente delimitati e a geometria variabile. È necessario muoversi nella direzione di assetti politico-istituzionali che vadano oltre le forme della modernità politica: lo Stato-nazione come dimensione standard dello spazio politico e la sovranità come descrittore della sua natura-qualità. Questo processo deve recuperare l'idea di democrazia come autogoverno delle comunità locali e superare il carattere generale della rappresentanza politica. È necessario, da un lato, andare oltre l'idea che la dimensione dello Stato-nazione costituisca il luogo unico della democrazia; dall'altro, mettere in discussione una caratteristica finora indiscussa della rappresentanza democratica a tutti i livelli, vale a dire la sua fisionomia generalista e unitaria. In altre parole, dobbiamo prendere le distanze dall'idea che la volontà popolare debba esprimersi nella forma di una volontà generale unitariamente rappresentata in un Parlamento. Il pensiero democratico ha talvolta cercato di superare il carattere rappresentativo della volontà popolare, ma non ha mai, nemmeno nelle sue versioni più radicali, dubitato che la sua forma di esistenza dovesse essere unitaria e generale.

[29] B. Latour, *Tracciare la rotta*, cit., p. 118.

[30] Papa Francesco, *Enciclica Laudato sì*, disponibile all'indirizzo: https://www.vatican.va/content/dam/francesco/pdf/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si_it.pdf2015, § 180, p. 162.

Serve, dunque, una maggiore integrazione europea, ma servono anche maggiori livelli di autonomia dei luoghi. Serve più integrazione nella regolazione delle grandi variabili macroeconomiche e nella garanzia dei diritti, ma anche più protezione delle società locali dal mercato globale, più autonomia alle comunità per progettare nuovi modi di vivere e di produrre.

English title: Climate change as a common threat. Cognitive and emotional aspects of a lack of perception

Abstract

This contribution proposes a discussion on the anthropological factors – emotional and cognitive – that hinder the adoption of suitable and effective responses to the threat that climate change poses to the present and especially future health and human condition. In other words, it reflects on what Amitav Ghosh has called the great blindness *and which we could describe in terms of an inadequate perception of the radical threat to future generations that results in the omission of a duty to them. The reasons for this failure to recognise this lie not so much in the intrinsic characteristics of the threat – the certain but indeterminate nature of the long-term outcomes of the phenomenon, the enormity of the effects, etc. – but rather, on the one hand, in the fact that the threat is a threat to the future generations, and on the other hand, in the fact that it is a threat to the future generations. – but rather, on the one hand, in the emotional and cognitive distortions stemming from the costs (in terms of consumption and freedom) that combating climate change would force us to bear; on the other, in a difficulty in imagining realistic and comprehensive solutions to the problem that translates into denial.*

Keywords: Climate change; future generations; risk; threat; imagination.

Dimitri D'Andrea
Università di Firenze
dimitri.dandrea@unifi.it

T

Marco Emilio

The Collective Challenge of Interlocked Risks

1. *New Risks and Collective Agency*

Risk overlapping is a peculiar challenge of global crises. Despite being spatially distant, local hazardous events are often interdependent and cannot easily be disjointed. The frequency of this kind of events is growing in scale and serves as a vector to investigate the various disciplinary notions of risk, in order to develop comprehensive theoretical frameworks for knowing, assessing, and deciding.

More specifically, international institutions, scholars, and practitioners tend now to pair, for instance, ecological[1] and social risk concepts. Nonetheless, this association is challenging. First, climate risks[2] are related to non-linear mechanisms, and pairing them with anything can deepen the complexity in decision-making processes to treat them. Second, the structural intersection of global and local factors varies contextually and engenders different hazards and vulnerabilities. In other terms, understanding how to inhabit the social impact of ecological crises can be conceived as a "wicked problem" that requires a thorough investigation of the different notions of

[1] The terms "ecological risks" and "environmental risks" (see § 3) have been used in similar ways in the scientific literature and institutional documents. The USA tends to prefer the former, and Europe the latter (see G.W.I. Suter, *Ecological Risk Assessment*, CRC Press, Boca Raton (FL) 2016). In the following, I will mainly adopt "ecological risk," broadly referring to risks pending onto non-human organisms.

[2] As it will be clarified below (§ 3), "climate risks" are risks that are engendered by global warming (see IPCC, *Climate Change 2023: Synthesis Report (Full Volume) Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Intergovernmental Panel on Climate Change, Geneva 2023).

 DOI: 10.4454/rmw1qs16

risks at stake with their epistemological and ethical interrelations. This is particularly salient in the effort of preventing global risks, as this demands several coordinated activities, usually developed by collective actors such as groups of researchers, public institutions, and NGOs[3]. Shared awareness regarding hazards is also routinely evoked as a crucial factor in coping with climate crises[4]. Following this line of thought, philosophical contributions on collective epistemology and responsibility might shed light on these issues[5] when different risks are considered in conjunction. However, investigations on risk and collective agency have yet to gain systematicity.

This article examines how we can account for contexts where ecological and social risks overlap. It will be argued that a coherent theoretical image of different kinds of risk should consider how knowledge production, knowledge sharing, and decision-making involve collective agencies.

The investigation will start by analyzing some standard accounts of ecological and social risks, and defining a few open theoretical issues recently highlighted within the scientific literature concerning climate policies. Hence, a critical examination of the links between risk assessment and decision-making processes will be sketched. In the third step, it will be suggested that the notion of collective epistemic responsibility[6] can play a crucial role in investigating risk communication and decision-making processes that involve non-expert laypeople. As a last move, a few implications will be outlined regarding interdisciplinary investigation[7] on risk and expertise.

2. *Interweaving of Risks*

Consider the following case. In April 2022, the regional governmental agency of Tuscany, Italy, claimed the compatibility of a new geothermal

[3] IPCC, *Climate Change 2023: Synthesis Report (Full Volume)*, cit.

[4] F.S. Khatibi *et al.*, *Can Public Awareness, Knowledge and Engagement Improve Climate Change Adaptation Policies?*, in «Discover Sustainability» 2 (2021) n. 1, pp. 1-24.

[5] See S.O. Hansson, *A Panorama of the Philosophy of Risk*, in S. Roeser (ed.), *Handbook of Risk Theory: Epistemology, Decision theory, Ethics, and Social Implications of Risk*, Springer Science & Business Media, Dordrecht 2012, pp. 27-54; S.O. Hansson, *Risk*, in E.N. Zalta, U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University Summer 2023, https://plato.stanford.edu/archives/sum2023/entries/risk/

[6] W. Fleisher, D. Šešelja, *Responsibility for Collective Epistemic Harms*, in «Philosophy of Science» 90 (2023) n. 1, pp. 1-20.

[7] J. Persson *et al.*, *Toward an Alternative Dialogue Between the Social and Natural Sciences*, in «Ecology and Society» 23 (2018) n. 4, pp. 1-11.

power plant on Mount Amiata[8], built by the multinational corporation Sorgenia, and the official energy transition strategy[9]. Nonetheless, recent sociological field investigations have observed rising social tensions linked to the project:

> Both the committees opposed to [geothermal] cultivation or some of its methods, as well as the plant operators and others in favor (most local governments, regional institutions, various experts, and technicians) use data and information to give strength to their arguments […]. For example, about the possible problems of dispersion and emission of chemicals and CO_2 and consequent impacts on health, the land, and the climate […]. In this context, expert opinions […] have become tools for developing conflicting plausible narratives[10].

At first glance, this case concerns a situation where a policy aimed at tackling global warming engenders different, overlapping risks. As a result, decision-making processes increase conflicts between parties with divergent knowledge sets and values, thus potentially slowing down the achievement of global emissions targets (i.e., the paramount goal of the policy itself). Social policy researchers[11] classify this situation as an instantiation of climate change's «superwicked problems»[12].

[8] «At Mt. Amiata (Italy) geothermal energy is used, since 1969, to generate electricity in five plants» (E. Bacci *et al.*, *Geothermal Power Plants at Mt. Amiata (Tuscany-Italy): Mercury and Hydrogen Sulphide Deposition Revealed by Vegetation*, in «Chemosphere» 40 (2000) n. 8, pp. 907-911, p. 907).

[9] Redazione t24, *Via libera alla centrale geotermica Sorgenia - Pronuncia positiva di compatibilità ambientale della Regione all'impianto sul Monte Amiata, ma la Soprintendenza potrebbe opporsi*, t24 Il quotidiano economico toscano, April 23, 2022, https://t24.ilsole24ore.com/art/geotermia-ok-della-regione-al-progetto-sorgenia-amiata (accessed December 11, 2023).

[10] «Nel caso geotermia, tanto i comitati che si oppongono alla coltivazione o alcuni suoi metodi, quanto i gestori degli impianti e altri soggetti favorevoli (gran parte delle amministrazioni locali, istituzioni regionali, vari esperti e tecnici) utilizzano dati e informazioni per dare forza alle proprie ragioni. […] Per esempio, in relazione agli eventuali problemi di dispersione ed emissione di sostanze chimiche e CO_2 e conseguenti impatti su salute, territorio e clima […]. In tale contesto, i pareri esperti e i differenti tempi e modi dei processi conoscitivi degli attori sono divenuti strumenti per l'elaborazione di storie plausibili contrastanti» (M. Villa, *Cambiare o traccheggiare? Politica e lavoro eco-sociale, transizione ecologica e la sfida della complessità: note di campo*, in E. Matutini (ed.), *Eco-social-work*, PM edizioni, Varazze (SV) 2023, pp. 33-88, p. 51 my translation).

[11] See M. Villa, *Crisi ecologica e nuovi rischi sociali: verso una ricerca integrata in materia di politica sociale e sostenibilità*, in G. Tomei (ed.), *Le reti della conoscenza nella società globale. Possibilità, esperienze evalore della mobilitazione cognitiva*, Carocci, Roma 2020, pp. 151-182.

[12] K. Levin *et al.*, *Overcoming the Tragedy of Super Wicked Problems: Constraining our Future Selves to Ameliorate Global Climate Change*, in «Policy Sciences» 45 (2012) n. 2, pp. 123-152, p. 123.

However, some authors, such as Catarina Dutilh Novaes[13], have recently claimed that the wickedness of these challenges may leave room for new conceptual inquiry and theoretical syntheses, suggesting a decisive role for philosophy in interdisciplinary investigation. In fact, the case underlines a potential conflict to scrutinize. On the one hand, preventing global climate risks demands more and more local policies to reduce GHG emissions, such as building new low-emissions geothermal power plants. On the other hand, local communities resist top-down energy transition policies that may change their economies and rural landscapes. This situation uncovers new hazards related to health conditions and unemployment.

Following this insight, it is worth noting that inquiry on eco-social work[14] is inclined to organically link the different notions of "climate risks"[15], "ecological risks", and "social risks"[16]. At the same time, this trend connects different scales of risk: global[17], related to climate change, and local, such as the worsening of living conditions of locals.

In line with IPCC documents, implementing transition policy at a local level is strongly related to mitigating climate change. However, as the geothermal plant case shows, identifying all the consequences of relevant decisions and acts seems to be entangled with uncertainty. Actors only partially know the future global outcomes of their decisions for local communities. Furthermore, institutions, power plant developers, experts, and citizens follow conflicting epistemic settings and values in assessing future implications of the energy transition project, and no "optimal solution"[18] seems to be in sight. In addition to this, preventing risks implies coordinated actions by many players. Simply aggregating individuals' deeds does not seem enough

[13] C. Dutilh Novaes, *A Plea for Synthetic Philosophy*, in «Daily Nous», May 30, 2023, https://dailynous.com/2023/05/30/a-plea-for-synthetic-philosophy-guest-post/ (accessed October 20, 2023).

[14] E. Matutini (ed.), *Eco-social work: politica e lavoro sociale nella crisi ecologica*, PM edizioni, Varazze (SV) 2023.

[15] H. Johansson *et al.*, *Climate Change and the Welfare State: do We See a New Generation of Social Risk Emerging?* , in M. Koch, O. Mont (eds.), *Sustainability and Political Economy of Welfare*, Routledge, London 2016, pp. 94-108.

[16] For instance, see T. Hirvilammi *et al.*, *Social Policy in a Climate Emergency Context: Towards an Ecosocial Research Agenda*, in «Journal of Social Policy» 52 (2023) n. 1, pp. 1-23, p. 13.

[17] As defined by World Economic Forum «as an uncertain event or condition that, if it occurs, can cause significant negative impact for several countries or industries within the next 10 years» (E.G. Franco *et al.*, *The Global Risks Report 2020*, World Economic Forum, Cologny-Geneva 2020, p. 86).

[18] C. Helgeson, *Structuring Decisions Under Deep Uncertainty*, in «Topoi» 39 (2020) n. 2, pp. 257-269.

to address global climate change and future social risks efficiently. As IPCC reports stress, "active involvement" of laypeople is keenly recommended. In brief, coping with climate change can be defined as a collective problem that requires new collective agents[19], not a mere sum of individual actions.

Against this background, at least three different research issues can be identified. First, the conceptual relationship between climate, ecological, and social risks. More specifically, global climate risks can elicit diverse contextual social risks depending on local institutional and normative frameworks. As highlighted by some authors and IPCC reports[20], climate and social risks relate to each other through the notion of vulnerability (see §3 below), which is context- and subject-dependent.

Second, connecting different kinds of risks in decision-making could engender "plural"[21] or "deep uncertainty"[22] The conflicting risks at stake not only make their interlocking more challenging to manage for institutional decision-makers but could significantly hinder bottom-up approaches[23] to risk-prevention.

Third, the growing number of situations where local and global, social and climate risks intertwine, coupled with the demand to exploit a closing window of opportunity[24], increase the circumstances where decision-making under deep uncertainty (DMDU) happens. In such cases, the effort to evaluate risk overlapping is increasingly resource- and time-consuming. Drawing on some contributions this tendency might push us to reconsider our traditional grasp of the distinction between risk and uncertainty[25].

That said, two broader topics of investigation should be taken into consideration. Recognizing that only collective actions can treat overlapping risks leaves room for a general question: who can be held responsible for

[19] R. Tuomela, *Social Ontology: Collective Intentionality and Group Agents*, Oxford University Press, Oxford-New York 2013.

[20] IPCC, *Climate Change 2023: Synthesis Report (Full Volume)*, cit.

[21] M. Ongaro, *Making Policy Decisions under Plural Uncertainty: Responding to the COVID-19 Pandemic*, in «History and Philosophy of the Life Sciences» 43 (2021) n. 2, pp. 56 (1-5), p. 56 (1).

[22] "Deep uncertainty" «refers loosely to contexts in which decision-makers lack complete information about (or cannot agree on) the probabilities for key contingencies, the availability of present and future actions, the outcomes to which available actions lead, or the value of these outcomes» (C. Helgeson, *art. cit*, p. 257).

[23] C. Costella *et al.*, *Can Social Protection Tackle Emerging Risks from Climate Change, and How? A Framework and a Critical Review*, in «Climate Risk Management» 40 (2023) n. 100501, pp. 1-6.

[24] IPCC, *Climate Change 2023: Synthesis Report (Full Volume)*, cit., p. 24.

[25] See D. Roser, *The Irrelevance of the Risk-Uncertainty Distinction*, in «Science and Engineering Ethics» 23 (2017) n. 5, pp. 1387-1407; S.O. Hansson, *Risk*, cit.

fostering the demanded collective agents? This insight might organically connect the recent lively debates on collective responsibility[26] and collective action in social ontology[27] to more concrete applications. More specifically, this link can shed light on some individualistic ontological premises in the social sciences[28], which could overshadow what a shared or «common understanding»[29] of risks may be. This point is relevant considering how risk-taking and risk-imposing[30] activities open the problem of who, individually or collectively, is assessing and deciding about risks.

Keeping in mind this rich framework of questions, I will first attempt to chart the different notions of risks most prevalently used in the literature. Hence, I will question some aspects of the traditional interpretation of the risk/uncertainty distinction, as suggested by some voices in the debate as well as by my above case study. Furthermore, drawing on the issues arising from DMDU, I will argue for a line of research about collective agency in risk prevention.

3. *Kinds of Risks*

At least three notions have emerged in the scientific literature regarding risks and ecosystems. In different ways, the concepts of "ecological risk" and "environmental risk" have played a role in identifying potentially damaging events or sources of hazards. By analyzing three different definitions formulated a few decades apart[31], it is possible to identify some evolutionary trends thereof.

[26] M. Smiley, *Collective Responsibility*, in E.N. Zalta, U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University Fall 2023, https://plato.stanford.edu/archives/fall2023/entries/collective-responsibility/

[27] B. Epstein, *Social Ontology*, in E.N. Zalta, U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University Winter 2023, https://plato.stanford.edu/archives/win2023/entries/social-ontology/

[28] For instance, see J.R. Searle, *Making the Social World: The Structure of Human Civilization*, Oxford University Press, Oxford-New York 2009; D. Tollefsen, *Social Ontology*, in N. Cartwright, E. Montuschi (eds.), *Philosophy of Social Science: A New Introduction*, Oxford University Press, Oxford-New York 2014, pp. 84-101; B. Epstein, *The Ant Trap: Rebuilding the Foundations of the Social Sciences*, Oxford University Press, Oxford-New York 2015.

[29] C. Taylor, *Philosophical Arguments*, Harvard University Press, Cambridge (MA) 1995, p. 139.

[30] S.O. Hansson, *A Panorama of the Philosophy of Risk*, cit.

[31] See S.M. Bartell, *Ecological Risk Assessment*, in S.E. Jørgensen, B.D. Fath (eds.), *Encyclopedia of Ecology*, Academic Press, Oxford 2008, pp. 1097-1101, p. 1097; G.W.I. Suter, *op. cit.*, p. 3; L. Na *et al.*, *Regional Ecological Risk Assessment Based on Multi-scenario Simulation of Land Use Changes and Ecosystem Service Values in Inner Mongolia, China*, in «Ecological Indicators» 155 (2023) n. 111013, pp. 1-13, p. 2.

"Ecological risks" and "environmental risks"[32] have initially been identified as the outcome of the impact on populations of natural organisms and ecosystems of specific factors, e.g. toxic chemical pollutants. Lately, the focus has increasingly turned on events, directly or indirectly, elicited by human activities. Therefore, according to the literature, the adverse repercussions of building a geothermal plant on local ecosystems can be identified as a "new ecological risk" engendered by a strategy of energy transition.

In keeping up with this recent line of thought, IPCC documents have introduced a new notion of risk related to climate change, or "climate risk", which has undergone some conceptual evolution in its own right. The 2023 IPCC *Full Report* indicates that:

> In the context of climate change […] risks result from dynamic interactions between climate-related hazards and the exposure and vulnerability of the affected human or ecological system to the hazards. Hazards, exposure, and vulnerability may each be subject to uncertainty in terms of magnitude and likelihood of occurrence, and each may change over time and space due to socio-economic changes and human decision-making[33].

It should be noted that an emphasis is placed on the dynamic interplay between hazards, exposure to hazards, the vulnerability of humans and ecological systems, and anthropic responses to climate change. Therefore, as some have recently pointed out, a progressively central role has been played by the newcomer notion of vulnerability[34].

With this in mind, risks associated with climate change have two main challenging features. First, the relationship between global causes and local effects is indirect, which points out that social risks are considered both certain and uncertain[35]. Second, traditional classifications of risks that affect human social life, i.e. social risks, were based upon the premise that a (national) community could bear the burden of risk-sharing. However, the distance between causal factors and effects makes the inequalities generated increasingly intractable due to the complex interplay of climate change and mitigation policies. This link seems apparent in the case of the geothermal

[32] G.W.I. Suter, *op. cit.*

[33] IPCC, *Climate Change 2023: Synthesis Report (Full Volume)*, cit., p. 128.

[34] C. Costella *et al.*, *art. cit.*, p. 5.

[35] «One of the IPCC's key conclusions is that the social risks associated with climate change are both certain and uncertain. […] since multiple climate hazards will occur simultaneously, and multiple climatic and non-climatic risks will interact, resulting in compounding overall risk and risks cascading across sectors and regions» (T. Hirvilammi *et al.*, *art. cit.*, p. 4).

plant above. It shows how "new social risks" engendered by energy transition strategies are imposed on local communities that, by having hosted geothermal energy plants in the past, have not significantly relied on fossil fuels and whose future social risks caused by climate change are partially unknown. In addition, the identification of "certain" and "uncertain" risks shows the ambiguities related to practical applications of the distinction between risk and uncertainty[36]. In the background, a more conceptual issue stands open: the plurality of interpretations of social risks.

As highlighted by interdisciplinary investigations[37], there is a wide variety of notions of social risks, ranging from the influential conception that they «represent the probability of some threats and uncertainties which have arisen as a result of modernizing the society, which imply irreversible damage for all forms of life (Beck 1992)» [38], up to the idea (influential in public policy) that «social risk represents the probability for a person to be affected by an unexpected, uncertain situation […] associated with loss of control over one's personal actions (Sirovatka, Winkler 2010)»[39]. Notwithstanding the different orientations, a crucial role is played by the notion of "vulnerability" everywhere. In this sense, vulnerability may be understood as «the degree to which an individual, a community, a system is exposed to the effects of a hazard based on some essential conditions»[40].

After this concise topography, some elements regarding the conceptual relations between climate, ecological, and social risks can be drafted. In broad terms, the scientific literature converges in recognizing that the vast plurality of losses and vulnerabilities due to climate change can affect both individuals and communities. Nonetheless, hazards affecting collective subjects do not seem to be treated specifically, and it seems that they are rather conceived as mere aggregations of individual ones.

This issue can be related to the unclear relation between climate and new social risks through the concept of vulnerability and «policy solutions»[41]. If new social risks are (directly or indirectly) related to global warming "and"

[36] S.O. Hansson, *Risk*, cit., par. 2.

[37] L. Lupu, *The Concept of Social Risk: A Geographical Approach*, in «Quaestiones Geographicae» 38 (2019) n. 4, pp. 5-13, p. 6.

[38] *Ibidem*; P.U. Beck, *Risk Society: Towards a New Modernity*, SAGE, London 1992.

[39] L. Lupu, *art. cit.*, p. 6; T. Sirovátka, J. Winkler, *The importance of new social risks in the current social sciences*, in «Sociální Studia» 2 (2010), pp. 7-21.

[40] L. Lupu, *art. cit.*, p. 8; UNISDR, *UNISDR Annual Report 2017 (2016-17 Biennium Work Program Final Report)*, United Nations Office for Disaster Risk Reduction, Geneva 2018, pp. 1-64.

[41] C. Costella *et al.*, *art. cit*., par. 2.1.

transition policies, they can seemingly be fully assimilated ontologically and epistemologically. Nonetheless, recent social ontology investigations have highlighted that "social kinds"[42], such as vulnerability, have different ontological properties than "natural kinds". In fact, the contrasting views about the vulnerabilities at stake in the mentioned case point to the complexity of the social risks engendered by the energy transition. These issues must be settled collectively as the definition of climate risks prevention then calls for a specifically collective treatment.

In fact, the relevant literature underlines that bottom-up and top-down approaches remain in tension[43], implying inefficacious coordination between parties that entails that new «risk can arise […] from the uncertainty in the implementation, effectiveness or outcome of climate policy»[44].

As previously mentioned, the misalignment between experts, laypeople, and policymakers can frequently take place in eco-social work[45], but it is also recurrent in other collective risk-management processes[46]. Thus, the literature suggests that this link between risk-knowledge corpus and decision-making in real-world contexts seems indirect and circular. Institutions and laypeople's judgments on risks (risk assessment, risk imposition, and risk prevention) are often fragmented, and the outcomes of their interactions are unpredictable. The "shared evaluation" of risks at stake seems strongly intricate[47]. Notwithstanding these apparent challenges, IPCC reports[48] and others in the literature[49] continue to stress the importance of bottom-up social processes in the picture.

[42] S. Haslanger, *Resisting Reality: Social Construction and Social Critique*, Oxford University Press, New York 2012.

[43] M. Villa, *Crisi ecologica e nuovi rischi sociali*, cit.; *Cambiare o traccheggiare?*, cit.

[44] IPCC, *Climate Change 2023: Synthesis Report (Full Volume)*, cit., p. 128.

[45] R. Cucca *et al.*, *Towards a Sustainable Welfare System? The Challenges and Scenarios of Eco-social Transitions*, in «Social Policies» 10 (2023) n. 1, pp. 3-26.

[46] P.A. Ebert, I.N. Durbach, *xpert and Lay Judgements of Danger and Recklessness in Adventure Sports*, in «Journal of Risk Research» 26 (2023) n. 2, pp. 133-146.

[47] D. Thorstad, *General-Purpose Institutional Decision-Making Heuristics: The Case of Decision-Making under Deep Uncertainty*, in «The British Journal for the Philosophy of Science» August 2022.

[48] IPCC, *Summary for Policymakers, Climate Change 2021* in IPCC, *The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge (UK)-New York 2021, pp. 3-32; IPCC, *Climate Change 2023: Synthesis Report (Full Volume)*, cit.

[49] See R. Cucca *et al.*, *art. cit.*; E. Matutini (ed.), *op. cit.*; M. Villa, *Crisi ecologica e nuovi rischi sociali*, cit.

4. *A Side Note on Risk and Uncertainty*

The IPCC documents stress that preventing climate risks and new social risks starts from the consideration that there are both certain and uncertain risks. However, this issue seems to be at odds with the risk/uncertainty distinction.

According to Roser[50], the contrast between risk and uncertainty is debatable for practical purposes. Starting from the premise that «the question whether we have probabilities is completely separate from the question how we ought to make use of them»[51], Roser underlines that, in everyday discourse as well as in applied and theoretical contexts, «there is no universally accepted distinction between risk and uncertainty based on whether we have probabilities or not»[52]. Moreover, he argues for action-guiding principles that yield justified decisions and claims that «low epistemic credentials are better than non-credentials», because «using more evidence is usually better than using less evidence»[53]. To sum up, he claims that «risk-uncertainty distinction is irrelevant because both high- and low-credentials probabilities should enter our decision-making»[54]. Hence «we always have subjective probabilities and epistemic probabilities»[55], and they should be used for guiding actions in tackling climate change.

An in-depth exploration of Roser's argument is out of the scope of the paper. However, two points can be underlined according to his theses. First, the distinction between risk and uncertainty has been understood and applied in several ways, and its implications for decision-making demand further investigation. Second, in coping with salient challenges, such as climate change, the point is to apply the best available probabilistic estimations we have, either epistemic or subjective. The last issue concerns "to whom" probability estimations are available, or who is the "we" tasked to manage the risk knowledge corpus and the decisions to be made.

In this vein, recalling the above example, the identification of the first plural "we" person is ambiguous, since there are many "we" with conflicting epistemic backgrounds and narratives. And therefore, given that risks are the product of hazard, exposure, and vulnerability, it can be stressed that

[50] D. Roser, *art. cit.*
[51] *Ivi*, p. 1389.
[52] *Ivi*, p. 1390.
[53] *Ivi*, p. 1404.
[54] *Ivi*, p. 1400.
[55] *Ivi*, p. 1406.

vulnerability evaluations are strongly subject-relative. This contrast entails that different actors could attribute varying salience[56] to different risks, due to diverging ethical and epistemic assessments. For instance, the locals in the Tuscany region could tend to pay more attention to cultural and psychological losses than private firms do.

In brief, two final points can be emphasized regarding the risk/uncertainty distinction.

To overcome these ambiguities two insights can be fruitful. Recently, the literature on risk has paid much more attention to the notion of "collective responsibility" [57], and this tendency can help highlight that most actors involved in the energy transition are collective ones. Thus, a comprehensive point can be outlined. Since no single actor can be held solely responsible for locally preventing climate and social risks, it is fair to suggest that only by developing some (new) forms of collective agency[58] and responsibility we can pave the way for effectively coping with climate-change-related challenges, as the debate about the accountability of plural subjects in the face of moral challenges posed by global warming indicates[59].

Considering these issues, some further research is in order. From the lens of collective agency, how can we account for decision-making under deep uncertainty? More specifically, can we better understand risk communication[60] between experts and laypeople, and the demanded iteration of decision-making under deep uncertainty[61]?

5. *Collective Responsibility and Interlocked Risks*

The question of how to account for collective responsibility has undergone a lively debate that has focused primarily on understanding whether

[56] F. Hindriks, F. Guala, *The Functions of Institutions: Etiology and Teleology*, in «Synthese» 3 (2019) n. 1, pp. 1-17.

[57] A. Placani, S. Broadhead, *Risk and Responsibility in Context*, Taylor & Francis, Abingdon (UK)-New York 2023.

[58] See N. de Haan, *Collective Moral Agency and Self-induced Moral Incapacity*, in «Philosophical Explorations» 26 (2023) n. 1, pp. 1-22.

[59] See S. Collins, *Corporations' Duties in a Changing Climate*, in J. Moss, L. Umbers (eds.), *Climate Justice and Non-State Actors*, Routledge, Abingdon (UK)-New York 2020, pp. 84-100.

[60] P.A. Ebert, I.N. Durbach, *art. cit.*, L. Zanetti, D. Chiffi, L. Petrini, *Epistemic and Non-epistemic Values in Earthquake Engineering*, in «Science and Engineering Ethics» 29 (2023) n. 3, pp. 18 (1-16).

[61] C. Helgeson, *art. cit.*

and how it is possible to speak of the moral agency of plural actors. That is, whether collections of people, groups, or institutions can develop joint intentions to act on their own and be held morally responsible for specific harms they may cause[62]. As it can be noted, the difference among the various positions revolves around the issue of ontological reducibility of collective agents[63].

Therefore, the topic has gained increasing traction also from the point of view of applied philosophical investigation[64]. A specific issue that is attracting rising consideration in this sense is that of forward-looking collective responsibility[65], understood as the power and accountability on the part of collective agents to implement a desired future reality. For instance, this issue asks for clarification on whether a collective subject, and which one, can be held responsible for building an environment in which climate and social risks will be significantly reduced, both at the foot of Mount Amiata and globally. Relatedly, the starting case immediately opens with a theoretical problem: a well-defined, singular collective subject that can cope with all the challenges posed by climate and new social risks does not (yet) exist. In other words, the current composition of the situation involves a bundle of heterogeneous collective subjects with different powers, risk assessments, preferences, and values. Therefore, the lack of coordination and cooperation shows that the preconditions of comprehensive collective agency as proposed by many social ontologists, such as group reasoning[66] and group ethos[67], are absent.

Nonetheless, according to Hindriks[68], the situation can be understood as a case where singular actors (in a broad sense) should join forces to cope with a challenging problem and avoid some potential shared threat. Since social and environmental harms cannot be prevented by a single player, as one single agent alone cannot possibly control the complex interactions of global and lo-

[62] M. Smiley, *op. cit.*, par. 1.

[63] S. Bazargan-Forward, D. Tollefsen, *The Routledge Handbook of Collective Responsibility*, Routledge, New York-London 2020, pp. 1-2; D.P. Schweikard, H.B. Schmid, *Collective Intentionality*, in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University Fall 2021, par. 4.2, https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality/

[64] S. Collins, *Collective Responsibility Gaps*, in «Journal of Business Ethics» 154 (2019) n. 4, pp. 943-954, p. 946.

[65] M. Smiley, *op. cit.*, par. 7.

[66] C. List, P. Pettit, *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford University Press, Oxford 2011.

[67] R. Tuomela, *op. cit*.

[68] F. Hindriks, *The Duty to Join Forces: When Individuals Lack Control*, in «The Monist» 102 (2019) n. 2, pp. 204-220.

cal factors, combining different efforts into forms of shared collective agency can counter potential hazards. Therefore, singular actors have «a duty to join forces: to approach others, convince them to contribute, and subsequently make a coordinated effort to prevent [eventual] harm»[69]. However, an issue emerges: how can random individual actors join forces and build a new collective subject capable of preventing climate and social risks? To put it simply, who in the geothermal plant case has an obligation to mobilize others and to help people join their forces to prevent risks for everyone? More specifically, regarding the epistemic issues in managing risks and uncertainty in complex cases, this line of thought investigates if any specific collective epistemic responsibility can be identified. Nevertheless, whether any epistemic collectives[70] do actually exist is a topic that needs to be clarified.

During the past few decades, there has been a growing interest in understanding whether groups can be identified as epistemic agents in a non-derivative sense, over and above individual members. This field of research, broadly defined as collective epistemology[71], can give relevant insights into the present issue. If joining forces to prevent climate and social risks may be a workable direction to go, the issue concerning what role risk experts[72] should play together in fostering the demanded cooperation and coordination should be addressed as well. Thus, can expert communities be held responsible for their risk communication? *Prima facie*, this problem seems quite challenging. However, recent contributions on the epistemology of seismic hazards[73] can shed some light on this. Some come to the rather counterintuitive implication that an aggregative procedure of different epistemic risk assessments can point to a «no one's model»[74] problem. All in

[69] *Ivi*, p. 204.

[70] According to Tollefsen, epistemic collective agents can be understood as «groups [that] have a rational point of view and are subject to the norms of rationality» (D. Tollefsen, *Collective Epistemic Agency*, in «Southwest Philosophy Review» 20 (2004) n. 1, pp. 55-66, pp. 62-63).

[71] See D. Tollefsen, *art. cit.*; H.B. Schmid *et al.* (eds.), *Collective Epistemology*, de Gruyter, Berlin-Boston (MA) 2011; J. Lackey, *The Epistemology of Groups*, Oxford University Press, Oxford 2021; P. Pettit, *Five Elements of Group Agency*, in «Inquiry» May (2023), pp. 1-21.

[72] M. Baghramian, C. Martini, *Questioning Experts and Expertise*, Taylor & Francis, Abingdon (UK)-New York 2022; F. Pongiglione, C. Martini, *Climate Change and Culpable Ignorance: The Case of Pseudoscience*, in «Social Epistemology» 36 (2022) n. 4, pp. 425-435.

[73] L. Zanetti, D. Chiffi, L. Petrini, *Epistemic and Non-epistemic Values in Earthquake Engineering*, cit.; L. Zanetti, D. Chiffi, L. Petrini, *Philosophical aspects of probabilistic seismic hazard analysis (PSHA): a critical review*, in «Natural Hazards» 117 (2023), pp. 1193-1212.

