ISSN 2280-7853

PHENOMENOLOGY AND MIND

20

Rosenberg & Sellier

Topics
Phenomenology and Social Ontology; Ethics and Political Theory; Cognitive Neurosciences, Philosophy of Mind and Language, Logic; Aesthetics, Metaphysics and History of Ideas; Gender Philosophies; European Culture and Politics.

Frequency
2 issues per year

Editors in Chief
Roberta De Monticelli, Università Vita-Salute San Raffaele
Francesca De Vecchi, Università Vita-Salute San Raffaele

Co-Editors
Claudia Bianchi (GENDER, Università Vita-Salute San Raffaele); Massimo Cacciari (Centro Culturale Europeo Palazzo Arese Borromeo, Università Vita-Salute San Raffaele); Vincenzo Costa (PERSONA, Università Vita-Salute San Raffaele); Massimo Donà (DIAPOREIN, Università Vita-Salute San Raffaele); Roberto Mordacci (IRCECP, Università Vita-Salute San Raffaele); Matteo Motterlini (CRESA, Università Vita-Salute San Raffaele); Roberta Sala and Massimo Reichlin (CeSEP, Università Vita-Salute San Raffaele); Andrea Tagliapietra (CRISI, Università Vita-Salute San Raffaele)

Vice-Editor
Stefano Cardini, Phenomenologylab

Managing Editor
Francesca Forlè, Università Vita-Salute San Raffaele

Web Site Editorial Board
Stefano Cardini, Francesca Forlé, Sarah Songhorian

Registrazione del Tribunale di Pavia, n° 6 del 9/07/2012
Direttore responsabile: Roberta De Monticelli
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

n. 20 - 2021

PHENOMENOLOGY AND MIND

DIGITAL IDENTITIES, DIGITAL WAYS OF LIVING: PHILOSOPHICAL ANALYSES

Edited by Greta Favara and Nicole Miglio

Rosenberg & Sellier

Phenomenology and Mind practices double blind refereeing and publishes in English.

SCIENTIFIC COMMITTEE

Ethics and Political Theory (CeSEP)
Giampaolo Azzoni (Università di Pavia)
Elvio Baccarini (University of Rijeka)
Stefano Bacin (Università di Milano)
Carla Bagnoli (Università di Modena e Reggio Emilia)
Antonella Besussi (Università di Milano)
Alberto Bondolfi (University of Geneva)
Patrizia Borsellino (Università di Milano-Bicocca)
Vittorio Bufacchi (University College Cork)
Ian Carter (Università di Pavia)
Emanuela Ceva (University of Geneva)
Antonio Da Re (Università di Padova)
Mario De Caro (Università di Roma III)
Corrado Del Bo (Università di Milano)
Emilio D'Orazio (POLITEIA – Centro per la ricerca e la formazione in politica e etica)
Adriano Fabris (Università di Pisa)
Maurizio Ferrera (Università di Milano)
Luca Fonnesu (Università di Pavia)
Rainer Forst (Goethe-Universität, Frankfurt)
Anna Elisabetta Galeotti (Università del Piemonte Orientale, Vercelli)
Benedetta Giovanola (Università di Macerata)
Barbara Herman (University of California, Los Angeles (UCLA))
John Horton (Keele University)
Andrea Lavazza (Centro Universitario Internazionale di Arezzo)
Neil Levy (University of Melbourne)
Beatrice Magni (Università di Milano)
Filippo Magni (Università di Pavia)
Susan Mendus (University of York)
Glyn Morgan (Syracuse University in New York)
Valeria Ottonelli (Università di Genova)
Gianfranco Pellegrino (LUISS, Roma)
Mario Ricciardi (Università di Milano)
Adina Roskies (Dartmouth College)
John Skorupski (University of St. Andrews)
Jens Timmermann (University of St. Andrews)
Nadia Urbinati (Columbia University)
Corrado Viafora (Università di Padova)

Cognitive Neurosciences, Philosophy of Mind and Language, Logic (CRESA)
Stefano Cappa (Institute for Advanced Study, IUSS, Pavia)
Claudio de' Sperati (Università Vita-Salute San Raffaele)
Michele Di Francesco (Institute for Advanced Study, IUSS, Pavia)
Francesco Guala (Università di Milano)

© The Author(s) 2021.
La presente opera, salvo specifica indicazione contraria, è rilasciata nei termini della licenza Creative Commons Attribution 4.0 Unported (CC BY 4.0: https://creativecommons.org/licenses/by/4.0/legalcode).

CC 2021 Rosenberg & Sellier
via Carlo Alberto 55
10123 Torino
www.rosenbergesellier.it

Rosenberg & Sellier è un marchio registrato utilizzato per concessione della società Traumann s.s.

Phenomenology and Mind, on-line:
http://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

Niccolò Guicciardini (Università di Bergamo)
Diego Marconi (Università di Torino)
Gianvito Martino (Università Vita-Salute San Raffaele)
Cristina Meini (Università del Piemonte Orientale)
Andrea Moro (Institute for Advanced Study, IUSS, Pavia)
Alfredo Paternoster (Università di Bergamo)
Marco Santambrogio (Università di Parma)
Andrea Sereni (Institute for Advanced Study, IUSS Pavia)
Achille Varzi (Columbia University)
Alberto Voltolini (Università di Torino)

History of Ideas (CRISI)
Claudia Baracchi (Università di Milano-Bicocca)
Simonetta Bassi (Università di Pisa)
Andrea Bellantone (Institut Catholique de Toulouse)
Giovanni Bonacina (Università di Urbino)
Enrico Cerasi (Università Vita-Salute San Raffaele)
Francesca Crasta (Università di Cagliari)
Stefano Cristante (Università di Lecce)
Amina Crisma (Università Alma Mater di Bologna)
Giulio D'Onofrio (Università di Salerno)
Catherine Douzou (Université François Rabelais de Tours)
Nicola Gardini (University of Oxford)
Sebastano Ghisu (Università di Sassari)
Simona Langella (Università di Genova)
Anna Marmodoro (University of Oxford)
Vesa Oittinen (University of Helsinki)
Gaetano Rametta (Università di Padova)
Vallori Rasini (Università di Modena e Reggio Emilia)
Francesca Rigotti (Università della Svizzera Italiana)
Hans Bernard Schmid (Universität Basel)
Homero Silveira Santiago (USP – Universidade de São Paulo)
Leonel Ribeiro dos Santos (Universidade de Lisboa)
Alexandre Guimarães Tadeu de Soares (Universidade Federal de Uberlândia, Brasil)
Attilio Scuderi (Università di Catania)
Emidio Spinelli (Università La Sapienza-Roma)
Pirmin Stekeler-Weithofer (Universität Leipzig)
Cristina Terrile (Université François Rabelais de Tours)
Italo Testa (Università di Parma)
Frieder Otto Wolf (Freie Universität Berlin)
Günter Zöller (Ludwig-Maximilians-Universität München)

Phenomenology of gendered personal identity, Gender and Political Normativity, Language and Gender, Philosophy of gender medicine, Women Philosophers (GENDER)
Saray Ayala (California State University, Sacramento)
Maddalena Bonelli (Università di Bergamo)
Antonio Calcagno (King's University College)
Cristina Colombo (Università Vita-Salute San Raffaele)
Luna Donezal (University of Exeter)
Massimo Filippi (Università Vita-Salute San Raffaele)
Sara Heinämaa (University of Jyväskylä)
Dan López De Sa (Universitat de Barcelona)
Anna Loretoni (Scuola Superiore Sant'Anna, Pisa)
Marina Sbisà (Università di Trieste)
Ingrid Vendrell Ferran (Goethe-Universität Frankfurt)
Lea Ypi (London School of Economics)

European Culture and Politics (IRCECP)
Petar Bojanic (Institute for Philosophy and Social Theory, Beograd University, Serbia and Center for Advanced Studies East South Europe, University of Rijeka, Croatia)
Mario De Caro (Università Roma Tre, Tufts University)
Helder de Schutter (Centre for Ethics, Social and Political Philosophy, Leuven)
Ioannis Evrigenis (Tufts University, Boston)
Maurizio Ferrera (Università di Milano)
Rainer Forst, (Goethe Universität, Frankfurt a.M.)
Benedetta Giovanola (Università di Macerata)
Simon Glendinning (London School of Economics, London)
Francesco Guala (Center for the Study of Social Action, Università di Milano)
Rahel Jaeggi (Humboldt Universität, Berlin)
Erin Kelly (Tufts University, Boston)
José Luis Martì (Universitat Pompeu Fabra, Barcelona)
Alberto Martinelli (Università di Milano)
Lionel McPherson (Tufts University, Boston)
Patricia Mindus (Uppsala University)
Philip Pettit (Princeton University)
Alberto Pirni (Scuola Superiore Sant'Anna, Pisa)
Matthias Risse (Harvard University, Boston)
Andrea Sangiovanni (King's College, London)
Thomas Shelby (Harvard University, Boston)
Francesco Tava (University of the West of England)
Antoon Vandevelde (Leuven University)

Phenomenology and Social Ontology (PERSONA)
Tiziana Andina (Università di Torino)
Lynne Baker († 2017)
Stefano Besoli (Università di Bologna)
Jocelyn Benoist (Université de Paris 1- Sorbonne)
Thiemo Breyer (Köln Universität)
Daniele Bruzzone (Università Cattolica Sacro Cuore, Piacenza)
Giovanna Colombetti (University of Exeter)
Amedeo G. Conte († 2018)
Paolo Costa (Fondazione Bruno Kessler, Trento)
Guido Cusinato (Università di Verona, Max Scheler Gesellschaft)
Paolo Di Lucia (Università di Milano)
Maurizio Ferraris (Università di Torino)
Elio Franzini (Università di Milano)
Shaun Gallagher (University of Memphis)
Vittorio Gallese (Università di Parma)
Margaret Gilbert (University of California, Irvine)
Vanna Iori (Università Cattolica Sacro Cuore, Piacenza)
Roberta Lanfredini (Università di Firenze)
Dieter Lohmar (Universität zu Köln)
Giuseppe Lorini (Università di Cagliari)
Anna Marmodoro (University of Oxford)
Verena Mayer (Ludwig Maximilian Universität München)
Gloria Origgi (EHESS, Paris)
Lorenzo Passerini Glazel (Università di Milano-Bicocca)
Jean-Luc Petit (Université de Strasbourg, Paris)
Sonja Rinofner Kreidl (Graz Universität)
Stefano Rodotà († 2017)
Alessandro Salice (University College Cork)
Corrado Sinigaglia (Università di Milano)
Paolo Spinicci (Università di Milano)
Massimiliano Tarozzi (Università di Trento)
Dan Zahavi (Institut for Medier, Københavns Universitet)
Wojciech Żełaniec (Uniwersytet Gdański)

CONTENTS

Greta Favara, Nicole Miglio
Digital Identities, Digital Ways of Living: Philosophical Analyses 12

SECTION 1. PERSONAL IDENTITIES – DIGITAL MINDS, BODIES AND PERSONS

Jean Du Toit and Gregory Morgan Swer
Virtual Limitations of the Flesh: Merleau-Ponty and the Phenomenology of Technological Determinism 20

Verbena Giambastiani
The Asymmetrical Relation Between Humans and Technologies 32

Sarah Pawlett Jackson
Three Bodies: Problems for Video-conferencing 42

Lorenzo Olivieri
Persuasive Technologies and Self-awareness: A Discussion of Screen-time Management Applications 52

SECTION 2. LANGUAGE AND MIND – SOCIAL MEDIA AND IDENTITY CONSTRUCTION

Andrea Addis, Olimpia G. Loddo, and Massimiliano Saba
Social Acts in Digital Environments 64

Silvia Donzelli
Countering Harmful Speech Online. (In)effective Strategies and the Duty to Counterspeak 76

Henk Jasper van Gils-Schmidt
Hilde Lindemann's Counterstories: A Framework for Understanding the #MeToo Social Resistance Movement on Twitter 88

SECTION 3. ETHICAL AND POLITICAL IMPLICATIONS OF DIGITAL TECHNOLOGIES

Helena de Preester
Life Is What You Fill your Attention with – The War for Attention and the Role of Digital Technology in the Work of Bernard Stiegler 102

Francesco Tava
Solidarity and Data Access: Challenges and Potentialities 118

Natalia Satokhina, and Yulia Razmetaeva
"The Loss of Experience" in Digital Age: Legal Implications 128

CONTENTS

SYMPOSIUM ON LUCIANO FLORIDI'S *PENSARE L'INFOSFERA. LA FILOSOFIA COME DESIGN CONCETTUALE*

Carlo Crosato
Introduzione: se la rivoluzione informazionale si rivela troppo conservativa 140

Maurizio Ferraris
Web tolemaico e web copernicano 146

Leonardo Manna
Il pensiero come relazione o intero semantico? Intorno alla filosofia di Luciano Floridi 164

Roberto Mordacci
Sapere, progetto, azione. Sul primato del pratico 176

Luigi Vero Tarca
Infosfera, apertura e verità 184

FREE CONTRIBUTIONS

Shiva Rahman
On the Existential Significance of 'Readiness Potentials' 204

Marco Viola
Three Varieties of Affective Artifacts: Feeling, Evaluative and Motivational Artifacts 228

INTRODUCTION

INTRODUCTION

Greta Favara, Nicole Miglio

GRETA FAVARA
Università Vita-Salute San Raffaele
favara.greta@unisr.it

NICOLE MIGLIO
Università Vita-Salute San Raffaele;
Università degli Studi di Milano
ninicole.miglio@gmail.com

INTRODUCTION
DIGITAL IDENTITIES, DIGITAL WAYS OF LIVING: PHILOSOPHICAL ANALYSES[1]

1 The topics collected in this special issue have been originally discussed during the San Raffaele School of Philosophy 2020, "Digital Identities, Digital Ways of Living: Philosophical Analyses". We wish to thank the organising committee of the San Raffaele School of Philosophy 2020 and all the participants, for having made possible a fruitful and stimulating discussion. We wish also to express our gratitude to Francesca Forlè, for her precious editorial support throughout the realisation of this special issue.

Phenomenology and Mind, n. 20 - 2021, pp. 12-16
DOI: https://doi.org/10.17454/pam-2001
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

This special issue seeks to problematize the role of digital technologies in the constitution of the self, taking up the phenomenological premise that experiential structures are shaped and renegotiated through interactions between subjects, environments, and the manipulation of both real and fictional objects. The articles herein address the *effects* of digital technologies on the human self and, conversely, the active, open, and plastic ways that the self experiences and shapes the digital world. Within contemporary phenomenological debate, macro questions concerning digital technology and identity constitution are increasingly popular topics, soliciting different approaches and theoretical perspectives. With this special issue, we hope to further contribute to this flourishing debate through its three main lines of investigations – corresponding to the three sections into which this issue is organized. We aim primarily to show how digital technologies are re-defining our activities and lives, while also affecting our senses of our identity and selfhood. Each of the three sections analyses interaction between digital technologies and the many layers of self constitution.
Section one interrogates the role played by digital technologies in shaping identity. The second section investigates how digital technologies shape our use of language, enabling new forms of communication and social struggle. Finally, the third section explores different ways in which new technologies might affect change in our socio-political world. In this way, the articles collected in this special issue offer an extensive overview of the implications of new technologies in our everyday experience, from the formation of our personal identities to their potential institutional impact.