[74] «Respect to the final hazard estimate, neither the individual proponent nor the integrator seem to have ownership of the final result» (L. Zanetti, D. Chiffi, L. Petrini, *Philosophical Aspects of Probabilistic Seismic Hazar*, cit., p. 1209).

all, this discussion, applied to the problem at hand, seems to suggest that focusing on "how" scientific communities decide what to communicate to laypeople can have a significant impact in improving the experts' sense of accountability of risk communication. (However, the question of how non-experts themselves can be involved remains open.)

A clever way to tackle such questions is to draw from the current research on individual and collective agency. The first insight relates to the idea that collective agency is "layered"[75] and dynamic[76]. Some investigations on the different "segments" and "strata" of human action embedded in social techniques (such as games) highlight that to achieve a complex goal (such as striving to play, e.g., enjoying playing a specific sport or game) demands intermediate coordinated actions and flexible sequences of sub-actions on the part of all the members (segments and strata). At the same time, it is necessary that players "submerge themselves" in each sub-action by zooming in on a specific task and, when required, zooming out to monitor whether the overall end of the game remains in focus. This intuition can show how risk assessment, risk-taking, and risk imposition can be framed as discrete levels of joint actions: only with reiterated joint zooming out and monitoring of the overall goals and strategies different actors[77] can build collective decisions about climate and social risk prevention.

Such steps indicate that building a collective perspective[78] on risk prevention requires subsequent stages of reflection, similar to discussions on we-reasoning[79] or group reasoning elsewhere[80]. All of this commands two more points concerning decision-making and acting together to prevent new social risks. The first point is about "how" risk communication and decision-making practices are structured. In fact, drawing on the geothermal plant case, assessing which methods are used for helping people with divergent understandings and perceived vulnerabilities to decide together about risk prevention appears to be crucial[81]. More explicitly, climate and social sci-

[75] C.T. Nguyen, *Games: Agency as Art*, Oxford University Press, New York 2020; L. Ferrero, *Games and the fluidity of layered agency*, in «Journal of the Philosophy of Sport» 48 (2021) n. 3, pp. 344-355.

[76] G. Thonhauser, M. Weichold, *Approaching Collectivity Collectively: A Multi-Disciplinary Account of Collective Action*, in «Frontiers in Psychology» 12 (2021) n. 740664, pp. 1-15.

[77] Joint zooming out can be roughly defined as a joint action that aims at building a shared understanding of the overall joint agency regarding a shared goal.

[78] G. Thonhauser, M. Weichold, *art. cit.*

[79] R. Tuomela, *op. cit.*

[80] P. Pettit, *art. cit.*

[81] As suggested by recent investigations on expertise, «trust being more a matter of com-

ence experts seem to have specific duties to avoid epistemic harms[82], such as a wrong estimation of the likelihood and vulnerability of salient future scenarios. As the case suggests, joining the epistemic forces of experts and laypeople to prevent climate hazards is not an easy task. Therefore, peculiar collective responsibilities of social experts (and practitioners) to structure decision-making under deep uncertainty seem to exist. Supporting different actors to mobilize themselves and join forces requires a specific social methodology[83] that can mobilize subjects with different epistemic risk understanding and build new collective agents.

To sum up, avoiding epistemic harms related to risk prevention may require collective epistemic duties on the part of experts to join their forces to help people mobilize, by responsibly structuring risk communication and decision-making under deep uncertainty. In addition, this collective framing of decision-making suggests we take into account the issue of collective learning. For instance, Helgeson[84] underlines that coping with embedded uncertainties requires iterative decision-making cycles; hence, inquiring how local collective agents learn together[85] could play a relevant role in developing a comprehensive view of preventing new social risks and navigating uncertain environments, as suggested by Doan[86] and generally by research on common pool resource management[87]. Therefore, a second future direction to explore

munication and emotional connection, and only to some extent secondarily a matter of credentials and certifications» (C. Martini *et al.*, *Knowledge Brokers in Crisis: Public Communication of Science During the COVID-19 Pandemic*, in «Social Epistemology» 36 (2022) n. 5, pp. 656-669, p. 666). See also S. Roeser, *Risk Communication, Public Engagement, and Climate Change: A Role for Emotions*, in «Risk Analysis» 32 (2012) n. 6, pp. 1033-1040; S. Roeser, *Risk, Technology, and Moral Emotions*, Routledge, New York 2017; F. Pongiglione, *Trust, Experts, and the Potential Side Effects of Critical Thinking*, in «Teoria. Rivista di Filosofia» 42 (2022) n. 2, pp. 163-174.

[82] «Epistemic harm: a harm affecting the epistemic status of a subject, group of subjects, or epistemically important social system» (W. Fleisher, D. Šešelja, *art. cit*, p. 8).

[83] M. Villa, *Cambiare o traccheggiare?*, cit.

[84] C. Helgeson, *art. cit*.

[85] A «theory of collective learning describes how the capacity to mentally represent objects, events, and minds as targets of firstperson plural attention facilitates cognitive collaboration in groups» (G. Shteynberg *et al.*, *Shared Worlds and Shared Minds: A Theory of Collective Learning and a Psychology of Common Knowledge*, in «Psychological Review» 127 (2020) n. 5, pp. 918-931, p. 926).

[86] M.D. Doan, *Collective Inaction and Collective Epistemic Agency*, in S. Bazargan-Forward, D. Tollefsen (eds.), *The Routledge Handbook of Collective Responsibility* Routledge, New York-London 2020, pp. 202-215.

[87] E. Ostrom, *Governing the Commons: The evolution of Institutions for Collective Action*, Cambridge University Press, Cambridge (UK) 1990; E. Ostrom, *A Polycentric Approach for Coping with Climate Change*, The World Bank, Washington (DC) 2009.

can be understanding how the notion of collective learning might improve the current debate on decision-making under deep uncertainty.

In conclusion, an account of collectively preventing climate-related risks demands focus on at least two areas of inquiry. One primary issue calls to investigate not only the ontology and epistemology of risks at stake but also how collective epistemic and policy decision-making processes unfold through different instances of risk evaluation, and what are the related collective responsibilities of experts, scholars, and scientific institutions. In addition to this, any comprehensive account should consider the procedural and dynamic evolution of collective agency, and how preventing (future) social risks demands collective learning by many plural subjects about the future collective affordances that can be made available by present choices.

6. *Interdisciplinarity and Collective Agency*

As a final point, some implications of this approach regarding the demand for interdisciplinary research on climate risks[88] can be outlined. Although inquiring risks and wicked problems demands interdisciplinarity, this tendency, as social risks communication shows[89], might increase laypeople's disorientation – and hesitancy to commit.

Nevertheless, investigating the role of experts through the lens of collective epistemic obligations[90] might reframe potential epistemic harms in many puzzling cases. First, inquiring about experts' duty to join forces to improve risk communication in decision-making under deep uncertainty might help avoid epistemic harms related to local climate risks, such as the "paralyzing effects" of some communication strategies. Second, experts can also help research communities to listen to stakeholders' values and their understanding of vulnerabilities, which is crucial for reframing shared problems that are at the center of interdisciplinary and transdisciplinary inquiry[91].

[88] M. MacLeod, M. Nagatsu, *What does Interdisciplinarity look like in Practice: Mapping Interdisciplinarity and its Limits in the Environmental Sciences*, in «Studies in History and Philosophy of Science» 67 (2018), pp. 74-84; J. Persson *et al.*, *art. cit*.

[89] See the notion of "linguistic uncertainty" in P. Döll, P. Romero-Lankao, *How to Embrace Uncertainty in Participatory Climate Change Risk Management-A Roadmap*, in «Earth's Future» 5 (2017) n. 1, pp. 18-36.

[90] A. Schwenkenbecher, *How We Fail to Know: Group-Based Ignorance and Collective Epistemic Obligations*, in «Political Studies» 70 (2022) n. 4, pp. 901-918.

[91] S. Efstathiou, Z. Mirmalek, *Interdisciplinarity in Action*, in N. Cartwright, E. Montuschi (eds.), *Philosophy of Social Science: A New Introduction*, Oxford University Press, Oxford 2014, pp. 233-248; J. Mittelstrass, *On Transdisciplinarity*, in «Trames» 15 (2011) n. 4, pp. 329-338.

From the standpoint of collective agency, experts might act as a two-way interface between laypeople and scholars in identifying which research problems should be tackled. Moreover, they could uptake the task to listen to laypeople's suggestions and concerns and thus foster a collective epistemic (sense of) agency; nonetheless, listening and deciding collectively, as suggested by research on "knowledge co-production" with indigenous people for instance[92], demand the conceptualization and design of new social processes that involve laypeople, experts, scholars, and decision-makers.

7. *Conclusions*

The investigative path I have sketched is not a conclusive analysis of the different issues at stake in theoretically understanding the interlocking of climate and social risks. On the contrary, my objective is to highlight how interdisciplinary research can substantially help improve the current debate on risk, uncertainty, and climate change.

To conclude, a brief overview of some tentative upshots of the analysis can be summarized. The different notions of ecological, climate, and social risks are related to each other through the conceit of vulnerability. This entails that identifying who, and why, should become more vulnerable due to an energy transition policy can become a highly debatable matter. This hint can help enlarge the current philosophical investigation on risk to inquire the issue of sharing moral and epistemic salience of vulnerability related to different risks. Moreover, recognizing the collective structure of climate challenges can suggest that epistemic failures inherent in risk-prevention demand an enhanced comprehension of the collective assessment of hazards, exposure, and vulnerability, as a joint effort of building collective agencies.

Therefore, it can be suggested that a duty to mobilize other people to join their forces to cope with climate change in local energy transition projects does exist, and it entails a specific role for scholars and experts to join their forces to avoid epistemic harms. The specific consequences of these harms are that unaccounted risks could impact on laypeople and, at the same time, that could lead astray in defining research problems within research com-

[92] N. Latulippe, N. Klenk, *Making Room and Moving over: Knowledge Co-production, Indigenous Knowledge Sovereignty and the Politics of Global Environmental Change Decision-making*, in «Current Opinion in Environmental Sustainability», 42 (2020), pp. 7-14.

munities. This duty can also be understood as an obligation to promote collective learning of plural subjects involved in energy transitions as a sort of "risk co-knowledge" building.

Finally, the discussed case is one of the many controversial examples of energy transition where social, technological, economic and ecological factors tend to conflict with each other[93]. Nonetheless, the research trajectory I have drafted shows the opportunity to improve the interdisciplinary dialogue between social sciences, climate sciences, and philosophical investigation on risk and collective agency, in order to better examine "how" decision-making under deep uncertainty and risk communication are structured to overcome the hazards of collective paralysis and powerlessness.

Abstract

Interweaving hazards in environmental crises can be framed as a wicked problem as well as an opportunity for the interdisciplinary contribution of philosophical analysis on risk. Due to nonlinear mechanisms and contextual variations, this shows the importance of inquiring about contrasting assessments of vulnerability and the demand for comprehensive collective actions in coping with climate risks. The article examines how to address overlapping ecological and social risks, focusing on decision-making in the context of local energy transition projects through the lens of collective epistemic responsibility. By analyzing disciplinary accounts and exploring the links between assessment and decision-making further research directions for collective risk prevention strategies will be outlined, and some implications for interdisciplinary investigation on risk and expertise will be sketched.

Keywords: climate risk; collective agency; deep uncertainty; social risk; vulnerability.

Marco Emilio
Istituto Universitario Salesiano di Venezia
m.emilio@iusve.it

[93] F.W. Geels, *et al., The Socio-Technical Dynamics of Low-Carbon Transitions*, in «Joule» 1 (2017) n. 3, pp. 463-479.

T

Žarko Paić

On the Navigation of Uncertainties: Chaos, Entropy, and Technological Singularity

Poetry has caught up. Moreover, we now have a machine
With its poetry as well as a new way of life,
Business, worldly, intellectual, sentimental,
With which the machine age has endowed our souls.
Alvaro de Campos, *Maritime Ode*

1. As stated in this article's title, it is about the possibilities of mastering what belongs to the coming future. The negative Cartesian concept of uncertainty cannot encompass the actuality of events, but emerging from the very logic of reality is what characterises reality in cognitive-theoretical insight, namely certainty as a certainty that being is what it is in the modality of its possibility as an actual necessity. Certainty, then, is *certitudo* and refers to human judgement about Being as such, speaking in Heideggerian terms. This judgement generally cannot be wrong because truth should be understood scholastically as the correspondence of opinion with things. It is certain, for example, that contemporary global capitalism represents the result of the technoscientific construction of reality as a network of events that appear cybernetically in the fourfold of information-feedback-control-communication. Nothing from this new trans-classical logic can be «efficient», «useful», or «pragmatic». Moreover, this certainty cannot be what exists by itself; it is a pure construction of events based on probability in science theory.

Chaos theory enters many areas of mathematics and focuses on the so-called deterministic laws of dynamic systems. The concept at the centre of this theory is not necessity but chance in the sense of disruption of order as a deviation from the usual course of cause and effect. Chaotic complexity sys-

 DOI: 10.4454/6jbgcd16

tems are based on interdependencies and the cybernetic notion of feedback loops, repetition, self-similarity, fractality, and self-organisation. Deterministic non-linear systems produce significant differences in the initial states of matter and energy. An accurate metaphor for that is when a butterfly flaps its wings in Brazil; suddenly, a tornado blows up in Texas. In contemporary philosophy, this is best creatively performed in Gilles Deleuze and Felix Guattari's book *What Is Philosophy?* in which the last chapter is entitled «From Chaos to Brain», and the first sentence of that conclusion is unforgettable:

> We require just a little order to protect ourselves from chaos[1].

Why is this chaos theory deterministic if its primary term denotes the opposite of necessity, i.e. chance, which means an event that escapes regularity and order in the sense of lawfulness? Because random cases necessarily happen in an evolutionary sense, the author of the theory, Edward Lorenz, claims that non-linear systems cannot be predicted. In other words, there is a rule of absolute contingency. This, in turn, does not mean any craziness regarding the non-determinism of all parameters that modern science takes into account. Instead, an understanding of chaos is at work in the sense that the present condition determines the future, but the proximate present does not approximately determine the future. The chaotic behaviour of parameters in non-linear systems should be characterised by the fluid flow of events and quantities' irreducibility, as in Poincaré's equations. We can find all this in philosophical terms not only in Deleuze but also in Gilbert Simondon, in Niklas Luhmann's cybernetic theories of systems in sociology and law, and in autopoietic models of events in which autonomous objects, initiated by artificial intelligence, function within the precisely realised plan of immanence. After all, my technosphere concept stems from the chaos, contingency, and emergence theory. If one wants to find a literary articulation of these ideas, the addressees are Thomas Pynchon's and Don DeLillo's novels. In addition to all that has been said, there is another term from physics: the second law of thermodynamics, which we call *entropy*[2].

Entropy in modern science theory includes two conceptions: Ludwig Boltzmann's and Claude Shannon's, which refer to statistical and informational entropy. According to the second law of thermodynamics, systems

[1] G. Deleuze, F. Guattari, *What Is Philosophy?*, Columbia University Press, New York 1994, p. 172.

[2] K.D. Bailey, *Social Entropy Theory*, SUNY Press, New York 1990.

tend to have maximum entropy as a balance in the probability of the system's sustainability disintegration. A sustainable system is in a state of order when all parts of the system relate to it as an autonomous unit to a higher order of energy and information regulation. *Entropy is, therefore, related to the concept of complex systems and is applied equally in physics, cybernetics, and social sciences. Boltzmann's understanding refers to the degree of probability by which order is brought to maximum entropy (chaos) by equalising the unavailability of energy and information within one system.* The measure used to achieve the exactness of the prediction of the maximum entropy of the system refers to the variables of the probability and improbability of events (physical and chemical) in the order of complexity of a state that goes from order to disorder. At the same time, statistical probability represents a measure that attempts to mathematically demonstrate the possibility of maximum entropy of the system within the framework of the physical data of a particular state. Information entropy represents an attempt to show the uncertainty of a system based on the production and distribution of information, which is necessary for the system to be effective. The Shannon-Weaver mathematical theory of communication, which enabled the computer age of information, presupposes precise information entropy in its foundations. *It is already clear from this that the concept of entropy does not refer to some «apocalyptic state» of the collapse of biological, physical, and social systems of complexity.*

On the contrary, with entropy, one tries to find a mental map for understanding the crossing of borders between two states of equilibrium-disequilibrium of the system where mass, energy, and information form the «essence» of the modern way of organising the technosphere. Mass society requires an order of high energy deliverability for survival. At the same time, information entropy represents its fundamental mode of communication, which is always on the verge of transitioning from one state to another. Therefore, the relationship between statistical and informational entropy is determined by the virtuality of actualising a chaotic order in which information itself assumes the properties of «mass» and «energy» in transforming social relations as technical relations between things. Global capitalism becomes a perfect model of total entropy in its state of maximum information-communication chaotic order.

Chaos and entropy are critical concepts for understanding our modernity. Instead of the necessity and pre-stabilised harmony of the cosmos and the world, on which classical and modern metaphysics still rest, everything «collapses», becomes «curved», and evaporates in «black holes», and from

everything, only *event horizons remain from an astrophysical point of view.* How is it possible that on these principles of complete fractalisation, the modern world of the rule of neurocognitive capitalism in the superintelligence of artificially created *homo kybernetes* sees its bright future without an essential human share in the form of capital as such? According to the great Italian thinker Emanuele Severino, and as I say in my five-volume work *Technosphere*, philosophers are inclined to return to Parmenides[3]. However, I do not consider a «great return to the beginning» a valid thinking alternative for the coming time. Severino claims that capitalism should be characterised by an outdated system of social relations concerning the boom of superintelligent technologies[4]. Of course, although it perfectly adapts to any new situation, including the rule of the principles of contingency, chaos, and entropy, what makes it obsolete in an idea is the essence of chaos theory. It is not about anything else but the possibility of collapse or the emergence of disorder, which begins what goes beyond the fundamental driving force of capitalism in general. It is about a dynamic procedure of desire as a thinking machine beyond any physical need. Like the current form of artificial intelligence, cognitive capitalism designates what absolute autopoiesis is not. Therefore, the will to Power must be included as a technopoietic desire to rule over Others outside the logic of primary, secondary, tertiary, and quaternary needs. When the system no longer needs anything from its environment, it is self-sufficient, like the Aristotelian God. It is no longer an immobile driver but a becoming or dynamic process of infinite techno-genesis of ideas as a polycentric information system. *Philosophically speaking, the synthesis of theoria, praxis, and poiesis is at work from the logic of what now makes them possible in the first place, and that can only be téchne.*

Why is ancient metaphysics, despite its realisation in the *technosphere*, still present as a regulative mechanism of thought in Kantian terms in circumstances where everything becomes chaotic and entropically placed in the nonlinearity of the world? Because *we require just a little order to protect us from chaos.* What else does it mean but a longing for human-too-human, animal-too-animal, plant-too-plant, or simply a longing for some form of rootedness and nativeness, for which Earth-earth is necessary, not heaven as an interplanetary space of wandering? We are, admittedly, beings of a wandering destiny, nomads and eccentrics, and this has been the fate of philosophy and art from mythic beginnings. But precisely because we are not

[3] Ž. Paić, *Technosphere,* vols. I-V, Sandorf and Mizentrop, Zagreb 2018-2019.
[4] E. Severino, *Capitalismo senza futuro*, BUR Rizzoli, Milano 2013.

angels or avatars, we need distance from the force and break of the 'infinite speed' with which everything goes into the abyss. We are losing what we need day by day, in the face of sceptical faith that there is still a possibility of overcoming metaphysics as a severe disease, that the antidote to this technologisation of thought exists in the mythopoetic vertigo of language, as in Fernando Pessoa's poetry.

The Cartesian Being-God-World-Human model of thought included the axiomatic of certainty (*certitudo*). The new era begins as a scientific picture of the world based on the proof that what we call reality means the pure certainty of thought in coincidence with the reality of external objects. The dispositive of such an opinion is the *cogito* as the pure subjectivity of the subject. Starting from thinking as a thing that unites the mind and the materiality of nature, *res cogitans* and *res extensa*, René Descartes was able to arrive at a proposition that represents the condition for the possibility of the emergence, in tendency-latency, of absolute subjectivity, which would receive the name «absolute» with Hegel. It is, therefore, absolutely nothing divine but the result of the synthesis of substance or Being and subject or thought. But this synthesis comes from a pure mind, that is, from the essence of thinking as unconditional subjectivity. Being becomes thinking only from the axiomatic that reads *cogito ergo sum.* That is why Martin Heidegger, in his lectures and discussions from the end of the thirties of the 20th century, such as *Besinnung, Vom Ereignis,* and especially «Die Zeit des Welbildes», advocated the position that the genesis of modern science denotes the emergence of modern technology that comes from the essence of metaphysics as nihilism[5]. This only means that the certainty of the opinion about Being-God-World-Human represents the result of the neutralisation and suspension of the Greek-scholastic image of the world, which, in its staticity and apology for the eternity of the universe, was based on the idea that being as such is *physis* and that the world and Human found in correlation with gods and God. *Metaphysical thinking from the beginning changes radically in the new age so that the cogito, the subject, and the absolute become a condition for the possibility of thinking «about» Being-God-World-Human.* What, then, should be a certainty other than the mental construction of the creation of a «new» world from a pure mind, but in such a way that between humans

[5] M. Heidegger, *Besinnung*, GA, Bd. 66, V. Klostermann, Frankfurt am Main 1997; M. Heidegger, *Beiträge zur Philosophie (Vom Ereignis),* BD 65, V. Klostermann, Frankfurt am Main 1989; M. Heidegger, *Die Zeit des Weltbildes*, in *Holzwege*, V. Klostermann, Frankfurt am Main 2003, pp. 75-113.

and nature in its objectivity, there is a necessary difference of «worlds», that of thinking, *res cogitans,* and that of bodily extension, *res extensa*?

The axiomatic power of the metaphysics of subjectivity is developed based on the scientific-technological structure of consciousness, and it begins with the concept of certainty, which needs to be mathematically and physically proven to be something that harbours no doubts regarding its ontological status. The proof can no longer be in God's hands but in human self-consciousness, articulating itself as the language of transcendental forms of thought and as a set of empirical facts. *What is certain comes from the self-certainty of thinking as a scientific-technological establishment of the world as a case. The language of modern metaphysics has already been mathematised by Descartes, Blaise Pascal, Baruch Spinoza, and Gottfried Wilhelm Leibniz, reaching its peak with Leibniz's idea of the infinity of the monad and the logic of sufficient reason in rationalism, with which the possibility of creating a thinking machine begins*. The reason for this lies in the calculating character of thinking as an analytical projection and construction of reality. The most significant saying of this constructive rationalism is Leibniz's *Cum Deus calculat etc cogitationem exercet, fit mundus!* What else should be a certainty than the management of the world as an a priori imposed set of Being, beings, and essence of humans that can be «programmed» at any moment only if he is also given what belongs to God as his determination? It is an intuitive knowledge that directly, suddenly, and instantaneously captures the essence of things without the mediation of evidence through mathematics and logic. Certainty, therefore, is already understood by Leibniz as a reestablished harmony of mind and body action with the help of rational and intuitive cognition. The first model determines philosophy as logic, mathematics, and physics, and the second belongs to art because it rests on the aesthetic power of imagination, without which rational cognition cannot be the unconditional power of absolute subjectivity. It is self-evident that the rationalism of the 18th century, when, after all, aesthetics as a philosophical doctrine of the beautiful and the sublime was born, is the most significant extension of the modern obsession with science and technology, what Michel Foucault called *mathesis universalis*[6].

I single out all this synthetically to show how, as Hans Blumenberg would say, *the myth of modernity* was created from the idea of the progress of science as a rational knowledge of Being. So, logically, it is equally certain that only reason, as the mind's fundamental structure of the world, determines what is

[6] M. Foucault, *The Order of Things: An Archeology of the Human,* Vintage, New York 1994.

certain and what is not. It is clear, therefore, that non-certainty as a negation of that *certitudo should be found only in the field of Cartesian res extensa* and not *res cogitans*[7]. Everything uncertain becomes chaotic and unordered, from Thomas Hobbes' horror of the «state of nature» to the revolutionary events in the realm of freedom and pure will as the essence of politics. Doubt that only thinking as logic and mathematics or the system of rationality of the modern world could have hidden within it something uncanny and rational and hence produce the structural uncertainty of the new world of fascinating reaches of contemporary technology such as automobiles, locomotives, hydropower plants, nuclear energy, etc. would only open entirely different perspectives of the so-called criticism of the Anthropocene after the Second World War. There is no doubt that Heidegger's thinking, which sees the unconditional progress of rationalism and technology in the 20th century as the greatest *danger* to the process of the destruction of Being, denotes a way for overcoming metaphysics as nihilism. *From the horizon of that constellation, uncertainty becomes a setup (Gestell) that is the essence of technology. It should only be understood as the origin of the abyss in the openness of modern metaphysics*[8]. Hence, this thinking with Cartesianism and Leibniz reaches the peak of *technodicy* and becomes the fundamental problem of the emergence of every possible risk, contingency, and chaos in the world as an apocalyptic event. Suppose the destiny and mission of the West genuinely emerge from metaphysics as a fundamental structure of thought that we inherited and continue to develop even in the age of the technosphere. In that case, even the end of technology in the idea of artificial intelligence means nothing «new» or unexpected but only corresponds to the path of technology.

The end of technology is predetermined and decisive from its very beginning because it corresponds to the decision about the supremacy of being over the primacy of Being. The end of technology does not simply mean no more further, but the opposite, because the end has already been decided for a long time and is always irrevocably and so on in its preliminary success[9].

In the «ontological» sense, the danger (*Gefahr*) that Heidegger says cannot be something external, as such, is necessarily located in the essence of

[7] H. Blumenberg, *The Legitimacy of the Modern Age*, MIT Press, Cambridge, MA-London 1985.

[8] M Heidegger, *Die Frage nach der Technik*, in *Vorträge und Aufsätze*, Klett-Cotta, Stuttgart 1954, pp. 9-40.

[9] M. Heidegger, *Leitgedanken zur Entstehung der Metaphysik, der neuzeitlichen Wissenschaft und der modernen Technik*, GA, Bd. 76, V. Klostermann, Frankfurt am Main 2009, p. 312.

nature as *physis* but is shown in the establishment of goals and plans for the transformation of nature into a modern system of information and energy delivery, such as the system networks of nuclear power plants or a communication system based on carbonised production and consumption that destroys a country's environment. *The danger denotes an apocalyptic event of risk, contingency, and chaos in the very essence of the rationalism of modern technology that drives progress in the core of science, not the other way around. If this is so, then the concepts of so-called cybernetic ontology after the 1960s and the introduction of the technosphere into everyday life as a triad of risk, contingency, and chaos so that the paradox is complete are no longer in the service of traditionally metaphysically understood uncertainty but a new order based on hybrid concepts such as Gilles Deleuze and Félix Guattari's chaosmos and Simondon's metastable equilibrium.* Heidegger's thinking about the event of openness as the «second beginning» of authentic history in the coming future can no longer be compared with a different logic of things that belong to the essence of the technosphere. Let us see what we have instead of «danger» and the openness of events [*Ereignis*]. We have a meta-theory of the uncertainty of events, which is based on the logic of cybernetics with the technosphere as an autopoietic way of unfolding reality in intervals of risk, contingency, and chaos[10].

Risk comes from the Italian *risco*, *rischio* and the French *risque*, which means a kind of danger that can be predicted to a certain extent by determining its intensity. In addition, risk is the ultimate loss or damage caused by war, natural disasters, and poorly assessed investments in the capitalist economy. This understanding is mainly reduced to the «profane» functioning of modern society, for which the decision of a free individual on the market represents a model of action. The negative Cartesian concept of uncertainty cannot encompass the actuality of events, but the very logic of reality characterises reality in cognitive-theoretical insight. Certainty, therefore, is *certitudo* and refers to human judgement about Being as such, speaking in Heideggerian terms. This judgement generally cannot be wrong because truth is understood scholastically as the correspondence of opinion with things. It is certain, for example, that contemporary global capitalism becomes the result of the technoscientific construction of reality as a network of events that appear cybernetically in the fourfold of information-feedback-

[10] Deleuze, Guattari, *op. cit.*; G. Simondon, *L'individuation à la lumière des notions de forme et d'information*, Jérôme Millon, Paris 2017; Ž. Paić, *Art and the Technosphere: The Platforms of Strings*, Cambridge Scholars Publishing, Newcastle upon Tyne 2022.

control-communication. *Nothing outside of this new trans-classical logic can be «efficient», «useful», or «pragmatic». Moreover, this certainty cannot exist by itself but represents a pure construction of events that rests on what we call probability in the theory of science*. The probability that something will happen just so assumes that the certainty of the occurrence of an event is not in the authority of God, nature, or man but in the authority of programming the course of events as a model of projection of reality in a specific shorter or longer period. The assessment of the risk of action corresponds to contemporary philosophical theories of probabilism, which have become scientifically binding and indelible in the media. Probability denotes the basic word for expecting the coming future as a risky event in the meteorological discourse of storm and hurricane forecasters; only after does so-called nice weather occur. The acclaimed sociological theory of Ulrich Beck's so-called risk society denotes, however, only one of several theories of negative probabilism, and it cannot «ontologically» deal with Deleuze's theory of the societies of control[11]. Why? Control denotes the third key concept of cybernetics as a *causa efficiens*. It signifies the possibility of a dynamic-active way of ruling over society as an object[12]. At the same time, the risk is only a consequence of what the cybernetic control system constantly produces. *Namely, nonhuman control produces risky consequences in society because it replaces human existential uncertainty with the rational order of risk, contingency, and chaos*. We used to be able to complain about lousy fate and curse God for such a fate. Nature took the place of the divine, and today, we have everything to blame for the system of new rules of this global planetary game of information capitalism; Deleuze says that anything should be rational except capitalism itself.

Contingency [*contingentia*] means the randomness of an event, then its uncertainty, the possibility of something being different than it is. In contemporary philosophy, especially in speculative materialism and postmodern pragmatism, Quentin Meillassoux and Richard Rorty, especially in logic, indicate the status of statements that are neither necessarily true nor necessarily false but depend on the context in which the statement about something appears. It seems evident that contingency cannot be a mere negation of necessity in the randomness of the event. A linguistic statement becomes true only from a specific situation, not a priori. For this reason, the term was used in the philosophy of pragmatism from William James to

[11] U. Beck, *Risk Society: Towards a New Modernity*, SAGE, London 1992.
[12] G. Deleuze, *Postscripts on the Societies of Control*, in «October» 59 (1992), pp. 3-7.

Hilary Putnam. However, it originates from late Wittgenstein and his theory of «language games» [*Sprachspiele*] as «forms of life». Language is not necessarily a universal signifier of thought but rather a contingent possibility of the occurrence of an event when it unexpectedly enters the horizon of thought as telling and perceiving and thus changes the order of the conceptual-categorical series. It is no coincidence that since the emergence of cybernetics, this term has also been expanded in the technosphere, politics, science, culture, and art. Everything suddenly became contingent precisely because the old metaphysical order with its ontological hierarchies of Being, beings and the essence of Human no longer works. Contingent means, therefore, the irreducible otherness of the event that one tries to think probabilistically, but in such a way that it is not appropriated and reduced to *object X*. Chance can no longer be the negation of necessity in the sense of Being-God-World-Human but rather the necessary contingency of the possibility that a third exists and that his logic is trans-classical like the technosphere. *Everything that can still be philosophically coherently said about contingency comes down to what is entirely different, unforeseeable and undetermined, uncertain and impossible from traditional, modern metaphysics from Descartes to Hegel. Let us be even more precise. Contingency means the opposite, not the negation, of necessity within Immanuel Kant's categories of modality so that neither possibility nor reality arises as a hierarchy of potentiality of thought but as that which gives a different meaning to the very possibility of the emergence of a new event starting from the absolute necessity of the Other in its autonomy and positivity.* What is the «function» of contingency in understanding cybernetic thinking? Nothing other than being the «essence» of an event that is not but happens in its contingency. Hence, a new event becomes contingent, not necessarily risky and chaotic[13].

Chaos in Greek means emptiness, boundlessness, a state without order and predictability, formlessness, indeterminacy, lawlessness, and disharmony. Greek mythology is about the deity by whom light and day, Earth and the underworld, and love were created. However, the modern understanding of chaos is entirely different. Chaos theory enters many areas of mathematics and focuses on the so-called deterministic laws of dynamic systems. *The concept at the centre of this theory is not necessity but chance in the sense of the disruption of order as a deviation from the usual course of cause and effect. Chaotic complexity systems are based on interdependencies and the*

[13] See J. Williams, *Gilles Deleuze's Philosophy of Time: A Critical Introduction and Guide*, Edinburgh University Press, Edinburgh 2011.

cybernetic notion of feedback loops, repetition, self-similarity, fractality, and self-organisation. Deterministic non-linear systems produce significant differences in the initial states of matter and energy, so it is an accurate metaphor when we say that when a butterfly flaps its wings in Brazil, suddenly, a tornado blows in Texas.

2. *Sed quid igitur fidem? Res cogitans. But still, what am I? A thing that thinks*[14]. Thing? Thinking? To think means to be present in thoughts as I. However, if artificial intelligence soon reaches this Cartesian position, we can freely say that apart from humans as a thing that thinks, there is also a thing that feels like what – human? If this is so by analogy with human thinking, then AI can have its subjectivity as a thing that thinks, or in other words, its I. *Self was a fundamental issue of modern philosophy. That Cartesian basis of thinking expresses itself in the language of thought and imagines the objects of its thinking in an image to the ultimate limit of thinkability, which we call absolute subjectivity*. At the same time, it is the key to understanding rationalism since Descartes was the thinker and mathematician, i.e. a scientist, who brought Aristotle's definition of human as *animal rationale* to the postulate of all Western metaphysics. This does not exclude that thinking takes place in the environment of the *cogito* that inhabits the human body and is endowed with animal desires, i.e. passions, and that between mind and body, there is what belongs exclusively to humans as a sphere of ethical mediation in society and community, which is always an expression of this hierarchically organised relationship between thinking, feeling and bodily automatism, that which belongs to the area of *res extensa*. This thing extends into infinity because it is about materiality, not spirituality. Finally, it follows from the Cartesian way of thinking that beings naturally absorbed by their instincts, i.e. animals, are necessarily automatons without a soul because their physicality within the given environment is the fundamental substance that determines the «meaning» of their existence and life. All the radical criticisms of modern Cartesianism as a contemporary metaphysics of subjectivity, paradigmatic among which are Heidegger's as well as Deleuze's, who therefore takes Spinoza and Leibniz and the thinking of immanence and «vitalistic materialism» as his true predecessors, still do not dispute something as a fundamental assumption for contemporary thinking.

[14] R. Descartes, *Meditationes de prima philosophia,* Demetra, Zagreb 1993, p. 94; translated from Latin by Tomislav Ladan.

Humans are *res cogitans* as things that think, and thinking necessarily presupposes individuation. However, this does not mean that some form of trans-individuation, as Simondon would say, cannot be the «cause» of necessary unification and human contingency. The only question is whether thinking as such should be universal and non-personal, though not in the sense of the Freudian One or *Es*, nor in the sense of Heidegger's conception as the openness of events (*Besinnung – Ereignis*), nor in the mind of the poetic expression of Arthur Rimbaud: is it wrong to say I think because I am someone else, truly just a «collective matrix» or, ontologically speaking, the totality of this already coincidental individuation or something even more original than the summing up of what is collected in thinking as telling? Before the individuation of the *cogito*, was there some pre-reflexive environment of thought that did not act in the way of absolute subjectivity but was organised in a completely different way against this I or Self? If AI thinks analogously to human thought, then «it» can call itself I and name itself in the language of human communication, like Stanley Kubrick's HAL 9000. But that is just an extension of human subjectivity to – what? Artificial intelligence, symbolically speaking, becomes the artificial brain in different bodies as devices and technical devices. Human subjectivity is embodied in one body, which, if we remove Descartes' definition of an automaton, has its passions, experiences, and imaginations, which suffers and enjoys, is born and dies as a *res extensa*. Does the same apply to AI, or is artificial intelligence a trans-individuation that can be «implanted» in English as an embedment into many devices and technically exist in them as a robot-cyborg-android in a posthuman condition? My notion of the technosphere goes beyond Cartesian dualism. Still, I do not dispute that the question of the individuation of thought is the ontological-epistemological dividing line that separates the human way of thinking from the non-human in the sense of the post-biological existence of superintelligent computers that, in addition to mind, also have a soul, i.e. tend to possess an artificial intuition. Therefore, thinking cannot be reducible only to *animal rationale*. Instead, thinking presupposes It and I as bodily individuation in the existential performance of a one-time life, not in the immortal substance of living, as Heidegger would say in *Being and Time* with the German word *Jemeinigkeit*[15]. However, the difference between «me» and «my power» as human existence versus «it» and «it», the technically created individuation of thinking as

[15] M. Heidegger, *Sein und Zeit*, GA, Bd.. 2, V. Klostermann, Frankfurt am Main 1977, pp. 153-173.

the calculation-planning-construction of a robot-cyborg-android, is that the «power» lies in the vulnerability and cunning of itself. Bodies appear on the horizon of thought, not through the projections of the world. What should be unique to human thinking cannot be a program but a vision[16].

With that in mind, what do we «see»? That there is no collective human mind, no Jungian collective unconscious, no quasi-mystical monolith from the pre-Stonehenge or Altamira era, which visionarily programmed us without, of course, being aware of it, so that we would be thought beings in our bodily-enactive subjectivity, as they would say in the language of neurophilosophy today. *No, instead of the trace of the divine Great Primordial, we have the mystery of individuation, which, in its pre- and post-state, presupposes the natural-and-technical as a synthesis of the creation of a thing that thinks, but that thing is not a thing in the sense of the creation of some stone or hardened lava, but a thing as the essence of thinking, which, in turn, is nowhere outside of thinking, but only happens when a person is a conscious being who thinks by being aware of himself, of his authentic Jemeinigkeit in one way or another at every moment – and let's be absolutely clear, that authenticity does not come from anything else, from any borderline situation of war or peace in society and politics, but only from the opinion of what is, what was, and what will be, the Self as the Self and the Self as the One that even in the posthuman condition of transindividuation is nothing but an issue about the meaning of the existential event that a thinking being with the will and desire to live leads from the beginning to the end or in the tendency to the infinity of what we call time.*

To conclude, individuation cannot be just the process of creating a fundamental and indivisible Self as an absolute subjectivity of thought. This is how one can think factually and contingently in «one's» body, be it living or artificial, Christlike or technological, imbued with the mysticism of suffering on the cross or the joy of life as an all-powerful Nietzschean affirmation of the will to power as an eternal recurrence of equality.Trans-individualised lies in the collective mind of something uncanny and inhuman. The problem lies in thinking as telling and visualising. Philosophy and art are mythopoetic sources of human thinking, while sciences are not; they are the technology of pure construction of the world as a realm of objects that think technogenetically. Philosophers and artists can be, and most often are, crazy and eccentric, outside the community and the mind of «common sense», solitary

[16] Ž. Paić, *Brain as a Vision and Program*, May 3, 2023, https://zarkopaic.net/blog-post/brain-as-a-vision-and-program/.

like Friedrich Nietzsche's «rare plant» or the nomad in Friedrich Hölderlin's parables about the poets in *Bread and Wine* who, like eternal wanderers, follow their stars «in the holy night». Scientists are never like that because they are driven by the so-called objective truths of their research «frenzy». Thinking becomes an event of individualised confrontation and struggle with chaos. And Deleuze is right when he claimed that

> The philosopher, the scientist, and the artist seem to return from the land of the dead. What the philosopher brings back from the chaos are variations that are still infinite but that have become inseparable on the absolute surfaces or in the absolute volumes that lay out a secant [sécant] plane of immanence: these are not associations of distinct ideas, but reconnections through a zone of indistinction in a concept. The scientist brings back from the chaos variables that have become independent by slowing down, that is to say, by the elimination of whatever other variabilities are liable to interfere so that the variables that are retained enter into determinable relations in a function: they are no longer links of properties in things, but finite coordinates on a secant plane of reference that go from local probabilities to a global cosmology. The artist brings back from the chaos varieties that no longer constitute a reproduction of the sensory in the organ but set up a being of the sensory, a being of sensation, on an anorganic plane of composition that is able to restore the infinite. The struggle with chaos that Cézanne and Klee have shown in action in painting, at the heart of painting, is found in another way in science and philosophy: it is always a matter of defeating chaos by a secant plane that crosses it[17].

Humans, as the governor of chaos on Earth, go to the sky as a figure of the historically created mystery of the creation of the One who is not a thing that thinks but a thing that makes new from itself and through the process of *autopoiesis* as techno-symbiogenesis, which means that both insects and wasps, ticks and flies think, but entirely *differently* than humans. Wittgenstein made the most enigmatic statement about this in modern philosophy in general:

> If a lion could speak, we could not understand him[18].

[17] Deleuze, Guattari, *op. cit.*, p. 173.

[18] L. Wittgenstein, *Philosophical Investigations*, 2nd ed., Basil Blackwell, London 1958, p. 223.

3. By analogy with Aristotle's concept of cause-purpose, cybernetics always acts in such a way that its «essence» exists in digital constructivism, which means that information creates a feedback loop. This effectiveness of system control reversibly produces interactive visual communication in the surrounding world of autopoietic states and not beings. *The cybernetic «fourfold» becomes a techno-poietic one, and the metaphysical one set by Aristotle is necessarily organological, which means that creation is always linked to the model of human-as-artisan*. This model determines the concept of art throughout the entire history of metaphysics, and it is interesting that both Heidegger and Deleuze took it over and transformed it in their own way. Hence, the primordial fourfold of Being-God-World-Man establishes the rule of thought as mimesis and representation of what already exists in the idea of the divine cause of all action – Aristotle's immovable mover. Cybernetics denotes the construction of what does not exist but is the techno-poietic creation of artificial reality. Artificial intelligence represents only a continuation of the cybernetic foursome for its model of autopoietic thinking-action; it can no longer be an artist-as-craftsman in all its transformations up to Deleuze's model of meta-film as a montage of living and non-living assemblies. The model for cybernetic creativity becomes inhuman, the black monolith from Kubrick's *2001: A Space Odyssey*, pure techno-genetic thinking that neutralises and suspends both the first and second metaphysical fourth, and the Being-God-World-Human and the formal, material, practical, and final cause. The fourfold of information-feedback-control-communication no longer has anything to do with metaphysics, although the technosphere is realised by analogy. But the problem is that the technosphere, and here is my essential difference with Deleuze, cannot be any form of «immanent transcendence» other than in the order of chaos and contingency that transcends the boundaries of ontology and cybernetics in general because it synthesises mind and intuition in a thought that is no longer divine or human, only thinking as an event of absolute creativity *of homo kybernetes* as a necessary stop on the way to the singularity of everything thinkable and possible as such. *Singularity no longer needs any fourfold because the essence of metaphysics, cybernetics, and transhumanism has been realised*[19].

In the tradition of metaphysics, the word-concept «Other» refers to the existence of the world beyond the limits of empirical knowledge. The second world is the one that refers in Christian theology to the kingdom of God

[19] See Ž. Paić, *War, Technosphere, and the Question of Evil*, in «Teoria» 43 (2023) n. 2, pp. 27-48.

beyond this world of materiality and the obsession of real Being and eternity and bliss belong to it. Therefore, God defines himself legitimately as the «Big Other» because he rules over this world, starting with the irreducibility of transcendence. To be over and beyond means to be concerning transcendence and metaphysics. Therefore, nothing in this world, from Plato through scholasticism to modern philosophies of spirit, happens autonomously because human existence represents the most significant reach of the freedom of this being whose essence is determined by his spirituality with its origin in that which resembles the divine, but not as a simulacrum of God. Instead, we can realise the five transcendentals as conditions for the possibility of all reality in general: *Unum, Bonum, Verum, Ens,* and *Pulchrum*. Both Platonism and Aristotelianism, as Idea and Energeia, are fundamental words for the meaning of Being, and they assume that the world was created by an act of divine will or the act of creation. Still, only the Thomism of scholasticism will equate God with the thought-concept-act of creative activity, which, in principle, corresponds to what humans have through the experience of all five senses. More transcendental are the ideas and forms in which the possibility of creating the human world in its perfection appears. *The five transcendentals encompass the five spiritual senses by which what arises from God's substance is realised here. The «Big Other» must necessarily be outside and beyond this world, and that other and different cannot be related to the characteristics of beings but rather Being as preceded by God*. In scholastic logic, as is known, which continues with Aristotelian logic, the term *tertium non datur* or «the third does not exist» indicates that it is not possible for a being to exist both here and there at the same time and that the truth of Being appears only in judgement in the sense of matching opinions and things themselves such that only one statement can be confirmed, the other being untrue or false. *The fundamental logic of metaphysics is either/or logic.*

But from Plato to Georg W.F. Hegel, Karl Marx, Sigmund Freud, Jacques Lacan, and cybernetics, we encounter something self-evidently uncanny: *Unheimlich*. Plato established the triad or trinity of theory, practice, and production. The latter is established as *poiesis* and signifies the production of beings from Being. Based on this assumption, Christian theology determines God as one who creates from nothing (*creatio ex nihilo*). *Poiesis* denotes a creation, work, and production by which man shapes his world as a work of necessity and freedom because production in the active and non-active sense, for example, poetry and sculpture, is something «innate» to him that determines his essence. Of course, production ranks third after theory and practice. Hegel's absolute is third in the highest rank and has the

characteristics of totality in historical-dialectical progress and development.

Thus, philosophy denotes a condition for the possibility of creating absolute science and is above art and religion. The «Big Third» appears for Hegel as the essence of absolute metaphysics and as the end of history in the sense of termination-overcoming [*Aufhebung*] the previous two stages of historical development: nature and society, subjective and objective spirit. Only at the third stage does the meaning of philosophy as metaphysics become severe, and history has the character of a theodicy of the world spirit that knows itself as the truth of the entire process of events from beginning to end. Marx, on the other hand, gets the «Big Third» through the so-called triple pattern of the rule of capitalism from the first stage of goods through money to capital or the pure idea of the actual process of the development of world history, which even prevails in what Vanja Sutlić calls the practice of work as scientific history, and that is nothing more than communism as the highest form of the meaningfulness of history in general[20]. Man, both in Hegel and in Marx, is a free individual who exists only in synthesising the «Big Third» and history as the conception of the idea's movement through the necessary characteristics of its appearance. *Communism denotes the «immanent transcendence» of the historical absolute in which man exists. Still, communism is not the goal of history but its end in the metaphysical sense of the word.* Freud talks about the subject's consciousness stage through the stages of *Id, Ego, and Superego*. Of course, the fundamental problem of psychoanalysis is the «Big Third», which commands and oppresses, liberates and oppresses, becomes an insurmountable obstacle for the free development of a person, or has the value of faith and hope in a future society of happiness, well-being, and all-round personal development. Lacan «deconstructs» this same scheme, so we have the imaginary, the symbolic, and the real, which is traumatic because it is a split between the first and the second and exists only as a desire to reach a sublime object. The «Big Third» within metaphysics is, therefore, the initiator of history, its goal in the sense of the synthesis of substance and subject, a moment that is found in the actuality of the «here» and «now» and not in the mythical past or the indefinite future.

To be third means to be that which transcends human existence in terms of form, but it is found in it in terms of content and determines the limits of its activity. Since humans do not have their own eternal and permanent «essence», the «Big Third» is always metaphysically thought of as Kantian regulative action in the sense of human historical perfection, self-conscious-

[20] V. Sutlić, *The Practice of Work as Scientific History*, Kulturni radnik, Zagreb 1974.

ness, self-organisation, and self-rule until the transition to a state where the character of the human-too-human no longer exists is complete. *God, history, the absolute, work, and what goes beyond the limits of metaphysics and appears in the trans-classical logic of the technosphere, i.e. the virtual actualisation of becoming, are simultaneously the transcendence and immanence of life itself, which, for Aristotle, was already synonymous with Being. However, the difference is that the technosphere, as the third order of cybernetics, signifies the rule of autopoiesis, which thinks and no longer goes through the stages or historical platforms of what belonged to the past. Still, its development indicates the irreversible hyperplasticity of artificial life in the eternal present (nunc stans).* From now on, we can only conditionally talk about the first, second, and third terms within the speculative and reflective triad because the idea of history as eschatology and soteriology, as theodicy and messianism in all imaginable versions from Karl Marx and Walter Benjamin to Emmanuel Lévinas and Jacques Derrida is no longer valid for the technosphere. We no longer value the «third» from dialectical logic as the highest in the rank of things. No, all that lies behind us belongs to the musealised past of thought.

The paradigm has three correlative meanings for understanding the technosphere. All three are at the same level of conceptual-categorical redundancy and are almost tautological. The reason for this is that the technosphere no longer has the metaphysical meaning of Aristotle's logic as an *episteme téchne*, but in the middle is an autopoietic construction of an artificial reality that does not exist in the so-called first or actual reality. When Simondon, the most crucial philosopher of cybernetics, explained that the concept of information cannot be by analogy the same as Aristotle's form in the sense of *eidos* and *morphé*, it was a reversal in the essence of metaphysics. The reason was that with the emergence of the thinking machine or computer, the model or matrix of the entirety of Western philosophy as an ontology changed. «Paradigm» should, therefore, be freed from its metaphysical meaning in Greek, even though the term is historically Greek, like almost all others that we still use in philosophy and science today, albeit with changed meanings after the Latin language dominated Western civilisation in the era of Rome and Christian scholasticism. Hence, these three meanings of the paradigm belong to an area that is no longer bounded by the relationship between philosophy, science, and art in the thinking of Being, traditionally speaking, but is open-closed in the «black box» of cybernetics, which, in Hegelian terms, presupposes the thinking of a thought. The paradigm, therefore, must be understood as (1) the transversal thinking of

the technosphere as a contingent event of the creation of a new artificial life; (2) a matrix or a cognitive-theoretical-pragmatic-production framework in which thinking is articulated throughout history as linguistic, the visual and numerical code thus becoming a universal tool for creating and retrieving reality in thoughts; and (3) a conceptual-categorical system of thinking is shown as a fundamental structure, such as idea in Plato, energy in Aristotle, the spirit in Hegel, work in Marx, event in Heidegger and Deleuze, theory of relativity in Einstein, theory of black holes in Hawking, and technological singularity in posthumanist Kurzweil[21].

The thinking of the technosphere is neither a thought «about» something nor an opinion «on» something but paradigmatic thinking in the form of a cybernetic circle of circles by which artificial intelligence constructs an artificial life from a trans-classical logic that combines the fourfold information-feedback-control-communication into a «transversal order» of meaning. «Transversal» in Latin denotes a transverse direction intersecting two other directions. In a figurative sense, we can talk about a crossroads that shortens the way from one place to another. Wavy motion in cosmology takes the transversal as a paradigmatic form in which thought appears through the above three meanings. That is why «paradigm» represents a term used in philosophy or the theory of science and refers to a reversal from the previous paradigm in understanding the universe's origin and essence concerning the relationship of matter, energy, and information. For example, the difference between Ptolemy and Copernicus is that they designated the best example of the emergence and operation of a paradigm in thinking as a universal communication system between actors in the human world. The same applies to the difference between Isaac Newton and Albert Einstein. However, the technosphere cannot be only about unconditional progress in understanding the scientific picture of the world; rather, it is also – and primarily – about an open-closed order of thought figures that are not a mirror of reality but conceptual holograms that are interconnected with other such tools of thought. Paradigm denotes a fundamental structure by which we think of matrices in a historical-epochal sense, such as Platonism and Aristotelianism in the Renaissance or Nietzscheanism and Heideggerianism in postmodern or 21st-century philosophy. It is the power of designing disorder and cuts in the historical development of thought rather than order, so the paradox is complete. Why? *Because the paradigm of the technosphere should*

[21] Ž. Paić, *Superfluity of the Human: Reflection on the Posthuman Condition,* Schwabe Verlag, Basel 2023.

be labelled as thinking in a transversal journey through emergent events in chaos and contingency. This journey becomes a post-metaphysical wandering without a first cause or a final purpose. What else does wandering mean besides an event of an entirely different matrix or paradigm of thought in the contemporary world?

Abstract

Chaos and entropy are critical concepts for understanding our contemporaneity. Instead of the necessity and pre-stabilised harmony of the cosmos and the world, on which classical and modern metaphysics still rest, everything «collapses», becomes «curved», and evaporates in «black holes»; only event horizons remain from an astrophysical point of view.

The assessment of the risk of action corresponds to contemporary philosophical theories of probabilism, which have become scientifically binding and indelible in the media. Probability became the basic word for the expectation of the coming future as a risky event in the meteorological discourse of storm and hurricane forecasters, only after which does so-called nice weather follow.

In this article, the author tries to articulate the fundamental assumption that the technosphere as autopoiesis becomes, at the same time, a matrix of new action in the system and environment of human-non-human communication and a model for the possible management of chaos, contingency, and technological singularity as the main concepts of contemporaneity.

Keywords: uncertainty; risk; chaos; contingency; technosphere; technological singularity.

Žarko Paić
University of Zagreb
zarko.paic@zg.t-com.hr

T

Veronica Neri

Intelligenza artificiale generativa, *deepfakes* e identità vulnerabile. L'etica dell'incertezza come risposta a un rischio (in)controllabile

1. *Premessa*

Il 30 novembre 2022 i mass media annunciano al grande pubblico la messa a punto di sistemi di intelligenza artificiale (IA) generativa. L'IA sta vivendo la sua primavera e ha acquisito in poco tempo un posto di primo piano nell'arena pubblica, non solo tra esperti e mondo della ricerca. Pochi lustri prima, intorno agli anni '90 del secolo scorso, si assiste, invece, in molteplici ambiti disciplinari, a un altro cambiamento che ha permeato la società, la così detta *iconic turn*, ancora in corso, e che ha interessato anche i sistemi di IA generativa.

La svolta algoritmica e la svolta iconica insieme hanno dato vita a sistemi di IA altamente performanti e performativi di generazione di immagini e *deepfake*[1].

Un simile scenario si incardina nella società globale del rischio come definita da Beck[2]. Le innovazioni nel campo dell'IA visiva, come ogni invenzione non solo tecnologica che si è susseguita nel corso del tempo, hanno generato opportunità, ma anche rischi, pericoli e minacce. L'incertezza sulle implicanze del loro impiego sugli individui, sulla comunità e sulla società in generale hanno aperto la porta a nuove sfide etiche.

Si tratta di sistemi in grado di agire simulando – senza comprenderne il senso – i processi mentali degli individui, dipendendo solo in parte dalla

[1] K. Hill, *La tua faccia ci appartiene*, Orville Press, Milano 2024; A. Pinotti, A. Somaini, *Teoria dell'immagine. Il dibattito contemporaneo*, Raffaello Cortina Editore, Milano 2009.

[2] Cfr. U. Beck, *World Risk Society*, Polity Press, Cambridge 1999 (*La società globale del rischio*, trad. di W. Privitera, Carocci, Roma 2013); Id., *Conditio humana. Il rischio nell'età globale*, Laterza, Roma-Bari 2011.

 DOI: 10.4454/5amq3312

volontà umana. Aprono a nuove preoccupazioni circa la salvaguardia della propria identità, la *privacy*, la tracciabilità dei dati, la manipolazione delle decisioni umane, discriminazioni strutturali, l'(in)transparenza delle procedure fino alla creazione e propagazione di nuovi immaginari sociali e/o al rafforzamento di vecchi intrisi di *bias*, incidendo sulle nostre scelte etiche, estetiche, pubbliche e informative[3].

Se il rischio – e il pericolo – connaturato a tali sistemi visivi allenta, fino a annullare, il nostro controllo, occorre ripensare il concetto di rischio alla luce di tali cambiamenti tecnologici. Sembra un rischio da intendersi non solo come un qualche cosa di (anche solo parzialmente) calcolabile[4] – sulla scia di quanto la più recente regolamentazione europea presuppone –, quanto di un pericolo, ovvero della probabilità che un evento, in un arco temporale definito, si verifichi indipendentemente dalla decisione umana, e di incertezze, eventualità non pronosticabili né misurabili. Come nella realtà oggettuale non possiamo calcolare la probabilità che il rischio di frodi o manipolazioni avvenga con una certa regolarità e frequenza statistica, così nella dimensione aperta dall'IA simili eventi possono solo estendersi in ragione dell'aumentare delle possibilità offerte dagli strumenti tecnologici che li consentono, ma non possiamo controllarli né quantificarli statisticamente; né possiamo calcolare l'impatto effettivo dell'IA generativa di immagini sulla creazione di nuovi immaginari sociali in termini probabilistici[5].

Il rischio svela pertanto gli stati di incertezza e di vulnerabilità – recepita come la predisposizione a subire un danno – ai quali è esposta l'umanità di fronte alle tecnologie artificiali. Non possiamo adottare dunque il principio di probabilità e calcolo razionale, riconducendo tutto a una impostazione positivista *tout-court* ovvero ai fattori tecnici insiti nel *design* stesso del sistema, peraltro anch'essi permeati dalla soggettività di chi lo realizza. Con il presente contributo si vuole mostrare come occorra ricalibrare il concetto di rischio dal punto di vista dell'etica non solo individuale, ma soprattutto pubblica e sociale, poiché pertiene la società nella sua complessità, includendo tutte le possibili categorie umane senza discriminazioni al medesimo tempo.

Alcune strategie di etica pubblica possono indirizzare l'opinione pubblica e la cittadinanza alla consapevolezza e alla co-responsabilità di tutti i soggetti coinvolti, dagli ideatori, ai governi (che debbono prendersi in carico

[3] https://www.agendadigitale.eu/cultura-digitale/cose-il-rischio-cosi-la-filosofia-ci-aiuta-a-capire-il-senso-dellai-act/

[4] G. Sturloni, *La comunicazione del rischio per la salute e per l'ambiente*, Mondadori, Milano 2018, pp. 5-10.

[5] C. Taylor, *Gli immaginari sociali moderni*, trad. it. P. Costa, Booklet, Milano 2005.

regolamentazioni specifiche) fino agli utilizzatori. Ciò per indebolire lo stato di incertezza e di paura nei quali il rischio getta i soggetti in generale e quelli più vulnerabili in particolare, cercando di renderli quantomeno più informati, consapevoli e critici.

Come scrive Lagadec rispetto ai primi dispositivi tecnologici con l'IA generativa di immagini siamo di fronte alla nozione del «rischio tecnologico maggiore» poiché la fragilità dei sistemi e i pericoli che fanno correre agli esseri umani aumentano il senso di vulnerabilità dei medesimi e diminuiscono il senso di fiducia nei confronti di una società ossessionata dalla sicurezza e dal controllo[6].

Il contributo si articolerà in tre parti. Dopo una introduzione sull'immagine artificiale e le *deepfakes* si affronterà il tema della vulnerabilità dell'individuo scontrandosi con *bias*, allucinazioni, protezione dei dati e il rischio di identità plurime, fasulle e incontrollabili; la terza e ultima parte affronterà il tema del rischio e dell'etica dell'incertezza rispetto alle immagini artificiali, cercando di proporre spunti di riflessione alle minacce etiche emergenti.

2. *Intelligenza generativa di immagini e il caso delle* deepfakes

La comunicazione visiva risulta sempre più supportata dall'IA. Questa tendenza ha, da una parte, potenziato alcuni ambiti disciplinari, dall'altra, ha imposto una verifica dei contenuti digitali a chi ne fruisce e, dunque, indebolito la fiducia nelle immagini in generale. Numerosi sistemi «che si comportano come se fossero intelligenti»[7], ad esempio ChatGPT per i testi e DALL-E per le immagini, solo per citare i sistemi di IA generativa più noti, hanno contribuito allo sviluppo della diagnostica clinica in ambito sanitario, alla semplificazione dei processi di traduzione, alla creazione di testi in diverse lingue adattati ai pubblici di riferimento[8], o, nel campo della giustizia, ai riconoscimenti facciali, fino all'ambito della creatività nel settore pubblicitario e artistico o al contesto giornalistico (per descrivere fatti e eventi)[9]. Tali strumenti hanno al-

[6] P. Lagadec, *La civilisation du risque. Catastrophes technologiques et responsabilité sociale*, Le Seuil, Paris 1981; G. Liuzzo *et al.*, *The Term Risk: Etymology, Legal Definition and Various Traits*, in «Italian Journal of Food Safety» 3 (1), 2014, p. 2269.

[7] M.R. Taddeo, *Costruire l'etica dell'intelligenza artificiale*, in *Il potere del pifferaio magico*, a cura di G. Fregonara, UPI, Pisa 2021 p. 166.

[8] D. Baidoo-Anu, L. Owusu Ansah, *Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning*, in «Journal of AI»7, 1 (2023), pp. 52-62; A. Barale (a cura di), *Arte e intelligenza artificiale. Be my GAN*, Jaca Book, Milano, 2020.

[9] L. Gaur (eds.), *Deepfake. Creation, Detection and Impact*, CRC Press, Boca Raton 2024,

tresì prodotto immagini e audiovideo tanto realistici quanto falsi indistinguibili da video realizzati con media tradizionali, manipolando l'opinione pubblica e/o ledendo la *privacy* e la reputazione di alcune persone.

Si tratta di sistemi che, attraverso interrogazioni sotto forma di stringhe testuali (prompt) da parte dell'utente, offrono risposte come se fossero individui in carne e ossa, mettendo in scena un vero e proprio dialogo persona-macchina.

Cristianini, in questo scenario, individua tre livelli di azione dell'IA generativa «*l'agente* che incontriamo nel mondo (per esempio ChatGPT), il *modello* interno che questo usa per prendere decisioni (per esempio GPT-3) e l'algoritmo che crea tale modello partendo dai dati (per esempio, il Trasformer). Per modello si intende il modello di mondo preso in considerazione che deve suggerirci quali eventi e situazioni sono probabili e quali non lo sono. Tale modello plasma solo una parte di mondo alla volta e consente di interagire con il mondo stesso divenendo una forma di comprensione del mondo stesso»[10]. Lo stesso accade quando la risultante della stringa è una immagine. Si tratta di una immagine generata, rispetto a quelle tradizionali frutto dell'immaginazione e della fantasia umane, «grazie a una specifica capacità di uni-formare» degli apparati[11]. Questa affermazione flusseriana, ancora attuale per la specifica tipologia di immagini tecniche che sono le immagini artificiali, mette in luce la capacità algoritmica di generare segni visivi efficaci, realistici o surreali come se fossero la risultante di una intensa attività immaginativa, senza però esserlo a pieno titolo, senza anzi conoscere come si sia arrivati al risultato e chi lo abbia prodotto, l'essere umano e/o la macchina. L'immaginazione algoritmica – se ad essa si può fare appello – risulta ad oggi ben lontana da quella umana[12]. L'immagine tecnica può definirsi, riprendendo Flusser, quale una immagine «generata da un apparato "artificiale", a seguito del predominio graduale del modello algoritmico basato su pixel, che ricostruisce una unità attraverso segni puntiformi»[13]. Occorre pensare alle immagini tecniche non tanto come «tentativi dell'es-

p. 91 ss.; M. Filimowicz (eds.), *Deep Fakes: Algorithms and Society*, Routledge, London 2022; C. Canali, R. Pedrazzi, *L'opera d'arte nell'epoca dell'intelligenza artificiale*, Jaka Book, Milano 2024.

[10] N. Cristianini, Machina sapiens. *L'algoritmo che ci ha rubato il segreto della conoscenza*, il Mulino, Bologna 2024, p. 35.

[11] V. Flusser, *Immagini. Come la tecnologia ha cambiato la nostra percezione del mondo*, Fazi Editore, Roma 2009, p. 15.

[12] E. Finn, *What Algorithms Want. Imagination in the Age of Computing*, MIT Press, Cambridge (MA) 2017; F. Restuccia, *Il contrattacco delle immagini, tecnica, media e idolatria da Vilém Flusser*, Meltemi, Milano 2021.

[13] V. Flusser, *Immagini*, cit., p. XIII.

sere umano estraniato dal mondo di farsi una immagine di questo mondo», quanto «di conseguenze del progresso scientifico»[14].

Attraverso *Generative Adversarial Networks* (GAN) e i Diffusion models, due dei sistemi più noti ed efficaci di generazione di immagini artificiali, è possibile realizzare immagini, anche artistiche, molto variegate per tipologia, per stile, per soggetti, per fine, ecc.[15]. Ciò che distingue una immagine digitale generata con programmi di computer grafica "tradizionali" dalle produzioni delle intelligenze artificiali è relativo, oltre al risultato estetico, al processo creativo. Il programma, dopo l'input dell'essere umano, assume un ruolo autonomo nella creazione dell'immagine[16].

Una GAN, in sintesi, è composta da due reti avversarie che hanno l'obiettivo di migliorarsi vicendevolmente. Da un lato, abbiamo una rete *generator*, il cui compito consiste nel produrre nuove immagini da un data set quanto più ampio possibile (in cui a ogni immagine è associata una etichetta testuale che ne descrive il contenuto); dall'altro, abbiamo una rete *discriminator*, che deve confrontarsi con i risultati del *generator* e segnalare se l'immagine si discosta (e in che misura) dalla richiesta dell'essere umano, se sembra eccessivamente falsa, se il risultato di una "allucinazione" del sistema, ecc. Nel mostrare un possibile errore o una incongruenza la rete avversaria apprende e permette anche alla rete *generator* di apprendere a propria volta. Come afferma Goodefellow:

> The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistiguishable from the genuine articles[17].

L'immagine finale si raggiunge quando il *generator* crea una nuova im-

[14] M. Menon, *Vilém Flusser e la «rivoluzione dell'informazione». Comunicazione, etica, politica*, Edizioni ETS, Pisa 2011, p. 42; V. Flusser, *Kommunikologie*, Fischer Taschenbuch, Frankfurt 1998, p. 102.

[15] L'invenzione delle GAN è tradizionalmente attribuita a Ian Goodfellow e ai suoi collaboratori: J. Goodfellow *et al.*, *Generative Adversarial Nets*, ArXiv:1406.2661 [Cs, Stat], June 2014: https://arxiv.org/pdf/1406.2661.pdf; M. Jovanović *et al.*, *Generative Artificial Intelligence: Trends and Prospects*, in «Computer» 55, 10 (2022), pp. 107-112. Relativamente ai Diffusion Models, cfr. A. Bordas, *What is generative in generative artificial intelligence? A design-based perspective*, in «Research in Engineering Design», 35 (2024), pp. 427-443 e H. Cao *et al.*, *A survey on generative diffusion Model*, arXiv:2209.02646

[16] A. Barale, *Arte e intelligenza artificiale: alcune domande*, in Ead. (a cura di), *Arte e intelligenza artificiale. Be my GAN*, Jaca Book, Milano, 2020, pp. 7-18.

[17] I.J. Goodfellow *et al.*, *Generative Adversarial Nets*, cit., p. 1.

magine che viene percepita dal *discriminator* "autentica"[18].

I più recenti Diffusion models sono invece fondati su un processo di diffusione che trae origine da un prompt testuale e da un dataset di coppie (image, caption). Viene applicato rumore casuale (noising process) alle immagini di addestramento e, successivamente, viene appresa la funzione inversa di denoising, la quale cerca di invertire il processo iniziale e di ricostruire un'immagine condizionata da un input testuale (plausibilmente compatibile ad esso e coerente con le sue richieste). Nel corso dell'addestramento il modello impara a prevedere il rumore aggiunto a una immagine a ogni passo del processo di diffusione per poterlo poi sottrarre correttamente nel reverse process. Il modello apprende pertanto, passo dopo passo, relazioni generali tra immagini da generare e prompt[19].