Section 1. Personal Identities – Digital Minds, Bodies and Persons

In this section, the authors engage with many ways that technology modifies our being in the world and the experiential structures of selfhood.
In the opening paper *Virtual Limitations of the Flesh: Merleau-Ponty and the Phenomenology of Technological Determinism*, the authors Jean Du Toit and Gregory Morgan Swer offer a theoretical reflection on Merleau-Ponty's phenomenology in order to reconsider the technological determinist outlook alternately in phenomenological terms, as an experiential response to the encounter with modern technology. They advocate recasting the instrumentalist-determinist debate in a phenomenological manner, since this makes possible a reconciliation between the apparent dualism of instrumentalist and determinist positions through Merleau-Ponty's ontology of the *flesh*. This ontology has recently been used to ground accounts of virtual embodiment. The authors argue that, in addition to moving beyond the classical form of technological determinism, it can also phenomenologically ground a novel

understanding of technological determinism—namely, a technological determinism of virtual embodiment.
Verbena Giambastiani's paper *The Asymmetrical Relation between Humans And Technologies* addresses the retroactive effects of technologies on the human self. By engaging with post-phenomenological thought (in particular in the works of Don Ihde, Peter Paul Verbeek and Evan Selinger), she addresses the question: "How does the experience of interacting with a specific technology mediate our experience of the world?". She examines the idea that technologies mediate the world such that perception of the self, world, and environment changes. As a case study, she extensively considers the phubbing phenomenon and its technological effects on social interaction.
In *Three Bodies: Problems for Video-conferencing*, Sarah Pawlett Jackson takes into account the specific way in which video-conferencing modifies structures of intersubjective awareness and interaction. In her paper, she focuses in particular on cases of multi-person interactions (involving more than two people) via video-call, arguing that this kind of interaction should not simply be conceived as "a mere linear extensions or additions of these dyadic interactions", and instead that the self is embodied and enactive. By unpacking some of the key features of multi-person intersubjectivity in cases of embodied co-presence, she shows where and how certain social affordances are strained or lost when multi-person interactions are transferred to the screen.
Lorenzo Olivieri's *Persuasive Technologies and Self-awareness: A Discussion of Screen-time Management Applications* considers persuasive technologies—namely, interactive systems designed to change and shape users' behaviour according to specific goals. By discussing the case of screen-time management applications, this paper explores how persuasive systems transform self-awareness and the self's cognitive architecture. Drawing on the notion of *tectonoetic awareness*, Olivieri illustrates how artefacts enable the transition from temporally bounded experience characterizing first-person perspective (*noetic awareness*) to the ability to self-reflect from a temporally extended third-person perspective (*autonoetic awareness*). He then argues that persuasive systems make possible new modalities of self-recognition and self-projection, while they simultaneously affect senses of agency by interfering with users' actions and intentions.

Section 2. Language and Mind – Social Media and Identity Construction

In their paper *Social Acts in Digital Environments*, Andrea Addis, Olimpia Loddo and Massimiliano Saba offer us a systematic analysis of the performance of social acts in digital environments, considering both fictional and real digital environments. They employ the notion of "organograms" as a key tool for unpacking the user's ability to perform different forms of social acts in digital environments. Interestingly, their analysis is premised on a new reinterpretation of Adolf Reinach's theory of social acts and its ideal development, as well as Czesław Znamierowski theory of the environment. By relying on such a theoretical framework, Addis, Loddo and Saba show how AI might be responsible for affecting both the performance and the perception of social acts in digital environments.
The papers by Silvia Donzelli and Henk Jasper van Gils-Schmidt also provide us with insightful analysis into how new forms of communication in digital environments can have a social impact. By doing so, Donzelli's and van Gils-Schmidt's papers set themselves on the boundary between the analysis of language and the philosophy of politics.
In *Countering Harmful Speech Online. (In)effective Strategies and the Duty to Counter-speak*, Donzelli's topic of analysis is counterspeech, which she defines as "a non-coercive and non-censoring method for reacting to harmful speech, with the aim of impeding or at least diminishing its damaging effects". Donzelli provides a novel philosophical examination into how counterspeech ought to be conducted, such that it might serve as an effective tool in

countering harmful speech and take full advantage of the communicative opportunities online speech dynamics make available.
Similarly, van Gils-Schmidt recalls, in his paper *Hilde Lindemann's Counterstories: A Framework for Understanding The #MeToo Social Resistance Movement on Twitter,* Hilde Lindemann's concept of "counterstories" to describe and assess online social resistance movements. Van Gils-Schmidt proposes a framework according to which we shape our identities in shared social spaces, structured around "master narratives" – narratives that "define the 'realm of possible identities' that we can assume, and form the basis for either recognizing or denying recognition to various social groups in specific roles that they might occupy". By relying on the notion of the "master narrative", van Gils-Schmidt explains that social oppression occurs every time a narrative forbids specific behaviours to their members, or determines the societal roles they might play. Paralleling Donzelli's argument concerning counterspeech, van Gils-Schmidt explains that counterstories are a powerful tool for fighting oppressive narratives. To this purpose, as a case-study, van Gils-Schmidt examines the #MeToo movement as an example of counterstory aimed at countering patriarchal narratives.

Section 3. Ethical and Political Implications of Digital Technologies

The papers by Helena de Preester, Natalia Satokhina, and Yulia Razmetaeva, and by Francesco Tava, further shed light on the political implications of new digital technologies. Helena de Preester examines, in *Life is what You Fill Your Attention with – the War for Attention and the Role of Digital Technology in the Work of Bernard Stiegler*, how marketized digital technologies are responsible for affecting and destructing attention. In support of such a thesis, de Preester draws from Bernard Stiegler's observations on the relationship between capitalism and the destruction of attention. In her contribution, de Preester identifies, and argues for, two possible counterforces for reinforcing attention: education and meditation. As de Preester compellingly argues, education and meditation are two powerful remedies to the detrimental effects of neoliberal capitalism: "if life is what you fill your attention with", de Preester claims, "then focusing or directing attention is one of the most valuable abilities for knowing how to live".
Francesco Tava offers a solidarity-based approach to data access and governance in his paper *Solidarity and Data Access: Challenges and Potentialities*. As Tava explains, discussions around the governance of data access fall within the boundaries of what has been defined as "infrastructure ethics"—namely, recalling a definition provided by Luciano Floridi, ethics which is focussed on analysing the "first-order framework of implicit expectations, attitudes, and practices that can facilitate and promote morally good decisions and actions". In his paper, Tava argues that solidarity ought to be included among "infraethical" practices, since solidarity stimulates "longer-term and risk-laden collective action aimed at addressing perceived injustices". By being so characterised, Tava argues that solidarity could be a useful tool for addressing the problem of digital data use and property. Indeed, analysing the issue of data access through the concept of solidarity allows us to grasp the collective interests that are involved in such practices and to develop new tools for their governance. Tava insightfully concludes his paper by applying his analysis to the pressing case of health data access.
Finally, Natalia Satokhina and Yulia Razmetaeva take inspiration from Hannah Arendt's *The Human Condition* in order to analyse what they define as the "loss of experience" in the digital age and its replacement with technology. Particularly, Satokhina and Razmetaeva reflect on our experience of the law, which they define as one of our modes of being-in-the-world, characterised by the experience of the mutual recognition of people's dignity. In their paper "*The Loss of Experience" in Digital Age: Legal Implications*, Satokhina and Razmetaeva offer an understanding of the legal aspects of experience through phenomenological hermeneutics, analysing their transformation in the digital age.

The symposium on Luciano Floridi's *Pensare l'infosfera. La filosofia come design concettuale*

In addition to the articles so far presented, this special issue includes a symposium on Luciano Floridi's *Pensare l'infosfera. La filosofia come design concettuale*—the Italian translation of an excerpt from *The Logic of Information. A Theory of Philosophy as Conceptual Design*. As Carlo Crosato explains in his introduction to the symposium, Luciano Floridi's *Pensare l'infosfera* offers a pivotal contribution to understanding how new digital means of communication have radically changed the forms of interaction between individuals. Floridi describes the rise of new information technologies as a 'revolution'; new technologies of information and communication have shaped human interactions in such a way as to structure them around communicative relationships, rather than political or economic relationships. How can philosophy contribute to the understanding of such a revolution? As Crosato clearly explains, Floridi's main interest is methodological—Floridi's proposal consists in an attempt to rethink philosophical practices and make them suitable for analysing, and possibly offering guidance on, the current transformation of communication. The contributions of Maurizio Ferraris, Leonardo Manna, Roberto Mordacci, and Luigi Vero Tarca included in this symposium critically examine the tenability and persuasiveness of Floridi's analysis. By considering the practical role of philosophy in contemporary times, the political implications of the Web, and the impact of computational technologies on individuals, Ferraris, Manna, Mordacci, and Tarca provide further enlightening insights into how the effects of new digital technologies upon everyday life ought to be understood and possibly governed.

SECTION 1

SECTION 1

PERSONAL IDENTITIES – DIGITAL MINDS, BODIES AND PERSONS

Jean Du Toit and Gregory Morgan Swer
Virtual Limitations of the Flesh: Merleau-Ponty and the Phenomenology of Technological Determinism

Verbena Giambastiani
The Asymmetrical Relation Between Humans and Technologies

Sarah Pawlett Jackson
Three Bodies: Problems for video-conferencing

Lorenzo Olivieri
Persuasive technologies and self-awareness: a discussion of screen-time management applications

JEAN DU TOIT
School of Philosophy,
North-West University
jean.dutoit@nwu.ac.za

GREGORY MORGAN SWER
Department of Social Sciences,
Walter Sisulu University
gregswer@gmail.com

VIRTUAL LIMITATIONS OF THE FLESH: MERLEAU-PONTY AND THE PHENOMENOLOGY OF TECHNOLOGICAL DETERMINISM

abstract

The debate between instrumentalist and technological determinist positions on the nature of technology characterised the early history of the philosophy of technology. In recent years however technological determinism has ceased to be viewed as a credible philosophical position within the field. This paper uses Merleau-Ponty's phenomenology to reconsider the technological determinist outlook in phenomenological terms as an experiential response to the encounter with the phenomenon of modern technology. Recasting the instrumentalist-determinist debate in a phenomenological manner enables one to reconcile the apparent dualism of the instrumentalist and determinist positions through Merleau-Ponty's ontology of the flesh. *This ontology has recently been used to ground accounts of virtual embodiment. We argue that in addition to explaining away the classical form of technological determinism, it can also phenomenologically ground a novel understanding of technological determinism. Namely, a technological determinism of virtual embodiment.*

keywords

Merleau-Ponty, technological determinism, philosophy of technology, phenomenology, virtuality

Phenomenology and Mind, n. 20 - 2021, pp. 20-31
DOI: https://doi.org/10.17454/pam-2002
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

The formative years of the field of the philosophy of technology are often characterised as a debate over the nature of technology between the opposing positions of instrumentalism and technological determinism. Broadly construed, technological determinism is the position that technology determines society. By this it is meant that firstly, technology is the prime determinant of the course of social change. And secondly, that technology is autonomous in the sense that it has its own internal developmental logic such that, if left to its own devices, there is a certain path that technological development will necessarily take independent of human intention. In short, technological determinists hold that there is something in the nature of technology that is beyond the intentions of its creators and the control of its users. Instrumentalist positions, on the other hand, eschew the possibility that technology might have its own teleology and powers of social determination. Such positions emphasise the instrumental nature of technology, stressing its relative value-neutrality and the key role played by the user's intentions in determining the ends towards which it is applied (Swer 2014).After the so-called empirical turn in the philosophy of technology, the field has been dominated by positions drawn from pragmatism, social constructivism and post-phenomenology (Brey, 2010). Modern forms of instrumentalist philosophy are in the ascendant, and determinism is no longer viewed as credible philosophical position (Du Toit & Swer, 2020; Peters, 2017; Swer, forthcoming). The tendency by contemporary philosophy of technology to dismiss technological determinism out of hand as uninformed or obscurantist has meant that little effort has been made to consider whether there is anything in the determinist account of technology that might be of philosophical merit (Du Toit & Swer, 2020, pp. 234-235). We suggest that, when considered from a phenomenological perspective, technological determinism can be seen as articulating a certain technological experience. When viewed in this phenomenological sense one can explore the experience that technological determinism expresses without endorsing its allegedly pessimistic attitude towards technological development or its essentialising tendencies. To this end, this paper draws upon the later philosophy of Merleau-Ponty, in order to reconsider instrumentalism and technological determinism from a phenomenological perspective. It suggests that both positions can be explicated as experiential aspects of complementary modes of being with technology, with instrumentalism relating to an embodiment mode, and determinism relating to a hermeneutic mode. The merit of viewing these positions from a Merleau-Pontian perspective is that it allows one to draw out the phenomenological bases of both positions. In addition, we argue, such an analysis reconciles these supposedly antagonistic positions

by demonstrating, through the example of digital technologies, that the hermeneutic mode is present in the embodiment mode, and vice versa. Instrumentalism and determinism then appear not as diametrically opposed theoretical positions, but as equally valid and complementary aspects of the phenomenological experience of human-technology relations. Having reinterpreted the classical technological determinist position as but one feature of broader technological experiential whole, we then suggest that another form of technological determinism can be developed from Merleau-Ponty's ontology of the *flesh*. Recent research into the phenomenology of virtuality (Irwin, 2014; Bailey, 2016; Hoel & Carusi, 2018; Ward, 2018; Du Toit, 2020; Du Toit, 2021) has drawn upon Merleau-Pontian phenomenology to explore the possibility of virtual embodiment. Such accounts have typically stressed the expansive features of virtual embodiment. We argue that the same Merleau-Pontian concepts that ground such accounts of embodiment in the virtual, can also ground a novel account of technological determinism. This determinism avoids the essentialist tendencies of classical technological determinism by remaining within the non-dualistic phenomenological framework provided by *flesh,* whilst also indicating the ability of the virtual to limit and determine the possibilities of the virtually embodied. Such an understanding, we suggest, can act as a corrective to the tendency to emphasise the expansive and liberating experiential aspects of digital technologies by drawing attention to the ability of technologically-mediated experience to diminish or constrain existential capacities.

The virtual should not be understood as related to the illusory, false, immaterial, or imaginary – nor as standing in opposition to the real, as Baudrillard's (1981) emphasis on the dichotomous relationship between reality and virtual reality suggests. As Pierre Levy notes, the virtual "is often meant to signify the absence of existence, whereas 'reality' implies a material embodiment, a tangible presence" (1998, p. 23). In contrast, following Levy, we view the virtual as "not a derealization (the transformation of a reality into a collection of possibles) but a change of identity, a displacement of the center of ontological gravity of the object considered" (Levy, 1998, p. 26). Thus, the virtual implies a detachment from the "here" and "now" (Levy, 1998, p. 27). In introducing the idea of embodiment we find a point of engagement, experiential in character, that serves to found the individual's engagement with the virtual by means of the digital technology artefact. Virtual embodiment is therefore suggestive of the idea that to thoroughly conceptualize the virtual, one must look towards both Philosophy of Technology (in terms of the digital technology artefact) and Phenomenology (in terms of the body-subject). There is thus, in describing the virtual, both recourse to the specific (in terms of individual artefacts) and to the general (in terms of the experiential encounter thereof by the individual).

2. Philosophy of Technology and Phenomenology

Early works in the philosophy of technology are frequently placed within instrumentalist or technological determinist schools of thought[1]. Instrumentalists contend that technology is best understood at the level of the individual artefact, which is purported to be value-neutral (value-considerations applying only to issues concerning application). Determinists typically reject the view of technology as a neutral instrument of human volition, arguing rather that technology determines the course of societal development (to some degree).

1 Lewis Mumford, Jacques Ellul and Langdon Winner are typically taken to be the three key technological determinist thinkers (Swer, forthcoming), although Merritt Roe Smith includes Herbert Marcuse, Martin Heidegger, Rene Dubos, Paul Goodman, Murray Bookchin, Kurt Vonnegut, David F. Noble and David Dickson as fellow travellers (1994, p. 28).

The recent turn[2] towards pragmatist and social constructivist modes of technological analysis has effectively enshrined the instrumentalist outlook as the default position for contemporary philosophy of technology. Technological determinism, on the other hand, was deemed pessimistic, totalising, or excessively substantive and is no longer viewed as a credible philosophical position. However, the rejection of technological determinism by mainstream philosophers of technology does not explain its former popularity or its hold upon the popular imagination. We argue that a Merleau-Pontian analysis of the phenomenological bases for both instrumentalist and determinist perspectives provides such an explanation, and can ground a phenomenological account of technological determinism as a form of inter-subjectivity, thereby broadening our understanding of the existential implications of the mediation of human experience by digital technologies.