Nel 2018 presso la casa d'aste Christie's viene venduto il ritratto di Edmond de Belamy, un'opera realizzata dall'IA generativa e da tre sperimentatori del collettivo parigino *Obvious*. Nel medesimo anno l'artista Klingemann realizza con l'IA generativa le *Memories of Passersby I*, una serie di ritratti di identità inventate. In tempi ancora più recenti la fotografa italiana Zanon ha pubblicato uno pseudo-reportage sulla guerra in Ucraina attraverso la piattaforma Midjourney, mentre il fotografo tedesco Eldagsen partecipa al premio "Sony World Photography Awards" con una immagine realizzata interamente con IA. Come scrive Cohen nel caso dell'IA «[l]a creatività non risiede né nel programmatore né nel programma, ma nel dialogo tra programma e programmatore»[20]. Emerge un sistema di co-autorialità non semplice da gestire dal momento che una collaborazione e cooperazione pienamente consapevole tra essere umano e macchina richiama «the lack of a common language between AI and humans»[21].

Il sistema genera immagini combinando insieme segni visivi di un data set

[18] Esistono diverse tipologie di GAN differenti per l'architettura della rete e per come vengono addestrate. Cfr. Z. Wang, Q. She, T.E. Ward, *Generative Adversarial Networks: A Survey and Taxonomy*, in «arXiv:1906.01529» 4 june (2019). Sul concetto di autenticità, cfr. C. Taylor, *The Ethics of Authenticity*, Harvard University Press, Cambridge (MA) 1992; C. Guignon, *On Being Authentic*, Routledge, London 2004; Id., *Authenticity*, in «Philosophy Compass» 3(2), 2008, pp. 277-290.

[19] Cfr. A. Bordas, *What is generative in generative artificial intelligence?*, cit.; O. Sanseviero *et al.*, *Hands-On Generative AI with Transformers and Diffusion Models*, O'Reilly Media, Sebastopol 2024; https://www.agendadigitale.eu/cultura-digitale/creare-immagini-dallimmaginazione-il-potere-dei-modelli-di-diffusione/.

[20] Cit. tratta nell'intervista pubblicata in V. Tanni, *Arte e intelligenza artificiale. Una storia che inizia negli Anni Cinquanta*, Artribune, 30/06/2023: https://www.artribune.com/progettazione/new-media/2023/06/arte-intelligenza-artificiale-storia/.

[21] Cit. di Ali Nikrang in G. Stocker, M. Jandl, A.J. Hirsch (eds.), *The Practice and Art and AI*, Ars Electronica. Ars, Technology, Society, Linz 2023, p. 30.

iconografico tanto ampio da non poter essere 'contenuto' da nessuna mente umana. Possono essere rappresentate relazioni iconiche impensate, unicamente frutto dell'autonomia del sistema stesso. Il risultato può sembrare molto efficace ed esteticamente convincente, per espressività, per l'iper-realtà e/o per bellezza, ma altresì fuorviante, risultato di allucinazioni[22]. Si profilano al contempo sfide di ordine etico su possibili rischi emergenti, come vedremo più avanti, inerenti la distorsione di immaginari sociali, la generazione di *bias* e pregiudizi di genere, etnia e professione, una certa uniformazione della 'creatività', fino ad usi dichiaratamente malevoli come il furto di identità, l'utilizzo (non accordato) dei dati personali e la perdita di privacy.

Nell'alveo del visivo artificiale si inscrivono, infine, le *deepfakes*. Il termine *deepfake* richiama l'unione di *deep*, che evoca i sistemi di *deep learning* (DL), e *fake*, falso. Il sistema che le realizza impiega algoritmi di DL per produrre e modificare immagini, video e audio e generare un media sintetico/falso[23]. Non pertiene più solo immagini statiche, ma video e/o audiovideo che si insinuano in circuiti "intrasparenti". Il sistema algoritmico che ne sta alla base consente in pochi passaggi di creare *ex novo* volti di persone o di alterarne di esistenti (modificando, ad esempio, il colore dei capelli, degli occhi o della pelle, solo per citare alcune possibilità) o ancora, di incrociare, sovrapponendoli, più volti insieme per ottenerne uno unico "nuovo". Può essere impiegata altresì per alterare o generare corpi, ambienti, spazi, luoghi, oggetti, ma anche animali e quant'altro si desideri. Può dare vita anche ad audio immaginari[24]. Se ne può fare, nel complesso, un uso benevolo, creando, per esempio, influencer virtuali e video didattici; può ben essere utilizzato per l'assistenza sanitaria e farmaceutica, per realizzare sfilate virtuali a basso costo da vedere su uno schermo o altre applicazioni di *entertainment*, ricostruire fatti e eventi nell'ambito della giustizia o, come accadeva con la VR, ambienti e personaggi storici; ma possono subentrare, di contro, approcci malevoli, non etici, come lo sviluppo di algoritmi di *deep learning* in grado di trasferire volti di celebrità in video pornografici, in atti di *cyberbullismo* o di violenza più in generale, per diffamare o propagandare idee e pensieri fasulli, scorretti, ecc.

[22] G. Finocchiaro, *Intelligenza artificiale. Quali regole*, il Mulino, Bologna 2024, p. 24.

[23] Le sue origini risalgono in realtà al 1997 e all'ideazione di un programma di riscrittura video: cfr. C. Bregler, M. Covell, M. Slaney, *Video Rewrite: Driving Visual Speech with Audio*, in «Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques» 24 (1997), pp. 353-360; N. Schick, *Deepfakes. The Coming Infocalypse*, Twelve, New York 2020.

[24] A. Tversky, D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, in D.J. Levitin (ed.), *Foundations of cognitive psychology: Core readings*, MIT Press, Cambridge (MA) 2002, pp. 585-600.

Esemplare il caso di *deepfake* sottoforma di audiovideo, datato 2017, in cui il Presidente degli Stati Uniti d'America Obama, con lo sguardo dritto nella telecamera, pronuncia frasi mai pronunciate (c.d. "effetto perturbante")[25].

Simili processi rendono indistinguibile il falso dall'originale. Il risultato dipende dalle regole delineate nel design del modello, dai big data di riferimento fino ai sistemi di controllo approntati. Appare chiara l'urgenza di generare nuovi algoritmi di rilevamento delle *deepfakes* per smascherare casi di uso malevolo dell'IA, i c.d. *deepfake detector*[26].

Simili contenuti artificiali possono essere ideati, generati e propagati da chiunque ne abbia l'intenzione. Per questo motivo può essere di ausilio una indagine sui possibili rischi e sui pericoli che l'IA generativa di immagini fa correre all'essere umano, attraverso la lente dell'etica, della comunicazione e delle linee di azione pubbliche e politiche.

3. *Manipolazione,* bias*, allucinazioni e il rischio di identità plurime*

Ogni essere umano può essere potenzialmente leso da immagini artificiali o *deepfake*. Impossibile prevedere le probabilità di tale eventualità in termini statistici. Nel nostro ecosistema massmediale, caratterizzato da infocrazia e disinformazione visive, l'IA generativa di immagini costituisce una minaccia in costante sviluppo. Ci espone a minacce difficilmente pronosticabili, che ci inducono a vivere nell'incertezza. È una esposizione che incide sulla nostra visione del mondo e sulla nostra identità e che ha implicazioni nelle diverse dimensioni di vita dell'essere umano, offline, *online* o, come teorizza Floridi, *onlife*[27].

Rischio, pericolo e incertezza appaiono parole chiave sulle quali appuntare brevemente l'attenzione prima di delineare le sfide etiche rilevanti che pertengono, più in generale, le immagini artificiali e più nello specifico, le *deepfakes* e che plasmano la *conditio humana* contemporanea.

[25] Cfr. M. Marini, *Video fake facilissimi da realizzare con un algoritmo: Obama "vittima" eccellente*, in «La Repubblica»: https://www.repubblica.it/tecnologia/2017/07/13/news/video_fake_facilissimi_da_realizzare_con_un_algoritmo_obama_vittima_eccellente-170703930/, 13 luglio 2017 (ultimo accesso 26 agosto 2024)

[26] Y. Li, S. Lyu, *Exposing DeepFake Videos By Detecting Face Warping Artifacts*, in «arXiv:1811.00656v3» (2019); L. Gaur (eds.), *Deepfake. Creation, detection and Impact*, CRC Press 2024, p. 91 ss.; M. Filimowicz, *Deep Fakes: Algorithms and Society*, Routledge, London 2022; N. Schick, *Deepfakes: The Coming Infocalypse*, Twelve, New York-Boston 2020.

[27] L. Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*, Oxford University Press, Oxford 2014, p. 4; D.J. Chalmers, *Più realtà. I mondi virtuali e i problemi della filosofia*, Raffaello Cortina Editore, Milano 2023.

Sulla scia di Le Breton, con rischio si intende «una conseguenza aleatoria di una situazione, ma non in termini di una minaccia, di un danno possibile». Si rimanda al termine italiano "risco", forma antica di "rischio", al latino *resecare*, con il significato di rimuovere tagliando e al latino classico *rixare*, "litigare" oltre che a *resecum*, colui che taglia. Anche la parola spagnola *riesgo*, roccia tagliata, scoglio, rimanda al medesimo concetto. Sembra cioè il momento in cui le strade si incrociano superando un prevedibile momento di pericolo, una 'roccia'. Il rischio sembra dunque appartenere alla categoria di incertezza quantificata, un pericolo possibile che può derivare da determinate circostanze ed eventualità, una sorta di misura dell'incertezza. Il rischio rispetto al pericolo lascia all'uomo ancora una responsabilità[28]. Come sottolinea Beck con la metafora della navigazione il rischio si corre ogni qual volta si naviga e ci si espone alla possibilità che uno scoglio possa danneggiare la nave[29]. Con l'AI generativa gli scogli sono frequenti e possono far addirittura affondare la nave.

Luhmann delinea, sulla scia di Knight[30], una distinzione di base tra rischio e pericolo: il primo termine implica una responsabilità soggettiva (dovuta alle scelte dell'essere umano), mentre il secondo è intrecciato a minacce che sfuggono al controllo dell'individuo[31], sebbene possa essere percepito in modo diverso da ciascuna e ciascuno.

L'incertezza si riferisce, invece, a qualche cosa di non prevedibile né quantificabile. Solo nello sviluppo degli eventi e delle circostanze si potrà rivelare o meno un pericolo. Sia Douglas che Luhmann affrontano il concetto sebbene con posizioni differenti. Se Douglas analizza l'incertezza come un fenomeno culturale, controllato attraverso la costruzione sociale delle categorie di purezza e rischio, Luhmann vede l'incertezza come una conseguenza della complessità sociale, che i sistemi devono ineludibilmente affrontare[32].

[28] Sul rischio, cfr.: M. Douglas, *Risk and Blame. Essays in Cultural Theory*, Routledge, London 1992 (*Rischio e colpa*, a cura di G. Bettini, il Mulino, Bologna 1996, pp. 33-34); N. Luhmann, *Risk. A sociological Theory*, Aldine de Gruyter, New York 1993 (*Sociologia del rischio*, trad. it. di G. Corsi, Mondadori, Milano 1996); M. Douglas, A. Wildavsky, *Risk and culture*, University of California Press, Berkeley 1982, D. Le Breton, *Sociologia del rischio*, a cura di A. Romeo, Mimesis 2017, pp. 14-15; P.L. Bernstein, *Against the Gods. The Remarkable Story of Risk*, Wiley, New York 1996.

[29] U. Beck, *La società del rischio. Verso una seconda modernità*, cit.

[30] Cfr. F.H. Knight, *Risk, Uncertainty, and Profit* (1921), Beardbooks, Washington 2002 (*Rischio, incertezza e profitto*, trad. it. di M. Giorda, La Nuova Italia, Firenze 1960).

[31] N. Luhmann, *Sociologia del rischio*, cit., p. 17.

[32] *Ivi*; M. Douglas, *Purity and Danger: An Analysis of Concepts of Pollution and Taboo* (1966), Routledge, London 2022.

Relativamente alle sfide etiche aperte dalle immagini in generale generate dall'IA esse possono, in primo luogo, riflettere lo stato di incertezza e di paura dell'uomo contemporaneo. In specie a causa di *bias* (o *contro-bias*) insiti nei dati di addestramento del sistema. Se l'IA viene alimentata da dati non bilanciati o parziali, potrebbe generare immagini distorte che rinforzano stereotipi o pregiudizi fondati su discriminazioni di genere, religione, etnia, cultura, professione, classe sociale sulla base dell'immaginario sociale di riferimento. Per *bias* [dal gr. "epikáros", dal fr. e provenz. ant. *biais* «obliquo»] non si intende tanto la deviazione sistematica da una norma, quanto le inclinazioni e la predisposizione al pregiudizio[33]. Di fronte a alcune immagini artificiali possono emergere *bias* cognitivi, automatismi o scorciatoie mentali dalle quali si generano credenze (non sempre eticamente orientate) e dalle quali si traggono decisioni veloci. Sono giudizi che, talvolta, riflettono le disuguaglianze sociali della realtà oggettuale e che impattano su decisioni, comportamenti e sviluppo del pensiero in contesti incerti, talaltra, ne generano di nuove. Sulla base del meccanismo dei *confirmation bias*, secondo cui le persone prediligono ricevere immagini che confermano le proprie preferenze e credenze, portando a negare qualsiasi evidenza contraria, tali meccanismi saranno ulteriormente corroborati e riproposti fino a che il sistema viene allenato con tale procedura. Nel mondo pubblicitario, ad esempio, sono spesso rappresentati individui che impersonano certe professioni, fondandosi su stereotipi e pregiudizi di genere, di etnia e sociali propri della realtà oggettuale, sedimentatisi nel tempo. E più questi personaggi sono permeati da stereotipi sociali che suggeriscono caratteristiche valoriali condivise più la pubblicità funziona e diventa efficace. Si contrae il nostro spazio di autonomia e aumenta lo spazio di azione dei 'difetti' algoritmici o pregiudizi (*bias*) iniqui, così perpetuando o esacerbando nuove, esistenti e/o passate logiche discriminatorie e forme di diseguaglianza. Un secondo aspetto, che discende dal precedente, riguarda la disinformazione e la manipolazione e persuasione sociali. Immagini false influenzano le decisioni individuali e l'opinione pubblica, giocando sulla credulonerìa, sulle emozioni e sulla vulnerabilità di chi guarda. Le identità personali degli individui *onlife*, finite nei data set di riferimento, vengono ridotte ad aggregati di dati in vendita.

Un ulteriore pericolo concerne il concetto di co-autorialità che contraddistingue le immagini artificiali. Se da una parte incide sui diritti d'autore e

[33] N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, il Mulino, Bologna 2023, p. 104.

sul copyright[34], intrecciandosi con aspetti di pertinenza giuridica, dall'altra artisti e creativi potrebbero sentirsi minacciati da sistemi che concorrono alla realizzazione di opere d'arte. Il rischio nel quale si può incorrere è di ottenere un prodotto visivo frutto di allucinazioni oppure immagini che si autocensurano sulla base dei valori guida del sistema (impostati durante il design dell'applicazione) oppure, ancora, cadere nel nichilismo, nel non credere più alle immagini che circolano nei massmedia e di ingenerare una crisi di fiducia nella capacità degli individui di distinguere tra realtà e finzione. Questo può avere conseguenze profonde sulla comunicazione visiva in generale, sul giornalismo e sul modo in cui interagiamo con il mondo digitale, ma anche in ambito di giustizia poiché difficile è stabilire l'autenticità delle prove visive.

In sintesi, le immagini create dall'intelligenza artificiale offrono grandi potenzialità comunicative e creative, ma sollevano altresì importanti questioni etiche (e legali) che richiedono una regolamentazione e una riflessione sui principi adeguati alla base dell'IA e di chi la utilizza per mitigarne i rischi e i pericoli.

Sulla generazione di immagini iperrealistiche e *deepfake* che falsificano la realtà rischi e pericoli si moltiplicano. Possono essere utilizzate per ingannare o manipolare l'opinione pubblica in modo più diretto e impattante di mere immagini statiche. Emergono al riguardo questioni etiche riferite alla violazione della *privacy* e dell'identità personale. L'IA può generare immagini di persone che non esistono, utilizzando dati provenienti da fotografie o informazioni personali tratte liberamente online senza consenso del soggetto ritratto. Ciò solleva questioni di *privacy*, poiché i dati visivi possono essere sfruttati per generare nuove identità o per comprometterne altre, senza che le persone coinvolte ne siano consapevoli. Inoltre, possono essere realizzate immagini o offensive e lesive della dignità dell'essere umano senza responsabilità diretta.

Come scrivono Thaler e Sunstein gli algoritmi, anche dunque quelli relativi al visivo, sono i nuovi «architetti della scelta», in grado di rimodellare e "architettare" i contesti e gli ambienti in cui formiamo i nostri gusti e compiamo le nostre scelte e, inoltre, formiamo (o inventiamo nel caso delle *deepfakes*) la nostra identità personale[35].

[34] Noto il caso del dicembre 2023 che vede il New York Time fare causa a OpenAI e a Microsoft per violazione del diritto di autore, ovvero per l'uso non autorizzato di milioni di suoi articoli per l'addestramento di chatbot. Cfr. https://www.ilsole24ore.com/art/new-york-times-fa-causa-openai-e-microsoft-uso-copyright-AFtsWfBC?refresh_ce=1.

[35] S. Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Carocci, Roma 2023, p. 67; R.H. Thaler, C.R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Penguin Books, London 2009.

L'identità dell'essere umano con l'IA si scontra con alcuni rischi che anche il Regolamento europeo sull'intelligenza artificiale (AI Act) ha cercato di affrontare adottando il *risk based approach*, ovvero la suddivisione di gruppi di sistemi di IA in livelli di rischio. Come scrive Dadà al riguardo[36] nell'*AI Act* europeo il termine 'rischio' «appare più di 350 volte»[37]. Si cerca sia di ottemperare alle esigenze di tutela contro le minacce provocate dai sistemi sia di far evolvere il progresso tecnologico. Il documento propone quattro livelli di rischi, dal più elevato al meno impattante: 1) sistemi a rischio inaccettabile, 2) sistemi ad alto rischio, 3) sistemi con rischio per la trasparenza e 4) sistemi a basso o a minimo rischio[38]. Relativamente ai rischi provocati dalle immagini artificiali e dalle *deepfakes* è il terzo livello quello deputato a tentare una forma di tutela. In ragione della difficoltà a comprendere se sono prodotti realizzati o meno da mano umana la regolamentazione impone l'obbligo di trasparenza sulla loro origine in vista di una IA più affidabile e, dunque, più etica[39].

Relativamente al rischio del terzo tipo le questioni etiche non riguardano tanto aspetti tecnici, quanto culturali, legati all'immaginario sociale e alla nostra identità individuale e sociale.

Rimodulare i data set di riferimenti dai quali l'IA attinge per creare immagini o *deepfake* è un problema pertanto sia culturale e sociale che tecnico. Sono gli esseri umani con i loro principi, valori, cultura e *bias* a fare una selezione dell'archivio iconografico dal quale il sistema attinge o a non farla, lasciando alla rete il ruolo di *hard disk* esterno di riferimento.

Da questa breve disamina affiora un'idea di rischio probabilistica ovvero fondata sulla probabilità del presentarsi di un evento dannoso e della gravità delle conseguenze che esso ha ingenerato ($R = P \times D$)[40].

Nel caso delle *deepfakes* appare chiaro che questa definizione sia difficile da applicare. Il controllo su certi dati e su determinate scelte appare utopistico pur limitando un sistema che ci sfugge sempre più di mano. Con

[36] S. Dadà, *Rischio e Intelligenza Artificiale. Un'analisi concettuale tra razionalità e incertezza*, in «Il pensiero critico» I (2014), pp. 47-66.

[37] https://artificialintelligenceact.eu/ (AI Act, approvato in ultima istanza il 12 luglio 2024).

[38] Approccio presente anche oltreoceano con l'US Algorithmic Accountability Act (2023), già proposto nel 2019 con l'obiettivo di fronteggiare i rischi relativi a possibili discriminazioni e violazioni della privacy in relazione all'utilizzo di sistemi di intelligenza artificiale ed il Canadian Directive of Automated Decision-Making (2020), fondato su principi etici cardine come la trasparenza, la responsabilità, la legalità e l'equità procedurale. Cfr. https://www.congress.gov/bill/118th-congress/senate-bill/2892/text; https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

[39] Sul concetto di rischio e IA, cfr. S. Dadà, *Rischio e Intelligenza Artificiale*, cit.

[40] Cfr. D.Lgs 81/2008; UNI EN ISO 12100-11; G. Sturloni, *La comunicazione del rischio*, cit.

le *deepfakes* si incide sui diritti fondamentali dell'essere umano legati alla tutela della propria identità e dei dati personali: la *privacy*, recepita non tanto come «il diritto a uno spazio in cui essere lasciati soli»[41], quanto come *privacy informazionale*, intesa sia come il diritto di impedire ad altri l'accesso alle nostre informazioni personali, sia come il diritto alla tutela della propria identità personale e della propria immagine *onlife*[42]. Il Regolamento (UE) 2016/679, ad esempio[43], ha affrontato in passato gli aspetti legati alla tutela del flusso dei nostri dati personali (*data protection*) e al controllo delle nostre informazioni (*privacy* informazionale).

Con le *deepfakes* siamo dunque di fronte a rischi di furto o distorsione malevola di identità incalcolabili, quindi all'incertezza estrema. Le minacce in corso possono ledere la dignità e i diritti fondamentali dell'essere umano[44].

Conclusioni. Dal rischio ad un'etica dell'incertezza

Per eludere la paura radicale, oramai ontologica, dell'individuo nella modernità[45] potremmo convenire con l'affermazione di Luhmann secondo il quale «se ci si astiene da una certa azione non si corre alcun rischio»[46]. Sulla base di questa opzione estrema l'inazione parrebbe la soluzione che mette a riparo da ogni possibile rischio e pericolo. Nel caso dell'IA generativa approccio luhmanniano potrebbe risultare anacronistico e poco efficace. La posizione di Giddens, secondo il quale «l'inazione è sovente rischiosa e vi sono alcuni rischi che, volenti o nolenti, noi tutti dobbiamo correre» rappresenta piuttosto la risposta consapevole ad una tecnologia che oramai pervade la vita degli esseri umani e della quale dobbiamo farci carico[47]. È nel dominio dell'incertezza e del pericolo ai quali siamo costantemente

[41] S. Warren, L.D. Brandeis, *The Right to Privacy*, in «Harvard Law Review» 4, 1980, pp. 193-194.

[42] S. Tiribelli, *Identità personale e algoritmi*, cit., p. 12; L. Floridi, *The Ontological Interpretation of Informational Privacy*, in «Ethics and Information Technology» 7, 2005, pp. 185-200; Id., *The Informational Nature of Personal Identity*, in «Minds and Machines» 21, 4, 2011, pp. 549-566; C. Koopman, *How We Became Our Data: A Genealogy of the Informational Person*, University of Chicago Press, Chicago 2019.

[43] https://www.garanteprivacy.it/il-testo-del-regolamento.

[44] S. Tiribelli, *Identità personale e algoritmi*, cit.

[45] A. Giddens, *Le conseguenze della modernità*, il Mulino, Bologna 1994.

[46] N. Luhmann, *Familiarità, confidare e fiducia: problemi e alternative*, in D. Gambetta (a cura di), *Le strategie della fiducia. Indagini sulla razionalità della cooperazione*, trad. it. di D. Panzieri, Einaudi, Torino 1989, p. 130.

[47] A. Giddens, *Le conseguenze della modernità*, cit. p. 41.

esposti che dobbiamo ricercare il senso del rischio al tempo dell'IA generativa, dal momento che l'autonomia morale e decisionale dell'individuo risulta indebolita e le procedure delle scelte algoritmiche appaiono opache. Certamente «calcolare l'incalcolabile» non offre risposte utili relativamente a come agire[48]. Non è possibile misurare la probabilità oggettiva che un dato evento accada o una certa immagine venga prodotta, poiché la tecnologia e i data set sono in costante evoluzione e procedono repentinamente.

Piuttosto nel caso delle *deepfakes* e delle immagini artificiali sembra più corretto riflettere sul ruolo del pericolo e, semmai, su un'etica dell'incertezza, nonostante i concetti di rischio e di incertezza storicamente non siano mai stati nettamente separati come teorizza Knight[49]. Rischio misurabile e pericolo non misurabile creano incertezza. Ciò che appare chiaro, invece, è che la società dell'incertezza baumaniana[50] che caratterizza la postmodernità ha reso l'umanità più vulnerabile, aumentando il senso di insicurezza esistenziale e personale[51]: «come in epoca premoderna, la base simbolica delle nostre incertezze è l'ansia creata dal disordine, la perdita di controllo sui nostri corpi, sui rapporti con gli altri, il necessario per vivere, e il grado di autonomia di cui possiamo godere nella vita quotidiana»[52].

Potremmo oggi aggiungere, altresì, la perdita di controllo sui propri dati, sui propri immaginari e della propria identità. Solo nello sviluppo degli eventi e delle circostanze si rivelerà o meno un pericolo arginabile attraverso, secondo Luhmann, sistemi sociali che mirano a ridurre la complessità ontologica della società[53].

Al riguardo Luhmann introduce, accostandolo alla nozione di rischio e di incertezza, anche il concetto di contingenza, con il quale indica la possibilità che una circostanza – diversa dalle proprie aspettative – accada nel corso del tempo e in un determinato ambiente, generando per l'appunto incertezza[54]. La comunicazione diventa, in questo frangente, un mezzo per far sì che la probabilità dell'attuarsi di determinati accadimenti diminuisca. Ma nel caso dell'IA generativa di immagini è la comunicazione (visiva) stessa a generare incertezza, vincolata dalla qualità e dalla tipologia delle immagini che il si-

[48] M. Dean, *Risk, Calcuable and Incalculable*, in «Soziale Welt» 49, 1998, pp. 25-42.
[49] F.H. Knight, *Risk, Uncertainty, and Profit*, cit., p. 205.
[50] Z. Bauman, *La società dell'incertezza*, il Mulino, Bologna 1999.
[51] Z. Bauman, *Modernità liquida*, Laterza, Roma-Bari 2002.
[52] D. Lupton, *Il rischio*, il Mulino, Bologna 2003, p. 9.
[53] N. Luhmann, *Sociologia del rischio*, cit.
[54] N. Luhmann, *Generalized Media and the Problem of Contingency*, in J.J. Loubser *et al.* (eds.), *Exploration in General Theory of Social Science*, Free Press, New York 1976, p. 509.

stema ha prodotto. Simile incertezza informativa è intrecciata alla (in)capacità (legittima) degli esseri umani di comprendere la veridicità dei messaggi visivi.

Come dunque arginare i pericoli emergenti e rispondere all'incertezza o, ancor meglio, al rischio dell'incertezza di fronte all'IA generativa di immagini?

Rischio, pericolo e incertezza rappresentano categorie sociali intrinseche alla società moderna.

Le stime degli esperti sono importanti e necessarie, ma non bastano per ricondurre il rischio e il pericolo sotto il nostro dominio[55]. Il controllo su basi matematiche può generare ulteriori conseguenze "irrazionali". Ecco perché, in questo quadro il rischio si trasforma in incertezza degli eventi, delle conseguenze degli eventi, «degli effetti collaterali e degli effetti collaterali degli effetti collaterali» degli stessi[56], venendo meno la possibilità di compensazione e la capacità di limitazione e controllo del danno poiché le immagini possono viaggiare ovunque senza sapere dove, quando e di fronte a quale sguardo.

Alla luce di quanto pretermesso l'etica sembra delinearsi quale strumento di decodifica e prospettiva di accettazione dell'incertezza, seppur non consenta di immaginarne le conseguenze[57]. Valori come l'equità, la libertà (consapevole) di scelta, la giustizia, la fiducia, il diritto alle informazione, la trasparenza, il rispetto dei diritti fondamentali degli esseri umani oltre alla valutazione (pur sempre relativa) della pericolosità (e dell'entità) del danno di alcune immagini costituiscono le basi per arginare se non almeno limitare minacce e pericoli. Sono standard morali che debbono essere impliciti al design e al data set di riferimento di ciascun sistema di IA generativa di immagini. Noi siamo esposti al rischio solo nel momento in cui decidiamo di utilizzare specifici sistemi algoritmici dei quali conosciamo i meccanismi e l'archivio iconografico di base di cui si avvalgono e, in questo senso, ci assumiamo un rischio in qualche modo calcolabile. Questa prospettiva però non può che risultare inattuabile se non illusoria, dal momento che l'opacità dei sistemi di IA e l'estrema ampiezza dei data set di riferimento rappresentano aspetti caratterizzanti dei sistemi algoritmici generativi.

Se dunque in questo quadro «i rischi sono sempre legati a decisioni, cioè presuppongono una possibilità di scelta»[58] da parte dell'essere umano, i sistemi, invece, compiono scelte autonomamente. Non vale l'assunto secondo cui «[l] e minace incalcolabili vengono trasformate dalla società industriale

[55] A. Giddens, *Il mondo che cambia. Come la globalizzazione ridisegna la nostra vita*, il Mulino, Bologna 2000, p. 40.

[56] U. Beck, *Conditio humana*, cit., p. 34.

[57] *Ivi*, p. 38.

[58] *Ivi*, p. 178.

in rischi calcolabili»[59]. Il rischio si colora di incertezza, contrariamente a quanto suggerito dal Regolamento europeo.

Possiamo dunque rispondere all'incertezza con un'etica sorretta dalla consapevolezza che miri a rafforzare le competenze degli esseri umani per avere le risorse (anche morali) per affrontare le conseguenze degli stati di incertezza e la paura di futuri distopici. Si tratta, da una parte, di una sorta di principio di precauzione che impone a ciascun individuo di avere una cassetta degli attrezzi efficace per fronteggiare e contenere i pericoli (e i danni) dell'IA visiva, dall'altra, di strategie di etica pubblica per sensibilizzare l'opinione pubblica e far agire i governi con regolamentazioni sempre più specifiche e in linea con lo sviluppo tecnologico, fino alla co-responsabilità di tutti i soggetti coinvolti, dagli ideatori, ai governi fino agli utilizzatori dei sistemi. Il vero rischio altrimenti consisterà nel trasformare il mondo in un mondo a dimensione dell'intelligenza artificiale generativa.

English title: Generative Artificial Intelligence, Deepfakes, and Vulnerable Identity: The Ethics of Uncertainty as a Response to an (Un)Controllable Risk

Abstract

The algorithmic turn has given rise to high-performance image generation systems (artificial images, deepfakes, fake images). This scenario is intertwined with the risks, dangers and threats posed by these systems, as well as the uncertainty surrounding their consequences for individuals and society as a whole. Bias, loss of control over one's own data, identity theft, opacity about data use, are just a few emerging ethical issues. Public ethics strategies can play a crucial role in guiding public opinion and citizenship towards awareness and co-responsibility, among all stakeholders from creators to governments (which must implement specific regulations) to users. Such strategies aim to mitigate the uncertainty and fear that these risks cause in individuals, equipping them to become more informed and critical in their engagement with these technologies.

Keywords: deepfake; ethics; generative artificial intelligence; risk; uncertainty.

Veronica Neri
Università di Pisa
veronica.neri@unipi.it

[59] *Ibidem*.

T

Anastasia Siapka

A Virtue Ethics Approach to AI-induced Risk

> Thinking clearly about risks and their acceptability in our lives is too important to be left to technical risk assessors and cost-benefit theorists[1].
>
> Carl Cranor

1. *From a technological risk society to risk-based technology regulation*

Once an uncontrollable force, tied to pre-determined natural or spiritual factors, risk has become a major societal preoccupation since the 20th century[2]. New technologies introduce risks that transcend spatial, temporal and social boundaries, ushering in a «risk society», where diverse yet incomprehensible futures are possible[3]. More recently, Artificial Intelligence (AI) systems – understood as (sets of) algorithms performing goal-oriented tasks that would otherwise require human intelligence – incur risks that necessitate attention and intervention[4]. Thus, self-regulatory frameworks addressed to AI compa-

[1] C.F. Cranor, *Toward a Non-Consequentialist Approach to Acceptable Risks*, in T. Lewens (ed.), *Risk: Philosophical Perspectives*, Routledge, London 2007, p. 51.

[2] J. van der Heijden, *Risk as an Approach to Regulatory Governance: An Evidence Synthesis and Research Agenda*, in «SAGE Open» 11, no. 3 (September 2021), pp. 1-12, https://doi.org/10.1177/21582440211032202.

[3] U. Beck, *Risk Society: Towards a New Modernity*, trans. M. Ritter, *Theory, Culture & Society*, Sage Publications, London 1992, p. 9; A. Giddens, *Risk Society: The Context of British Politics*, in J. Franklin (ed.), *The Politics of Risk Society*, Polity Press, Cambridge 1998, pp. 23-34.

[4] A. Siapka, *The Ethical and Legal Challenges of Artificial Intelligence: The EU Response to Biased and Discriminatory AI*, SSRN Scholarly Paper, Social Science Research Network, New York, 11 December 2018, https://dx.doi.org/10.2139/ssrn.3408773.

 DOI: 10.4454/wj35mz94

nies and developers are conceived specifically for or adapted to AI risk[5].

This pervasiveness of risk, hitherto confined to private practices, affects legally binding regulation. Under the General Data Protection Regulation (GDPR), AI developers acting as data controllers consider «risks of varying likelihood» and perform impact assessments for high-risk processing, but are provided with minimal guidance on how to do so[6]. The Artificial Intelligence Act (AIA) moves further than the GDPR does, adopting a «proportionate risk-based approach» as a core feature of its architecture[7]. AI developers acting as providers adhere to different obligations (e.g., conformity assessments, monitoring, risk management systems and voluntary codes of conduct) based on the system's risk level[8]. Despite the AIA's expansive material and territorial scope, including its possible role as a «benchmark» for other jurisdictions given the «Brussels effect», guidance on the risk-based approach remains vague, leaving AI developers «to their own devices»[9].

[5] Examples of the former include the NIST AI Risk Management Framework and ISO/IEC 23894, while an example of the latter is COSO ERM 201715: J. Schuett, *Risk Management in the Artificial Intelligence Act*, in «European Journal of Risk Regulation» 15, no. 2 (2024), pp. 368-369, https://doi.org/10.1017/err.2023.1.

[6] Articles 24 (1), 25 (1), 35 (1) and Recital 75. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*, Pub. L. No. 32016R0679, OJ L 119 (2016), http://data.europa.eu/eli/reg/2016/679/oj/eng.

[7] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, Pub. L. No. COM/2021/206 final (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206; Directorate General for Communication, *EU AI Act: First Regulation on Artificial Intelligence*, Article, European Parliament, Strasbourg (France), 19 December 2023, https://www.europarl.europa.eu/pdfs/news/expert/2023/6/story/20230601STO93804/20230601STO93804_en.pdf. In December 2023, the EU's Parliament and Council reached a provisional, political agreement on the contents of the long-awaited AIA. However, as at the time of writing the revised text has not been released, I take into account its 2021 version. Given that the paper does not go into detail about specific provisions and instead uses the AIA for illustrative purposes only, any subsequent changes to the text of the law are not expected to affect my arguments therein.

[8] A precursor to the AIA's approach is found in the work of the German Data Ethics Commission. In 2019, this Commission put forward a «risk-adapted regulatory approach», suggesting a classification of AI systems into five levels of criticality. Datenethikkommission, *Opinion of the Data Ethics Commission*, Data Ethics Commission of the Federal Government, Berlin, December 2019, pp. 173-182, https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/datenethikkommission-abschlussgutachten-lang.pdf;jsessionid=789B1C3D1FC30ACF12B067AD01FDFD38.live881?__blob=publicationFile&v=5.

[9] Schuett, *op. cit.*, pp. 367-380. The «Brussels effect» describes to the EU's power to influence rules and regulations in other jurisdictions beyond its Member States. As for the AIA's

This rising prominence yet concurrent under-specification of risk-based approaches to AI constitutes the first reason for this paper's focus on AI risk[10].

The second reason concerns the multiple material and immaterial forms that AI risk takes[11]. Examples of the former include adverse outcomes to health and safety by AI-embedded products; impeded access to essential services by AI-based credit scoring; and deprivation of liberty by AI used in law enforcement[12]. Examples of the latter include discrimination by AI-based social scoring; surveillance and hindered freedom of assembly by AI used for remote biometric identification; and diminished career and education prospects when AI determines access to educational or employment opportunities[13]. Therefore, AI risk spans individuals, groups and society as a whole. These risks are posed by Narrow AI, which outperforms humans in specific tasks yet lacks the versatility of human intelligence[14]. Contrariwise, General AI (or Artificial General Intelligence) would be endowed with broad cognitive abilities tantamount to those of humans[15]. The mere possibility of General AI, especially after developments in Generative AI, has sparked concerns about longer-term, existential risks to humankind by systems unaligned with human values[16]. This paper does not further examine AI risks, but targets the approaches used for their assessment. Rejecting technocratic approaches, it evaluates AI risk through two contrasting normative theories: consequentialism and virtue ethics.

scope, it covers AI systems across multiple application domains and encompasses all providers placing AI on the market or putting it into service in the EU.

[10] This focus does not exclude the applicability of the paper's arguments to other types of technology. In addition, regulation is here understood in a broad sense, comprising regulatory acts by actors that may or may not have a legal mandate, in line with Black and Murray's definition (Section 4): «By regulation (and regulatory governance) is meant sustained and focused attempts to change the behaviour of others in order to address a collective problem or attain an identified end or ends, usually but not always through a combination of rules or norms and some means for their implementation and enforcement, which can be legal or non-legal». J. Black, A.D. Murray, *Regulating AI and Machine Learning: Setting the Regulatory Agenda*, in «European Journal of Law and Technology» 10, no. 3 (30 December 2019), https://www.ejlt.org/index.php/ejlt/article/view/722.

[11] Artificial Intelligence Act, *op. cit.*

[12] *Ibidem*.

[13] *Ibidem*.

[14] Siapka, *op. cit.*, pp. 17, 22.

[15] *Ibidem*.

[16] Generative AI implies AI systems that «generate brand-new, unique artifacts». Gartner, *Definition of Generative AI*, in «Gartner Glossary», Information Technology Glossary, accessed 25 August 2023, https://www.gartner.com/en/information-technology/glossary/generative-ai. For an overview of approaches to the existential risk (or x-risk) of AI, see, PauseAI, *The Existential Risk of Superintelligent AI*, in «Pause AI», accessed 16 December 2023, https://pauseai.info/xrisk.

2. *The technocratic approach to risk*

a. *Objectivity vs. normativity*

By and large, (self-)regulatory instruments divide risk-based approaches into two stages: (i) risk assessment, implying the identification of risks and evaluation of their acceptability and (ii) risk management, including the selection and adoption of measures to mitigate the previously identified and evaluated risks[17]. The first stage is considered an objective, neutral process, in which technical expert advice leaves little to no room for normative judgement[18]. The normative character of the second stage is more straightforward, since decisions about risk mitigation involve not only scientific and technical but also ethical, societal, political, financial, practical and other qualitative considerations.

However, this distinction between a value-free process of risk assessment and a normative one of risk management is artificial[19]. Far from being discovered by experts in an exclusively empirical way, risks are identified also on the basis of norms, values and often subjective perceptions, while being «strongly involved with social relations and meanings»[20]. As argued in science and technology studies and in the foundational report *Taking European Knowledge Society Seriously* specifically, «questions of risk can be recognised intrinsically to be shaped and framed by social values, sometimes embodied in routinised habitual ways of institutional thinking, and political interests»[21]. Indicatively, selecting the forms of risk relevant to the assessment, the measurement criteria to be employed, the weight to be placed on possible effects, and the thresholds of risk acceptability is a value-laden process[22]. Focusing on certain dimensions of risk privileges some normative

[17] A third stage of risk communication may also be distinguished but is not strictly relevant to the arguments of this paper.

[18] U. Felt *et al.*, *Taking European Knowledge Society Seriously*, Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate, Directorate General for Research, European Commission, Directorate General for Research and Innovation (European Commission), Belgium, January 2007, pp. 32-42, https://op.europa.eu/en/publication-detail/-/publication/5d0e77c7-2948-4ef5-aec7-bd18efe3c442; C.F. Cranor, *The Normative Nature of Risk Assessment: Features and Possibilities*, in «RISK: Health, Safety & Environment (1990-2002)» 8, no. 2 (March 1997), pp. 123-136; N. van Dijk, R. Gellert, K. Rommetveit, *A Risk to a Right? Beyond Data Protection Risk Assessments*, in «Computer Law & Security Review» 32, no. 2 (April 2016), pp. 286-306, https://doi.org/10.1016/j.clsr.2015.12.017; van der Heijden, *op. cit.*

[19] Felt *et al.*, *op. cit.*, pp. 32-42.

[20] van Dijk, Gellert, Rommetveit, *op. cit.*, p. 289.

[21] Felt *et al.*, *op. cit.*, p. 34.

[22] *Ibidem*.

perspectives or commitments while occluding others. Therefore, risk assessments are, at least in part, normatively construed.

This omission of normativity matters beyond risk assessment. It affects risk management, which consecutively builds upon and reflects the types of risk identified during assessment. It also affects the risk-based approach more broadly, given its function in facilitating decision-making about risks. Granted that risk-based approaches aim to eliminate or reduce risks, if these are not accurately identified and evaluated in the stage of assessment, given its disregard for normativity, the measures adopted for such elimination or reduction in the subsequent stage of management will be correspondingly misguided. This interconnection between risk assessment and management is so strong that the possibility of their separate treatment is doubted[23]. Hence, risk-based approaches, be they in voluntary or binding regulation, fail to achieve their aims unless they incorporate both objective and normative considerations throughout.

b. *Technical vs. ethical understanding*

Risk is broadly a «technique for creating knowledge and certainty about future events that are uncertain by definition»[24]. EU legal instruments associate it with the notions of «likelihood» or «probability» of harm and its «severity»[25]. These notions point to a technical understanding of risk, nu-

[23] Cranor, *Normative Nature of Risk Assessment*, *op. cit.* p. 128.

[24] van Dijk, Gellert, Rommetveit, *op. cit.*, p. 301.

[25] See, respectively, «[a] risk is a scenario describing an event and its consequences, estimated in terms of severity and likelihood» and «severity and likelihood of this risk should be assessed» in Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679*, European Commission, Brussels, 4 April 2017, p. 6, http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236; *Opinion 05/2014 on Anonymisation Techniques*, European Commission, Brussels, 10 April 2014, p. 7, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. Similar wording is used in Recitals 75-76 GDPR: General Data Protection Regulation, *op. cit*. Likewise in the AIA, «the AI systems pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence» and «taking into account both the severity of the possible harm and its probability of occurrence»: Artificial Intelligence Act, *op. cit*. As for other EU legislation, «"risk" means a function of the probability of an adverse health effect and the severity of that effect, consequential to a hazard» in *Consolidated Text: Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 Laying down the General Principles and Requirements of Food Law, Establishing the European Food Safety Authority and Laying down Procedures in Matters of Food Safety*, Pub. L. No. OJ L 031 (2002), http://data.europa.eu/eli/reg/2002/178/2019-07-26. Like-

merically representing the outcome of the probability of a possible harm multiplied by the severity of said harm.

From this perspective, risk is distinguished from uncertainty. In decisions under risk, the probabilities of different adverse outcomes materialising are available and part of the calculation of risk[26]. In decisions under uncertainty, the different possible outcomes might or might not be available, but their probabilities are definitely not[27]. This distinction is, however, contested. The exact probability of a risk occurring is known solely in artificial cases (e.g., rolling a dice), compared to the more frequent real-life cases of uncertainty, where the probabilities of possible outcomes are unknown[28]. Relying on statistics and probabilities, this understanding of risk simplifies «the full range of uncertainties to the more comforting illusion of controllable, probabilistic but deterministic processes»[29].

Conversely, ethicists invoke risk in its ordinary usage, denoting the possibility that an adverse or undesirable outcome, such as harm, injury or loss, will occur[30]. This broader view of risk illuminates nuances that the focus on probability and severity overlooks. Given that ethics examines the attribution of praise and blame, it approaches risks differently depending on whether they are apt for such an attribution. Hence, risks that we face differ from risks that we take[31]. The first type includes risks whose occurrence we cannot control but whose management is to a certain degree under our control (e.g., risks caused by natural disasters). The second type includes risks to which exposure is chosen and over which there is a dimension of control

wise, «"risk" means the probable rate of occurrence of a hazard causing harm and the degree of severity of the harm» in *Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the Safety of Toys*, OJ L 170 § (2009), https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009L0048.

[26] M. Hayenhjelm, J. Wolff, *The Moral Problem of Risk Impositions: A Survey of the Literature*, in «European Journal of Philosophy» 20 (June 2012), p. E30, https://doi.org/10.1111/j.1468-0378.2011.00482.x.

[27] *Ibidem*.

[28] S.O. Hansson, *Philosophical Perspectives on Risk*, in «Techné: Research in Philosophy and Technology» 8, no. 1 (Fall 2004), pp. 11-12, https://doi.org/10.5840/techne2004818.

[29] B. Wynne, *Uncertainty and Environmental Learning: Reconceiving Science and Policy in the Preventive Paradigm*, in «Global Environmental Change» 2, no. 2 (June 1992), p. 123, https://doi.org/10.1016/0959-3780(92)90017-2.

[30] K. Shrader-Frechette, *Risk*, in *Routledge Encyclopedia of Philosophy*, 1st ed., Routledge, London 1998, https://doi.org/10.4324/9780415249126-L088-1. On the criticism against technocratic approaches to risk, see van der Heijden, *op. cit*.

[31] N. Rescher, *Risk: A Philosophical Introduction to the Theory of Risk Evaluation and Management*, University Press of America, Washington 1983, pp. 6-7. Although in practice there might be overlapping or borderline cases, the distinction adds nuance to the moral picture.

lacking from the first type (e.g., risks caused by human-made products). Risks of the first type approximate incidents of luck, impeding ascriptions of responsibility. It is the second type, risks to which we decide to expose ourselves and others, that matters for responsibility. Within this second type, we can differentiate between risks to which we decide to expose ourselves (self-imposed) and those imposed on us by others (other-imposed)[32]. In the latter case, the roles of those imposing the risk, their motivations for doing so, and the voluntary or not acceptance of these externally imposed risks matter from an ethical standpoint yet are captured by neither severity nor probability.

Based on the foregoing, the technocratic approach to risk, comprising an objectivist perspective on risk assessment and a scientific conceptualisation of risk, is rejected as artificial and overly narrow. In this paper, AI risk is not a free-floating, objectively accessible and measurable entity whose assessment exclusively relies on the properties of the AI system in question. Instead, it is conceived in its ethical usage, denoting the possibility of a future undesirable event occurring because of AI development/deployment, and particularly in its second type, denoting risks that involve the exercise of choice by AI developers. The process of its assessment is likewise considered imbued with normativity.

3. *The consequentialist approach to risk*

a. *Overview*

Although ethicists acknowledge that a complete absence of risk is impossible, they seek to evaluate the extent to which risk is acceptable. In most cases, they do so by appealing to consequentialism[33]. Consequentialism is the strand of normative ethical theory that evaluates actions as morally right or wrong based on a comparison of their overall beneficial and harmful consequences.

Consequentialist approaches to risk are premised upon the assumption that all consequences are comparable and aggregable[34]. The standard form they take is the Risk Cost Benefit Analysis (RCBA)[35]. Regulatory agencies

[32] Cranor, *Toward a Non-Consequentialist Approach to Acceptable Risks*, *op. cit.*, p. 50.

[33] Hayenhjelm, Wolff, *op. cit.*, pp. E28, E32.

[34] S.O. Hansson, *Risk and Ethics: Three Approaches*, in T. Lewens (ed.), *Risk: Philosophical Perspectives*, Routledge, London 2007, p. 26.

[35] T. Lewens, *Introduction: Risk and Philosophy*, in T. Lewens (ed.), *Risk: Philosophical Perspectives*, Routledge, London 2007, pp. 1-20.

employ the RCBA to assess the desirability of varying technological interventions, «from building a liquefied natural gas facility to adding yellow dye number two to margarine»[36]. RCBA encompasses «decision-aiding techniques» that seek to identify all likely good (benefits) and bad (risks/costs) consequences of an option and, by employing numerical terms, to «add up the likely overall good consequences of a decision option and to subtract from that figure the likely overall bad consequences»[37]. If the resulting overall good/bad consequences ratio is favourable – i.e., if the former outweigh the latter – risk is acceptable. Upon repeating this process for all available options, the one maximising net benefits or minimising net risks/costs is chosen.

Comparisons between good and bad consequences are straightforward when these are of the same type – e.g., if AI decreases the jobs available in a certain domain but increases those available in another. This is not often the case, though, rendering such comparisons difficult. For example, AI deployment in healthcare may be concurrently linked to the benefit of faster access to treatment and to the risk of biased diagnoses. For this reason, such approaches convert consequences that may differ a lot from each other into a single, usually monetary attribute[38]. To achieve this conversion, RCBA identifies «how much people would be willing to pay to have (or to avoid) these consequences»[39]. Following the previous example, individuals' hypothetical willingness to pay more for faster AI-enabled medical treatment than for avoiding a racially biased AI-enabled diagnosis would suggest the acceptability of AI risk.

However, not all RCBA techniques are single-attribute ones. Multi-attribute risk benefit analysis suggests that, as a first step, each consequence should be measured separately using the scale appropriate for it[40]. In the previous example, the number of hours from admission to treatment might be appropriate for measuring the consequences of AI-enabled healthcare

[36] K. Shrader-Frechette, *The Real Risks of Risk-Cost-Benefit Analysis*, in «Technology in Society» 7, no. 4 (1985), p. 399, https://doi.org/10.1016/0160-791X(85)90007-7.

[37] Lewens, *op. cit.*, p. 7. See also S.O. Hansson, *Risk*, in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2018, https://plato.stanford.edu/archives/fall2018/entries/risk/; Shrader-Frechette, *The Real Risks of Risk-Cost-Benefit Analysis*, *op. cit*.

[38] Hansson, *Risk*, *op. cit.*, section 7.4.

[39] Lewens, *op. cit.*, p. 5.

[40] M. Peterson, *On Multi-Attribute Risk Analysis*, in T. Lewens (ed.), *Risk: Philosophical Perspectives*, Routledge, London 2007, pp. 68-83. If we, however, consider the presence of a single scale to be a defining feature of risk/cost benefit analyses, then multi-attribute analysis can be deemed a distinct kind of consequentialist approach.

services, whereas the number of lives saved as a result of accurate AI-enabled diagnoses might be appropriate for measuring the risks of AI bias. As a next step, these measurements are aggregated to formulate an overall ranking for each action and then used to compare alternative actions based on their overall rankings[41]. Although multi-attribute analyses include more dimensions than single-attribute ones do, they resemble the latter in expressing consequences in aggregated, numerical terms.

b. *Objections*

Where legal instruments refer to weighing the risks of an option against its benefits, they allude to some sort of RCBA, as the seemingly rational and rigorous guidance of this approach has rendered it the dominant choice[42]. Despite their popularity and ostensible precision, however, consequentialist approaches to risk face objections.

First, accurately identifying the consequences brought about, for instance, by AI is possible only in hindsight, after these have come to fruition. Such approaches cannot assist developers in evaluating the system's outcomes beforehand. Alternatively, consequentialists appeal to expected or hypothetical, instead of actual, consequences. However, the novelty of emerging technologies, including AI, poses difficulties in predicting their consequences – even just the expected ones – and the probability of their occurrence. Comparing AI's risks and benefits demands considerable information; yet, it is questionable whether such information is at the developers' disposal for the time being and, even where that is the case, whether such information is sufficiently complete or reliable. Due to its unpredictable nature and fast pace of development, the consequences of AI are often unintentional and unexpected; asking developers to identify risks that are by their nature hard to foresee is admittedly a tall order.

Second, by maximising benefits over costs in the aggregate, consequentialist approaches are impersonal in terms of their distributive effects. They would favour AI systems that incur more benefits than costs, without examining who would bear these. In this way, they leave open the problematic possibility that risks/costs are piled up in one part of the population and

[41] *Ibidem*.

[42] For example, the proposed AIA prohibits remote biometric identification, except for narrowly defined cases in which the public interest benefits outweigh the risks. Artificial Intelligence Act, *op. cit*.

benefits in another[43]. Even if the risks shouldered by the first population group were extremely severe, they would be justified by consequentialism as long as the second population group, which would bear the benefits, were larger[44]. AI developers would thus evaluate risks and benefits to users generally, without distinguishing among the needs and characteristics of different individuals or groups as such or in comparison with each other. Nonetheless, not all individuals or groups experience risk in the same way. Certain groups (e.g., children) are considered more vulnerable and require particular attention, which consequentialism could not justify. Apart from differences across individuals/groups at a given time, differences might exist across generations (e.g., present ones embracing the benefits while future ones bear the risks), for which consequentialist calculations cannot account. The use of willingness-to-pay indicators likewise overlooks that these depend on one's income: those with lower incomes are inevitably able and thereby willing to pay less to avoid risk than those with higher incomes without this implying that the former are actually less risk averse[45].

Third, consequentialist approaches to risk are not merely impersonal but even crude or cruel. Examining solely the outcomes of actions, they overlook the means employed to reach said outcomes. If an AI system promised benefits that significantly overrode its costs, its development/deployment would be acceptable even if, for example, relevant decision-making occurred through authoritarian procedures. Relatedly, consequentialist approaches do not differentiate between risks we face and those we take nor do they account for risks imposed, the agent(s) imposing these risks, their motivations and the (in)voluntary acceptance by risk bearers, all aspects that section 2 considered morally significant.

The cruelty of consequentialism additionally emerges in its effort to homogenise all values, rights, goods and moral commitments by translating them into commensurable terms[46]. For instance, AI risks to human life

[43] This would be reminiscent of Beck's claim that «wealth accumulates at the top, risk at the bottom»: Beck, *op. cit*., p. 35.

[44] We might even conceive of a three-party relationship, in which one group is subject to risks/costs owing to transactions between two other benefitting groups.

[45] Shrader-Frechette, *The Real Risks of Risk-Cost-Benefit Analysis*, cit., p. 403. For a more detailed version of Shrader-Frechette's assessment of the RCBA, see K.S. Shrader-Frechette, *Assessing Risk-Cost-Benefit Analysis, The Preeminent Method of Technology Assessment and Environmental-Impact Analysis*, in *Science Policy, Ethics, and Economic Methodology: Some Problems of Technology Assessment and Environmental-Impact Analysis*, D. Reidel, Dordrecht 1985, pp. 32-64, https://doi.org/10.1007/978-94-009-6449-5_2.

[46] Shrader-Frechette, *The Real Risks of Risk-Cost-Benefit Analysis*, cit.

or the environment would be placed on the same (monetary) scale as AI's potential benefits in efficiency and would even be acceptable if that scale tilted towards the latter. There is something corrupting about the mere act of subjecting goods, such as human life or the environment, to such calculations. This crude approach to values or goods that are commonly considered «priceless or sacred» alters their perceived worth in harmful ways, converting them into tradeable commodities[47]. Put simply, «some goods are cheapened when we try to attach a price to them»[48].

A fourth and broader objection is metaphysical. It challenges the adequacy of consequentialist approaches in capturing the breadth of «human situational understanding»[49]. By focusing on «allegedly transparent rationality and scientific know-how», analytic, formal and economic frameworks of thought upon which RCBA draws are reductionist and detached from reality[50]. In practice, human decision-making (especially in policy) resists such formalisation. It relies on intuitions and judgements that, akin to wisdom, are shaped by expertise and skills beyond algorithmic ways of thinking[51]. Even attempts to engage in such formal ways of thinking are unlikely to succeed, as humans' perspectives on what might be the consequences of an action differ substantially, as do their perspectives on which of these consequences are good or bad. For instance, if AI deployment in logistics is likely to increase the number of product deliveries achieved within a certain timeframe, this likelihood might be classified as a cost/risk by an environmentalist but as a benefit by an economist.

4. *The case for an alternative approach*

a. *Overview*

Traditional ethical theories are geared towards evaluating actions with more or less certain or knowable outcomes (*deterministic bias*)[52]. When extended to non-determinate settings, meaning to actions whose outcomes are

[47] D. MacLean, *Cost-Benefit Analysis and Procedural Values*, in «Analyse & Kritik» 16, no. 2 (1994), p. 171, https://doi.org/10.1515/auk-1994-0205.

[48] MacLean, *op. cit.*, p. 168.

[49] Shrader-Frechette, *The Real Risks of Risk-Cost-Benefit Analysis*, cit., p. 400.

[50] *Ibidem*.

[51] *Ibidem*.

[52] S. O. Hansson, *Ethical Criteria of Risk Acceptance*, in «Erkenntnis» 59, no. 3 (2003), p. 291, https://doi.org/10.1023/A:1026005915919.

uncertain, or to mixed determinate and non-determinate settings, the result is unsatisfactory, as the preceding objections to consequentialism demonstrate.

If, however, as Cranor cautions in the epigraph and following my argument thus far, risk acceptability cannot be entrusted to technical or cost-benefit risk assessors, to whom should it be assigned? This paper suggests an examination of risk acceptability from a normative perspective that does not focus on the certain or uncertain outcomes of actions and might thereby evade the deterministic bias of mainstream ethics. I refer here to (Aristotelian) virtue ethics. Redirecting ethical enquiry from the question of «*what should I do?*» to «*what sort of person should I be?*», virtue ethics concentrates on one's character and specifically on whether it manifests virtues. Virtue is a stable disposition of a person to do the right thing for the right reasons, in the right way and with the right emotion. The right thing to do is a mean state between two possible reactions, an excessive and a deficient one, and differs according to the situation at hand.

A virtue-ethical approach, then, shifts the focus from the actual or expected consequences of the risk-inducing situation to the «the risk-taker as an intentional agent and, in particular, on said *agent's attitude towards risk-taking* and *sensitivity to the context* in which risks are taken, all of which will reflect her moral character»[53]. Although AI developers may not control the outcome of risk-inducing decisions, they do control the decisions to take risks, so they should be deemed responsible for these decisions. Following the distinction in section 2 between risks we face and those we take, given that AI is deliberately developed and deployed by humans, its risks approximate those to which we decide to expose ourselves and others, compared to, say, risks incurred by natural disasters. Hence, focusing moral evaluation and responsibility attribution on AI developers' character and their decision to risk, rather than on the consequences resulting from such a decision, seems justified. From this perspective, the fact that the adverse consequences of a risk-inducing AI system did not materialise (e.g., because of luck or other external factors) would not suffice to retrospectively absolve AI developers from their responsibility if their decision-making was vicious. Conversely, that their actions eventually led to the imposition of risk or harm would not suffice to affirm their responsibility if their overall attitude was virtuous.

[53] N. Athanassoulis, A. Ross, *A Virtue Ethical Account of Making Decisions about Risk*, in «Journal of Risk Research» 13, no. 2 (March 2010), p. 218, https://doi.org/10.1080/13669870903126309. Emphasis added.

b. *Relation to consequentialism*

Because of this shift towards developers' character and decision-making, virtue ethics evades the first objection to consequentialism about AI's uncertain consequences. Being preoccupied with the «*what should I do?*» question, consequentialism is exclusively act-centred. By contrast, as a predominantly (yet not exclusively) agent-centred theory, virtue ethics embraces an open-ended reflexivity, which takes into account the situational particulars of normative problems and thereby of emerging technologies. By engaging the agent's reasoning, virtue ethics is better placed to address diverse and borderline ethical issues that are inadequately subsumed under binary comparisons or inflexible calculations.

Concerning the objections about consequentialism being impersonal and cruel, again virtue ethics is better situated. Dispensing with aggregate evaluations, it takes into account contextual considerations about AI developers, risk bearers and beneficiaries. Such contextual considerations encompass procedural as well as outcome-oriented aspects. This emphasis on context (especially through the virtue of practical wisdom) likewise places virtue ethics in a better position than consequentialism concerning the objection about the inaccuracy of formalised ways of thinking.

A welcome corollary is that, while avoiding these objections, virtue ethics does not exclude the costs or benefits of AI from being factored into developers' reasoning. That an ethical approach considers consequences does not necessarily mean that it is a consequentialist one[54]. The difference is that in consequentialism, which aims at the maximisation of good over bad consequences, comparisons between consequences are the sole or primary means of evaluation. In virtue ethics, which aims at a good, flourishing life more broadly, considerations of consequences are included among other, more important factors, particularly the agent's virtuous/vicious dispositions and reasoning. Instead of maximising for a single criterion, virtue ethics considers plural, heterogeneous and incommensurable values that are not salient in consequentialism. For example, it would not place AI developers' dispositions on the same scale, allowing the lack of one virtue to be compensated by the increased presence of another. In addition, a virtue ethics consideration of consequences would take into account the context of development/deployment, opposing predictions of AI bringing about certain

[54] Otherwise, almost any ethical theory would be «consequentialised». Conversely, any ethical theory that included some consideration of virtue would be converted into a virtue-ethical one.

consequences independently of social context or use. Such contextualisation allows virtue ethics to adapt to emerging technologies and different cultures in a way in which consequentialism cannot, despite being important to AI as a technology that crosses national, regional or cultural frontiers. Thus, virtue ethics preserves yet addresses the epistemic and moral uncertainty as well as complexity of the situation at hand, rather than reduce them to a simpler, fixed picture or mask them behind quantification as consequentialism does.

c. *Objections*

Although a virtue ethics approach to risk avoids the pitfalls of consequentialism, it can be criticised for failing the role expected of ethical theories, which is to provide a decision procedure, namely «an organized and systematic way of telling us what is the right thing to do»[55]. Just as technical manuals should do the intellectual heavy lifting for us, clarifying the steps we should follow to operate machinery, ethical theories should do the moral heavy lifting for us, issuing instructions we should follow to perform the morally right action in each circumstance. As the steps indicated – for machinery or right actions – are equally available to everyone, this «technical manual model» remains attractive, setting success standards for ethical theories[56].

Consequentialism abides by this model: «[i]t isolates one simple principle behind the directives of our everyday ethical discourse, and then tells us how to formulate this principle and apply it to tell us, systematically and specifically, what to do»[57]. Conversely, by not focusing on the traits of an action but the «qualities of agency» displayed therein, including the risk-taker's motivation, disposition, capacities and reasoning, virtue ethics struggles to identify in advance and in a manner applicable to everyone what a right action or morally acceptable risk would be[58]. It leaves agents without instructions precisely in morally fraught cases when identifying right action becomes critical[59].

[55] J. Annas, *Being Virtuous and Doing the Right Thing*, in «Proceedings and Addresses of the American Philosophical Association» 78, no. 2 (November 2004), p. 62, https://doi.org/10.2307/3219725. See also R.B. Louden, *On Some Vices of Virtue Ethics*, in «American Philosophical Quarterly» 21, no. 3 (July 1984), pp. 227-236.

[56] Annas, *op. cit*.

[57] Annas, *op. cit.*, p. 63.

[58] D. Cox, *Agent-Based Theories of Right Action*, in «Ethical Theory and Moral Practice» 9, no. 5 (October 2006), p. 506, https://doi.org/10.1007/s10677-006-9029-3.

[59] *Ibidem*.

Even if virtue ethics identified right-making features of an action, it is objected that deliberating along their lines would be impermissible[60]. Cox argues that if it is morally right to do x, it must be morally permissible to accurately deliberate about doing x[61]. Virtue ethics might consider that an act manifests the virtue of courage – and is thereby morally instead right – without allowing agents to explicitly deliberate performing said act *because* it manifests courage, lest they exhibit the vice of moral narcissism[62]. Breaking the link between performing and deliberating right action (specifically rendering the latter a violation of the former), the virtue-ethical theory of right action appears contradictory[63].

Two responses are plausible against this criticism. The first accepts that ethical theories should be action-guiding but questions whether virtue ethics fails to be so[64]. Virtue ethics suggests that an action is right if and only if it is what a virtuous agent would characteristically do in the circumstances[65]. Albeit considered under-specified, this theory of right action exhibits the same structure as consequentialism. For the latter, an action is right if and only if it promotes the best consequences, which is not action-guiding until one specifies what counts as the best consequences. Therefore, virtue ethics cannot be less action-guiding solely because it requires specification.

This theory of right action is additionally considered circular: it identifies the right action by reference to the virtuous agent, who, in turn, might be defined as one who performs right actions. Hence, one cannot know what a virtuous agent would do, unless one is already virtuous, in which case guidance is unnecessary. However, an agent may find and consult exemplars in their environment, a practice intuitively used yet unaccounted for in consequentialism. Alternatively, as this paper does, one may focus on canonical virtues/vices.

The second response challenges the very need for ethical theories to provide action guidance, in the form of a decision procedure, in order to be complete. The virtue-ethical theory of right action reproduces the manual

[60] *Ibidem*.

[61] *Ibidem*.

[62] *Ibidem*.

[63] Cox, *op. cit.*; J. Hacker-Wright, *Virtue Ethics without Right Action: Anscombe, Foot, and Contemporary Virtue Ethics*, in «The Journal of Value Inquiry» 44, no. 2 (March 2010), pp. 209-224, https://doi.org/10.1007/s10790-010-9218-0.

[64] R. Hursthouse, *On Virtue Ethics*, 1st ed., Oxford University Press, Oxford 1999; R. Hursthouse, *Normative Virtue Ethics*, in R. Crisp (ed)., *How Should One Live?*, Oxford University Press, Oxford 1996, pp. 19-33.

[65] Annas, *op. cit.*, p. 67.

model, albeit via a proxy, namely, the technical/virtuous expert, whose instructions are treated authoritatively[66]. However, the desirability of being «told what to do» by manuals or experts is questionable[67]. If what matters is merely the application of a theory that tells agents what to do rather than let them make their own moral decisions, praise and blame are more fitting to the theory itself than the agents' character, undermining the need for the latter's improvement[68].

Instead, virtue ethics offers a developmental and aspirational account. On this account, instructions and exemplars are starting points, but agents gradually develop an independent and critical understanding of what virtue requires, an understanding that might not only transcend but further oppose received learnings. Praise and blame are attributed to the agents' actions and decisions, which reflect their character, rather than the agents' application of a theory[69]. Hence, an all-purpose decision procedure available to anyone, regardless of their learning stage, background or character, would both be unrealistic and confine agents to the beginner's state[70]. Contrary to following instructions, agents must do the moral heavy lifting on their own.

Returning to Cox's objection, virtuous agents thus do not ask how an act would reflect on their character but how «experienced people of good character» would act in these circumstances[71]. These people are not necessarily fully virtuous but are better than us (more generous, temperate, and so on). As discussed in sub-section 4.b, such deliberation encompasses the consequences of one's actions on others, rendering the charge of moral narcissism void[72].

Overall, even if virtue ethics is not deemed sufficiently action-guiding, it offers more important guidance on how to improve one's reasoning, considering where agents themselves and their role models stand in their moral development. It recognises that «moral life is not static; it is always developing. When it comes to working out the right thing to do, we cannot shift the work to a theory, however excellent, because we, unlike the theories, are always learning, and so we are always aspiring to do better»[73]. This aspirational, developmental approach of virtue ethics renders it apt for grappling with risk.

66 Annas, *op. cit*., p. 68.
67 Annas, *op. cit.*, pp. 64-65.
68 Annas, *op. cit.*, p. 65.
69 Annas, *op. cit*.
70 Annas, *op. cit*.
71 Hacker-Wright, *op. cit.*, p. 220.
72 Hacker-Wright, *op. cit*.
73 Annas, *op. cit*, p. 74.

d. *Virtuous AI risk-takers*

By focusing on agents, the suggested approach moves from an analysis of *risk* as a noun to an analysis of *(to) risk* as a verb[74]. The morality of risking, then, depends on the risk-taker's dispositions and responsiveness to contextual features of the situation[75]. While several virtuous dispositions are supported in the literature, four are considered 'cardinal' by philosophers in antiquity and later: courage, temperance, justice and (practical) wisdom. Next to these, a fifth one, friendship, is key in Aristotelian virtue ethics.

Whereas thin concepts (e.g., right/wrong, good/bad) denote evaluation only, virtues are thick concepts that combine evaluative with non-evaluative descriptions and thereby are more information-rich. The content of virtues can be specified to bring out dimensions appropriate or important to each context. As part of such a specification effort, and adopting the tenets of Aristotle's virtue ethics as these are fleshed out in his *Nicomachean Ethics*, the following questions are suggested for AI risk-takers' self-assessment[76].