Swer (2014), drawing on Ihde (1979), suggests that instrumentalism and technological determinism have phenomenological roots in forms of human-technology relations that can be described as either embodiment or hermeneutic relations. Embodiment relations describe the extension of the subject's perceptual field and are characterised by transparency as the subject experiences the world through the technology (and not technology itself). With hermeneutic relations technology-use becomes opaque, i.e. technology becomes the terminal point of the subject's experience (in contrast to extension of the subject's perceptual field). Swer suggests that the instrumentalist position is rooted in the phenomenological and embodied experience of technology. This human-technology relation is typified by the use of individual technologies, such as a cane or hammer. In such technologically-mediated praxis the artefact 'withdraws' and the user experiences technology as an extension of their bodily intentionality. Swer links technological determinism to hermeneutic human-technology relations that typify human interfaces within larger-scale technological systems. Technology at this level appears not as a neutral carrier of human volition but as a limit to it, an external object that must be interpreted – here the conception of technology as opposing human intentions becomes meaningful.

This analysis suggests that instrumentalism and determinism are grounded in equally valid, though opposing, experiences of technological praxis. In the next section we argue that Swer's (2014) phenomenological characterisation of the instrumentalism-determinism dichotomy may be enriched through engagement with a Merleau-Pontian perspective. Merleau-Ponty's phenomenology, we argue, can circumvent the agonistic dichotomy present in Swer's account by demonstrating that embodiment and hermeneutic relations are present in all forms of technological praxis.

3. Technology and the *Flesh*: Beyond the Instrumentalism/ Determinism Dichotomy

The proto-theory[3] of technology presented in *Phenomenology of Perception* focuses on five specific technological artefacts: a feathered hat, a car, a blind man's stick, a typewriter, and an organ. The account of the blind man's stick is descriptive of both a motor habit (as one learns to use the cane) and a perceptual habit:

> Once the stick has become a familiar instrument, the world of feelable things recedes and now begins, not at the outer skin of the hand, but at the end of the stick. .. . the

2 See Achterhuis (2001), Brey (2010), Van den Eede, et al (2017), and Du Toit & Swer (2020).

3 The account of technology given in *Phenomenology of Perception* does not constitute an encompassing argument on the nature of technology, as do the full-fledged theories of technology presented by philosophers of technology such as Mumford and Ellul. Rather, in the early work of Merleau-Ponty, the use of examples of technological artefacts (such as the blind man's cane and a woman's feathered hat) present an explication of his phenomenological reasoning concerning embodiment (Ihde and Selinger, 2004, p. 361–367).

> stick is no longer an object perceived by a blind man, but an instrument with which he perceives. It is a bodily auxiliary, an extension of the bodily synthesis (Merleau-Ponty, 1962, pp. 175–176).

The blind man's stick reveals an instrumentalist perspective that is propounded throughout Merleau-Ponty's early technological thought, whereby a tool is incorporated into the body schema to become transparent while allowing for expanded perceptual and motor potentiality. This instrumentalist conceptualization of technology extends to Merleau-Ponty's examples of the car and the organ, whereby the artefact described becomes in some sense useful in the extension of the body-subject. The feathered hat and typewriter, in turn, illustrate how technology relates to the extension of the body through embodied skills, and that skilful use of an artefact is needed to utilize said artefact as an instrument for human use with specific ends. Again, this embodied mode of being with technology accords with the instrumental view of technology – the latter being sketched against a technological determinism that is rooted in the hermeneutic mode.

This changes in Merleau-Ponty's later work, with Carusi and Hoel noting that instruments, tools, and technologies seem to become a constant preoccupation of Merleau-Ponty's later thought, particularly in *The Visible and the Invisible* and his unfinished manuscripts and lecture notes (2015: 73). The instrumentalist perspective shifted as Merleau-Ponty's phenomenology matured, and we argue that his later concept of the *flesh* suggests a means to overcome the instrumentalist and determinist dichotomy regarding technology-relations. The *flesh* is an ontological concept that aims to overcome dualisms through the reconceptualization of such 'dualisms' as intertwining, chiasmically associated, and interdependent. Examples include world and consciousness, sensing and sensible, or technological artefact and the body (Merleau-Ponty, 1968, p. 123). The *flesh* suggests a reaching across of the world to the body-subject, and the body-subject to the world, via a 'space' of connection – a co-implicity.[4]

It is argued that, in accounting for the co-implicity that always resides in the body-subject and in the world simultaneously, Merleau-Ponty may ground a new account of technologically-mediated experience. Such a novel account of technology, which will be illustrated through the example of digital technologies, suggests that ontologically the hermeneutic mode of technology relations is also present in the embodied mode, and that the embodiment mode is present in the hermeneutic mode. The later Merleau-Pontian account of technologically-mediated experience dissolves the dichotomy between technological instrumentalism and technological determinism that typified classic philosophy of technology. Furthermore, this account unifies both positions by arguing that the *flesh* explains our engagement with technological artefacts (objects in the world) as a mutual constituted experiential field.

Understanding the co-implicity of body-subject and world by means of the *flesh* is also crucial for understanding the equating of tools and symbols in later Merleau-Ponty (in *Eye and Mind*). Merleau-Ponty describes both tools and symbols as technical objects, stating that: "Like all other technical objects, such as tools and signs, the mirror has sprung up along the open circuit *between* the seeing and the visible body. Every technique is a 'technique of the body,' illustrating and amplifying the metaphysical structure of our flesh" (1964, p. 6). Both tools and symbols are means by which experience of the world occurs, with a comparable

4 Co-implicity is thus suggestive of the *sensorium commune* described in *Phenomenology of Perception* as the space of the intertwining of the senses "sometimes affected from one side, sometimes from the other" (Merleau-Ponty, 1962, p. 244). Whereas Merleau-Ponty's initial account of embodied perception is characterized by a movement towards incarnated meaning, his later work moves beyond sensory perception to extend into and to be constitutive of intellectual life (De Saint Aubert, 2008, p. 10, 14).

capacity to decentre the perceiving body (Hoel & Carusi, 2018). On the basis of this mutuality, Merleau-Ponty's *flesh* grounds a new account of technologically-mediated experience that, ontologically, consolidates and unifies the hermeneutic and embodied modes of technology relations in the body-subject through a mutually constituting experiential field.

4. Technological Mediation as experiential and behavioural restriction

While the *flesh* offers considerable resources for a more nuanced understanding of technologically-mediated experience, recent analyses of digital technologies that utilize Merleau-Ponty's phenomenology tend to focus on the seemingly utopian capacity of the virtual to generate expanded horizons and potentialities for action and experience – often disregarding the possibility of technological mediation as experiential restriction.[5] These analyses take up early Merleau-Pontian instrumentalism (artefacts as value-neutral tools), while disregarding the possibilities presented by the *flesh* for hermeneutic modes of being with technology. Instrumentalist accounts sketch the use of the technological artefact as tool, but disregard the deterministic influence of the technological artefact itself. However, as the *flesh* shows, there is a hermeneutic mode of being with technology that should be emphasized – again, a crossing, a chiasma, an intertwining of sense and world. The *flesh* represents for Merleau-Ponty the genesis of sensibility, for "he who sees cannot possess the visible unless he is possessed by it, unless he *is* of *it* ..." (1968, pp. 134-135): just as there is encroachment between the two poles of the 'dualisms' that Merleau-Ponty identifies, so the world encroaches upon us and alters us – still, while we are *of* the world, we are paradoxically not *the* world (1968, p. 127).[6]

The mode in which technology mediates the body both expands and constrains the body-subject through the intertwining of body and the technological artefact. Thus, while one encounters the digital technology artefact through a bodily sensitivity and sensibility, through a sensory receptivity and a spontaneity, it should also be considered that the technological artefact can constrain and delimit the ways in which these aspects of the body schema may find expression in the virtual. In this paper we highlight how the hermeneutic mode, considered in relation to the virtual, reveals the manner whereby technological artefacts may exert a deterministic influence of concurrent expansion and constraint on the body-subject – particularly through its influence on individual behaviour.

Merleau-Ponty rejects purely mechanical, physiological, behaviouristic, and reflex-arc explanations of behaviour as restrictive and reductionist from *The Structure of Behaviour* onwards (Corriveau, 1972, p. 19). Behaviour is not, for Merleau-Ponty, merely the sum of its parts but relates to a certain milieu characteristic of the human being, a specific 'being-in-the-world' that is situational (Merleau-Ponty, 1963, p. 127). Both *Phenomenology of Perception* and *The Visible and the Invisible* expand on the intertwined relation between body-subject and world to suggest that behaviour relates to sensibility – a giving sense and a making sense of the world that precedes thought, representation, and formal symbolic activity and which takes place through intentionality to allow the body-subject to achieve a maximal perceptual grip on the world (Corriveau, 1972, pp. 22-23).[7]

5 See Irwin, 2014; Bailey, 2016; Hoel & Carusi, 2018; Ward, 2018; Du Toit, 2020; Du Toit, 2021. Other descriptions of the virtual, such as Pierre Levy's, are similarly suggestive of such expanded horizons for action and experience. Levy suggests that the virtual "is a fecund and powerful mode of being that expands the process of creation, opens up the future, injects a core of meaning beneath the platitude of immediate physical presence" (1998, p. 16).

6 The idea that the world is not merely an object "does not mean that there was a fusion or coinciding of me with it: on the contrary, this occurs because a sort of dehiscence opens my body in two, and because between my body looked at and my body looking, my body touched and my body touching, there is overlapping or encroachment, so that we may say that the things pass into us, as well as we into the things" (Merleau-Ponty, 1968, p. 123).

7 Behaviour thus transcends the merely physiological aspects of the body, while remaining bound within the limitations of the body.

Merleau-Ponty's description of behaviour (particularly as habitually learnt) links to Skinner's concept of operant conditioning, but we carefully note that there exists a fundamental incompatibility in terms of these two thinkers. However, both views find agreement in that they are based in the relational character of behaviour that does not presuppose strict linearity (Pompermaier, 2018). Learning, for Merleau-Ponty, occurs through the process of trying to achieve maximal grip on the world through intentional actions in embodied and socially contextualized situations (Jing & Ejgil, 2017). Learning as a being-in-the-world relates to sense-making of the structures of the world – it is the process of increasing grip (sensibility) on the world, whether in *actuality* or in *appearance*.

The idea of achieving *apparent* maximal grip on the world is important for analysing engagement with the virtual because the process of achieving maximal grip in the virtual differs from that used to achieve maximal grip within the spatiality of the material world. Rather, the structuring of digital technologies in line with rewards (through captological programming and Pavlovian mechanisms, for instance) convinces the individual of having achieved maximal grip on the virtual – certain ways of behaving in the virtual lead to a sense of reward, which in turn suggests that one has made sense of the virtual (achieved maximal grip). An example of such a technological positing of having achieved maximal grip is the captological technique of designing software applications in such a way as to continually notify users of content (such as recurrent Facebook notifications). These notifications might be a reminder of a birthday, an indication of someone having liked one's post, updates on public forums, or a direct message from a friend. The variance in meaning of each of these notifications allows the device to take on the characteristics of a slot-machine – one is never sure of what one is being notified, whether the notification is indicative of a public message on a topic one is interested in (small win), or a direct message from a friend (jackpot). In either case, once a specific behaviour has been enacted (the phone unlocked, the app opened) the potentiality of a notification is replaced with a certainty.

While the concept of *flesh* illuminates the hermeneutic mode of human-technology relations and resolves ontologically the determinism/instrumentalism dualism, the concordant embodied account of behaviour being reinforced through the mechanistic and predictable functioning of the digital technology artefact underlines the technologically deterministic capacity of the virtual to structure human behaviour – both fields of action and modes of perception are deterministically transformed and distorted. In the next section we expand our discussion of individual experience of the virtual towards a description of the social conditioning wrought by the virtual on a larger scale to show how values, systems of rationality, and thought are conditioned and limited in concrete ways.

5. The Uniformity of the Virtual Bodily Schema

In the philosophy of technology, analysis of the value-ladenness of technology tends to focus on the role that values play in technology design, or the ways that the operation of a technology enforces or restricts certain forms of behaviour. The classic example of an inherently political technology is the Moses bridge where a particular technological design was chosen to restrict access to the beaches by certain racial/socioeconomic groups (Winner, 1986, p. 21).[8] Analyses of virtual technologies rarely explore such 'deterministic' issues and

8 Winner describes how the low-hanging overpasses in Long Island were designed by architect Robert Moses to realize a specific social function. The specifications of these low-hanging overpasses deliberately prevented passenger buses (which typically served African-Americans and low income individuals) from passing underneath, while expensive automobiles (such as those owned by the wealthy) could traverse the roads with ease. Evidence from the period suggests that these bridges gave manifestation to Moses's racial bias and social-class discrimination (Winner, 1986, pp. 123-124).

tend instead to explore virtuality's expansion of experiential possibilities.[9] One reason for this neglect might be the apparent immateriality of the virtual. Instances of constraining technologies like the Moses bridge are decidedly material, in the sense of being tangible, and operate deterministically in an equally tangible way, by physically preventing the passage of certain types of vehicle along the bridge. The ostensible immateriality of the virtual appears to render it ontologically different from older technological forms and thereby immune to deterministic analyses that seem rooted in the concrete.

Merleau-Ponty's later phenomenology serves as a corrective to such dualistic treatments of the technological through the application of his ontology of the *flesh*, a matrix of intertwined and reciprocal relations that serves as the foundation of the body's relational engagement with its environs, to the virtual. A Merleau-Pontian analysis of virtual technologies grounds our experience of the virtual world in the body and connective lattice of the *flesh*. It unifies the virtual and the material, for the experience of the virtual arises in the space between the embodied subject and digital technology. The virtual and the 'real' are the same in that they are both experienced by an embodied subject who perceives and navigates within them through the employment of bodily frames of reference. Consequently, questions concerning technology's ability to constrain or limit asked of older, material technologies can also be asked of 'immaterial' virtual technologies. The individual's experience of technology is as embodied in the 'real' world as it is in the virtual. If the virtual is to be described as concrete in this embodied manner, then the virtual is as concrete as the material (i.e. no dichotomy exists) and the concepts utilized in philosophy of technology (such as technological determinism) are equally applicable.

Phenomenological analyses of digital technologies that employ Merleau-Ponty's work (e.g. Irwin, 2014; Bailey, 2016; Hoel & Carusi, 2018; Ward, 2018; Du Toit, 2020; Du Toit, 2021) tend to focus on the ways in which the virtual extends perceptual capacities. However, we argue that in order to remain true to Merleau-Ponty's principle of reversibility one is committed to considering the technologically-mediated experience of the virtual world both ways. Not just with regards to the experiences of the embodied subject 'touching' the virtual, but also with regards to the ways in which the virtual 'touches' the embodied subject. In considering the virtual as enmeshed in the *flesh*, we recognise an interfusing of specific properties that relate to one's embodied encounter with the world and the virtual.

The virtual world, understood as a 'concrete' space, is one that is layered with values. The subject embodied in the virtual finds their range of possibilities thoroughly circumscribed, sometimes by physical limitations of the technology and sometimes by constraints designed by those who create and maintain the technology. Online behaviour is as norm-governed as offline behaviour.[10] And for all that the virtual can radically augment human perceptual and expressive capacities, the procedures for such augmentation are rule-governed. Navigating the virtual requires an implicit grasp of technological rationality. The virtual is "a distributed system where bodies, symbolic systems, technologies and environments are intertwined and in which phenomena are articulated in characteristic ways, according to particular styles" (Hoel & Carusi, 2018, p. 62). And praxis in the virtual world is as formative of patterns of thought and behaviour as praxis in the material world. The analysis of measurement and the body can flesh out an account of the deterministic effects of the virtual on the embodied subject.

Paul Virilio sketches the phenomenological dimensions of the transferal of standards of measurement from the virtual to the material world. Analysing the shift in spatial and

9 Cf. Pierre Levy, *Becoming Virtual* (1998).

10 Floridi (2015, p. 264) argues that in the post-digital era no separation exists between the two dimensions.

temporal experience that occurred with the rise of what he terms "computer-assisted representation", Virilio stresses the role of measurement in founding our ability to experience the world (1991, p. 51). Being embodied in the virtual, Virilio suggests, means acquiring a system of orientation radically different from that of material spatio-temporality, one that obliterates distance and positionality. He further argues that the spatio-temporal orientation acquired in the virtual is then transported into the non-virtual, facilitated and reinforced by the increasing incorporation of virtual technologies into work and leisure activities.