Courage

Courage (*andreia*) is a mean state between fear and over-confidence, distinguished by its motivation. Taking risks to avoid another evil (e.g., repercussions or reproach) indicates cowardice, whereas courage stems from a motivation to achieve what is noble and good[77]. Questions to consider include:

– Do you strive to strike a balance between risk-averse and risk-seeking behaviour?
– Are you disposed to put yourself in harm's way (e.g., to confront internal/external pressures) to promote users' flourishing? Are you disposed to speak up about errors, limitations or blind spots, whether yours or those of the AI system?
– Are you disposed to embrace external criticism, divergent scientific views and other sources of knowledge/expertise to develop scientifically excellent systems?

[74] Hansson, *Philosophical Perspectives on Risk*, cit., p. 30.

[75] Athanassoulis, Ross, *op. cit*. I interpret these two conditions as conjunctively (rather than disjunctively) required.

[76] Aristotle, *Nicomachean Ethics*, in J. Barnes (ed.), *The Complete Works of Aristotle: The Revised Oxford Translation*, trans. W.D. Ross and J.O. Urmson, vol. 2, Bollingen, 71: 2, Princeton University Press, Princeton 1985, pp. 1729-1867. Hereafter, *NE*, with references in Bekker numbering.

[77] *NE*, III.7, 1116a10-15; NE, III.8, 1116a25-30.

Temperance

Temperance (*sophrosune*) regulates pleasure[78]. Intemperate agents take pleasure in things that are wrong or take pleasure in things that are right but do so in a wrong way[79]. Temperate agents seek pleasures that promote health, well-being or nobility and do not exceed the available means. Questions to consider include:

– Are you disposed to forgo pleasurable returns (e.g., economic rewards) or self-indulgent goods when deciding about AI development/deployment?
– Are you disposed to develop systems that promote users' health, safety and overall good lives?
– Are you disposed to strike a sustainable balance in terms of the (environmental) resources used as a means to AI development/deployment?

Justice

Justice (*dikaiosune*) describes lawful and equal agents, whereas unjust agents are greedy, unfair and unlawful[80]. By benefitting those interacting with the just agent (e.g., fellow citizens), justice is strongly relational[81]. It comprises two types: distributive justice concerns the distribution of honour, wealth or anything shared among citizens; corrective justice concerns the correction of voluntary (e.g., commercial) or involuntary (e.g., mandatory) relations/transactions[82]. Questions to consider include:

– How will AI risks and benefits be shared among users (present and future ones)? Which are the different stakeholders and what are their particular status and needs?
– How voluntary or involuntary will the acceptance of the AI system and its risks be by users? Are these risks self- or other-imposed?
– How will you correct for possible harms? Are you open to taking responsibility for this AI system? Have you established chains of accountability?

Friendship

Friendship (*philia*) requires that (i) parties should express mutual goodwill and (ii) this goodwill should stem from a pursuit of the noble, pleasant or useful, with (ii) determining the kind of friendship and content of the good

78 *NE*, III.10, 1117b25-30.
79 *NE*, III.11, 1118b22-27.
80 *NE*, V.1, 1129a31-1129b1.
81 *NE*, V.1, 1130a2-5.
82 *NE*, V.2, 1130b30-1131a9.

that parties wish for each other[83]. Friendships also vary on the basis of association (*koinonia*) among the parties[84]. All joint undertakings foster friendship, with the broadest one being political association between citizens and accordingly political/civic friendship[85]. Questions to consider include:

– Do your decisions to risk demonstrate goodwill (e.g., care and empathy) towards users?
– Is the decision to risk undertaken jointly with other stakeholders? Are you disposed to engage non-experts, particularly citizens possibly affected by the AI system, in the decision-making process?
– Does the decision to risk serve a mutual pursuit of a noble, pleasant or useful objective?

Practical wisdom

Practical wisdom (*phronesis*) is an intellectual virtue that shapes the aforementioned moral ones[86]. It is a practical disposition that involves right reasoning about what is good or bad for humans[87]. Its practicality means that it does not focus on theoretical or abstract goods but on the actions that bring about the practical or moral good[88]. As such, it concurrently assesses the means and ends of a particular action. Questions to consider include:

– What are the broader end(s) that risk-taking does or should serve in this case?
– What are the most suitable means to achieve these ends? Are there less risky means available?
– What are the morally salient features of this situation? What risks does the AI system pose (e.g., on users' health, safety, social and psychological states, rights)? Have you attempted to imagine how these risks might be perceived from the perspective of users?

Briefly put, in AI-related decision-making, developers should aspire to take the risks that a courageous, temperate, just, friendship-promoting, and practically wise agent would accept. As stable dispositions, such virtues require practice to become part of one's way of life. This self-assessment should be iteratively performed throughout the AI lifecycle, while the risk-

[83] *NE*, VIII.2-VIII.3.
[84] *NE*, VIII.9, 1159b25-32.
[85] *NE*, VIII.9, 1160a9-14; NE, IX.6, 1167b1-5.
[86] *NE*, VI.3, 1139b15-17; NE, VI.13, 1144b30-32.
[87] *NE*, VI.5, 1140b1-5.
[88] *NE*, VI.7, 1141b10-15.

taking in which AI developers habitually engage over time is of greater interest than is risk-taking in extreme or high-profile instances. Reliable access to training, role models and virtue-friendly environments is thereby necessary for developers' ongoing self-cultivation.

Moreover, unlike the consequentialist approach concentrating on agents' actions at the expense of their motives, virtues are dispositions to act for the right reasons, implying that AI developers should justify their answers to these self-assessment questions. Without an understanding of the developers' reasoning, third parties can neither evaluate whether developers are virtuous risk-takers nor hold them responsible or trust them[89]. However, virtues are also dispositions to act with the right emotional responses. Illuminating the relevance of emotions to decision-making, virtue ethics challenges conventional portrayals and ideals technologists as purely analytical, impassive professionals.

Although none of these questions determines on its own the acceptability of AI risk, those highlighting considerations of means and ends are particularly important, albeit often neglected. Developers should justify the riskiness of their AI system against the backdrop of not only other systems or digital solutions but also non-technical options. Doing so will counter the «entrenched assumption that the mere advancement to market of a new product, process or technology is demonstration of social "benefit"»[90]. Additionally, comparing AI systems with both technical and non-technical means will illuminate whether their adoption is voluntarily chosen among multiple other options, merely the best among a limited range of alternative options, or even the sole option suitable for achieving the desired end.

Overall, virtue ethics provides a heuristic that does not face the same challenges as consequentialism, as it does not depend on outcomes, but manages to capture more of the ethically relevant dimensions of risk-taking, furnishing a broader *and* more tailored viewpoint. This set of questions is put forward as a preliminary framework. Far from painting a complete picture, they can be supplemented with other virtues or altogether different considerations[91]; still, they may extend the range of normative questions factored into developers' decision-making and serve as ideals to which developers may aspire.

[89] A. Ross, N. Athanassoulis, *Risk and Virtue Ethics*, in S. Roeser *et al.* (ed.), *Handbook of Risk Theory*, Springer Netherlands, Dordrecht 2012, pp. 833-856, https://doi.org/10.1007/978-94-007-1433-5_33; Athanassoulis, Ross, *op. cit*.

[90] Felt *et al.*, *op. cit.*, p. 84.

[91] See, e.g., S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press, New York 2016, https://doi.org/10.1093/acprof:oso/9780190498511.001.0001.

5. *Where to now?*

While AI risk appears like a novel threat to specific aspects or the entirety of human life, concerns about risk have permeated the societal and legal fabric over the past years. Within this «enhanced risk apparatus» of society in general and technology regulation in particular, risk is largely approached in an objectivist, technical manner[92]. Instead, this paper highlights its normative dimensions and illustrates that technocratic approaches to its governance are incomplete or even inaccurate without normative, particularly virtue-ethical, ones.

My illustration has not been exhaustive, especially since the contextual nature of virtue ethics resists codification; it does, however, lay the foundations for further fundamental and applied research. Future applications of the virtue-ethical approach could tailor the (cardinal or other) virtues to risks or challenging situations identified in the AI development literature. It would also be promising to examine whether these virtues could be exercised not only by AI developers as individual risk-takers but in the form of group or institutional virtues exercised by organisations as collective risk-takers. Policymakers are likewise urged to acknowledge the normative dimensions of risk and integrate them into risk-based regulation. Such dimensions may accordingly be embedded in efforts to audit AI.

At the same time, considerations of risk refine ethical reasoning itself, as in practice we operate in far less certain environments than the ones assumed by consequentialism. This uncertainty and its nuances are better captured by virtue ethics. While risk-takers cannot guarantee that the consequences of their actions will eventually occur, they have greater control over the quality of the decision-making that results in said consequences, and this distinction bears on ascriptions of responsibility. This is why virtue ethics attributes primacy to risk-takers' attitudes and their attuned responsiveness to context rather than cost-benefit calculations. To grossly simplify, the goodness (or not) of risk is deduced from the goodness (or not) of one's character. As such, virtue ethics is applicable to risks posed by emerging technologies, including AI, and the ever-changing context that these shape. The aim is not that the actions performed or systems built by AI developers be faultless or that their benefits score higher than their costs in relevant calculations but that AI developers prove to be the courageous, temperate, just, friendship-promoting, and practically wise risk-takers that our society

[92] van Dijk, Gellert, Rommetveit, *op. cit.*, p. 288.

needs. This might prove a more realistic aim, dispensing with modern illusory perceptions of absolute risk objectivity and controllability.

AI practitioners broadly speaking could operationalise the virtue-ethical framework proposed herein as a complementary to or integral facet of AI risk governance, for instance, through the integration of its questions into their codes of conduct. Additionally, policymakers/legislators, employers/managers, and educators are urged to foster organisational cultures and broader environments conducive to the development and exercise of virtues by AI developers. Virtue ethics may thus serve as a compass for navigating AI-induced risk and discerning the moral needs of our messy world writ large.

Acknowledgements

I am grateful for valuable comments from an anonymous reviewer. The research for this paper has been funded through a PhD Fellowship for Fundamental Research by the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek), grant no. 1151621N/1151623N.

Abstract

Risk increasingly permeates technology regulation, as exemplified by the EU's General Data Protection Regulation and Artificial Intelligence (AI) Act. Nonetheless, contrary to common distinctions between an objective stage of risk assessment and a normative one of risk management, I argue that risk governance is normative throughout; hence, it should accordingly integrate normative considerations. To achieve this integration, this paper adopts a normative perspective on AI risk governance in particular. It examines AI-induced risk from the dominant approach of consequentialism, highlighting its limitations in conditions of uncertainty. It suggests virtue ethics as an alternative yet overlooked approach to AI-induced risk and concludes with implications of this approach for research, policy and practice.

Keywords: Virtue ethics; risk-based approach; Artificial Intelligence; AI ethics; AI risk.

Anastasia Siapka
Centre for IT & IP Law (CiTiP), KU Leuven
anastasia.siapka@kuleuven.be

T

Leopoldo Sandonà

L'*ethically informed risk management* in sanità come caso paradigmatico di integrazione etica

> "Rischio" è non solo una nozione che viene impiegata in un argomento centrale da discipline assai diverse, è il modo in cui la "società ibrida" guarda, descrive, valuta e critica la sua stessa ibridità. Questa complessa "e", che resiste al pensare in termini di "o…o", è ciò che costituisce il dinamismo culturale e politico della società globale del rischio[1].

La riflessione di Ulrich Beck sul rischio, definita a livello internazione e globale più che su temi definitamente puntuali, rappresenta una cornice di riferimento per affrontare la questione del rischio in ambito sanitario nello specifico dell'*ethically informed risk management* che si offre come caso tipico, nel contesto della complessità dei sistemi, per evidenziare una domanda di etica non astratta né indeterminata, ma organizzativamente e proceduralmente cogente.

In chiave di premesse sono necessarie due precisazioni. Da un lato la prospettiva statistico-quantitativa, messa in atto dalle scienze sociali e dall'ambito matematico-informatico, si offre come base fondamentale per definire il concetto di rischio, che tuttavia sottende una combinazione di conoscenza degli eventi che non è possibile leggere solo in chiave oggettiva. Il rischio infatti presuppone delle scelte e decisioni che derivano da opzioni personali o comunitarie che rendono l'analisi "scientifica" del

[1] U. Beck, *Risikogesellschaft. Auf dem Weg in eine andere Moderne*, Suhrkamp Verlag, Frankfurt am Main 1986 (trad. it. di W. Privitera, C. Sandrelli, G.C. Brioschi e M. Mascarino, *La società del rischio. Verso una seconda modernità*, Carocci, Roma 2000, p. 340).

 DOI: 10.4454/5rbjjp42

rischio inscindibile dai valori e dai giudizi pratici[2].

Dall'altro lato, come seconda premessa va sottolineato che a livello medico l'azione di terapia e cura può essere letta dentro il quadro offerto da Ivan Illich di iatrogenesi clinica, sociale e culturale[3]. Se nel primo caso abbiamo dei rischi connessi al danno che il medico può fare come "effetti collaterali" ma anche per tutelarsi da un'eventuale conseguenza indesiderata – anticipando il concetto di medicina difensiva –, iatrogenesi sociale è fenomeno in cui si assiste ad una medicalizzazione della vita in termini sempre più estensivi, per arrivare alla iatrogenesi culturale, in cui viene distrutta la capacità potenziale dell'individuo di far fronte in modo personale e autonomo alla propria umana debolezza, vulnerabilità e unicità: «il paziente in preda alla medicina contemporanea non è che un esempio dell'umanità in preda alle sue tecniche perniciose»[4]. L'analisi illichiana in termini critico-negativi – paragonabile a quella foucaultiana in termini biopolitici e come in quel caso carente di elaborazioni propositivo-generative – permette tuttavia di leggere il fattore "rischio" come elemento sempre più presente nella pratica medica e per certi versi autoriproducentesi all'infinito. Se uniamo tale inevitabilità con la prima premessa di un necessario allargamento etico-antropologico alla pura misurazione-previsione quantitativo-statistica, l'analisi del *risk management* si offre come caso specifico di approfondimento non insignificante.

Il presente contributo parte dunque proprio da una contestualizzazione del *risk management* in ambito sanitario, per definire nel secondo passaggio un'analisi di tipo etico-applicativo e riallacciare nell'ultimo passaggio dell'itinerario tale caso specifico ad una cornice più ampia della società del rischio e della bioetica intesa in ottica globale e integrale, nella connessione tra caso singolo e prospettiva globale, nella centralità del principio di giustizia, nella risemantizzazione di concetti chiave della bioetica classica e con un'apertura propositiva e non solo critico-negativa.

[2] P. Sacco, *Rischio*, in *Enciclopedia Filosofica*, vol. 10, Fondazione Centro Studi Filosofici di Gallarate-Bompiani, Milano 2006, pp. 9771-9772. Paradossalmente una mera analisi quantitativa, anche in medicina, conduce all'azzardo più che al rischio, come riconosciuto anche in tempi vicini alla nascita della bioetica, E. Pochin, *Risk and Medical Ethics*, in «Journal of Medical Ethics» 8 (1982) n. 4, pp. 180-184.

[3] Cfr. I. Illich, Limits to Medicine. Medical nemesis, Penguin Books, Harmondsworth, 1977 (traduzione italiana, *Nemesi Medica - L'espropriazione della salute*, Mondadori, Milano 1977).

[4] *Ivi*, p. 28. Su questo tema per quanto non recente rimane di interesse l'intervento del Comitato Nazionale per la Bioetica, *Scopi, limiti e rischi della medicina*, 14 dicembre 2001.

La sanità tra svolta aziendale, professionisti della cura e centralità della persona assistita

Nel contesto dello sviluppo dei sistemi sanitari, che affrontano dopo e oltre la pandemia una sfida decisiva per la propria sopravvivenza[5], si è sviluppato negli ultimi venticinque anni un crescente interesse per il *risk management*, dapprima compreso come semplice rischio clinico e poi allargato anche alle componenti organizzative del sistema sanitario.

La salute, compresa dentro il duplice riferimento dell'art. 32 della Costituzione[6] e delle definizioni dell'Organizzazione Mondiale della Sanità[7], rappresenta per un verso un diritto fondamentale, per altro verso, quando negato o limitato, tale diritto è capace a livello sociale e contestuale di essere «il principale bene e la maggior risorsa per la società in quanto capace, alternativamente, di produrre benefici o disperdere potenzialità in tutti i settori»[8]. Se da un lato l'introduzione di modelli di *governance* per il contenimento della spesa sanitaria ha migliorato la qualità organizzativa di istituzioni sovente carenti di impostazione efficiente, dall'altro lato un'eccessiva aziendalizzazione può far correre il pericolo di mettere a repentaglio la centralità della persona assistita. Sposare invece un modello *patient-centered-care* significa «riconoscere il ruolo preminente degli utenti dei servizi sanitari, adeguando di conseguenza le proprie decisioni strategiche ed operative»[9]. Un passaggio culturale fondamentale è rappresentato dal *community engagement*, cioè dallo sviluppo reticolare delle dinamiche sanitarie, non semplicemente demandate alla scelta dell'autorità politica e amministrativa in una prospettiva *top-down*, ma caratterizzate da una comunicazione costante con la popolazione che si fa prevenzione e comparteci-

[5] Cfr. M. Nefeli Gribaudi, *La sanità nell'emergenza COVID-19: profili di responsabilità sanitaria, aspetti organizzativi e di risk management, riflessioni etiche,* Wolters Kluwer, Milano 2020.

[6] «La Repubblica tutela la salute come fondamentale diritto dell'individuo e interesse della collettività, e garantisce cure gratuite agli indigenti. Nessuno può essere obbligato a un determinato trattamento sanitario se non per disposizione di legge. La legge non può in nessun caso violare i limiti imposti dal rispetto della persona umana».

[7] Anche se è più celebre la definizione di salute del 1948, non priva di elementi critici: «uno stato di completo benessere fisico, mentale, psicologico, emotivo e sociale», nel 2011 in modo più bilanciato la salute è stata definita come «la capacità di adattamento e di autogestirsi di fronte alle sfide sociali, fisiche ed emotive».

[8] F. Rotondo, *Nuovi modelli di governo nel settore sanitario secondo la prospettiva "patient-centered care"*, in L. Marinò (a cura di), *Modelli di* management *nel sistema sanitario: criticità e prospettive*, Giappichelli, Torino 2016, p. 69.

[9] *Ivi*, p. 71.

pazione delle persone assistite sia a livello di macrodecisioni che di singole scelte sulla propria salute, in una cornice reticolare pur nella presenza di necessarie asimmetrie organizzative e gestionali. Tale contestualizzazione nel quadro attuale della sanità ha implicazioni pragmatiche di rilievo[10]. Il terzo tassello del sistema sanitario, oltre all'istituzione e alle persone assistite con i familiari, è rappresentato dagli operatori che, dentro un'alleanza terapeutica rinnovata e un clima di benessere organizzativo, possono ridurre il livello di *stress* operativo e morale, con conseguente miglioramento della qualità delle prestazioni e minor ricorso alla medicina difensiva.

In tale scenario si inserisce la prospettiva del *risk management*, che non riguarda solo alcuni elementi dell'istituzione sanitaria ma «pervade l'intera organizzazione e pertanto tutti i soggetti ad essa partecipanti devono essere sensibilizzati alla gestione dei rischi ed orientare la loro attività in tal senso»[11]. Sarebbe quindi scorretto attribuire a questa funzione una mera dimensione di tecnicalità operativa. Il rischio è intrinseco ad ogni attività, si potrebbe dire che il rischio rappresenta nel contesto globale una caratteristica antropologicamente originaria per i singoli come per le comunità e la dimensione ecologica, definendo dunque un'impossibilità di eliminazione totale del rischio. Già nella letteratura di tipo economico-aziendale – che distingue tra rischi generici e intrinseci ma anche prevedibili e rischi puri legati a situazioni estemporanee[12] – si può rilevare dunque una duplicità intrinseca del rischio, fonte di pericolo e di danno ma anche, quando riconosciuto, accettato e gestito, fonte di opportunità per i singoli come per le comunità. Se sul primo versante è possibile almeno una parziale previsionalità, nel secondo caso, come verificatosi nell'emergenza pandemica, la capacità di calcolo previo appare ridotta, chiamando in causa invece tutte quelle virtù dell'organizzazione che, se debitamente allenate, permettono una pronta risposta nel momento del pericolo. Pensare di annullare questa dimensione intrinseca dell'esistere rappresenta quindi una forma di separazione dal reale, che va invece accompagnato e accolto nella sua complessità sfidante.

[10] *Ivi*, p. 93: «solo attraverso la creazione di vere e proprie *healthy communities* è possibile ridurre evitabili accessi ai servizi di emergenza-urgenza e i tassi di ospedalizzazione».

[11] K. Corsi, *Il* risk management *nelle aziende sanitarie: profili tecnici e culturali*, in L. Marinò (a cura di), *Modelli di* management *nel sistema sanitario: criticità e prospettive*, cit., pp. 99-127.

[12] Cfr. M. Del Vecchio, L. Cosmi (a cura di), *Il risk management nelle aziende sanitarie*, McGraw-Hill, Milano 2003.

Il quadro integrato delle aziende sanitarie definisce così una dimensione che è insieme sociale, legale ed etica del rischio clinico. Appare degno di sottolineatura il fatto che, a dispetto di una crescita dei contenziosi in ambito sanitario, l'alleanza sociale ed etica del momento di cura stia divenendo sempre meno centrale a vantaggio della mera prospettiva legale[13]. Come già osservato, anche la bioetica rischia in molti casi di assumere una dimensione difensiva.

Dal punto di vista degli operatori, la prevenzione o rimozione della malattia (principio di beneficialità) e il non arrecare danno al paziente (principio di non maleficità), manifestano il polo professionale dei princìpi bioetici, accanto a quello dei pazienti (principio di autonomia) e a quello socio-istituzionale (principio di giustizia):

> è forse partendo dalla dimensione etica, e dalla responsabilizzazione del personale sanitario che ne deriva, che si può apprezzare più facilmente la poliedricità del rischio clinico, poiché se tale responsabilità viene elusa, si genera inevitabilmente un danno, una penalità di tipo economico, sociale o legale[14].

La dimensione etica, quando non presente, pone dunque in essere elementi critici che hanno implicazioni variegate. Per converso ciò manifesta la non accessorietà e la non delegabilità del momento etico per l'organizzazione sanitaria.

Tradizionalmente la dimensione del rischio in campo aziendalista viene affidata all'analisi degli *incident reporting* rispetto all'accadere di errori e violazioni dei protocolli che si presentano nonostante le difese messe in atto e le procedure preventive. Le situazioni critiche non devono essere obliate e nascoste in nome di presunte eccellenze, ma analizzate a fondo a livello organizzativo attraverso i dati clinici, le segnalazioni dell'utenza, le indicazioni dei professionisti. Anche in questo caso lo spazio etico di interlocuzione e discernimento rappresenta un fattore decisivo per l'implementazione di determinati *standard* e il mantenimento di una qualità adeguata. Nell'analisi dei dati l'emersione sempre più forte dell'utilizzo

[13] Un esempio significativo e per certi versi paradossale è dato, nel momento legislativo, dalla Legge n. 24/2017, su *Disposizioni in materia di sicurezza delle cure e della persona assistita, nonché in materia di responsabilità professionale degli esercenti le professioni sanitarie*, 8 marzo 2017, che, mentre richiama «l'insieme di tutte le attività finalizzate alla prevenzione e alla gestione del rischio connesso all'erogazione di prestazioni sanitarie e l'utilizzo appropriato delle risorse strutturali, tecnologiche e organizzative», a. 1 c. 2, ha generato una risposta tendenzialmente solo difensivistica.

[14] K. Corsi, *art. cit.*, p. 103.

dell'intelligenza artificiale può avere risultati positivi, se programmata e definita anche con criteri etici[15].

Sinteticamente la cultura della prevenzione del rischio passa da una continua integrazione di «processi e cambiamenti dell'intero sistema organizzativo e che contemplano interventi di semplificazione, supporti alla memoria e all'attenzione, standardizzazione delle procedure, "*checklist*", nonché impegno della "*leadership*", creazione di "*team work*", eliminazione della paura»[16]. Questi snodi consentono di uscire dal circolo vizioso della medicina difensiva per mettere in atto tutti quei comportamenti dei professionisti, degli utenti, delle direzioni che prevengano errori, segnalino anomalie, suggeriscano miglioramenti.

Uno strumento bioetico centrale in questo senso è il consenso informato, in cui si realizza la comunicazione come tempo di cura, inteso come luogo dell'incontro delle autonomie del professionista e della persona assistita. Non va negato in questo senso, tra gli altri servizi aziendali interessati (uffici relazioni con il pubblico, comitato unico di garanzia), anche un possibile ruolo del Comitato etico per la pratica clinica, laddove presente, che ha tra le sue funzioni quella di consulenza sull'allocazione delle risorse[17]. Non è nascondendo le criticità che le organizzazioni possono crescere, alimentando una cultura della colpevolezza e della punizione ma assumendo le crisi come opportunità di sviluppo nello specifico riconoscimento delle competenze di ognuno.

Da un modello basato sull'alternatività tra interessi dei pazienti e dei professionisti, un autentico *risk management* si situa nella prospettiva di un'alleanza per la protezione delle persone assistite e dell'organizzazione sanitaria.

[15] D'altro canto l'aumento esponenziale della potenza di trattamento dei dati può portare ad una dinamica di rischio aggiunto: J. Banja, *How Might Artificial Intelligence Applications Impact Risk Management?*, in «AMA Journal of Ethics» 22 (2020) n. 11, pp. 945-951. Tale prospettiva può riprendere, nel campo delle tecnologie sanitarie, anche il prezioso cammino nel campo *dell'Health Technology Assessment*, che si è arricchito nel 2023 di novità sia in campo nazionale che europeo.

[16] K. Corsi, *art. cit.*, pp. 119-120.

[17] L'allegato B della DGR 938 del 2014 della Regione del Veneto, in materia di definizione delle funzioni dei Comitati etici per la pratica clinica, recita così: «5.d. d) Contributo alla riflessione sull'allocazione e sull'impiego delle risorse nel Servizio Socio Sanitario Regionale 1. Il tema dell'appropriata ed equa allocazione delle risorse nel Servizio Socio-Sanitario Regionale e la valutazione del loro impiego costituisce un ambito di riflessione etica del Comitato in un contesto complesso, caratterizzato da problemi di sostenibilità del sistema a fronte di risorse sempre meno consistenti».

La prospettiva etica per il risk management

La prospettiva del *risk management*, dunque, che intuisce la dimensione integrata dell'esperienza medico-clinica nel contesto delle organizzazioni complesse, mostra oggi il proprio limite se non conciliata con un *ethically informed risk management* in grado di unire sia la componente di etica professionale nelle organizzazioni sanitarie come la gestione etica del rischio aziendale.

Il fatto di avere una missione sociale per sviluppare il diritto alla salute, fanno dell'organizzazione sanitaria un'organizzazione non solo economica ed aziendale. I princìpi della bioetica possono bene illustrare la loro efficacia anche in ottica di *ethically informed risk management*, sul versante pratico-applicativo ma senza negare una dimensione fondativa[18].

Il principio di beneficità, che obbliga a perseguire il bene della persona assistita, bilanciando i benefici con i rischi, diviene nel *risk management* un bilanciamento non solo sul piano medico-clinico o sul piano economico, ma in tutti gli ambiti che integralmente riguardano la persona assistita, compreso l'ambito psicologico e spirituale.

Il principio di non maleficità, che mette in guardia dall'evitare attivamente che procedure messe in atto si rovescino nel loro contrario, è particolarmente chiamato in causa perché si allarga ad una prevenzione e valutazione anche etica dei danni possibili, sempre intesi in chiave integrale e non solo clinica o economicamente quantitativa. I rischi infatti possono sorgere proprio da una mancanza di sguardo complessivo sull'organizzazione e sulle implicazioni di determinati interventi.

Il principio di giustizia, che si pone al livello della società e delle istituzioni di riferimento, assume la funzione di snodo strategico tra i diversi partecipanti all'intervento terapeutico e di cura. Da un lato infatti ci troviamo di fronte alla giustizia distributiva, che vuole evitare le disuguaglianze nell'erogazione e nella qualità di cura, ma anche l'attribuzione di colpe quando avviene un evento avverso e un "incidente". Si parla però in questo contesto anche di una cultura della giustizia, che pone l'accento sulla capacità del *team* e della comunità istituzionale di assorbire gli errori e i fallimenti, di contro ad una cultura punitiva e della colpa. Così la giustizia non è solo distributiva ma anche procedurale (*procedural justice*). Infine tale principio schiude alla prospettiva di una giustizia riparativa (*restorative justice*) delle

[18] A.J. Card, *What Is Ethically Informed Risk Management?*, in «AMA Journal of Ethics» 22 (2020) n. 11, pp. 965-975.

organizzazioni, nella misura in cui essere sanno dischiudere attraverso dei programmi interni possibilità ricostruttive e generative, "compensando" le vittime degli errori. Tra l'altro l'esperienza dimostra che, al di là dell'indubbio valore sociale e culturale della giustizia riparativa, essa non manca di positive implicazioni sul piano economico-finanziario con la riduzione delle compensazioni monetarie.

Il principio di autonomia sembra apparentemente limitato da un *risk management* che riporta la centralità sul piano strutturale e sistemico. In realtà i diversi attori in campo ritrovano nella prospettiva indicata le autonomie in relazione, rendendo meno cogente la prospettiva legale e difensivistica. Il bilanciamento tra i princìpi, come tradizionalmente accaduto per la bioetica, non può essere garantito da un meccanismo automatico: «sfortunatamente non esiste una *checklist* o un algoritmo che garantisca il "giusto" bilanciamento»[19]. Il *risk manager* è colui che si prende in carico le decisioni per soddisfare al meglio, o almeno sufficientemente, i requisiti.

A questi princìpi, come accade anche in altri ambiti dell'etica applicata[20], vengono associati altri elementi fondamentali; alcuni di essi sono soprattutto collegati ai pazienti: la pratica clinica come *patient-centered*, poiché il *risk management* si pone all'intersezione dei quattro princìpi classici e quindi essi vanno orientati nella direzione del bene dei pazienti nell'intreccio di lavoro dei medici, degli amministratori e dei familiari.

In altri casi la centralità è degli operatori: una dinamica di competenza e di professionalità di eccellenza che, in linea con uno sviluppo delle virtù dell'organizzazione, migliori costantemente le prestazioni; va notato in questo senso che un *ethically informed risk management* non è alternativo ad una medicina basata sull'evidenza, anzi permette di dar corpo non solo a valori percepiti da visionarie *leadership* ma renderli obbligazione organizzativa, oltre che morale, per il buon funzionamento delle istituzioni.

Per la gran parte tali elementi ricadono nell'ottica della giustizia: anzitutto il richiamo all'equità, all'onestà e alla trasparenza; inoltre con il vorticoso aumento delle tecnologie informatiche non è fuori luogo sottolineare il ruolo etico del *rispetto per la privacy*, finora purtroppo declinato prevalentemente in un'ottica legale e procedurale; la prospettiva di un *participatory design*[21],

[19] *Ivi*, p. 968: «unfortunately, there is no checklist or algorithm to ensure the "right" balance is struck».

[20] L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina editore, Milano 2022, pp. 91 ss. parla esplicitamente di cinque princìpi dell'intelligenza artificiale in cui, ai quattro classici, si aggiunge l'esplicabilità.

[21] A.J. Card, *art. cit.*, p. 968. Appare degno di nota che una *ethics by design* è sempre più

nella logica di sistemi complessi e adattativi che non si situano anzitutto nella prospettiva dell'esplorazione dei problemi e della risoluzione tecnica degli stessi. Si tratta in altri termini di integrare costantemente la prospettiva di un sistema complesso, che funziona in ottica di costante miglioramento e raffinamento delle procedure. Perché tale sviluppo non sia lasciato alla mera dinamica tecnico-procedurale, l'indicazione etica appare cogente in questa evoluzione, permettendo, in chiave di giustizia riparativa, di dar voce non solo ai soggetti "soddisfatti" dal sistema socio-sanitario, ma anche a quei soggetti vulnerabili che sono stati "feriti" da una situazione di danno e di cura non riuscita.

La prospettiva qui indicata, che raccoglie delle pratiche ancora non riconosciute nei Codici professionali[22], ha il pregio di sottolineare una questione centrale, nella forma tuttavia quasi di un codice etico più che di una pratica e sistematica implementazione organizzativa. Per molti versi siamo ancora nella fase di una descrittività di alcune buone pratiche che possono diventare normative entro il quadro di una bioetica – non solo di una biogiuridica – che radica culturalmente ed estensivamente quanto è ancora presente in forme aurorali. Per realizzare questo passaggio sono necessari luoghi di presenza etica nelle organizzazioni, nella forma di un servizio di bioetica o di Comitati etici di riferimento[23].

Viceversa, l'introduzione di elementi etici attraverso i servizi di qualità e di *risk management* – quando non divengano delega specialistica di una tematica altrimenti assente dalla quotidianità dei professionisti –, può divenire un portale di accesso per la presenza delle questioni etiche in ambito sanitario e organizzativo, anche laddove non siano presenti servizi di bioetica.

richiesta dalle prospettive delle tecnologie emergenti e convergenti, accompagnando "a monte" i processi di programmazione e di sviluppo, di contro ad un'etica medico-clinica e ad una biogiuridica casistica che intervengono "a valle" su singole storie con rilevanza (mediatica ed) etica.

[22] S.J. Schweikart-D.M. Eng, *AMA Code of Medical Ethics' Opinions Related to Risk Management Ethics*, in «AMA Journal of Ethics» 22 (2020), n. 11, pp. 940-944. L'elemento più rilevante rispetto ai codici deontologici riguarda le dimissioni dei pazienti e i possibili rischi connessi a diagnosi errate.