On Virilio's account the virtual and non-virtual form a continuum with no clear boundary. As more forms of praxis are mediated by virtual technologies, the application to the material world of norms and orientations derived from the experience of virtual embodiment proliferates. And the result is a "delirium of interpretation due to the excess of mediation of experiences", a state in which it becomes increasingly common to experience social reality within the interpretative framework of the virtual (Virilio, 1991, p. 48). Thus, praxis becomes increasingly technologically-mediated even when it does not directly involve the use of technology.[11] Returning to Merleau-Ponty, we further suggest that alterations in the experiential framework by which we measure and engage the world, virtual and material, in turn produce alterations in the self-understanding and capacities of the embodied subject, which itself is the ground of all measurement.

As Hoel and Carusi suggest, regarding Merleau-Ponty's thought in *The Visible and the Invisible* (1968) and *Nature* (2003), the body is considered the "standard or measure of things" whereby "'measurement' [becomes] an ontological concept that concerns the inner scaffolding of the existential field, the 'invisible armature' of the perceived" (2018, p. 48). The body is intertwined and complicit in phenomena, which is of particular relevance to any analysis of the virtual that explores the expansion or constraint of body schemas via technological mediation.[12] The measuring body emphasizes "the mediated nature of knowledge and being, by more radically integrating mediating artefacts into the perceptual/conceptual complex" (Hoel & Carusi, 2018, p. 49). Hereby a relative autonomy is intuited with regards to symbols and tools, which through their operation "serve to decentre and displace the interrogating capacities of the perceiving body in productive ways" through the expansive dynamic of the *flesh* (Hoel & Carusi, 2018, p. 49).

The body as measurement (i.e. the perceiving body reconfigured into a symbolically and technologically disturbed and mediated measuring body) highlights how both symbols and tools serve as 'measures of being' (Merleau-Ponty, 1973, p. 124). Merleau-Ponty argues "my body is not only one perceived among others, it is the measurant (*mesurant*) of all, *Nullpunkt* of all the dimensions of the world" (1968, pp. 248–9). It is against this concept of the measuring body that the relation of the body-subject to the world makes sense, and whereby behaviour becomes understandable as meaningful in a larger unified whole (rather than as momentary and fleeting). Movement and behaviour, through the world and through the virtual, are elucidated through the *flesh* wherewith the measuring body is known – and due to the delimiting and structuring effects of the virtual, behaviour is structured according to the confines of the digital technology artefact's functioning.

In this regard it is sensible to, in a manner of speaking, turn Hoel and Carusi's postulation of the measuring body around to trace not the influence of the body-subject on the world, but

11 Cf. Mark B.N. Hansen, particularly *Bodies in code* (2012) and *Feed-forward* (2015).

12 Hoel and Carusi describe the measuring body as a "new conceptual tool" that "neither privileges nor coincides with sensory perception. It acknowledges that technoscientific interrogations of the world involve distributed and displaced agencies of observation that engage in a two-way formative exchange between observer and observed" (2018, pp. 47-48).

rather to trace the measuring effect of the technological artefact. The technological artefact, in delimiting and funnelling being, functions in a technologically deterministic manner.
In other words, the measuring artefact, rather than merely broadening the body-subject's horizons of engagement by means of the virtual, generates a uniformity of body schemas by structuring experience and behaviour along predictable and linear lines of influence – the digital technology artefact functions in a manner similar to technological artefacts in a broad sense (as suggested by technological determinism). This means that engagement with the virtual happens in a consistent manner because the specific digital artefact can only ever function in a similar manner regardless of the wide variety of individuals across a society, or between societies, that make use of it – thus, such devices structure human perception and behaviour in an expected and predictable manner. Often, the structuring effects of the measuring artefact entail capitalistic, advertising and consumerist agendas – and it is these criteria that are emphasized, rather than the expansive and creative potentialities inherent in said technology.
Merleau-Ponty's phenomenology makes allowance for a plurality of lived world(s) and suggests that these worlds are not monadic (1968, p. 48, 62, 84, 269; 2002, p. 210, 214). The body of the body-subject is where sense is made of space and things, but the body is also already interrelated to other body schemas by means of the *flesh* as an interworld. The *flesh* of the virtual, the point at which the body-subject and the digital technology artefact intermesh, makes this interworld concretized and linked; lived bodies are unfurled unto each other by means of the structure of technological mediation (which allows expansion of the body, while also curtailing and delimiting this unfurling along capitalistic lines). From the individual's behaviour one may therefore explicate a collective element to behaviour across society – for "as soon as something is taken as expressive there is a carnal communicability to it, and with that an 'opening to generality'" (Hoel & Carusi, 2018, p. 56). Technological systems are thus "standards or measures of being, which have the capacity to transform the metaphysical structure of the interrogating apparatus and hence to displace the horizons of the perceptible/intelligible world" (Hoel & Carusi, 2018, p. 62). The virtual, in concretizing intersubjective relations, serves to structure and delimit (whilst expanding and opening up) human perception and behaviour according to the strict criteria of the 'measuring artefact'.
Furthermore, the rethinking of the virtual presented here allows one to explicate the formative and transformative capabilities of technological tools, symbols, and cultural forms of expression by means of the virtual (at the individual and broader cultural level). Bodily schemas that are contorted by the virtual are linked to wider intellectual/cultural patterns to such an extent that the technological flattening of schemas is likely to have an impact at a cultural level too: Shared symbols in the virtual lead to a uniformity of bodily schemas and suppresses (but never extinguishes) the body's capacity to displace established horizons of meaning.

6. Conclusion

An opening up of Merleau-Ponty's early instrumentalist account of technology to the broader experience of technological praxis by means of the *flesh* reveals that both embodied and hermeneutic relations are concurrently found in the individual's encounter with technological artefacts. Such a recognition allows one to phenomenologically validate the experiential insight into different aspects of human-machine relations that underlies both instrumentalist and technological determinist positions, whilst resolving their apparent antagonism by showing both positions to be correct but partial descriptions of technologically-mediated experience.
We also argued that a Merleau-Pontian account of virtual embodiment opens up the possibility of developing a novel form of technological determinism. This new determinism,

like the classic form of technological determinism, is grounded in the bodily experience of the technologically-mediated subject. It, however, does not reinforce the instrumentalist-determinist dichotomy that Merleau-Ponty's ontology of the *flesh* was used to dissolve, but rather uses this ontology as the foundation of its account which focuses on the relations between the Merleau-Pontian body-subject and digital technology. The elucidation of such a technological determinism reveals a capacity of certain technologies to constrain or restrict the experiential and behavioural potentialities found in the technologically-mediated existential field. To this end we have reconsidered the relation of the body-subject and the virtual, illustrating the ways in which digital technology artefacts may at once open up new horizons of experience while also concretely determining values, systems of rationality, and thought.

REFERENCES

Achterhuis, H. (2001). *American Philosophy of Technology : The Empirical Turn*. Bloomington: Indiana University Press;
Bailey, J. (2016). The body in cyberspace: Lanier, Merleau-Ponty, and the norms of embodiment. *Christian Scholar's Review*, 45(3), p. 211;
Baudrillard, J. (1981). *Simulacra and simulation*. Michigan: University of Michigan press;
Brey, P. (2010). Philosophy of Technology after the Empirical Turn. *Techné: Research in Philosophy and Technology* 14(1), 36–48. https://doi.org/10.5840/techne20101416;
Corriveau, M. (1972). Phenomenology, psychology, and radical behaviorism: Skinner and Merleau-Ponty on behavior. *Journal of Phenomenological Psychology*, 3(1), 7-34. DOI: 10.1163/156916272X00029;
Du Toit, J. (2020). The (oh-so-queerly-embodied) virtual. *South African Journal of Philosophy* 39(4), 398-410. DOI: 10.1080/02580136.2020.1846106;
Du Toit, J. (2021). Living in the age of the embodied screen. *Indo-Pacific Journal of Phenomenology*, DOI: 10.1080/20797222.2021.1876895;
Du Toit, J. & Swer, G.M. (2020). A Manifesto for Messy Philosophy of Technology: The Past and Future of an Academic Field. *Theory of Science*, 42(2), 231-252. doi.org/10.46938/tv.2020.491;
De Saint Aubert, E. (2008). *Maurice Merleau-Ponty*. Paris: Hermann;
Floridi, L. (2015). *The Onlife manifesto: Being human in a hyperconnected era*. Basingstoke: Springer Nature;
Hansen, M.B. (2012). *Bodies in code: Interfaces with digital media*. London: Routledge;
Hansen, M.B. (2015). *Feed-forward: On the future of twenty-first-century media*. Chicago: University of Chicago Press;
Hoel, A.S. & Carusi, A. (2015). Thinking technology with Merleau-Ponty. In Rosenberger, R. & Verbeek, P.P. (Eds.), *Postphenomenological investigations: Essays on human-technology relations*. Maryland: Lexington Books, pp. 73-84;
Hoel, A.S. & Carusi, A. (2018). Merleau-Ponty and the measuring body. *Theory, Culture & Society*, 35(1), 45-70. DOI: 10.1177/0263276416688542;
Ihde, D. (1979) *Technics and praxis: A philosophy of technology*. Dordrecht: D. Reidel;
Ihde, D., & E. Selinger. (2004). Merleau-Ponty and epistemology engines. *Human Studies*, 27, 361–376. DOI: 10.1007/s10746-004-3342-4;
Irwin, S.O. (2014). Embodied Being: Examining Tool Use in Digital Storytelling. *Tamara Journal for Critical Organization Inquiry*. Vol 12. (2);
Jing, H.E. & Ejgil, J. (2017). Habitual learning as being-in-the-world: On Merleau-Ponty and the experience of learning. *Frontiers of Philosophy in China*, 12(2), 306-321. DOI: 10.3868/s030-006-017-0022-9;
Levy, P. (1998). *Becoming Virtual*. New York: Plenum Trade;

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge and Kegan Paul. Reprint, Routledge, 2002;
Merleau-Ponty, M. & Fisher, A.L. (1963). *The structure of behavior*. Boston: Beacon Press;
Merleau-Ponty, M. (1964). *Signs*. Illinois: Northwestern University Press;
Merleau-Ponty, M. (1973). *The prose of the world*. Illinois: Northwestern University Press;
Merleau-Ponty, M. (1968). *The visible and the invisible: Followed by working notes*. Illinois: Northwestern University Press;
Merleau-Ponty, M. (2003). *Nature: Course notes from the Collège de France*. Illinois: Northwestern University Press;
Peters, John Durham (2017). "You Mean My Whole Fallacy Is Wrong": On Technological Determinism. *Representations* 140, 10-26;
Pompermaier, H.M. (2018). Skinner and Merleau-Ponty's criticism of causality. *Trends in Psychology*, 26(1), 111-123. DOI: 10.9788/tp2018.1-05pt;
Smith, M.R. (1994). Technological determinism in American culture, in Smith, M.R., Marx, L. (ed.), *Does Technology Drive History?: The Dilemma of Technological Determinism*. Cambridge, Massachusetts: MIT Press, 1–36;
Swer, G.M. (2014). Determining Technology: Myopia and Dystopia. *South African Journal of Philosophy*, 33(2), 201-210. doi.org/10.1080/02580136.2014.923696;
Swer, G.M. (forthcoming). A Surfeit of Sociophobia?: Digital Society and Technological Determinism. *Online Civilization*;
Van Den Eede, Y., Goeminne, G., and Van den Bossche, M. (2017). The Art of Living with Technology: Turning Over Philosophy of Technology's Empirical Turn. *Foundations of Science*, 22(2), 235-246;
Virilio, P. (1991). *The Lost Dimension*. Brooklyn, New York: Semiotext(e);
Ward, D. (2018). What's Lacking in Online Learning? Dreyfus, Merleau-Ponty and Bodily Affective Understanding. *Journal of Philosophy of Education*, 52(3), 428-450;
Winner, L. (1986). *The Whale and the Reactor*. Chicago: University of Chicago Press.

VERBENA GIAMBASTIANI
Independent Researcher
verbena.giambastiani@gmail.com

THE ASYMMETRICAL RELATIONSHIP BETWEEN HUMANS AND TECHNOLOGIES

abstract

The aim of my proposal is to address the following question: "How the experience of interacting with a specific technology mediate our experience of the world?". I will do this by exploring the ideas of the postphenomenological theorists: Don Ihde, Peter Paul Verbeek and Evan Selinger. Postphenomenological studies tend to focus on understanding the roles that technologies play in the relations between humans and world.
I would examine the idea that technologies mediate the world in such a way that perception of the self, world and environment changes. I am going to deepen the phubbing phenomenon, because it reveals the effect of technologies on social interaction.

keywords

Postphenomenology, Phubbing, technosphere, quasi-otherness, placeless

Phenomenology and Mind, n. 20 - 2021, pp. 32-40
DOI: https://doi.org/10.17454/pam-2003
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

The central question of my analysis is: "How the experience of interacting with a specific technology mediate our experience of the world?". Studies of human-technology interaction have traditionally positioned the human and the technological as opposites, or at least as ontologically distinct. Rather than enabling people to achieve their own intentions, technologies are sometimes interpreted as an oppositions against human realization. In this technophobic view, technology's development is centrally responsible for the inhuman oppression that characterized our society. One version of this approach is articulated by Andrew Feenberg. Andrew Feenberg is a political philosopher who investigates the relationship between neo- Marxist tradition and phenomenological approaches to technology. His reflection on power is inspired by the late Foucault's thoughts. He affirms that political power and the cultural appropriation of technology are closely connected. According to Feenberg, the current technological milieu of our society is profoundly problematic. He states that industrial society "orients technological development toward disempowering workers and the massification of the public" (Feenberg 2005, p. 53). Due to these circumstances, the autonomy of the owner or his representative "reproduce the conditions of their own supremacy at each iteration of the technologies they command" (ibid.). This situation ultimately results in technocracy, when "technology and management spread to every sector of social life" (Feenberg 2005, p. 55).

Another version of this technophobic approach is the one sustained by Neil Postman. Postman, in his book *Technopoly: the surrender of culture to technology* (Postman 1993), argues that human culture can be divided in three periods with respect to technology: the tool-using, the technocratic and the technopolistic. In the last period, the technopoly, the logic of industrial production will control not just economic thought, as it was in the technocratical time, but cultural and philosophical thought as well.

However, these points of view are problematic because human beings cannot be understood in isolation from technology, just as technology cannot be understood in isolation from humanity. Technology helps to shape what it means to be human, they are fused in a manner which makes it impossible to understand one without reference to the other. This does not mean, of course, that all kind of involvement of technology in human life are equally pleasant. Instead, it does mean that the paradigm of oppression and resistance might not be the most appropriate paradigm if we want to investigate the relationship between human beings and technologies.

The emerging technologies reshape the relationship between human beings and the world (De Preester 2010). Don Ihde refers to our technological environment as a "technosphere" in which

we find ourselves and which involves all dimensions of our relations: "This environment has become so comfortable to us that we can hardly imagine our lives outside these cocoons" (Ihde 1973, pp. 13-14). The term ecosystem should be replaced with *technosystem* because it more correctly describes the texturing "cocoon" in which most of us now live.
We are nowadays citizens of two spaces, the biosphere and the infosphere. Infosphere is a neologism made using the words information and sphere, coined by Sheppard in 1971, and employed by Luciano Floridi to denote the whole informational environment constituted by all informational entities (thus including informational agents as well), their properties, interactions, processes and mutual relations. In this sense the natural world is being absorbed into the virtual in a great ontological twist and reversal. Rather than the virtual becoming actual, the actual is becoming virtualized.
Floridi maintains that we are living through an informational turn (Floridi 2014) or fourth revolution, following the scientific revolutions of Copernicus, Darwin and Freud, in which the unification of technology and human beings is defined *inforgs*: informational embodied organisms, embedded in an informational environment. In the infosphere, boundaries between online and offline environments merge, so that we live in a manner termed *onlife* (Floridi 2015). To put it dramatically, the infosphere is progressively absorbing any other space. In the onlife world, artifacts have ceased to be mere machines simply operating according to human instructions. They can change states in autonomous ways and can do so by digging into the exponentially growing wealth of data, that currently are increasingly available, accessible and processable (Bibri 2015).
In our time, with the advent of new technological innovations, the rate of information production has rapidly accelerated. The emergence of social media has made this a global phenomenon where hundreds of millions of people are publishing information on a wide range of social media. It seems very difficult to make decision and to act in this kind of situation. Then, it is very familiar the situation of Information Overload, or infobesity or infoxication. We are not good in dealing with all the data and the info that are communicated to us. We have a paradoxical problem of agency in the face of the changes: we at the same time create the problem and are unable when it comes to solving it.

2. The reciprocal human–technology interrelationship

Technologies mediate the world in such a way that the perception of self, world and environment changes. While technology transforms our experience of our lifeworld, it simultaneously reveals the world in a transformed manner. Therefore, it is necessary to consider human-technology relationships in its complexity and mutual correspondence or constitution. As Thompson (Thompson 2006, p. 116) remarks, technologies mediate both perception (through amplification and reduction) and interpretation (through concealing and revealing). Technologies necessarily provide agents with value-laden access to the world, and, more important, remake the agents and the world in the process. Postphenomenology describes the relation between human beings and world as being changed or at least influenced by technologies and technological artifacts. The term "postphenomenology" is Don Ihde's (1990; 2009) variant of Husserlian phenomenology purged of its transcendental and essentialist tendencies and injected with American pragmatism. Don Ihde recognizes the legacy of phenomenology but takes distance from its perspective. Ihde defines postphenomenology as "a nonfoundational and nontranscendental phenomenology which makes variational theory its most important methodological strategy" (Ihde 2015, p.vii). It is a "particular mode of science-technology interpretation" and that "its arrival coincides with a late-twentieth- to twenty-first-century radical shift in science-technology analysis" (Ihde 2015, p.viii).
Postphenomenology places its philosophical and cultural context in the continental philosophy developed at the beginning of the 20th century, more precisely in the phenomenological

movement. Don Ihde (1990) has called the special relations between human beings and such artifacts with the espressions "embodiment relations", since such artifacts seem to become part of our embodiment. This perspective is influenced by Merleau-Ponty. Already in 1990, Don Ihde argued that in Merleau-Ponty's thought, and especially his phenomenology of the perceiving body, there is an implicit "latent phenomenology of instrumentation" (Ihde 1990, p. 40), specifically, a theory of the process by which the body runs both a technical exteriorization of its functions and an incorporation of technical tools. In Ihde's words, in an embodiment relation "I take the technologies into my experiencing in a particular way by way of perceiving through such technologies and through the reflexive transformation of my perceptual and body sense" (Ihde 1990, p. 72). For Ihde artifacts mediate perception: the experience of one's environment through an artifact. These intentional human-technology relations transforms our perceptions of the world and ourselves. Hence, embodied artifacts change our relation to the world, sometimes even profoundly. The microscope, for example, has forever changed the way we understand our world. These artifacts are not normally perceived as objects in one's environment, but instead are used as means through which the environment is experienced and acted on.

Postphenomenology is influenced by some of the central themes of American pragmatism: anti-foundationalism, mediation, materiality, concreteness, practice, and the enlargement of meanings. Inspired by the rise of science and technology studies (STS) in the 1980s and 1990s, Peter-Paul Verbeek (Verbeek, 2005, 2011) has embraced what Achterhuis (Achterhuis 2001) calls the "empirical turn". His characterization of postphenomenology emphasizes its bottom-up empirical turn away from top-down "classical" approaches to philosophy of technology that tended to privilege abstract theoretical concerns over serious consideration of artifacts. Postphenomenology adapts several aspects of pragmatism. Don Ihde and Peter-Paul Verbeek (Ihde 1990; Verbeek 2003) have criticized classical phenomenology for being too subject-centered and have argued that it overlooks the mediating role that technology plays in human-world interactions. Postphenomenology shifts the constitutive element of subject and object to the middle: neither subject nor object are primary or self-sufficient. Practitioners of postphenomenology address issues such as "scientific and medical imaging, computer interface, virtual reality, traffic safety, robotics, educational technologies, sustainable design, wearable computing, and bodily implants" (Rosenberger, Verbeek, 2015, p. 2).

Postphenomenology remains experiential, but not *subjectivistic*. This inter-relation ontology is well-matched with an experiential version of the interaction between a living organism and its environment (Dewey's notion), as a human experiential interaction with a lifeworld. The postphenomenology perspective conceptualizes the role of the emerging technological environments in human existence and it remarks the transformative dimension of technical artifacts on human experience (Borgmann 1984).

Don Ihde, Evan Selinger, and Peter-Paul Verbeek effort is revolving around the concepts of technical mediation and relational ontology. For Verbeek (Verbeek, 2005) the mediating role of things takes place in this middle and relates to the very ontological status of subjectivity and objectivity:

> Things are not neutral intermediaries between humans and world, but mediators: they actively mediate this relation. (...) Mediation does not simply take place between a subject and an object, but rather co-shapes subjectivity and objectivity. [...] Humans and the world they experience are the products of technological mediation. (Verbeek 2005, p. 114)

According to Verbeek (Verbeek 2011) proper engagement involves neither rejecting technology nor embracing it enthusiastically but asking which technologically mediated

subjects we want to be, as a preference, a question of value, about the good life. This is a method not oriented to humans or to technology alone, but to "human–technology associations".

The postphenomenological approach is grounded on an empirical turn that culminates in the concept of the non-neutrality of technologies. Postphenomenology investigates the complexity of a technologically environment and its impacts on the body, self/identity and environment itself. The "penetration" of lived experience by technology is a common denominator in the ontologies of time, mobile places, and fluid spaces and the intensifying collapse of time and space. Technologies, therefore, mediate any authenticity or naturalness alleged of the human body, its experience and its various "relational" manifestations with others. Technologies simultaneously form the lifeworld and mediate the embodied (and discursive) human experiences of that world (Gouch, 1999; Rickinson, 2001).

Rather than starting with a subject and object correlate, postphenomenologists start with the relations themselves. There are various ways in which technologies help to shape relations between human beings and the world: embodiment relations, hermeneutic relations, alterity relations, and background relations (Rosenberger, Verbeek, 2015, p. 13). Embodiment relations are those in which a user's experience is reshaped through a device. Peter-Paul Verbeek introduced cyborg relations as a radical variant of embodied relations in which "technologies actually merge with the human body" (Verbeek 2011, p. 144). Hermeneutic relations, such as looking at and interpreting a wristwatch, are those in which "the user experiences a transformed encounter with the world via the direct experience and interpretation of the technology itself" (Rosenberger, Verbeek, 2015, p. 17). According to Don Ihde, hermeneutic relations are a specific kind of technologically mediated I-world relations in which the technology must be "read" and "interpreted" to access the world (Ihde, 1990). Alterity relations are those in which we relate to technical devices in ways that are like to the ways in which we relate to other humans. iPhone users, for example, tend to relate to Siri as a quasi-other. Finally, background relations with appliances such as refrigerators and air conditioners make up our environmental context.

Postphenomenology's approach is crucial because it holds on to an asymmetrical approach to humans and nonhumans. According to Don Ihde, the background of asymmetrical relation between human beings and technological systems or artifacts is based on intentionality. Ihde has contributed to this debate by declaring how humans maintain an exclusive and nontechnically reproducible stance in the world. He maintains that even when humans and technology co-construct one another, such reciprocity always contains an ineliminable asymmetrical dimension. Specifically, he insists, only human agency has intentionality. However, this does not mean that it would imply an *a priori* separation of subject and object. It sees subjectivities and objectivities as the outcomes of processes of mediation. Without separating subjects and objects, it keeps up a distinction between them. As Ihde (1979, p. 4) states: "human-machine relations are existential relations in which our fate and destiny are implicated, but which are subject to the very ambiguity found in all existential relations."

If it is true that our use of technologies has changed the way we interact with the world and each other, then it would be interesting to closely examine the argument explained. The *phubbing* phenomenon is a test case that shows how the feature of these connections remains ambiguous.

3. The "Phubbing" Phenomenon

Jesper Aagaard presents a postphenomenology study on digital mediation and attention. Aagaard (2015) borrowed the term *wall-window* from Wellner (2011). He wants to demonstrate how digital devices *open* their user up to a nonlocal level of social interaction while at the same time producing obstacles to the local level of potential exchanges. The *wall-window* metaphor suggests that digital technologies create an imaginary *wall-window* between you and person/object you

engage with, dislocating your attention to a virtual space, while still allowing you to be bodily engaged in the physical space. Aagaard suggests that the screen wors as a portal through which the users enter in another realm. Thus, the digital mediation makes one both absent and present at the same time. This is the same situation of the phenomenon known as *phubbing*.

Phubbing is a combination of the words phone and snubbing. To be phubbed is "to be snubbed by someone using their mobile when in your company" (Roberts and David 2016, p. 134). Phubbing is a state of altered relations and it is the practice of ignoring one's companion or companions in order to pay attention to one's phone or another mobile device. Phubbing phenomenon addresses how we communicate and relate with others. It shows a paradox: we are physically present but we are absent, occupied with our smartphones. Phubbing is the state of being present absent. Present absence can be observed when people are physically present but effectively absent because their focus is set somewhere else (MacCormick, Dery, & Kolb 2012). Meredith David and James Roberts (2017) demonstrate the harmful effects of phubbing.

Being phubbed is an experience of being ignored. It is a threat to human interaction because it determines social exclusion (Lee, Jaehoon, and L. J. Shrum 2012). It appears that the use of mobiles is contagious, because the phubbed individuals turn to social media to gain attention and to restore his/her sense of inclusion.

It is helpful to compare phubbing with the concept of idle talk developed by Martin Heidegger. Heidegger observes that idle talk is the possibility of understanding everything without previously making the thing one's own. It forces *Dasein* within its inauthentic existence. In a similar way, phubbing discourages any new inquiry and any disputation, it suppresses them and holds them back. Both idle talk and phubbing share the characteristics of superficiality, gossiping, wide publicity.

Nevertheless, what is even more peculiar of phubbing is the fact that the screen has replaced the human face. New technologies are interactive as they maintain a certain dialogue with the human being. Don Ihde (1990) defines new technologies as ''quasi-otherness''. He explains (1990, p. 100) that technological otherness is a quasi-otherness, stronger than mere objectness but weaker than the otherness found within the animal kingdom or the human one. The construct of quasi-other is located mid-way between inanimate objects and living beings. Quasi-otherness is that quality in objects which positions the human and the object in a human-like relation. If technology is a quasi-other, the screen is a quasi-face (Wellner 2014). The screen is a subset of a face. It shares a few elements with the human face.

Following Levinas, if interaction with others is based on face, what kind of interaction phubbing phenomenon aims to? In Levinas' thought, things, technologies included, are faceless. The face is uniquely reserved for human beings. Moreover, the face demands an ethical response, the screen not. Screen-face serves to display certain human functions, henceforth it is not designed to be ethical (Verbeek 2009). Turkle (2012) argues that because people in interpersonal social situations, particularly young people, are distracted by their phones, they will pay insufficient attention to one another, creating increasingly shallow relationships. It seems that people have become so accustomed to technologies that it is a new way of being alone together. We have lost the skills of conversation and replaced it with connection through text and social media.

Social media connections, as in the phubbing phenomenon, lack communication. According to Jean Baudrillard (1981), we can affirm that social media creates a world of simulation, because when we post a text online, we generate what Baudrillard named a *simulacrum* of conversation, a *simulacrum* in evidence. In this world of simulation, you can be *other* than you are in the physical world because the signifiers of your identity— avatars, profile picture, cover picture, etc.—do not necessarily match your physicality. Cyberspace, for Baudrillard, is a realm of shining surfaces where the meanings float and can be joined to virtually any signified.

Consequently, meanings ultimately mean nothing. Simulation is production—the production of increasingly self-referential sign systems that are reality or, to use Baudrillard's term, the real is hyperreal. Hyperreal denotes more than real, it is a special kind of social reality in which reality is created and simulated on the surface. In the hyperreality, the difference between reality and imaginary collapses. How to understand this situation and how to distinguish between the virtual experiences and the physical one is a crucial question to pose. The ambiguity between these options has the potential to create a placeless location, more real than physical world itself.

In 1976, the humanistic geographer Edward Relph wrote *Place and Placelessness*. In this book, he argued that places were becoming placeless. He gives many reasons for this – mass production or an increasingly mobile world. One of the supposed major causes of placelessness is increased mobility. This mobility has led some to go as far as to declare the "end of geography". The vanishing of a sense of place is interlaced with the lack of values of identity, relations and history. Therefore, it is not just that places have become standardized, but, likely, our relationship to the environment and to the others become distanced and disconnected. In a dystopic perspective, Leroi-Gourhan figures out a human race living in a prone position. Their only ability is to push buttons (Leroi-Gourham 1964).

Marc Augé (2008) formulates the neologism "non-place" to address to spaces with a similar purpose (transport, commerce, leisure), and to characterize the relations between individuals and these spaces. Non-places are non-relational and unhistorical. They create an anonymous and solitary identity: the customer. According to Augé, the concept of non-space is part of the category of Supermodernity. Supermodernity is characterized by excess, a charged surplus in the three domains of philosophical and anthropological cornerstone: time, space, the individual. One of the characteristics of excess with regard to Supermodernity concerns space. It produces non-places: spaces which are not themselves anthropological places and which do not integrate the earlier places.

Mankind has a world, while animal has an environment. Man, as opposed to other animals, has no specific habitat, but a specific perception of space. Man is bound by nature to shape his own *oikos*: he is world-forming, a technological being. In the near future, environments will be probably of interconnected sensor networks of sensors, managed using artificial intelligence and analyzed by big data-enabled automation systems. Smart Environments promise a hyper-personalized kind of environment. However, the communication among people has reached an unprecedented intensity and scale, but it has inaugurated a kind of loneliness not known before, in which the collective, historical community rooted in the traditional heritage disappears. It is necessary to rediscover the importance of place for both individuals and cultures, especially because contemporary society was seeing a surge in placelessness.