[23] Non è fuori luogo anche un ripensamento di base della formazione accademica che tenga conto di queste nuove frontiere entro le discipline tecnico-scientifiche, come testimoniato per esempio da Y. Guntzburger, T.C. Pauchant, P.A. Tanguy, *Empowering Engineering Students in Ethical Risk Management: An Experimental Study*, in «Science and Engineering Ethics» 25 (2019) n. 3, pp. 911-937 e Id., *Ethical Risk Management Education in Engineering: A Systematic Review*, in «Science and Engineering Ethics» 23 (2017) n. 2, pp. 323-350.

L'integrazione etica nel contesto della società del rischio e della bioetica globale

L'integrazione sperimentata in un caso specifico permette di delineare un utilizzo del concetto di rischio particolarmente interessante nelle etiche applicate. Lasciando una prospettiva puramente statistico-formale del rischio stesso, la cui gestione viene semplicemente proceduralizzata sul piano organizzativo, la gestione etica del rischio consente anzitutto una ritraduzione dei princìpi classici della bioetica (autonomia, beneficità, non-maleficità, giustizia). Se il modello principialista è stato messo in crisi dopo i primi decenni della bioetica, il riferimento ai princìpi garantisce un ancoraggio rispetto all'ipertrofia delle procedure diversificate o alla negazione delle procedure in nome di un generico e indeterminato intervento etico. Parallelamente a quanto accade nel mondo dell'intelligenza artificiale, tale ancoraggio ai quattro princìpi classici può garantire un punto di riferimento fondativo ma anche normativo ed educativo nella traduzione etica a livello individuale come sociale e comunitario. Non va infine negato, per questo primo punto emergente, la sempre più cogente centralità del principio di giustizia, che nell'età secolare sembra sovrastare, senza annientarli ma allargandone la prospettiva, i princìpi di beneficialità e non maleficità tipici della tradizione medica in particolare moderna e il principio di autonomia dirompente nella dinamica novecentesco-contemporanea. L'alleanza terapeutica è oggi alleanza non di singoli ma di servizio sanitario offerto da un *team*, dentro una logica non solo clinica ma anche aziendale e che fa vedere gli effetti decisionali sul breve come sul lungo termine, nello stato di salute del singolo paziente come nell'uso delle risorse sanitarie che riguardano tutti e ciascuno, nella gestione dell'ordinaria urgenza clinica come nelle sfide delle straordinarie emergenze comunitarie-epidemiologiche.

Un secondo elemento emergente è quello dell'allargamento della bioetica verso l'etica organizzativa. Le virtù dei professionisti della salute, proprio perché non sono virtù accessorie ma sono competenze etiche definite, non possono essere lasciate nel campo delle *skill* naturali e individuali, che un professionista può avere come no. Non tutti i professionisti della salute possono avere un *expertise* morale, cioè quello di persone che sappiano distinguere problemi nei vari contesti, esperti di etica che possono identificare e soppesare una vasta gamma di rischi etici, che hanno familiarità con concetti e distinzioni che sono preziosi nelle diverse scelte etiche; ci troviamo dunque di fronte ad una competenza che riguarda la conoscenza delle teorie morali per risolvere i conflitti morali, le alternative dilemmatiche o meno. Anche nelle

professioni sanitarie è però fondamentale che alcuni professionisti abbiano tali competenze specialistiche, per esempio maturate facendo parte di un Comitato etico. Tale *moral expertise* non sarebbe dunque riferito a tutti i professionisti della cura, ma sarebbe una competenza di punta utile per l'intera organizzazione. Attraverso l'*expertise* morale di questi professionisti possiamo arrivare ad una competenza morale dell'organizzazione, non demandata ad alcuni ma come lievito che fermenta in tutta l'istituzione. Come si può facilmente evincere, per arrivare a questo traguardo sarà fondamentale la formazione, la creazione di spazi deputati per il confronto su tali temi e una verifica appropriata e doverosa. In questo trittico assume fondamentale importanza la presenza di chi detiene l'*expertise* morale in chiave individuale per divenire lievito di una *competence* etica condivisa. Anche se non tutti i professionisti sanitari saranno esperti di etica, è fondamentale che tale competenza venga fatta crescere e ci abitui a riconoscere determinate problematiche.

Così è possibile disegnare, anche a partire dal caso affrontato, una circolarità generativa tra dimensione fondativa dell'etica, attenta ai princìpi di riferimento, e dinamica applicativa, non deduttivisticamente intesa. In questo senso la dimensione integrativa dell'etica pratica si situa in particolare sul versante formativo e parenetico, che consente di evitare una riconduzione alla pura normatività, a rischio precettistico e moralistico, come alla pura fondazione, a rischio di astrazione, o alla pura descrittività, con un forte rischio di situazionismo. L'etica integrale dei sistemi complessi può divenire etica integrativa perché, accogliendo la sfida della contestualità, mette in circolo gli elementi fondativi entro un quadro descrittivo e aperto alle nuove esperienze ma anche normativo, facendo dell'esperienza un campo aperto di inveramento dei princìpi e non un campo chiuso di applicazione degli stessi senza sviluppi possibili.

Ancora emerge in questo quadro una necessaria riformulazione transdisciplinare di concetti chiave della bioetica come consenso (e decisione), comunicazione, responsabilità, prevenzione. Dalla pretesa storica di definire una strada autonoma per la disciplina bioetica, con la conseguente frammentazione di significati dei termini utilizzati, la bioetica può in queste frontiere pragmatiche recuperare la sua dimensione di ponte transdisciplinare, presente in differenti saperi epistemologici più che formando una dimensione a sé stante. Non si potrà dunque, per esempio, fornire una lettura solo giuridica o psicologica o etico-morale del tema del consenso, ma esso dovrà essere osservato, nei luoghi deputati, in tutte queste differenti sfaccettature. Per quanto riguarda temi centrali come la comunicazione e la responsabilità il discorso può essere parallelo, dal momento che essi sono spesso ridotti

al momento giuridico e burocratico, di fronte ad una declinazione plurale e inevitabilmente transdisciplinare. Infine il tema della prevenzione da elemento epidemiologico e tecnicamente sanitario può farsi momento educativo, civile e culturale, come è stato evidente nella dinamica pandemica, già relegata in un quadro di passato remoto.

In quest'ottica l'*ethically informed risk management* si offre come caso tipico, nel contesto della complessità dei sistemi, per evidenziare una domanda di etica non astratta né indeterminata, ma organizzativamente e proceduralmente cogente. Il concetto di rischio è così individuato attraverso il caso singolo ma anche ricollegato al contesto più generale di analisi globale. Tra le implicazioni che emergono si può notare infatti una necessaria configurazione epistemologicamente transdisciplinare in cui, senza negare le specifiche appartenenze, le singole discipline si misurano dialogicamente. In particolare appare degno di nota il parallelo tra la dimensione della bioetica globale[24] e il cosmopolitismo metodologico[25] con cui è stata descritta la società del rischio. Un caso specifico, come quello analizzato, può inserirsi dentro un'analisi globale, intendendo tale analisi sia sul piano estensivo che intensivo, a livello geografico come contenutistico ma anche metodologico e normativo, connettendo diverse discipline e analizzando le forme di vulnerabilità. Non a caso nel campo della bioetica globale si parla di una necessaria traduzione locale dei princìpi cosmopoliti definiti anche attraverso documenti internazionali[26]. La vulnerabilità[27] diviene qui concetto chiave sia nella forma descrittiva e diagnostica che in quella propositiva e prognostica, ricollegandosi così alla matrice duplice del rischio come pericolo/ferita e opportunità/apertura. All'interno della globalità riflessiva e critica possono svilupparsi le risposte che la società secolare post-contemporanea definisce rispetto alle minacce presenti e future.

[24] Cfr. H. Ten Have, *Global Bioethics. An Introduction*, Routledge, London-New York 2016 (trad. it. di L. Mariani, *Bioetica globale. Un'introduzione*, Piccin, Padova 2020).

[25] Cfr. U. Beck, *Weltrisikogesellschaft. Auf der Suche nach der verlorenen Sichereheit*, Surkhamp Verlag, Frankfurt am Main 2007 (trad. it. di C. Sandrelli, *Conditio humana. Il rischio nell'età globale*, Laterza, Roma-Bari 2008).

[26] *Dichiarazione Universale sulla Bioetica e i Diritti Umani* (2005). In questo documento i tradizionali quattro princìpi della bioetica si ampliano in quindici princìpi fondamentali: dignità umana, beneficio e danno, autonomia e responsabilità, consenso, mancanza di capacità decisionale in determinati soggetti, rispetto della vulnerabilità umana, *privacy* e riservatezza, uguaglianza ed equità, non discriminazione, rispetto per la diversità culturale e pluralismo, solidarietà e cooperazione, responsabilità sociale e salute, condivisione dei benefici, salvaguardia delle generazioni future, protezione dell'ambiente.

[27] Cfr. H. Ten Have, *Vulnerability. Challenging Bioethics*, Routledge, London-New York 2016; S. Dadà, *Etica della vulnerabilità*, Morcelliana, Brescia 2022.

Infine l'allargamento propositivo promosso dall'integrazione etica, che vede una centralità del momento educativo, spinge ad una rilettura dei compiti dell'etica applicata, la cui domanda si pone al centro del confronto globale delle società del rischio, tra ricerca di stabilità – con la tentazione dell'appiattimento *senza* rischio delegato alle tecnologie onni-potenti – e avventure speculative slegate dai dati – in cui l'*iper*-rischio mette a repentaglio il futuro –, nella consapevolezza dell'evoluzione del contesto globale nelle prime due decadi di millennio dominate da crisi geopolitiche, ambientali e sanitarie.

In campo personale come comunitario appare in aggirabile il compito indicato da un noto chirurgo e saggista statunitense:

> è assurdo pensare che le *checklist* finiranno per rendere superflui il coraggio, l'ingegno, l'improvvisazione ispirata. L'attività medica è troppo intricata, troppo legata alla personale individualità, per seguire una simile evoluzione: i buoni clinici non potranno mai fare a meno dell'audacia competente. Ma dovremo anche predisporci a riconoscere le virtù di un corretto inquadramento[28].

English title: Healthcare ethically informed risk management as paradigm of integrative ethics

Abstract

In the context of the development of health systems, which after and beyond the pandemics show a decisive challenge for their own survival, we consider the development over the last 25 years a growing interest in risk management, at first understood as simple clinical risk and then extended to the organisational components. This perspective, which integrated the dimension of the medical-clinical experience in the context of complex organisations, today shows its limits if it is not reconciled with an ethically informed risk management capable of uniting both the component of professional ethics in healthcare organisations as well as the ethical management of corporate risk. This integration makes it possible to outline a particularly interesting use of the concept of risk in applied ethics. The concept of risk is thus identified through the individual case but also linked to the more general situation of global analysis.

[28] A. Gawande, *The Checklist Manifesto. How to Get Things Right,* Metropolitan Books, New York 2010 (trad. it. di D. Sacchi, *Checklist. Come fare andare meglio le cose*, Einaudi, Torino 2011, pp. 165-166).

Keywords: risk management; patient-centered-care; restorative justice; global bioethics.

Leopoldo Sandonà
Facoltà Teologica del Triveneto
leopoldo.sandona@fttr.it

T

Ilaria Malagrinò

Adolescents and the New Culture of Risk On-line: a Conceptual Framework for an Ethical Training Pragmatics

Young people, risk and uncertainty: a complex relationship in a complex society

It is widely accepted in public discourse that youth are prone to taking risks[1]. In recent years, there has been heightened concerns for young people as a social group at risk due to their risky behaviors[2].The reason for this is that when young people partake in risky behaviours, they jeopardize not only their own futures but also, by extension, that of society as a whole.

Adolescence has long been considered as a period of immense transition, a time for exploring new identities, and a shift in focus from parents to peers in relationships[3]. The tendency of young people to engage in seemingly reckless behaviour has been categorized as extremist conduct, which may stem, at least partially, from their struggle with identity uncertainty and their desire to resolve it[4]. Consequently, uncertainty is a fundamental aspect in the understanding of youth risk-taking[5].

In recent literature, studies have documented a strong correlation between the search for popularity, risk behaviours, and uncertainty. Adoles-

[1] J.O. Zinn, *Understanding Risk-Taking*, Palgrave Macmillan, Switzerland 2020, p. 182.

[2] L.E. Ponton, *The Romance of Risk: Why Teenagers Do the Things They Do*, Basic Books, New York 1997, p. 2.

[3] J.T. Siegel, W.D. Crano, E.M. Alvaro, A. Lac, D. Rast, V. Kettering, *Dying to Be Popular A Purposive Explanation of Adolescent Willingness to Endure Harm*, in M.A. Hogg, D.L. Blaylock (eds.), *Extremism and the Psychology of Uncertainty*, Blackwell Publishing Ltd, Oxford 2012, pp. 115-130, p. 118.

[4] J.O. Zinn, *op. cit.*, p. 184.

[5] J.T. Siegel, W.D. Crano, E.M. Alvaro, A. Lac, D. Rast, V. Kettering, *op. cit.*, p. 115.

 DOI: 10.4454/ve3y3b33

cents who prioritize popularity may be more willing to harm themselves as a means of gaining acceptance and minimizing uncertainty[6].

According to this perspective, it has been argued that the root cause of adolescent risk-taking behaviour is not a lack of rationality, impulsiveness, or because they perceive themselves as invincible. Rather, purposeful risk-taking is motivated by a specific intention, and the goal of the apparently reckless behaviour is not self-destruction but the reaching a desired objective.

Therefore, risk-taking has been interpreted as a practice that adolescents engage in as a response to a specific state of vulnerability. Risk-taking requires skills, is linked to one's identity and the desire for social recognition and may be a last resort when basic needs or feelings of ontological security have been threatened or already affected.

Nowadays we live in a society of increasing complexity. This is contextualized by considering the current period as an age of ongoing and increasing uncertainty, where the very definitions of young and youth become problematic[7]. Young people today seek to establish their identities in a complex globalized socio-economic environment, where new ecological and technological challenges have arisen. The current state of global affairs, from the 2008 global crisis to economic and pandemic uncertainties, the Russian-Ukrainian and Israeli-Palestinian conflicts, and climate change, only serve to heighten feelings of instability and uncertainty. Thus, the notion of risk is one of the most significant and timely concepts that contextualizes young people's uncertain lives[8]. In this context, the experience of identity faces

[6] As Hogg argues groups reduce uncertainty because they provide their members with a clear, unambiguous, and distinct sense of self and social identity built around the group's prototype (M.A. Hogg, *Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes*, in «European Review of Social Psychology» 11 (2000), pp. 223-255; M.A. Hogg, *Uncertainty-identity theory*, in M.P. Zanna (ed.), *Advances in experimental social psychology*, Academic Press, San Diego (CA) 2007, pp. 69-126). Thus, as popular people are seen as being socially knowledgeable, then by becoming popular, the adolescents may believe they will become socially knowledgeable. As popular people are seen as attractive, then by becoming popular, adolescents likely believe they will become attractive as well. As popular people have certain desirable personality characteristics, and more favorable levels of self-concept, then by becoming popular, adolescents likely believe they will gain a more favorable personality, and a more positive sense of self.

[7] R. Huq, *Beyond Subculture: Pop, Youth and Identity in a Postcolonial World*, Routledge, Abingdon 2006.

[8] Two of the most important contributions to the issue of "risk society" are those of Beck and Giddens (U. Beck, *Risk Society: Towards a New Modernity*, SAGE Publications Ltd, London 1992; A. Giddens, *Modernity and self-identity: self and society in the late modern age*, Wiley, Hoboken 1991).

new risks. Old certainties are challenged; the crisis young people are experiencing is not only socio-economic, but also a crisis of values. What has emerged is a less predictable world characterized by insecurity. One of the most prevalent features of the risk society is that dangers have become globalized to the point where the consequences of risk surpass the limitations of time and space. At the same time, people's sense of risk has become more personalized. In this sense, as Sørensen[9] argues, the new risks reinforce the general individualization process characterizing modernity.

In an uncertain globalized environment, individuals must find their own ways to manage the challenges posed by these risks.

Yet, according to Giddens, it is the wide number of possibilities to which individuals are exposed, and the associated returns which they have to calculate, that generates further anxiety. In this context, personal decisions are intertwined with worldwide social changes, making it even harder to deal with any feelings of uncertainty. It could be argued that in today's culture, self-identity is formed in a way that seems to offer young people more options, but doesn't always shield them from the risks associated with those choices.

Thus, in the context of fluid experiences, the feeling of uncertainty is particularly intense for young people. According to Bauman and Raud: «For better or worse, uncertainty has become our fate: for worse, because uncertainty is an un-drying fount of our misery, and for better, because it is also the prime cause of our glory – of human inventiveness, creativity, and our capacity of transcending one by one the limits it sets to human potential»[10]. In addressing uncertainty in today's socio-economic context, young people come to accept the need to embrace or at least live with precarity. Uncertainty has been normalized. It seems that young people attempt to regain some control by "embracing uncertainty", but this approach actually amplifies the feeling of uncertainty. As a result, they seek a way out from everyday lives by emigrating or escaping into the digital world in search of what they hope to be a better life.

The digital space thus becomes crucial when trying to analyse the lives of young people today.

The internet is significant as it provides a platform for social experimentation and the development of interpersonal relationships. It's a unique and

[9] M.P. Sørensen, *Ulrich Beck: exploring and contesting risk*, in «Journal of Risk Research» 21 (2018) n. 1, pp. 6-16, p. 14.

[10] Z. Bauman, R. Raud, *Practices of Selfhood*, Polity, Cambridge 2015, p. viii.

inclusive social outlet, reaching far and wide. A number of studies have shown that adolescents' internet usage can lead to increased identity exploration, self-expression, and positive development. Online interactions have a positive impact on reducing social anxiety and loneliness. They create an environment where adolescents can express their true selves and gain acceptance in positive ways. For this reason, social media practices allegedly bring a sense of connection and belonging.

The advent of the internet has undoubtedly resulted in a profound change in people's life experiences[11]. But it is precisely this transformation that never ceases to worry. Indeed, it is possible that social media is changing not only our personal identity, but also our broader sense of moral responsibility towards our fellow humans. In this way, social media is at the centre of many of our greatest public policy debates; however, its role in shaping the future of humanity is still uncertain. Nevertheless, it is important to evaluate some of the ethical consequences of moving our personal identity online and how we might create the conditions for moral responsibility and the new forms it takes on through and because our virtual reality.

This study aims to explore how risk is expressed and understood online, as well as how social media acknowledges and responds to the connection between risk and uncertainty. The ultimate goal is to comprehend how social media influences the way we perceive and accept moral responsibility and its impact on younger generations. It is important to question the inherent moral value of social networking technologies not just based on their political significance, but also to rethink how we understand and define our own moral values. Technologies don't just shape our perceptions, but also our praxis, introducing a novel set of considerations to the moral issues that are central to the study of technology and its impact on our lives.

Risk online as a shared narrative

In a world filled with captivating visuals, such as in the digital realm, risk is often depicted and shared narratively through images and representations. Risk is carried out in real time for an audience that observes the performer. In the context of being showcased online, risk can be seen as an artifact, a lasting result of a performance for others to observe at their own

[11] As Verbeek explains technology acts upon us as we act with it (P.P. Verbeek, *Moralising Technology: Understanding and Designing the Morality of Things*, University of Chicago, Chicago 2011, p. 8).

pace. One can intentionally present a certain image to a particular audience and monitor their responses through direct replies, even if the communication is not happening in real time.

In this way, when participating in social media young people may become part of the spectacle instead of just being passive spectators.

In the world of social media, people create and consume their own content, whether it's text-based or visual. Thus the notion of prosumption, i.e. the integration of production and consumption[12], is especially important for examining the nature of online risk. As per the findings of Gabriel and Lang, «consumption becomes substantially a consumption of images or a consumption for the benefit of generating images»[13]. The intersection of digital spaces is where young people both create and consume content, and it is here that they define their identities among their peers.

Since a large portion of online self-expression involves visual elements, it is important to examine how camera phone usage contributes to the perception of risk. Second-generation camera phone practices involve the use of social media sharing applications (apps) such as Instagram and Snapchat. This has resulted in an increase in the number of images and videos shared by people, capturing moments of their lives. As noted by Hjorth and Hendry, the use of camera phones results in the emergence of alternative visual modes[14]. Pictures and videos captured on a modern smartphone are edited in apps that can then be shared almost instantly across multiple platforms. Photo and video sharing apps make it easy for prosumers to produce and consume the risk exhibited.

Digital platforms have introduced new ways for young people to consume. The possibility of prosumption intensifies the performative aspect of consumption.

A performance involves not only doing, but also pointing, underscoring, and displaying the act of doing. Goffman defined a performance as «all the activity of a given participant on a given occasion which serves to influence in any way any of the other participants»[15].

[12] A. Burns, *Blogs, Wikipedia, Second Life, and Beyond*, Peter Lang Publishing, New York 2008.

[13] Y. Gabriel, T. Lang, *New Faces and New Masks of Today's Consumer*, in «Journal of Consumer Culture» 8 (2008) n. 3, pp. 321-340, p. 330.

[14] L. Hjorth, N. Hendry, *A Snapshot of Social Media: Camera Phone Practice*, in «Social Media + Society» 1 (2015) n. 1, pp. 1-2, p. 1.

[15] E. Goffman, *The Presentation of Self in Everyday Life* (1959), Penguin Books, London 1990, p. 15.

Performance conjures expectations of theatre. Performativity is linked to preparation, presentation, script, symbolism, props, drama, and most importantly, an audience, whether real or envisioned. In this sense, Goffman's dramaturgical approach is frequently considered a useful foil for understanding risky behaviours exhibited online by young people.

Adolescents thus engage in performances where risk-taking is not part of everyday life, but rather a state of exception that experimenters enjoy and manage carefully.

The performative aspect of risk-taking and sharing conveys meanings related to a sense of belonging in groups.

In cycles of presenting risks and forming impressions, individuals perform on multiple stages. Patterns of action that unfold during a performance are known as parts or routines. In subsequent work these are referred to as restored behaviours. Restored actions encompass the mechanical and conscious activities that become part of the performative repertoire marking one's identity. Language is essential to performativity, as it both describes and presents a form of doing. Whereas restoration and repetition of behaviours reproduce "the Other as the Same", performativity enables a reproduction of the Other in which "the Same is not assured".

The role of risk, like in a play, allows individuals to experiment with different roles and identities by combining, remixing, and practicing various behaviours. These playful practices combine language and aesthetics to construct narratives that support a storytelling of the self, ever in progress and unfinished. Autobiographical performances use performativity to transition from private to public and back, sustaining self-storytelling. Performances such as these often produce staged personal narratives, shaping how audiences understand them and then reinterpret them.

By incorporating strategies of play, adolescents take risks in blending their public and private identities, deconstructing and reconstructing performances in their journey towards an authentic sense of self. Performances thus enable individuals to move from private to public. Sedgwick[16] clarifies, however, that such traversals are further supported by affective processes, which infuse new meaning into the texture of a performance, frequently through linguistic play or reversal of norms. Emotional release and affect are important aspects of the expressive and connective gestures available

[16] E.K. Sedgwick, *Touching feeling: Affect, pedagogy, performativity*, Duke University Press, Durham (NC) 2003.

through social networks[17]. Potentialities for being, then, are both reproduced and multiplied through play and interpretation.

Social media platforms expand the array of performative props, offering a heightened potential for theatricality and drama, which individuals find appealing. Boyd[18] explains that persistence, replicability, scalability, and searchability are important affordances of networked publics. The message conveyed is that self-presentations, captured in video and images, endure and cannot be entirely eliminated, can be effortlessly duplicated, are accessible to both familiar and unfamiliar groups, and can be readily found through searching. Architectures that prioritize default information sharing enhance these capabilities.

Young people usually use tags to mark their videos and posts. Tagging is the act of signing an art performance. Artists develop specific tags to represent their works among known crowds. Tagging categorizes the performance and makes it accessible to wider audiences. It provides more visibility to performative statements of the self, effectively making them the namesake.

Posting risky behaviour can be seen as a way to seek attention and elevate one's social status by presenting oneself in a high-status position. This is known as "aspirational production". The power of the image in this respect is undeniable: individuals curate their pictures and experiences in a way that may provide some sort of affirmation. The subject that arises from this practice involves not only an ontological state but also inevitably involves a politics of visibility, both at the personal level and within the technological infrastructure. It is this visibility that leads to labels of narcissism and vulnerability assigned to young people as the "Look-at-me generation"[19].

A managed self must appear flawless to others. However, this reflection is also influenced by the feedback from peers or the intended audience. Thus, the progressive co-dependence between impressions and the audience has become a defining feature of the social media sphere, highlighting their integral roles within this digital realm. For Leary, «the process of controlling how one is perceived by other people is called self-presentation or impres-

[17] Affective gestures infuse the risk narrative with emotive impressions that enhance performances of the self but may also entrap the self in a continuous loop of mediated affect.

[18] D. Boyd, *Social network sites as networked publics: Affordances, dynamics, and implications*, in Z. Papacharissi (ed.), *A networked self: Identity, community, and culture on social network sites*, Routledge, New York 2010, pp. 39-58.

[19] K. Mallan, *Look at me! Look at me! Self-representation and self-exposure through online networks*, in «Digital Culture & Education» 1 (2009), pp. 51-66, p. 52.

sion management»[20]. He argues that what truly matters is how others react to people's efforts to control their impressions, regardless of whether these responses align with their expectations, irrespective of whether these perceptions are positive or negative.

The power of the gaze can play a role in reinforcing the idea of tailoring digital impressions for an audience-centric strategy. In an online setting, the desire to express different parts of oneself aligns with the belief that the audience holds great influence, sometimes even dictating how the performance unfolds[21]. Conceptually, the audience plays a pivotal role within the changing landscape of internet-related media and advancing technologies. Within the dramaturgical approach, the audience refers to those who observe a specific actor and monitor their performance. These are the people for whom one "puts on a front". This front consists of the specific details that one presents to create the desired impression, as well as the unintentional details that are given off during the performance. In the context of social media, the audience is a conceptual audience. According to Litt[22], the imagined audience is the mental conceptualization of the people with whom we are communicating. An imagined audience may not exactly match the actual viewers, but it also includes users from broader social media community. In effect, Marwick and Boyd[23] have identified an audience, known as a "networked audience", which combines real and imagined viewers to form a wide audience based on social media mass and users' own social networks.

The tragic nature of the risk exhibited online

In online risk narratives, the concept of "seeing" plays a central role, allowing for the exploration of the complex dynamics concerning the am-

[20] M.R. Leary, *Self-Presentation: Impression Management and Interpersonal Behavior* (1996), Routledge, New York 2018.

[21] It has been highlighted by Frison and Eggermont, that the process of "liking" a photo is a significant and daily part of users' engagement with social media platforms, and that it can have an impact on the poster's self-esteem and satisfaction (E. Frison, S. Eggermont, *The impact of daily stress on adolescents' depressed mood: The role of social support seeking through Facebook*, in «Computers in Human Behavior» 44 (2015), pp. 315-325).

[22] E. Litt, *Knock, Knock. Who's There? The Imagined Audience*, in «Journal of Broadcasting & Electronic Media: Socially Mediated Publicness» 56 (2012) n. 3, pp. 330-345, p. 331.

[23] A.E. Marwick, D. Boyd, *I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience*, in «New Media & Society» 13 (2011) n. 1, pp. 114-133, p. 129.

biguity of the drama. This leads to the shaping of identity through tragedy and emerges as a paradigmatic moment of an impending catastrophe. The tragedy of online risk permeates every aspect of the narrative, including both the extrinsic act of posting and the behaviours that ensue. This is due to the specific way social media allows the negotiation of the bond between risk and uncertainty.

Social media is a crucial way for young people to adapt to such uncertain futures. Through the digital realm, young people can seize control over their life paths, but the escape that it offers is quintessentially ephemeral. Social media platforms offer the "illusion" of dynamism to young people, allowing them to construct their identities in the digital space.

Self-management is promoted through social networking sites, providing young people with a semblance of control. Nevertheless, when risk-taking practices become the norm, the ensuing experience may not necessarily be one of safety and predictability. When young adults partake in risky activities on a regular basis, it can cause instability in their daily lives and uncertainty about their future.

Social media offers a sense of stability while also perpetuating instability. The issue at hand is whether young people can find genuine stability in a space that is heavily influenced by the opinions of others. As Ricoeur states, in the "city of opinion", greatness depends on fame and the esteem of others. Each person has no existence and is only considered great in the eyes of others[24]. The most important concept here would be demonstration, which is a key aspect of the pragmatics of judgment.

In their pursuit of fame, individuals who strive to become micro-celebrities, a term that refers to modern celebrity culture on social media, take risks in order to stand out and be one-of-a-kind. According to Senft, «micro-celebrity is best understood as a new style of online performance that involves people 'amping up' their popularity over the Web using technologies like video, blogs and social networking sites» [25]. A key concern for Marwick and Boyd[26] is the apparent transformation of celebrity culture: the fragmented and widely applicable social media self-presentation practices reach and influence more people so that celebrity is formed by what an individual does and not who he or she is.

[24] P. Ricoeur, *Responsabilité et fragilité*, in «Autres Temps. Cahiers d'éthique sociale et politique» (2003) n. 76-77, pp. 127-141; pp. 132-133.

[25] T.M. Senft, *Camgirls: Celebrity and Community in the Age of Social Networks*, Peter Lang, New York 2008, p. 25.

[26] A.E. Marwick, D. Boyd, *op. cit.*

Users trying to raise awareness and become famous has led to a new type of micro-celebrity seeking recognition on social media platforms. The phenomenon observed here could be attributed to the fact that many people, both individuals and non-professionals, are inclined to experiment with building celebrity and its performative element.

In this way, young people utilize social media as a means to create a digital representation of themselves that is perceived as likable. Maximizing "likes" may indicate the desire to increase reputation and rewards on social media platforms.

Social peer influence for acceptance becomes more significant due to the platforms' architecture, which prominently displays the total number of likes or hearts under each shared post. In a way, "likes" can serve as an indicator of peer status and popularity, as well as a method for young people to gauge their personal image. This process impacts the online strategies of young people.

Young people's plans and decisions about posting are linked to the number of "likes" they attract. In general, such constant competition for "likes" is connected to the concept of comparing oneself to others and evaluating appearance.

The pursuit of "likes" by young people is a way of seeking attention that creates an idealized but ephemeral self. It has emerged as their response to the normalized uncertainty they face. Engaging with social media seems to create added pressure for young people to constantly and instantly reshape their identity. This new sense of selfhood is characterized by its ephemerality. It reflects a de-standardization of young people's biographies by a fragmentation of the self into discontinuous ephemeral moments. As Nelson argues[27], the trouble from a moral perspective is that the narrative self and its composition are inevitably dependent on context. Social media, which often lacks context beyond the temporary interactions, highlights the complexities of shaping a digital identity based solely on a "here is where I stand" approach.

Furthermore, individuals need to improve themselves to stay competitive in a "market" where the key is to attract attention. Young people need support to remain focused in a competitive environment. In this way, individuals need to enhance their sense of self by engaging in and posting acts that are increasingly risky in order to reinforce their own feeling of invulnerability.

[27] L.S. Nelson, *Social Media and Morality: Losing Our Self Control*, Cambridge University Press, Cambridge 2018, p. 175.

An invulnerability that Le Breton[28] would define as tragic as young people's shared posts are increasingly characterized by innovation. In this way the element of tragedy lasts only a few seconds before being overshadowed by another post, video or image. These publications are released in abundance, like a litany, but they no longer have an impact on the audience. They do not reciprocate any emotional impact on our personal thoughts and feelings.

And the tragedy reaches its peak in the repeated act of posting that never ends. Social media is rooted in the premise and the expectation that young people will post relentlessly.

In this way social media promotes and encourages the commercialization of personal life by supporting well-packaged and promoted commodified self. The focus of the social media experience is to give young people a leading role in their self-promotion while promoting a reassuring sense of belonging.

The marketing potential of social media lies in its ability to facilitate self-branding, which is predominantly cantered on visual and performative aspects. The notion of branding encompasses not only the consumption of products and services, but also the consumption of meanings that can complement or improve the consumer's self-perception. In this way, self-branding strategies no longer include expensive purchases but involve choreographing a version of oneself. The abundance of photos and videos showcasing risky behaviour allows individuals to visually promote themselves as faultless. The social media arena is a confusing environment for showcasing and consuming one's identity.

Through this visualization, the intersection of social media and consumption results in a changing experience. It offers a hyper-real environment, an idealized version of reality and self that young people perform and present to their peers to be "consumed". Social media is no longer restrained by temporal and spatial limitations. Virtual platforms provide young people with the resources they need to negotiate their identities in a new way. The unfortunate reality of this process is that it prioritizes consumption by young people, even at the cost of their own identities and vulnerabilities.

In this regard, it seems correct to say that young people have become the victims of their own success.

[28] D. Le Breton, *Jeux de mort à l'adolescence*, in «Empan» 97 (2015), pp. 29-38.

From tragedy to fragility to responsibility

Yet it's not all bad. Still, it's unrealistic to live without technology or to believe we will ever eschew our mobile devices and return entirely to face-to- face communication. Therefore, it's necessary to consider how we might evoke a sense of moral responsibility towards each other when we engage online, and perhaps address our moral shortcomings offline as well.

Verbeek suggests that technologies should not be destructive to humanity, but rather they should be explicitly designed to help shape the morality of subjects. This approach to social media is influenced by Heidegger, who discusses the "saving power" within the danger of technology[29].

The tragic nature of the risk narratives shared online has had a twofold direction. One is focused on the immanence that traps in the despair of a repeated action without a final resolution, and the other moves towards the meaning announced in the ambiguity of the "vision", both granted and denied at the same time. Therefore, we must read the drama as teleologically oriented forward, towards the search for practical truth.

The online world's tragedy, pushed to its breaking point, can evoke a longing for resolution in the form of a melancholic realization that develops into a sense of wisdom, a wisdom tinged with tragedy. In light of this the question becomes: What lesson can we learn from the tragic online narrative? This last reminder provides a thread that deserves to be followed in this regard. It is, a call to "think more and differently" as a response is sought to "this terrible risk shown" without simply giving in to defeat. Rather, let us strive for a transformation and a different conjunction, not only in our thinking, but also in our emotions and conduct, especially in terms of morality and politics, towards others. As stated by Ricoeur, this path take us "from the guilty man to the capable man". In this last theme, the portrayal of human fragility takes precedence over dwelling on the tragedy of the action.