REFERENCES

Aagaard, J. (2015). Drawn to distraction: A qualitative study of off-task use of educational technology. *Computers & Education*, 87, 90–97. https://doi.org/10.1016/j.compedu.2015.03.010;

Augé, M. (2008), *Non-Places: An Introduction to Supermodernity*. London, London: Verso;

Aydin, C., & González W. M. & Verbeek, P.-P., (2019), Technological Environmentality: Conceptualizing Technology as a Mediating Milieu. *Philosophy and technology*, *32*(2), 321-338. https://doi.org/10.1007/s13347-018-0309-3;

Baudrillard, J. (1981) Requiem for the Media. *For a Critique of the Political Economy of the Sign* (pp. 164-184) .New York: Telos Press Publishing;

Borgmann, A. (1984), *Technology and the Character of Contemporary Life*. Chicago / London: University of Chicago Press;

Brey, P. (2000). Technology and Embodiment in Ihde and Merleay-Ponty. *Metaphysics,*

Epistemology and Technology (Research in Philosophy and Technology, vol.19 pp. 45-58, Elsevier/JAI Press;
Bibri S. E. (2015), *The shaping of ambient intelligence and the Internet of Things*, Paris: Atlantis Press;
Castells M. (2003), *The Internet Galaxy*, London: Clarendon Press.
De Preester, H. (2010). Postphenomenology, Embodiment and Technics. *Hum Stud* 33**,** 339–345. https://doi.org/10.1007/s10746-010-9144-y;
Dusek, V. (2006). *Philosophy of Technology: An Introduction*. London: Blackwell;
Feenberg, A. (1999), *Questioning technology*, New York: Routledge;
Feenberg, A. (2005), Critical Theory of Technology: An Overview. *Tailoring Biotechnologies* 1(1): 47–64;
Floridi L. (2014), *The fourth revolution: how the infosphere is reshaping human reality*, Oxford: Oxford University Press;
Floridi L. (2015), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, London: Springer;
Hayles, N. K. (2012), *How we think: digital media and contemporary technogenesis*. London: University of Chicago Press;
Han B.C., (2017) *In the swarm. Digital prospects*, Boston: MIT Press;
Hoq K. M. (2014) Information Overload: Causes, Consequences and Remedies: A Study. *Philosophy and Progress*, Vols. LV-LVI (1-2), 50-68. https://doi.org/10.3329/pp.v55i1-2.26390;
Ihde, D. (1990) *Technology and the Lifeworld: From Garden to Earth*, Indianapolis: Indiana University Press;
Ihde, D. (2008) Introduction: Postphenomenological Research. *Hum Stud* 31. 1–9. https://doi.org/10.1007/s10746-007-9077-2;
Ihde, D. (2009). *Postphenomenology and technoscience: The Peking university lectures*. Albany: State University of New York Press;
Ihde D., (2010) *Embodied Technics*, Automatic Press/VI;.
Jasanoff S., Kim, H. (2015), *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: The University of Chicago Press;
Kalthoff, H., & Roehl, T. (2011). Interobjectivity and interactivity: Material objects and discourse in class. *Human Studies*, 34, 451–469. https://doi.org/10.1007/s10746-011-9204-y;
Kaplan D.M., (2009). What Things still don't Do. *Human Studies*, 32, 229–24. https://doi.org/10.1007/s10746-009-9116-2;
Kiran, A. H., & Verbeek, P.-P. (2010). Trusting ourselves to technology. *Knowledge, Technology & Policy*, 23(3-4), 409–427. https://doi.org/10.1007/s12130-010-9123-7;
Leroi-Gourhan A., (1964). *Le geste et la parole*, Paris: Albin Michel ;
Merleau-Ponty, M., (1962). *Phenomenology of Perception*, London: Routledge;
Postman, N., (1993). *Technopoly: the Surrender of Culture to Technology*, New York: Vintage Books;
Turkle, S. (2012). *Alone Together: Why We Expect More From Technology and Less From Each Other*. New York, NY: Basic Books;
Van Den Eede, Y. (2011). In between us: on the transparency and opacity of technological mediation. *Foundations of Science*, 16(2/3). 139–159. https://doi.org/10.1007/s10699-010-9190-y;
Verbeek P.-P., (2005). *What things do: philosophical reflections on technology, agency, and design*, University Park: Pennsylvania University State Press;
Verbeek P.-P., (2011). *Moralizing Technology. Understanding and Designing the Morality of Things*, Chicago: The University of Chicago Press;
Verbeek, P.-P., (2003). Material Hermeneutics. *Techné: Research in Philosophy and Technology* 6 (3), 181-184. https://doi.org/10.5840/techne20036325
Verbeek, P.-P., (2008). Disclosing Visions of Technology. *Techné: Research in Philosophy and Technology* 12 (1), 85-89. https://doi.org/10.5840/techne200812116;
Verbeek, P.-P., & Vermaas P. E. (2012). Technological Artifacts. In: Jan Kyrre Berg Olsen Friis,

Stig Andur Pedersen & Vincent F. Hendricks (Eds.), *A Companion to the Philosophy of Technology*, (165-171) Wiley-Blackwell.
Wylie J., (2003) *Landscape, performance and dwelling. A Glastonbury case study*, Cloke P (Eds.) Country Visions, Harlow: Pearson.

SARAH PAWLETT JACKSON
St Mellitus College and University of London
sarah.pawlett-jackson@stmellitus.ac.uk

THREE BODIES: PROBLEMS FOR VIDEO-CONFERENCING[1]

abstract

In this paper I examine a specific way that video-conferencing modifies structures of intersubjective awareness and interaction. I focus on multi-person interactions (involving more than two people) via video-call. By unpacking some of the key features of multi-person intersubjectivity in cases of embodied co-presence, I will show where and how certain social affordances are strained or lost when multi-person interactions are transferred to the screen.

keywords

intersubjectivity, phenomenology, second-personal awareness, video-chat

1 The author wishes to thank two anonymous reviewers for their helpful comments and additional suggested reading which have helped to improve the paper.

Phenomenology and Mind, n. 20 - 2021, pp. 42-50
DOI: https://doi.org/10.17454/pam-2004
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Digital platforms shape, scaffold and modify our identities and experiences in many ways. Of these, modification of our *intersubjective interactions* with others are of real significance. With the increasing ubiquity of digital forms of interaction, particularly in the light of the COVID-19 pandemic and new norms of physical distancing, it is important to reflect on some of the specific ways that digital interfaces shape intersubjective and collective experiences. The meanings and possible meanings of these modifications need to be analysed and understood. In this paper my interest is specifically in the effects of video conferencing on multi-person forms of intersubjectivity. Multi-person intersubjectivity refers to personal interaction between more than two people. These are interactions that involve more than a single self-other dyad but are of a small enough size that interpersonal awareness between all parties can be maintained, namely interactions in small groups. Video-conferencing platforms, such as Skype or Zoom, make provision for multi-person conversations of this kind, with multiple people present and interacting on the same call. My objective is to draw attention to some specific ways that video-conferencing modifies forms of interpersonal awareness in these multi-person cases. I argue that social affordances that are ordinarily available in multi-person cases are not available in the same way via video.

This phenomenological interrogation is particularly important as multi-person intersubjectivity has been underexplored as a phenomenon in its own right. It is often implicitly (or explicitly) assumed that multi-person intersubjectivity can be understood as though it were the mere addition of various self-other dyads. Similarly, existing literature which considers the nature of modification to intersubjective communication via video has tended to consider factors that apply across dyadic and multi-person interactions, not looking specifically at the phenomenological modifications to small group video interactions *qua* small group.

I start by laying out some of the 'social affordances'[1] that are ordinarily available to us in small group contexts of embodied "co-presence" (Gilbert 2014, p. 325) which involve more than two people. I will highlight how these affordances are grounded in embodied and enactive awareness. I will then look at how the structure of these forms of intersubjective awareness is typically modified when interacting via video calls. I will conclude that the inability to engage

1 The language of affordances originally comes from Gibson (1986/2015). I will explain this phenomenon in more detail below.

with interactions-between-others is the most significant curtailment to organic small group interaction in online contexts.

2. Multi-person intersubjectivity

The philosophy of intersubjectivity has tended to focus on forms of interaction within a self-other dyad. Typically, this has focused on either the first-person plural phenomena of joint attention and joint action (Hobson 1989; 2008; Eilan *et al* 2005; Seemann 2012; Zahavi 2014; Reddy 2015), or on second-person singular forms of direct "mutual recognition," (Gilbert 2014, pp. 329-331), "mutual attention" (Reddy 2005. cf. Darwall 2006; De Jeagher & Di Paolo 2007: Reddy 2008: Stawarska 2009: Eilan 2014, Zahavi 2014: 2015), or shared emotion (Reddy 2008). Elsewhere, I have argued that cases of multi-person intersubjectivity, involving more than two people, should not be understood as mere linear extensions or additions of these dyadic interactions (Pawlett Jackson 2018; 2019). Rather, the complexities that emerge when interactions occur in the context of *other* interactions need to be taken into account. For example, I have argued that the *second-person plural* should be understood as a distinct form of interpersonal awareness in its own right, rather than a linear extension of the second-person singular. The second-person plural is not simply a grammatical category, but a specific form of intersubjective awareness in which I perceive and engage with multiple others *as a relational unit.*

For example, I might say to two friends of mine, a couple who have hosted me for the evening – 'thank you so much for your hospitality'. All three of us, it seems, are aware of one another as subjects, and not only this, but this awareness is reciprocal in all directions. It is transparent to all that we are all aware of one another and all aware of one another's awareness. This is not the whole story, however, because there is also an *asymmetry* in our awareness of one another that we are all aware of: I am also aware of them *as a pair* in a particular kind of way.

In this situation I am aware of both of my friends as distinct – they do not blur into a generic mass in my awareness, as might happen at a football match. Neither do they appear to me as isolated from one another. I experience them as individuals, but I *also* experience them *as a couple*. This is of course framed and clarified by an understanding of their existing historical relationship over time, but this is brought into and includes my awareness of them as *in relation to one another* in this moment in time and space, around the breakfast table. Minimally, and importantly for understanding the intersubjective structures in play, I am aware of them as peripherally but reciprocally aware of one another in their joint attention (as a 'we') *directed towards me*. I perceive a "conjoint availability" (Fivaz-Depeursinge & Corboz-Warnery 1999, p. 6) in the two of them which makes it possible and meaningful to address them as a pair. My awareness of their reciprocal awareness of one another is required if my address is to be understood by all parties as a "double address" (Fivaz-Depeursinge & Corboz-Warnery 1999, p. 114). This is precisely what seems to be in play when it is clear to all parties that I am addressing the two of them *together* with thanks. In such cases, I am able to address a pair or a small group as 'you' – that is *you plural.* All parties can recognise and understand what is happening in this kind of interaction – the second-person plural awareness is – in a phrase used by both Naomi Eilan (2005) and Margaret Gilbert (2014) – "out in the open" between us, "manifest to [all] participants"" (Eilan 2005, p. 1).

This kind of intersubjective awareness, understood from the 'other side' of the interaction, as it were, reveals that it is possible to *be addressed* or to *address as* a relational unit. As part of a first-person plural – a "plural subject" to use Gilbert's (2014) term[2] – we can jointly attend

2 My use of this term does not mean that I am necessarily committed to every aspect of Gilbert's account. I simply use it to refer to groups structured by joint attention.

to and address a third person together. A key principle here is that intersubjectivity is not just about interacting with others, but also about interacting with the *relationships between others* as well. This is why there are many forms of multi-person interaction which cannot be sufficiently understood as the mere addition of multiple second-personal dyads.
As I will consider further in the next section, our ability to recognise and participate in complex networks of interactive and asymmetrical jointness and mutuality requires sophisticated co-ordination. This does not mean, however, that this complexity of parts is salient in our lived experience. Indeed, we find that quite the opposite is true: we typically experience this reciprocity of awareness without needing to put in much attentional effort. This shows up in the way we perform complex coordination of our activity and communication with others without needing to think much about it. As Ackerman & Bargh (2010) note, "for many complex tasks...the ability to effectively coordinate with others requires intensive training. However, social coordination also occurs automatically, nonconsciously and effortlessly throughout our daily encounters with other people" (p. 335). Our ability to co-ordinate attention and action with others is grounded in pre-reflective forms of embodied awareness. I turn now to elaborate the role of 'intercorporeality' in multi-person interactions, before turning to look at video-conferencing in the light of this understanding.

3. Embodied and enactive

In recent years, phenomenologists and cognitive scientists have converged in understanding the human subject as inescapably '4E' – *embodied, embedded, enactive and extended* (Varela *et al* 1991; Newen *et al* 2018). For the purposes of this paper I focus on subjectivity as embodied and enactive. As enactive beings the world presents itself to us not just as data for comprehension but as possibilities for action – or 'affordances'.[3] An understanding of attention as embodied and enactive is significant in our thinking about intersubjective interactions. Here, the explosion of recent work on enactive social perception underlines ideas originally laid down by Merleau-Ponty that 'intercorporeality' is the foundation of intersubjectivity (Morganti *et al* 2008; De Jaegher & Di Paolo 2007; Gallagher 2014; 2017; De Jaegher *et al* 2017; Fuchs 2017). This fundamental level of connection to others is, Merleau-Ponty (1945/2012) says, "achieved through the reciprocity between my intentions and the other person's gestures, and between my gestures and the intentions which can be read in the other person's behaviour" (pp. 190-192). In this vein, Francesca Morganti (2008) notes that traditional models "conceive of communicative interaction between two people as a process that takes place in a set of mental operations that end up being expressed (externalised) in speech, gesture and action" (p. 8). Embodied and enactive approaches, on the other hand, understand gestures, postures, bodily orientation and facial expressions to be substantive parts of intersubjective attention and intersubjective affectivity itself.
Ordinarily, I am aware of some components or combination of the other(s)' posture, gesture or expressions. I experience the other's attention directed at me in her shift in bodily orientation, for example. That (and how) I am held in the other's attention might be clear to me in the openness of her posture, the hesitancy of her gestures or the thoughtfulness of her facial expression, as well as – and sometimes in spite of – what is communicated verbally.
This is the basis of the pre-reflexive coordination of bodies in space. Thomas Fuchs (2017) identifies this in the "inter-bodily resonance" that provides the basis of intersubjectivity, highlighting that when "mutual incorporation" is in play, "each lived body with its sensorimotor body schema reaches out, as it were, to be extended by the other" (p. 8). Fuchs highlights that this includes the "inter-affective" resonance between bodies. This can be seen

3 See fn1.

most explicitly in high-level coordination in joint kinaesthetic activities like sport or dance, but it is also found in the micro-exchanges of glances, movements and posture.

To communicate mutuality of awareness verbally is already 'too late' for ordinary forms of second-personal interaction, as it were. This communicative mutual awareness is the condition of the possibility of addressing second-personally at all. The embodied and extra-linguistic components of second-person mutuality make up what Stawarska (2009) calls the "deep grammar" or "protogrammar" (pp. 72; 130) of second-personal interactions.

To return to multi-person interactions specifically, we find that extra-verbal, embodied and enactive elements of recognition and communication are crucial in creating the conditions for the kinds of complex interlocking intersubjective awareness identified above. The ability to form a 'joint address' or to receive a 'double address' requires that I am able to coordinate my own actions, attentiveness and responses to and with multiple others at the same time. Significantly, I am able to recognise intersubjective interaction *between others* in *their* bodily coordination with one another. I can 'see' inter-affective resonance between others – I am able to recognise that the conversation between a couple is becoming tense, for example, or I can see that they are at ease between themselves and are inviting me to join their conversation. Where multiple perspectives within and between subjects and plural subjects are co-ordinated in the right way, seamless bodily and conversational coordination is possible between myself and multiple other subjects.

Unpacking this multiplicity, we can note that a single action, such as a gesture, can be a communication to more than one person at once. Not only this, but the body-language with which I communicate (and by which others communicate to me and to other others) is not limited to one kind of action. Gesture, posture, tone and facial expressions all play a role, as do multiple sensory modalities. Elizabeth Fivaz-Depeursinge and Antoinette Corboz-Warnery (1999) identify the orientation and coordination of *pelves*, *torsos*, *gazes* and *expressions* as relevant in multi-person communication. They identify that these different forms of bodily orientation and expression can be coordinated in multiple ways, producing different kinds of three-way interaction (p. 57).[4] I can also employ different bodily actions in a coordinated way to target more than one other person at once. When faced with multiple others I might, at any one time, be listening to one while looking directly at the other, for example, or looking at one while reaching out to the other with an inclusive gesture.

We can note specifically that I am able to coordinate attention *with* someone *to* another by coordinating different modes and actions. As *we* face *you*, I may gesture to my co-attendee, shifting my posture to implicate her in my attentiveness – whilst looking at the one we are attending to together. There are some cases in which trying to coordinate attention to more than one other person with more than one modality could be confusing or distracting, causing a split in my or our attention. In other cases, however, these multiple sensory inputs may be integrated into a smooth whole – as part of an interaction that all three of us are attuned to as participants.

As someone receiving a joint address, I can 'read' the intersubjective awareness *between the others* in the plural subject which I face. The reason I can do this is because it is enacted between them bodily. Their joint attention is not locked away from me as a private affair between their two (or more) disembodied minds but is out in the open for me to perceive and engage. Indeed, their attention to and with one another presents itself to me as a particular kind of social affordance for my second-personal engagement with them (together).

That the others' interlocking activity presents itself to me as an affordance or an invitation for engagement is important for understanding how various forms of asymmetrical multi-person interaction can emerge. It is not just a third-person plural observation of others' behaviour

4 This study looks at mother-father-infant triads as they play together.

that is in view here. I can respond and participate, not just recognise. I watch two people both writing at the same table in the café, for example. I would like to use a seat on the table and want to ask if I can take it. Are they there together? Should I address them separately or jointly? It is hard to tell at first, but as I take a moment to watch them, I recognise that they are there together by the way they are both holding their bodies, their orientation and their posture at the table. (Or indeed, I might recognise that they are there independently, functionally oblivious to one another.) This transforms how I address them with my request. Having laid out some key features of multi-person intersubjectivity in typical cases of co-presence, I turn now to look at multi-person interactions via video-conferencing. In considering the phenomenology of interactions via video I hope to illuminate specific social affordances that are strained or even lost in these contexts.

4. Others apart: multiplicity via video

In this section I focus on video-conferencing as a specific way that our intersubjective landscape can be mediated and therefore modified by digital technology. Literature already exists looking at a number of different ways that video technology modifies the phenomenology of intersubjective communication along axes such as empathy, affectivity, privacy, intimacy or objectification (Fuchs 2014; Ferencz-Flatz 2019; Krueger & Osler 2019; Osler Forthcoming). However, the significance of modifications to interpersonal structures of awareness in specifically multi-person contexts remains underexplored. My focus here is to note that there is a dyadic paradigm built into the architecture of the video call, and that certain key characteristics of organic multi-person dialogue are therefore impeded.

When contrasting video calls with conversation that takes place in embodied co-presence, we can note that spatial and physical proximity is modified. Participants on a video call are temporally co-present to each other, but not spatially so.[5] It is worth noting that there are other forms of technologically mediated multi-person interaction which are modified in different ways due to their different form. Group text messaging, for example, can involve multiple others who are dispersed not only in space but also in time. Some of the same principles of modification may apply to both video-conferencing and to group messaging, but there will be further issues to explore on the phenomenon of group messaging which I do not have space to look at here. I focus here on the way that spatiality is modified on video calls, and how this constrains embodied and enactive social affordances that we take for granted in co-located multi-person contexts.

In the words of product designer and software engineer John Palmer (2020), "video calls are faces inside static rectangles". Not only this, but these static rectangles are stacked around each other on my screen in a way that does not and cannot take into account any relation between each of these others. The spatiality of the video call is a literal flattening out of those that I am speaking to – no one is nearer or further away from me or from anyone else. No-one can orient themselves to gesture at any other particular person, as we are not arranged together in shared space. To quote Palmer again:

> When in real life are you ever looking at a grid of faces? Never...In real life, you look at people around a dinner table, or around a circle at...a meeting, or at someone beside you as you go for a walk (Palmer 2019).

5 As an aside, we can note that the prevalent phenomenon of glitches in internet connection often and typically modifies the temporal structure of a conversation over video as well. This is an interesting and important issue to explore for understanding modifications to interpersonal exchanges via video more generally. I will not focus on this issue here, staying with issues facing modified multi-person exchanges specifically. See Osler (Forthcoming) Section 5.2, for a discussion of the temporality of video-calls.

As a result of this modified spatiality, it is not possible to shift one's posture or orient one's body towards a particular person or sub-group of others within the group as a whole. Where gestures are made, they can only be made as a way of communicating to all interlocutors in the same way. If I try to orient towards or gesture to a particular person on the call, it is not transparent to this person, or anyone else on the call, that it is *this* particular person who I am gesturing to, save if I use language explicitly to single them out. I cannot catch a particular person's eye, not only because video calling makes this inherently difficult, even in dyadic cases, but because it is not possible for others to tell who it is on the call that I am trying to directly engage by my gestures or expression. This means that I cannot form the awareness of jointness or mutuality with particular people or sub-groups within the group – or at least not in the more organic way I would be able to do in a fully embodied context.

Likewise, because each of the others is cut off from one another in the same way, there is no bodily interaction between others in the group for me to 'read' and engage with. Whereas ordinarily I can 'pick up' the intercorporeal, inter-affective interactions between others in the ways that they orient and gesture towards one another, or share looks and facial expressions – none of this is available on a video call. I therefore cannot interact with others as a relational unit (from which I am distinct) in the same pre-reflective way. This particular form of social affordance is shut down when interacting via video.

While all participants share a joint awareness that everyone else is on the call, this is in a much more symmetrical or 'flattened' way. While in 'real life' there are various asymmetries to our awareness of each of the others – asymmetries that other others can recognise and then in turn interact with – this is not the case on a call. The 'conjoint availability' presented to me by others is far more limited. Each 'other' is atomised from the other others. On video calls we therefore experience a first-person plural experience between all of us participating, but this 'jointness' is the same in all directions, so to speak.

This is not to say that there is no scope for interacting with others as a relational unit whatsoever. We can still pick up tone of voice and detect conversational patterns between others. Particularly in established group relationships, with habituated forms of interaction in these relationships, it may be easier for us to pick up when two friends are sharing an in-joke together on a call, for example. Opportunities to recognise inter-affectivity between others are not completely shut down then, but they are significantly muffled. Video-calls need not involve a complete or absolute 'flattening' of asymmetrical dynamics. The point remains, however that the (non)spatial (non)arrangement of subjects means that I cannot organically form interactional sub-units within the wider group as I would ordinarily be able to do through my embodied and enactive co-ordination. It is not possible to align bodies with another person to speak as a 'we' to a third 'you', for example. Ordinarily in a group of three or four, conversation may naturally shift in its form from each participant speaking as an individual subject to parts of the conversation where two are addressed and jointly reply to the rest of the group as a plural subject. This kind of shift in intersubjective awareness, rooted as it is in embodied and enactive awareness, is much harder to access via video. As the key extra-verbal ways that we indicate jointness or mutuality are strained or lost, so forms of joint or mutual awareness can only be communicated explicitly and verbally. While this means that forms of joint address are still possible, ("X and Y, can I get your opinion on Z based on your experience of this together last year?"), this serves to make salient that the embodied and enactive scaffolding that usually grounds them is missing. More work is needed to make explicit what would otherwise be more immediately evident.

5. Conclusion

In the above I have argued that there are specific forms of multi-person interaction which we ordinarily experience and participate in. These forms of interaction should not be understood reductively, as the mere addition of adjacent or overlapping dyads. This non-reductive understanding of multi-person interaction allows us to see that there are specific social affordances which video-conferencing curtails. Having laid out these specific structural inhibitions, it is hopefully clear that the claim that video-conferencing does not sufficiently enable forms of multi-person intersubjective awareness is not a vague or misplaced romanticism about 'really being together'. Rather, this is an observation about the loss of multiplicity which is embedded in embodied co-presence. Going forward, this observation presents an opportunity for interface designers: what might it look like to build video-conferencing which builds shared spatiality and the possibility of asymmetrical intersubjective reciprocity into the technology itself? Future design work which takes seriously the non-reducibility of many forms of multi-person intersubjectivity may open up social affordances not yet possible via video technology.

REFERENCES

Ackerman, J.M. & Bargh, J.A. (2010) Two to Tango: Automatic Social Coordination and the Role of Felt Effort. In Bruya, B. (ed.) *Effortless Attention* (pp. 335-371). Cambridge, Mass: The MIT Press;.
Darwall, S. (2006) *The Second-Person Standpoint.* Cambridge, Mass; London: Harvard University Press;
De Jaegher, H. & Di Paolo, E. (2007) Participatory sense-making: An enactive approach to social cognition, *Phenomenology and the Cognitive Sciences,* 6(4), 485–507. doi: 10.1007/s11097-007-9076-9;
De Jaegher, H, Pieper, B., Clénin, D., & Fuchs, T. (2017) Grasping intersubjectivity: an invitation to embody social interaction research, *Phenomenology and the Cognitive Sciences*, 16(3), 491-523. doi: 10.1007/s11097-016-9469-8;
Eilan, N., Hoerl C & Roessler, J. (2005) (eds) *Joint Attention: Communication and Other Minds.*Oxford: Oxford University Press;
Eilan, N. (ed.) (2014) 'The Second Person' Special Issue, *Philosophical Explorations,* 17(3), 265-387;
Ferencz-Flatz, Christian (2019) Ten theses on the reality of video-chat: A phenomenological account, *Communications*, 44(2), 204–224. doi: 0.1515/commun-2018-2012;
Fivaz-Depeursinge, E. & Corboz-Warnery, A. (1999) *The Primary Triangle.* New York: Basic Books;
Fuchs, T. (2014) The Virtual Other: Empathy in the Age of Virtuality, *Journal of Consciousness Studies,* 21(5-6), 152-173
Fuchs, T. (2017) Intercorporeality and Interaffectivity. In Meyer, C; Streeck, J & Jordan, J. Scott, *Intercorporeality: Emerging Socialities in Interaction* (pp. 3-24). New York: Oxford University Press;
Gallagher, S. (2014) 'Phenomenology and Embodied Cognition' in L. Shapiro (ed.) *The Routledge Handbook of Embodied Cognition* (pp. 9-18). London; New York: Routledge;
Gallagher, S. (2017) *Enactivist Interventions: Rethinking the Mind.* London: Oxford University Press;
Gibson, J.J. (1986/2015) *The Ecological Approach to Visual Perception.* New York: Routledge;
Gilbert, M. (2014) *Joint Commitment: How We Make the Social World.* New York: Oxford University Press;
Hobson, P. (1989) On sharing experiences, *Development and Psychopathology,* 1, 97–203. doi: 10.1017/S0954579400000390Hobson, P. (2008) Interpersonally Situated Cognition, *International Journal of Philosophical Studies* 16(3), 377-397. doi: 10.1080/09672550802113300
Krueger, J. & Osler, L. (2019) Engineering Affect: Emotion Regulation, the Internet, and the Techno-Social Niche, *Philosophical Topics,* 47(2), 205-231, doi:10.2307/26948114;
Merleau-Ponty (1945/2012) *Phenomenology of Perception,* trans. D.A. Landes. London; New York: Routledge;
Morganti, F., Carassa, A. & Riva, G. (2008) *Enacting Intersubjectivity.*Amsterdam: IOS Press;

Morganti, F. (2008) What intersubjectivity affords: paving the way for a dialogue between cognitive science, social cognition and neuroscience (pp3-16). In Morganti *et al* (2008);
Newen, A., De Bruin, L. & Gallagher, S. (eds.) (2018) *The Oxford Handbook of 4E Cognition.* Oxford: Oxford University Press;
Palmer, J. (2019, August 30) Spatial Interfaces. [Web log post]. Retrieved from: https://darkblueheaven.com/spatialinterfaces/
Palmer, J. (2020, April 9) Spatial Software. [Web log post]. Retrieved from: https://darkblueheaven.com/spatialsoftware/
Pawlett Jackson, S. (2018) Gestalt structures in multi-person intersubjectivity, *Synthese*. doi: 10.1007/s11229-018-02001-y
Pawlett Jackson, S. (2019) Exploring different intersubjective structures in relation to dialogue. *Arts and Humanities in Higher Education.* 18(1) 22-33;
Osler, L. (Forthcoming) 'Taking Empathy Online', *Inquiry.*
Reddy, V. (2005) Before the Third Element: Understanding Attention to Self (pp. 85-109). In Eilan (2005);;
Reddy, V. (2008) *How Infants Know Minds.* London: Harvard University Press;
Reddy, V. (2015) Joining intentions in infancy, *Journal of Consciousness Studies,* 22(1–2), 24–44.
Seemann, A. (2012) *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience.*Cambridge, Mass: The MIT Press;
Stawarska, B. (2009) *Between You and I: Dialogical Phenomenology.* Athens, Ohio: Ohio University Press;
Varela, F., Thompson, E. & Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience.* USA: The MIT Press;
Zahavi, D. (2014) *Self and Other.* New York; Oxford: Oxford University Press;
Zahavi, D. (2015) You, Me, and We: The Sharing of Emotional Experiences. *Journal of Consciousness Studies,* 22(1–2), 84–101.

LORENZO OLIVIERI
University of Bologna
lorenzo.olivieri3@unibo.it

PERSUASIVE TECHNOLOGIES AND SELF-AWARENESS: A DISCUSSION OF SCREEN-TIME MANAGEMENT APPLICATIONS[1]

abstract

Persuasive technologies are interactive systems designed to change and shape users' behaviours towards specific goals. By discussing the case of screen-time management applications, this paper explores how persuasive systems transform self-awareness and the self's cognitive architecture. Drawing on the notion of tectonoetic awareness, I will illustrate how artefacts enable the transition from the temporal bounded experience characterizing first-person perspective (noetic awareness) to the ability of reflecting on oneself from a third person and temporally extended perspective (autonoetic awareness). I will argue that persuasive systems make possible new modalities of self-recognition and self-projection while they simultaneously affect the sense of agency by interfering with users' actions and intentions.

keywords

persuasive technologies, tectonoetic awareness, material agency, self-recognition, sense of agency

1 I am grateful to Marta Caravà for her feedback, and to Margoth González Woge, Ciano Aydin and Michael Nagenborg for their comments and suggestions to an extended version of this paper. I wish to thank two anonymous reviewers for their helpful comments.

Phenomenology and Mind, n. 20 - 2021, pp. 52-60
DOI: https://doi.org/10.17454/pam-2005
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Persuasive technologies are interactive systems designed to change and shape users' behaviours towards specific goals. The variety of cognitive and psychological stimuli and feedback enabled by digital technologies offers new means for guiding human choices and promises to improve or radically transform previous modalities of persuasion (Fogg, 2003). Persuasive systems invite users to behave in supposedly morally good or healthy ways; they perform autonomous actions and they adopt emotionally or intellectually engaging persuasive strategies in order to constrain and nudge users' actions. Due to this deliberate interference with human agency, persuasive technologies have received, over the last years, considerable attention from a moral perspective as they challenge the traditional understanding of freedom, autonomy and responsibility (Brey, 2005; Guthrie, 2013; Nagenborg, 2014; Spahn, 2012; Verbeek, 2006). However, the goal of this paper is to show that the effects of persuasive systems on human life are not limited to their possible unethical consequences or to the ways they might modify, and potentially inhibit, the ability for moral reasoning. I will suggest that persuasive technologies might re-model the self's cognitive system and thus lead to new modalities of experience and perception of the self. The notion of *tectonoetic awareness* (Malafouris, 2008a), articulated upon the relation between *noetic awareness* and *autonoetic awareness*, will allow us to stress the twofold implications of persuasive technologies on the human self. First, these technologies modify the self's sense of agency, through cues and feedback interfering with users' actions and designed to steer them towards certain behavioural patterns; second, persuasive technologies aim to make users more aware of their past and current actions and to shape a "future self" who is liberated from "harmful" behaviours.

The paper contributes to the debate about the implications of digital technologies on the human self and cognition (Clowes, 2015; Clowes, 2018; Heersmink, 2016; Smart, 2017; Smart, 2018). However, none of these works explicitly address persuasive technologies. Clowes (2018) describes how some persuasive systems - like *Fitbit* or *Stay Focusd* - have the potential for self-shaping and for constructing "new forms of self-regulation practices", but his discussion is overall more concerned with the enhancing effects of Internet technologies on human agency rather than with persuasive technologies and self-awareness.

2. Persuasive technologies: the case of screen-time management applications

The label "persuasive technology" refers to an heterogenous set of computing technologies (applications, websites, videogames, smart and virtual environments, wearable devices) which are applied to a likewise heterogenous set of target domains and behaviours (health care, education, environmental sustainability, e-commerce, social networks). Due to the vastness of the field, I will discuss a specific subset of persuasive systems, namely self-imposed persuasive technologies designed for helping and motivating people to adopt beneficial behaviours and to avoid harmful ones. This choice cuts out persuasive systems imposed by third parties without the explicit consent of the users, such as social networks, and it cuts out persuasive systems imposed by third parties (governments, institutions, companies) to steer the behaviours of citizens, students or employees toward supposedly good behaviours. Self-imposed persuasive technologies are less problematic from an ethical perspective, at least *prima facie*, because they presuppose users' deliberate decision to deploy them and hence they seem to better preserve freedom of choice.