As Ricoeur argues in *Responsabilité et fragilité*[30], there is an important relationship between the phenomena of fragility and that of tragedy: this consists in the fact that both the fragile and the tragic are born from the conflict between quality human beings, i.e. both those who post and those who see, whom their very greatness confronts. Moreover, fragility, like tragedy, demonstrates a sort of obstinacy in finitude, a tendency to be closed off to others, from the very forces that the action encounters. The big difference,

[29] M. Heidegger, *Basic Writings*, M.A.D.F. Krell (ed.), HarperCollins, London 1993, p. 297.
[30] P. Ricoeur, *op. cit.*

however, between the fragile and the tragic lies in their different relationship to responsibility. The tragic scenario evokes a situation where man becomes painfully aware of a destiny or a fatality which weighs on his life, his nature or his very condition. The presence of the "fatal" or "destiny" dimension of the posting and that of the action signifies an irreparable conflict that ultimately results in the destruction of the protagonist who is forced to risk his life more and more to stay cool. The fragile does not include this faculty by virtue of the fact that the latter contribute to their downfall by the very efforts they deploy to avoid the disastrous outcome. On the contrary, fragility calls for action which is inherently linked to the concept of responsibility.

As Hans Jonas argues in *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*[31], responsibility has as its specific counterpart fragility.

Ricoeur recalls that Jonas refers to it as a principle because it is immediately expressed as an imperative with nothing preceding it. This principle is enveloped in a feeling that we discover, a feeling that affects us at the level of a fundamental mood in which we place ourselves first of all. We are compelled, enjoined by the fragile, to take action, not just to offer aid, but more importantly to promote growth and enable fulfilment and flourishing. The intensity of emotion lies in its ability to make us experience that exists, yet should not. The imperative is associated with what we perceive as deplorable, unbearable, unacceptable, and unjustifiable. We are made responsible by the fragile. Now, what does it mean: made responsible? This: when the fragile is not something but someone, this someone appears to us as entrusted to our care, placed in our charge. We are responsible for it.

The fragile person is someone who is counting on us; he awaits our help and our care; trusting that we will be there for him. This bond of trust is fundamental. It is important that we encounter it before suspicion arises, as it is intimately linked to the request, to the injunction, to the imperative. It follows that in the feeling of responsibility we feel that we are made responsible for and by.

The question then is: what will we do with this fragile being shown on the screen, what will we do for him?

It is the future of this being, the future of the new generation displayed on

[31] H. Jonas, *Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation*, Insel-Verlag, Frankfurt am Main 1979 (transl. by H. Jonas and D. Herr, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*, The University of Chicago Press, Chicago 1985).

the screen, that we must help to survive and grow. That is our focus. And this future is already our present. The imperative to act now in order to safeguard humanity for the future indicates the far-reaching effects of our technological interventions.

It is in the midst of otherness that we actually become responsible. At its core, this is about acknowledging each other as equals and seeing beyond differences. It's about viewing the other person as a fellow human being, rather than a stranger. We are responsible for a new entity, one that is still in the process of being created rather than recognized. This entity represents a fragile humanity that has lost the absolute values that modernity was based on, and is now defining itself by merging with the virtual world. The strictly political and social question is certainly unprecedented. However, it is to the extent that it is unprecedented that it calls for an anthropological reflection on the ethical behaviours likely to guide responsible behaviour in what Ricoeur would have described as the new fragile region.

It is enough for the recognized fragile to summon us here and now, with no guarantee of success, or even immediate effectiveness. We are at a pivotal moment in history where it is a question of recognizing that the figures of tragedy or, as we said more precisely, the sources of the fragile are also the sources of history, in the sense of "making history", to quote Ricoeur. We are at a turning point if we want to continue "being human in an Hyperconnected era"[32].

Abstract

Traditionally, adolescence has always been associated with a culture of risk. Nowadays, the Internet has radically changed the contexts, opportunities, and ways of expressing risk. In literature we find many studies discussing the assumption of risky behaviors on social media from a cognitive-behavioral point of view, but we lack a suitable conceptual framework to analyze it. In addition, a philosophical reading is also important in order to design an ethical training pragmatics. This analysis is complex, as risk is ambivalent and always refers back to uncertainty. Furthermore, as Verbeek argues, digital platforms are not simple tools, but have become means through which the subjective perceptive experience is created and mediated, with important ethical implications.

[32] L. Floridi (ed.), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Springer 2014.

Therefore, the aim of this paper is to investigate the expressions and meanings of risk exhibited online and how social media responds to the relationship between risk and uncertainty. The final objective is to understand how social media changes the perspective we take on moral responsibility and its impact in relation to new generations. Not only do we need to question whether social networking technologies are inherently moral or immoral because of the political significance we attach to them, but we also need to reconsider how we understand and define our own moral sensibilities. Technologies mediate not only our perceptions, but also our praxis, introducing a novel set of considerations to the question of morality that animates the study of technology and its effect on our lives. We are at a turning point if we want to continue "being human in an Hyperconnected era".

Keywords: risk; uncertainty; adolescents; fragility; moral responsibility.

Ilaria Malagrinò
Università di Messina
ilaria.malagrino@unime.it

premio di studi

Vittorio Sainati

2023-2024

Edizioni ETS

www.edizioniets.com/premiosainati

Premio Vittorio Sainati 2023-2024

T

Giulia Bernard

That which necessarily interests everyone? Writing philosophy in the "age of Enlightenment"

> I know that there are many people who find philosophy
> a great deal easier than higher mathematics!
> But what such people understand by philosophy is
> simply what they find in books which bear that title[1].

Introduction

That philosophers share an odd devotion to grappling with fundamental questions that often lack definitive answers, is something known. The question of *what philosophy is* belongs squarely within the domain of those questions that never cease to haunt philosophy. Radically evolving through the history of philosophy, the issue of differentiating 'philosophy' – its methods, aims, objects, literary genre – from other inquiries requires significant effort. Kant's engagement with this issue serves as a prime example thereof.

One of the main loci where Kant engages in defining philosophy is *The Architectonic of Pure Reason*, where he contrasts philosophy with mathematics, presenting both as rational, rather than historical, engagements with concepts. While historical cognition arises according to facts, external material that is given, rational cognition derives from principle and requires

[1] I. Kant, *Untersuchung über die Deutlichkeit der Grundsätze der natürlichen Theologie und der Moral*, in *Akademie-Ausgabe*, 2, Berlin 1905, pp. 273-302, p. 283 (transl. by D. Walford, *Inquiry concerning the distinctness of the principles of natural theology and morality*, in *Theoretical philosophy, 1755-1770*, Cambridge University Press, Cambridge 1992, p. 255). Kant's works are cited according to the *Akademie-Ausgabe* (AA), with an indication of the volume and page number, except for the *Critique of Pure Reason* (KrV, A and B). Translations are by the author when not otherwise noted.

 DOI: 10.4454/7z2xfx61

to be active with one's own reason. In contrast to mathematics, which, according to Kant, constructs concepts from a priori, non-empirical intuition and essentially operates 'synthetically' (i.e., combining separate elements both intuitively and conceptually), philosophical cognition relies fundamentally on the analysis of concepts that guide human understanding. Hence, philosophy can be understood as a rational activity that is not conducted historically and aims to determine the limits of human knowledge.

In accounting for the specificity of philosophy as discursive cognition from concepts, Kant famously differentiates between a scholastic and a cosmic concept of philosophy. In the scholastic sense, philosophy strives to achieve scientific form through systematic unity. In the cosmic sense, philosophy is not pursued merely for logical perfection, nor is it indifferent to its ends. Rather, it is «the science of the relation of all cognition to the essential ends of human reason (*teleologia rationis humanae*)»[2]. In this sense, philosophy is inextricably linked to the vocation of human beings and is understood as wisdom.

By giving the scholastic endeavour directions toward the highest practical ends of reason, philosophy in the cosmic sense concerns «that which necessarily interests everyone»[3]. Yet the very possibility of philosophy being accessible to everyone, allowing them to exercise their own reason, has long appeared disputable. According to Kant, among the rational cognitions only mathematics can be learnt rationally; philosophy as rational discursive cognition cannot be learnt rationally, but only historically. In that case, it can be learnt as other disciplines (i.e. in a non-exceptional way), but at the cost of becoming something other than itself: i.e. a historical, non-rational (non-philosophical) discipline.

Considerable literature has explored philosophy's exceptionalism with regard to the understanding of philosophy as a discipline among others, i.e. as a cognition that could be taught and learned. While much emphasis has been placed on philosophy transcending its mere discipline-being[4] to

[2] I. Kant, *Kritik der reinen Vernunft*, A 839/B 867 (transl. by P. Guyer and A.W. Wood, *Critique of Pure Reason*, Cambridge University Press, Cambridge 1998, pp. 694-695).

[3] *Ibidem*. On this see J. Stolzenberg, *»Was jedermann notwendig interessiert«. Kants Weltbegriff der Philosophie*, in R. Barth, C.-D. Osthövener, A. von Scheliha (eds.), *Protestantismus zwischen Aufklärung und Moderne*, Peter Lang, Frankfurt an Main 2005, pp. 83-94.

[4] Cfr. L. Illetterati, *Sul concetto di filosofia. Le aporie della scientificità*, in «Giornale di metafisica» XL (2018) 2, pp. 448-471; T. Tuppini, *Immanuel Kant. Nachricht von der Einrichtung seiner Vorlesungen in dem Winterhalbenjahre, von 1765-1766*, in S.-K. Lee *et al.* (eds.), *Philosophical academic programs of the german enlightenment. A literary genre recontextualized*, frommann-holzboog, Stuttgart-Bad Cannstatt, 2012, pp. 251-264; G. Micheli, *L'insegnamento*

be a critical pursuit[5], there remains a significant aspect that has received comparatively little attention: the role of *writing* in delineating philosophy's exceptionalism, particularly in terms of how participation in philosophical discourse is made possible[6]. This paper seeks to fill this gap by highlighting the crucial significance of writing not only within the scholastic concept of philosophy, as might be expected, but also within its cosmic concept, and most notably, in their complex interrelation. To substantiate this claim, the argument unfolds in three stages, each addressing a distinct pair of concepts: private/public (1), scholastic/cosmic (2), and *Bildung*/science (3). In examining each case, I will show that the presence of *writing* is one of the main elements that complicates the clear demarcation between the concepts under consideration. I argue that such complications do not signify incoherence but rather underscore the nuanced exceptionalism inherent in philosophy's concept as rational endeavour capable of enhancing freedom.

1. *Private vs Public*

Philosophy maintains a complex relationship with writing. One might even assert that the entire history of philosophy is a history of experimentation with styles and genres of writing, and, above all, reflections on how writing constitutes a key element to differentiate philosophy from other inquiries. Kant is no exception. In the *Conflict of the Faculties* he identifies substantive debate in print exchange, as the place where scholars engage in reasoning with those who can place themselves in the same position. Writing is the place, according to Kant, where philosophy is philosophy in the proper sense, articulated as science. In writings – polemically asserted against

della filosofia secondo Kant, in L. Illetterati (ed.), *Insegnare filosofia. Modelli di pensiero e pratiche didattiche*, Utet Università, Novara 2007, pp. 136-159.

[5] Cf. N. Hinske, *Kants Verankerung der Kritik im Weltbegriff. Einige Anmerkungen zu KrV B 866 ff.*, in M. Ruffing *et al.* (eds.), *Kant Und Die Philosophie in Weltbürgerlicher Absicht: Akten des XI. Kant-Kongresses 2010*, De Gruyter, Berlin 2013, pp. 263-276; M. Lewin, *Kant's Metaphilosophy*, in «Open Philosophy» 4 (1) (2021), pp. 292-310; A. Ferrarin, *The Powers of Pure Reason. Kant and the Idea of Cosmic Philosophy*, The University of Chicago Press, Chicago-London 2015; C. La Rocca, *La saggezza e l'unità pratica della filosofia kantiana*, in Id., *Soggetto e mondo. Studi su Kant*, Marsilio, Venezia 2003, pp. 217-242.

[6] On the concept of *Bildung* I draw upon the perspectives articulated by G.F. Munzel in *Kant on Moral Education, or 'Enlightenment' and the Liberal Arts*, in «The Review of Metaphysics», 57 (2003) no. 1, pp. 43-73. However, I expand upon these insights to explore how philosophy shapes its own definition.

popular-inspired instructions, combining feelings, inclinations, and rational concepts – philosophy progresses and is «completed»[7].

However, such statements are not without difficulties. To start framing the significance of writing for the way in which philosophy justifies its concept, Kant's portraying of the use(s) of reason in his 1784 Essay *An Answer to the Question: 'What is Enlightenment?'* is of relevance. In the 1784 Essay Kant distinguishes between two uses of reason: private and public. Under private use of reason, i.e. the use of reason a person may make «in a particular *civil* post or office with which he is entrusted»[8], Kant addresses the issues of participation in knowledge, learning and teaching. The private use of reason experiences some structural limitations, because of the need to respect a duty as instructor within the state. To adhere to established, canonical written doctrines is thus paramount. Conversely, the public use of reason, whose nature is instead unrestricted, is defined as the use which anyone may make «as scholar in front of the entire public of the *reading world* [*Leserwelt*]»[9] – the «reading public»[10], which is indeed for Kant the public «in the truest sense of the word»[11]. However much reasoning scholars do in print, they will still obey their boundaries in private use of reason.

Writing and reading are inherently entwined in both these realms, presenting a structural complexity. For writing, in print-exchange fostering the public use of reason, is according to Kant the arena where scientific advancement and rational debate should take place and be fostered. Nevertheless, precisely the book-form is cited among the instruments that mostly prevent the emergence from immaturity. «If I have a book to have understanding in place of me […] I need not make any efforts at all. I need not think»[12] – one reads at the beginning of the Essay. Book-form is not *per se* an emancipatory tool, but rather may stifle individual effort and hinder the development of critical thinking.

The non-autonomous exercise of judgment is what must be prevented in the Age of Enlightenment to enhance critical thinking. This task, which pertains to philosophy, raises certain vexing questions: how can the cultiva-

[7] I. Kant, *Philosophische Enzyklopädie*, AA 29, p. 30.

[8] I. Kant, *Beantwortung der Frage: Was ist Aufklärung?,* AA 8, pp. 33-42, p. 37 (transl. by H.S. Reiss, *An Answer to the Question: 'What is Enlightenment?'*, in *Political Writings*, Cambridge University Press, Cambridge 1989, pp. 54-60, p. 55).

[9] *Ibidem*, transl. modified.

[10] *Ibidem*.

[11] *Ivi*, p. 37 (transl. p. 56).

[12] *Ivi*, p. 35 (transl. p. 54).

tion of critical thinking be promoted? Is the cultivation a matter confined solely to the private use? Given Kant's assertion that the public use of reason «alone can bring about enlightenment among human beings»[13], then what happens in the public sphere cannot be divorced from the (albeit allegedly only private) question of how critical thinking can be learned and performed. Thus, the issue of *Bildung* cannot be dismissed as merely 'private' but instead resurfaces on the level of the public, explicitly written use of reason – where a certain amount of training is already required.

2. *Scholastic vs Cosmic*

The issue of writing poses intricate challenges for philosophy. While the tension inherent in written communication – serving as both a tool for emancipation, allowing individuals to exercise their own reasoning, and yet not guaranteeing per se a successful emergence from immaturity – impacts various fields, its implications are more profound and radical for philosophy and the participation in it.

That Kant understood writing to be related to the issue of philosophy in its exceptional (non-)learnability may be gleaned from some statements he makes in his *Announcement of the Programme of his Lectures for the Winter Semester 1765-1766*, and in *The Architectonic of Pure Reason* (A 836/B 864 - A 840/B 868). In both cases, the role of writing is anything but incidental.

In his *Announcement*, Kant contends that to learn philosophy «is impossible»[14]. All the sciences which can be learned in the strict sense can be either historical or mathematical. In everything historical, one's own experience or the testimony of other people constitute what is given and which can be assimilated and is therefore available for use. In everything mathematical, on the other hand, there is still something given, though *toto coelo* different: the self-evidence of the concepts and the infallibility of the demonstration. In both cases it is possible to impress either on the memory or on the understanding that which can be presented as an already complete discipline. Precisely this is problematic for philosophy. For, says Kant, «to be able to learn philosophy as well there must already be a philosophy which

[13] *Ivi*, p. 37 (transl. p. 55). Transl. modified.

[14] I. Kant, *Nachricht von der Einrichtung seiner Vorlesungen in dem Winterhalbenjahre, von 1765-1766*, AA 2, pp. 303-313, p. 306 (transl. by D. Walford, *Announcement of the Programme of his Lectures for the Winter Semester 1765-1766*, in *Theoretical philosophy*, *op. cit.*, pp. 287-300, p. 292).

actually exists in the first place»[15]. Kant makes this point claiming that

> [i]t must be possible to produce a book and say: 'Look, here is wisdom, here is knowledge on which you can rely. If you learn to understand and grasp it, if you take it as your foundation and build on it from now on, you will be philosophers'[16].

In this passage, Kant underscores the absence of a definitive philosophical text comparable to historical narratives or mathematical treatises, which one could rely on as a foundational resource for practicing philosophy. Two aspects stand out. First, the question of learning pertains not solely or primarily to the scholastic concept of philosophy but extends to its cosmic dimension, where philosophy is referred to as «wisdom». This highlights an intricate connection between philosophy and written discourse, that does not resolve in the alleged opposition between scholastic and cosmic, as if the latter were not embedded in the shortcomings of writing. Before delving further into this issue, it is worth mentioning a second point. The use of writing and the book-form seem at first to imply a conception of philosophy that aligns with non-exceptionalism, suggesting that philosophy, if encapsulated within a definitive text, would be learned *as philosophy* in a way akin to other disciplines, i.e. to the historical ones. Kant's remarks do not exclude such a view when he adds that

> until [*bis*] I am shown such a book of philosophy, a book to which I can appeal, say, as I can appeal to Polybius in order to elucidate some circumstance of history, or to Euclid in order to explain a proposition of mathematics – until I am shown such a book, I shall allow myself to make the following remark: […] one would be betraying the trust placed in one by the public, if […] one were to deceive them with a philosophy which was alleged to be already complete and to have been excogitated by others for their benefit[17].

A complex scenario arises for philosophy here. On the one hand, without a definitive text, learning philosophy *rationally* (i.e. as philosophy) is impossible. In this *interregnum* – where philosophers must be wary of presenting incomplete works as comprehensive knowledge, for this would mean to deceive with an «illusion of science»[18] – philosophy cannot be learned *as philosophy*.

One could argue that this does not mean that philosophy is entirely unlearnable: one can study Wolff's system and, in a way, learn philosophy, but

[15] *Ibidem*.
[16] *Ivi*, p. 307 (transl. p. 293).
[17] *Ibidem*.
[18] *Ibidem*.

only historically. In this case, philosophy would be non-exceptional, akin to learning other historical sciences. However, it should be counter-argued that 'learning philosophy historically' means not learning philosophy in its true rational form, but rather distorting it into something else. Therefore, it holds true for philosophy that it – as *rational* science – cannot be learned as such, and its exceptional nature must be preserved.

On the other hand, it seems that this exceptionalism – its non-learnability *as philosophy* – would cease once a comprehensive text exists[19]. With such a text, it would become possible to learn philosophy *rationally*. Since this is already possible for mathematics, philosophy would, in this case (as in the previous case with historical learning), no longer be exceptional. Against this view, a radical difference between the two should be highlighted. The learnability of mathematics relies on the fact that, for Kant, the sources of cognition on which the teacher draws lie in the «principles of reason, and consequently cannot be derived from any where else by the student, nor disputed in any way»[20], since reason is founded «in pure and therefore error-free intuition»[21]. Philosophy is different: it does not construct concepts nor is it founded in intuition. It can be objectively rational but subjectively historical. Thus, even in this case, philosophy's exceptionalism persists. Consequently, it becomes clear that writing does not imply the non-exceptionalism of philosophy, aligning it with other sciences like mathematics. This raises questions: How can engagement with written philosophy promote critical thinking and enable rational learning? How can writing uphold its rational essence?

This harsh tension, I contend, is further explored in *The Architectonic of Pure Reason*, introducing an additional layer of complexity. Here, Kant posits that

> philosophy is a mere idea of a possible science, which is nowhere given *in concreto*, but which one seeks to approach in various ways until [*so lange, bis*] the only footpath, much overgrown by sensibility, is discovered, and the hitherto unsuccessful ectype, so far as it has been granted to humans, is made equal to the archetype. Until then [*bis dahin*] one cannot learn any philosophy; for, where is it, who has possession of it, and by what can it be recognized?[22].

[19] My point is not that such a possibility is achievable. My interest lies in highlighting the pivotal challenges posed by the issue of writing in Kant's efforts to delineate the concept of philosophy, which reveals problematic (and exceptional) in both cases, whether such a book would ever be written or not.

[20] KrV A 838/B865 (transl. p. 694).

[21] *Ibidem*.

[22] KrV A 838/B866 (transl. p. 694).

Seen *objectively*, philosophy is a model (*Urbild*), the idea of a science that can never be given *in concreto*. While the initial reference to the idea seemingly suggest that philosophy cannot be definitively encapsulated in a definitive book, the subsequent thoughts paradoxically echo the sentiments of the *Announcement*, presenting analogous challenges. Until (*so lange, bis*) the «only footpath» is discovered, it is not possible to learn philosophy *as philosophy* – where is it deposited, where is it written down? How could an ectype equal the archetype? This dilemma presents a further complication in *The Architectonic of Pure Reason* that goes beyond the issues raised in the *Announcement*. After deliberating on the impossibility of learning philosophy *as philosophy* 'until the discovery of the only footpath', Kant adds in the following paragraph a further condition: «until now [*bis dahin*] the concept of philosophy has been only a scholastic concept»[23]. The issue of 'bis dahin' linked to the aforementioned 'so lange, bis' presents a significant challenge. One might interpret 'bis dahin' either (i) temporally, marking the current state where Kant discovers 'the only footpath' in contrast to his predecessors, or alternatively, as proposed by Alfredo Ferrarin[24] (ii), as a pivotal moment in the argument's transition from the scholastic to the cosmic concept. Both interpretations raise profound questions. If the 'bis dahin' is (i) temporal, one wonders about the fate of philosophy's exceptionalism once the path to a cosmic concept is attained. Similar to the questions raised in the *Announcement*, one might inquire about what would become philosophy: could one at that point finally 'appeal' to a philosophical book, as we did with Polybius or Euclid? On the other hand, if 'bis dahin' signals (ii) a shift towards addressing the cosmic concept in Kant's argumentation, one might question how the non-learnability of philosophy transforms when wisdom becomes the focus. Does the exceptional (non-)learnability – meaning philosophy being learnable only historically and thus non-rationally – apply solely to the scholastic concept, thereby creating a permanent division between scholastic and cosmic?

3. *Bildung vs science*

One might be tempted to evade these questions by recalling Kant's distinction between philosophy and something that, for Kant, can indeed

[23] *Ibidem.*

[24] A. Ferrarin, *The Powers of Pure Reason,* cit., p. 75.

be effectively taught and learned: *philosophizing*. To philosophize means knowing how to use the tools of reason, logic and the rules of reasoning. To strengthen this argument and circumvent the challenges posed by writing, one could emphasise the dimension of *Bildung* and argue that the responsibility for fostering rational engagement in the science lies not within philosophy itself, but within learning (and teaching) *how to philosophize*. In this context, a different, pedagogical, and critical approach to writings could be advocated as part of philosophizing – something that could be termed a *spiritual reception*, borrowing terminology from later usage[25]. Kant's *Lectures on the Philosophical Encyclopedia* illustrate this perspective, viewing books not merely as models to emulate but as opportunities to exercise reason and judgment.

This perspective, consistent with Kant's *Announcement*[26] and revealing a notable continuity in the significance of the issue of writing, does not, however, conclusively resolve the question. Instead, it redirects itself back to philosophy as a task: How does philosophy, when properly understood – whether this coincides with the discovery of the only secure path or not – articulate itself philosophically? Should philosophy, in distinguishing its method from both historical cognitions and mathematics, account for its correct reception?

To suggest that accounting for rational participation in philosophy is not inherent to its concept but is rather the responsibility of teachers presents at least two problematic points.

Firstly, it would require a strong endorsement of orality and a robust justification of pedagogy's capacity to promote freedom, both in Kant's writings and in his lecturing practice. While Kant aims to avoid merely teaching 'cognitions' in his lectures, focusing instead on critically displaying modes of thinking, his lecturing style vividly demonstrates an unresolved tension between oral instruction as providing opportunities for critical engagement (albeit within strong institutional constraints) and the praise for substantive scholarly discourse, facilitated by print exchange. As Sean Franzel's analy-

[25] On the issue of 'spiritual reception' as opposed to the approach of historical cognition, I would like to refer to my contribution *Mitteilung of the Absolute: Performing Knowledge in the Philosophy of Religion*, in «Verifiche» LII (2023) n. 2, pp. 207-238.

[26] Cfr: «The philosophical writer […] upon whom one bases one's instruction, is not to be regarded as the paradigm of judgement. He ought rather to be taken as the occasion for forming one's own judgement about him, and even, indeed, for passing judgement against him» (AA 2, p. 307; transl. p. 293).

sis of Kant's nuanced approach to learning to philosophize reveals[27], Kant fails to «present any positive account of oral instruction, more comfortable relegating philosophy's critical potential to the exchange of mature scholars in print»[28]. The absence of a robust consideration of the emancipatory potential of oral instruction, coupled with the strong emphasis on written communication among scholars (as in the *Conflict of the Faculties*), argues against viewing philosophizing and *Bildung* as definitive solutions to the tensions inherent in writing.

Connected to this argument, there is a second reason: the concept of philosophy is framed not neglecting the issue of participation and the necessity of rethinking it to foster a correct, critical attitude against non-philosophical or dogmatic approaches. This holds true also for its *cosmic* sense, not only for its *scholastic* concept, i.e. as wisdom, which is inextricably linked to the destination of human beings. Contrary to what the reference to 'wisdom' might initially suggest, Kant identifies one of the crucial criteria for distinguishing philosophy from non-philosophy in its discursiveness. This perspective can be discerned in *On a Recently Prominent Tone of Superiority in Philosophy* (1796). Here, Kant aims to define philosophy against the misconception of it as an intuitive ability to grasp what concepts cannot attain. By emphasizing philosophy's rigorous «labor on resolving and again compounding its concepts according to principles»[29] which implies «many steps to make advances in knowledge»[30], Kant cautions against the dangers of popularization of a discursive science requiring much labour. Anyone claiming knowledge from mystical revelations or intuitions, without engaging with concepts, according to Kant, is not engaged in philosophy; such activity constitutes the «death of all philosophy»[31]. This perspective, which partly explains Kant's hesitance to wholly endorse oral instruction, demonstrates that Kant's stance, rather than reflecting elitist scepticism toward participation, underscores the imperative of reimagining a correct, critical participation as intrinsic to philosophy itself.

[27] S. Franzel, *A 'Popular', 'Private' Lecturer?: Kant's Theory and Practice of University Instruction*, in «Eighteenth-Century Studies» 47 (Fall 2013) no. 1, pp. 1-18.

[28] *Ivi*, p. 14.

[29] I. Kant, *Von einem neuerdings erhobenen vornehmen Ton in der Philosophie*, AA 8, p. 398 (transl. by G. Hatfield, M. Friedman, *On a recently prominent tone of superiority in philosophy*, in *Theoretical Philosophy after 1781*, Cambridge University Press, Cambridge 2002, pp. 425-446, p. 438).

[30] *Ibidem.*

[31] *Ibidem*.

This necessity, although not explicitly developed by Kant, emerges as a task even after the discovery of the cosmic concept of philosophy. The question of how critical thinking can be learned once philosophy is conceived as a discursive and conceptual endeavour remains pertinent.

If this is indeed the case, then philosophy must reconsider its form to avoid succumbing to misconceptions. This is where Franzel's insightful observation proves crucial, even though it falls short in addressing structural shortcomings. Even if Kant had developed a positive account of oral instruction (which, as Franzel notes, he never convincingly did), the issue of writing would still remain central for philosophy as a discursive science. Orality, within Kant's framework, could never provide a fully explored or entirely satisfactory solution to the contradictions inherent in the written articulation of philosophy and the risks of hindering critical thinking.

If this sounds plausible, then Kant's position seems to be near to some issues that emerge in Plato's critique of writing in the *Phaedrus*. As Derrida elucidates[32], the critique contains an underlying polysemy, where writing is seen as *pharmakon*, both as a 'cure' and a 'poison'. This duality – to which Plato try to remedy by emphasising the role of orality – manifests within the act of writing itself. In a quasi-platonic manner, one could see in Kant a portrayal of philosophy always articulating itself through writing, thereby constantly susceptible to the risk of deviating from its true essence, becoming a historical doctrine not to be a non-discursive enthusiasm[33]. However, if one takes seriously the inherent contradictions of writing, sostituire con: Kant's challenges may be even more profound than Plato's, not because writing is the sole dimension available, but because the 'other' of writing can never fully resolve the contradictions inherent in writing as the medium where philosophy is philosophy. In differentiating its methods and aims from other cognitions, philosophy is always exposed to the risk of reverting into some-

[32] Cf. J. Derrida, *Plato's Pharmacy*, in B. Johnson (ed.), *Dissemination*, University of Chicago Press, Chicago 1981, pp. 61-171. Notably, Ferrarin writes that «in themselves cognitions are dumb, like writing for Socrates in Plato's *Phaedrus*. It is always our judgment that brings them to life» (*op. cit.*, p. 71). Nevertheless, he does not delve deep into the problems that arise for *philosophy*, reproducing the divide between (written) cognitions and the livingness of our rational activity. The problem is not put aside by noting that learning is learning a method (AA 29, p. 6), for philosophy is *also* the science of the relation *of all cognition* to the ends of reason. On the relation between science and wisdom cf. L. Illetterati, *Sul concetto di filosofia*, *art. cit.*, pp. 467-468.

[33] The role of intuition, albeit different from that played in mysticism, signals the distance of mathematics from philosophy as a *discursive* endeavour that is not *constructive*. On the polemic with Plato and the possibility of an intellectual intuition cfr. AA 8, p. 389 ff. (transl. p. 431).

thing historical. The crux of the matter is that this risk can neither be faced nor avoided except *in* writing itself.

Conclusion

Delving into the intricate tapestry of writing offers a re-examination of the delicate balance between contrasting concepts: private versus public, scholastic versus cosmic, and *Bildung* versus the realms of science. This exploration lays the groundwork for a deeper examination of Kant's account of philosophy's exceptionalism, entrusted to the ambiguous nature of writing. Moreover, it unveils the profound connection between this exceptional nature and the quintessential query of the Enlightenment: the engagement in matters of intrinsic, necessary interest to all. The analysis demonstrates that the problematic tension between written philosophical works and participatory engagement remains not only problematic, as shown by Franzel, but is fundamentally intertwined with the very nature of philosophical inquiry. This, I argue, epitomizes the essence of philosophy itself, with writing serving as both a symbol and a conduit for its exceptionalism – not the non-exceptionalism of becoming something historical (always possible) or akin to mathematics (never possible). The challenges posed by writing reveals thus essential to (i) shaping the concept of philosophy *as philosophy* and (ii) setting the basis for redefining participation in philosophy, a task that is crucial and unavoidable once its fundamental discursive nature has been delineated against misconceptions.

Writing thus reveals as the unavoidable arena where such tensions are contested. Against them, two perspectives emerge. One could view philosophy as a science among others and «not think that philosophy should be 'written', any more than science should be. Writing is an unfortunate necessity»[34]. In this case, 'form' does not matter, and the focus is put only on the subjects using their own reason and/or their teachers – facing though some of the shortcomings pointed out by Franzel. Alternatively, one could take seriously the question of *form* of a science that should bear the burden

[34] R. Rorty, *Philosophy as a Kind of Writing: An Essay on Derrida*, in Id., *Consequences of Pragmatism*, University of Minnesota Press, Minneapolis 1982, pp. 90-109, p. 94. Interestingly, such a claim – according to Derrida, on whom Rorty is writing – is uttered by «the Kantian tradition». For, according to Derrida, «no matter how much writing it does» the unique interest of the Kantian tradition would be «to exhibit, to make one's interlocutor stand at gaze before the world» (*ibid*). The present paper can be considered a problematisation of this claim.

of enabling the subjects exercising their own reason without reverting into something historical – and thus allowing to learn a form of spiritual receptivity that cannot be taken for granted in the age of Enlightenment. When the issue of participation arises as intrinsic to the task of defining philosophy as philosophy, as Kant contends, then the second possibility seems the only viable one.

Abstract

In this paper, I aim to investigate the nuanced role that writing, i.e. the act of articulating and conveying arguments and concepts through textual communication, plays within Kant's philosophy, particularly in shaping the concept of philosophy and elaborating ways to participate in it. Central to my argument is the assertion that the issue of writing stands as a linchpin for comprehending the exceptionalism inherent in philosophy: the way in which inquiry in philosophy is somehow epistemologically different from inquiry in other disciplines. I show how such a specificity is inextricably intertwined with the ways philosophy is (not) learnt. By scrutinizing Kant's project to substantiate philosophy's capacity to foster freedom, I contend that a comprehensive exploration of the role of writing is imperative. Through this lens, the interplay between dichotomous concepts – private/public, scholastic/cosmic, and Bildung*/science – emerges as pivotal in elucidating philosophy's exceptionalism and showing how it is linked to the broader Enlightenment-era inquiry concerning participation in critical, emancipatory endeavours.*

Keywords: philosophical writing; Kant; Enlightenment; philosophical exceptionalism.

Giulia Bernard
Università degli Studi di Padova
giulia.bernard@unipd.it

Edizioni ETS
Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa
info@edizioniets.com - www.edizioniets.com
Finito di stampare nel mese di novembre 2024