To understand the functioning of a self-imposed persuasive technology, consider applications like *Rescue Time, Freedom* or Apple's *Screen Time,* which are designed to both manage the time spent on laptops or smartphones and to support people to remain focused on their work-related activities. In addition to these general goals, these systems aim to mitigate Internet addiction (Montag *et al.*, 2015) and the so-called checking habits (Oulasvirta *et al.,* 2012), namely those relatively short but repeated sessions in which emails, social network updates or news headlines are revisited. Checking habits threaten and hinder the achievement of a permanent and deep level of attention and concentration (state of flow), and they have negative effects both on work productivity and on personal life (Duke & Montag, 2017). To support the development of healthier digital habits, persuasive technologies for screen-time management enable users to set up a time limit for websites and apps or to completely block them; they provide users with historical data showing the time spent on productive or entertainment activities, and they allow to create customized schedules of internet usage. Despite small differences, the logic underlying the functioning of these systems is the same and evokes those of other self-imposed persuasive technologies, such as persuasive mirrors (Nakajima & Lehdonvirta, 2011; Nakajima *et al.*, 2008)[1]. A lifestyle tracking component collects information on user behaviour by automatically tracking and recording screen-time activities. Feedback information is then presented to the users in order to persuade them to come back to working activities or to not get distracted. Alerts or notifications pop up on the screen blaming the users for not being working or, on the contrary, praising them for having reached a pre-established goal. The feedback might also have a more disruptive modality, for instance by completely blocking the possibility to access a website once the time limit is achieved or during a "focus time" session. Besides these immediate types of feedback, screen-time management applications also provide users with accumulated feedback, usually through reports, graphs and charts showing the amount of time spent daily, weekly, or monthly in leisure or work-related activities or through a resume of the "challenges" won or lost. Screen-time management applications are thus interactive and individualized systems which transform how people make decisions about their digital behaviour. The use of these systems amounts to a delegation of decision-making processes, to an "outsourcing

1 The goal of persuasive mirrors is to support behaviour change by providing users with personalized visual feedback that reflects the progress toward a beneficial lifestyle. Like a traditional mirror reflects a person's physical appearance, a persuasive mirror reflects back to the person her alignment to the desired target behaviours in order to increase awareness about her choices and actions. Persuasive mirrors can alter the reflected image either through augmented reality or by displaying virtual scenarios, and can extract and process diverse sources of users' data. See Verbeek (2009) for a discussion of the ethical implications of persuasive mirrors.

of conscientiousness" (Guthrie, 2013). As Guthrie (2013) notes, a controversial aspect of persuasive systems is that the actions elicited by them "may carry the form but not always the power of the virtues of thoughtfulness or order" (p. 328). On the other hand, these technologies can be treated as a form of "potential agentive enhancement" which do not only help to comply to our own policies, but also offer new possibilities for self-reflection (Clowes, 2018). These views are inspired by diverse interpretations (and concerns) of the effects of persuasive technologies on human thinking and action. However, the more fundamental problem underlying these different viewpoints concerns the role played by artefacts in shaping not only human agency and cognition but also in transforming the experience of oneself. The next section will thus be dedicated to outlining, through the notion of *tectonoetic awareness* (Malafouris, 2008a), how the material agency of artefacts re-models the self's cognitive system and brings forth new possibilities for perceiving and reflecting upon oneself.

3. Tectonoetic awareness: artefacts and the extended self

Malafouris (2008a; 2008b) grounds the concept of tectonoetic awareness on the distinction between two levels of awareness: *noetic awareness* and *autonoetic awareness*. Noetic awareness refers to "the basic sense of oneself as acting in and on the environment at a time, according to one's first-person perspective" (Malafouris, 2008b, p. 407). Autonoetic awareness refers to the ability of reflecting upon oneself from a third person perspective; it introduces a temporal dimension in the perception of the self, namely the process of self-recollection (the mental reinstatement of past events and experiences) and self-projection (the ability of thinking, imagining and planning about the future) (Malafouris 2008a; 2008b). As Malafouris acknowledges, the notions of noetic and autonoetic awareness were originally introduced by Tulving (2002), for whom the former "was used to describe the conscious state that accompanies thinking about (knowing) the world", while the latter "was used to describe the experiential 'flavor' of remembering, or recollection" (p. 4). However, Malafouris's use of these two concepts shares stronger affinities with the concepts of *minimal self* and *narrative self* (Gallagher, 2000). More specifically, noetic awareness corresponds to the minimal self, namely the consciousness of oneself as an immediate subject of experience, un-extended in time and characterized by the first-personal givenness of experiential phenomena (Gallagher, 2000; Zahavi 2007). It includes the sense of ownership (the sense that is my body which is undergoing an experience) and the sense of agency (the sense that I am the initiator or source of action). On the other hand, autonoetic awareness shares important features with the narrative self, a more or less coherent self (or self-image) that is evolving and extended in time to include memories of the past and intentions toward the future and which is understood in the light of one's own self-interpretation (Gallagher, 2000; Gallagher & Zahavi, 2012)[2]. Malafouris's thesis is that the passage from noetic to autonoetic awareness is made possible, both ontogenetically and phylogenetically, by the use of and interaction with artefacts. Artefacts enable humans to extend into their material surroundings and simultaneously to detach themselves from the temporal and spatial contingencies of first-person experience. Objects, in other words, liberate the self from the here and now of ordinary experience and allow the minimal self to be anchored into its social surroundings (Malafouris 2008a; 2008b). Tectonoetic awareness has thus to be understood as "a scaffolding process of ongoing

2 I am aware that the two notions - autonoetic awareness and narrative self - are not identical. First of all, unlike autonoetic awareness, the narrative self is not necessarily conscious or reflective. Also, whereas the narrative self is mostly conceived in terms of narrative structures and personality (including the endorsement of values and beliefs), Malafouris's autonoetic awareness stresses the embodied and ecological dimensions of the self. However, in the context of this paper what I think is relevant is that they both insist on the diachronic nature of selfhood, on its persistency over time and on the awareness of such temporal structure.

structural coupling that grounds in action and integrates the noetic and autonoetic aspects of selfhood" (Malafouris, 2008a, p. 1998). Different objects illustrate this process: beads for self-decoration (Malafouris, 2008b), a Mycenean golden ring (Malafouris, 2008a), the sword of Myceanean warriors (Malafouris, 2008c). Individually and culturally invested with memories and events associated to their use and ownership, these objects helped humans "to move across the scales of time and to construct bridges between temporal phenomena that operate at different experiential level" (Malafouris, 2013, p. 247). In this way, objects embody a "dynamic cognitive biography" which redefines the boundaries of biological memory and that brings forth a new kind of autonoetic awareness, making possible explicit self-recognition through objectification (Malafouris, 2008a, p. 1999).
It is worth noting that any modality of interaction with tools and objects might enable the transition to autonoetic awareness. Tool use and manufacture played, before body decoration, a decisive role in the development of human self-awareness (Jeffares, 2010). However, what is crucial is that different forms of autonoetic awareness emerge as an effect of the specific epistemic qualities of the material mediums: beads, being permanently attached to the body, made possible the emergence of forms of self-knowledge and self-recognition that cannot result from the interaction with and production of stone tools (Malafouris, 2008b). Things impose their own agency on human cognition and modify the self's cognitive architecture in virtue of their material properties. The notion of tectonoetic awareness is thus grounded upon a process of reciprocal causation: the self extends into its environments and emerges through objects which enable novel and specific forms of cognition and self-recognition; in turn, the material agency of objects re-organizes and re-models the self's cognitive system. What needs to be analysed in the next section is how this process of reciprocal causation between objects and the self applies in the case of screen-time management applications: how do they shape users' first-person experience, and their sense of agency? What are the modalities of self-knowledge and self-recognition emerging from using them?

4. Self-awareness and persuasive technologies

Two preliminary considerations are needed in order to properly analyse the type of self-awareness brought forth by screen-time management applications. First, there is a substantial qualitative leap between the archaic artefacts discussed by Malafouris and current digital technologies (Aydin *et al.*, 2019). Digital technologies are, in fact, increasingly defined by interactivity – which allow them to interfere with, and even to anticipate, human actions – and by their hidden functioning – which allow them to operate without requiring attention. The alerts and notifications as well as the self-tracking features of screen-time management applications capture precisely these two properties of digital technologies. Second, what is at the stake in the case of persuasive technologies for screen-time management is not the transition from the noetic to the autonoetic awareness as discussed by Malafouris, who, as a cognitive archaeologist, is mainly interested in the phylogenetic trajectory of human cognitive development. In fact, the decision of adopting a self-imposed persuasive technology already presupposes a narrative self who is conscious of its past behaviours and who willingly undergoes a persuasive process in order to shape a "future self". However, the notion of tectonoetic awareness and the distinction between noetic and autoneotic awareness are crucial because, as I shall show below, persuasive technologies modulate and affect the relation between those two levels of awareness.
Having made these considerations, I suggest treating the immediate feedback and the accumulated feedback of screen-time management applications as the specific epistemic qualities of these systems. To begin with, by tracking and recording users' activities, these technologies are part of that E-memory revolution (Clowes 2013; 2015) that has considerably enhanced the possibility to access information about oneself, since E-memory technologies

can potentially record everyday activities on a scale and with a fidelity and completeness previously unknown. The data about smartphone or laptop activities shown on graphs give access to qualitatively and quantitatively new information about users' actions compared to the information which can be remembered by the brain. As digital repositories of our behaviours, screen-time management applications can consolidate and digitalize moments of our past that we did not remember and, also, those that we do not want to remember. The accumulated feedback provides users with a diachronic representation of their screen-time activities, enabling new ways for perceiving themselves as temporally extended subjects. In this respect, screen-time management applications strengthen and enhance the ability of self-recollection because they allow users to remember, with unprecedented degrees of accuracy, their past actions. Moreover, they offer new possibilities for self-projection because, by being confronted with the amount of time spent on productive or entertainment activities, users can reflect upon the consequences of their behaviours, and thus they can adjust their digital habits in order to reach their future goals.

While accumulated feedback affects autonoetic awareness, the alerts and notifications popping up on users' screen interfere with the noetic dimension of the self, especially the sense of agency. The sense of agency is a phenomenologically ambiguous concept characterized by different elements (Gallagher, 2013). The immediate feedback of screen-time managements systems does not involve the pre-reflective aspect of agency connected to motor control, but rather "a more reflective sense of intention, involving attention directed toward the project or task that we are engaged in, or toward the means and/or end that we aim for" (Gallagher, 2013, p. 12). When I wish to stop working and create an online chess match, but the screen-time management system blocks the website and prevents me from accessing it, my sense of agency is significantly compromised. The persuasive system interferes with my intention of playing chess and re-directs myself toward my work-related activities. Similarly, when the system notifies me that I have reached the daily limit of usage and invites me to put the phone down, it triggers thoughts and reactions which influences my behaviour. Importantly, these immediate feedbacks deeply affect the motivational and agential character of cognition (Walsh, 2017). As Walsh notes, there are crucial phenomenological differences between cognitive processes supported by artefacts and those limited to the skin-and-skull boundaries. Such differences have to do with how relevant information becomes phenomenally conscious and with the role of intellectual virtues such as understanding and self-reliance. In the case of screen-time management applications, there is a difference between *understanding* that I should stop browsing the web and merely *knowing* it because a notification told me so.

Putting all this together, screen-time management applications make possible the emergence of a highly personalized form of autonoetic awareness, one in which new modalities of self-recognition and self-projection are possible. At the same time, however, they deeply affect and modify the noetic aspect of self-awareness, by interfering with people's intentions, preventing them from behaving in certain ways and redirecting their patterns of action. My conclusion applies not only to screen-time management applications, but it can also be extended to other self-imposed persuasive technologies relying on the same feedback logic. Moreover, focusing on the noetic and autonoetic levels of awareness enables the assessment of alternative design approaches to persuasive technologies, such as mindless computing (Adams *et al.*, 2015) and reflective computing (Munson, 2012).

The idea of mindless computing is to design systems that do not rely on users' motivation and ability but rather aim at influencing users' behaviour in subliminal, subconscious ways. Whereas most of the persuasive technologies depend on conscious awareness and imply reliance on motivation and capacity of self-control, mindless computing aims to operate below the threshold of conscious awareness, in order to automatically trigger the desired behaviours

while relieving the users from the burden of motivation and reflection (Adams *et al.*, 2015). The triggers deployed in mindless computing, in fact, produce immediate and automatic responses which affect users' behaviour without being noticed or perceived by them. A user of mindless computing device is thus aware that it is her body which is acting in a certain way, but she does not control the source of that action. The noetic self is thus deprived of the sense of agency and it is left only with the sense of self-ownership, while the absence of any forms of accumulated feedback and the subconscious modality of persuasion do not allow for the emergence of new forms of self-recognition.
On the other hand, reflective computing (Munson, 2012) has a radically different goal and it shares strong affinities with the boosting approach (Hertwig & Till Grüne-Yanoff, 2017). The aim is to provide users with relevant data about their behaviours, but without having the system prescribing what to do, inviting to action or setting specific goals. The design of these systems thus does not include any persuasive feedback mechanism, but rather it merely reveals data to the user in order to make them reflect. In this case, due to the lack of any type of immediate feedback interfering with human action, the effects of the persuasive system on noetic awareness are very weak, whereas the persuasive process is directed to strengthen autonoetic awareness, to make it thicker, by giving users' the possibilities to learn about their past behaviours but without the introduction of interactive persuasive elements.

5. Conclusion

In this paper I argued that persuasive technologies have a double effect on self-awareness. First, they strengthen and thicken autonoetic awareness: new possibilities of self-recognition and self-projection emerge due to the possibility to access data about one's own screen activities. Second, persuasive technologies erode and weaken noetic awareness by interfering with the self's sense of agency. The notifications and cues (or blockage) to actions appearing on the screen redirect our intentions and shape our actions. Lastly, I suggested that focusing on these two levels of awareness allows us to analyse the design and functioning of other approaches to persuasive technologies and their effects on the self's cognitive system.
Before concluding, I wish to suggest that the notion of tectonoetic awareness might represent a precious analytical entry point for discussing digital technologies because it stresses the dynamic and developmental dimension of the human mind. Far from being a static feature of human cognition, self-awareness is constantly re-shaped and restructured by material culture. As shown by the case I discussed, the relation with digital devices has become particularly redundant. On the one hand, phones, websites and social networks ask for attention through their visual or acoustic signals, they afford fast access to a vast array of informational rewards and they are purposefully designed to make people spend increasingly more time with them. On the other hand, digital systems can become persuasive agents which record and store significant portions of people's lives and to which self-control is delegated in order to reach goals and desires. These apparently opposite tendencies are however mediated by the same material support, or, in Gibsonian terms, by the same surface. For this reason, framing the discussion about the consequences of digital systems in terms of enhancement or diminution might reiterate an essentialist understanding of the human self and mind, and it might prevent us from conceiving of them as historically situated contingencies shaped by the social and technical environment.

REFERENCES

Adams, A.T., Costa, J., Jung, M.F., & Choudhury, T. (2015). Mindless computing: designing technologies to subtly influence behavior. *UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 719–730 https://doi.org/10.1145/2750858.2805843;

Aydin, C., González Woge, M., & Verbeek, P.-P. (2019). Technological Environmentality: Conceptualizing Technology as a Mediating Milieu. *Philosophy & Technology*, 1–18. https://doi.org/10.1007/s13347-018-0309-3;
Brey, P. (2005). Freedom and privacy in ambient intelligence. *Ethics and Information Technology*, 7(3), 157–166. https://doi.org/10.1007/s10676-006-0005-3;
Clowes, R. W. (2013). The Cognitive Integration of E-Memory. *Review of Philosophy and Psychology*, *4*(1), 107-133. https://doi.org/10.1007/s13164-013-0130-y;
Clowes, R. W. (2015). Thinking in the cloud: The cognitive incorporation of cloud-based technology.;
Philosophy and Technology, 28(2), 261–296. https://doi.org/10.1007/s13347-014-0153-z;
Clowes, R. W. (2018). Immaterial engagement: human agency and the cognitive ecology of the internet. *Phenomenology and the Cognitive Sciences*, 1–21. https://doi.org/10.1007/s11097-018-9560-4;
Duke, É., & Montag, C. (2017). Smartphone addiction, daily interruptions and self-reported productivity. *Addictive Behaviors Reports*, 6(April), 90–95. https://doi.org/10.1016/j.abrep. 2017.07.002;
Fogg. B.J. (2003). *Persuasive Technology Using Computers to Change What We Think and Do*. Morgan Kaufmann Publisher;
Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci*. 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5;
Gallagher, S., & D. Zahavi. 2012. *The Phenomenological Mind*, 2nd ed. New York: Routledge.;
Gallagher, S. (2013). "Ambiguity in the sense of agency," in *Decomposing the Will*, eds J. Kiverstein and T. Vierkant (Oxford: Oxford University Press), 1–17.;
Guthrie, C. F. (2013). Smart Technology and the Moral Life. *Ethics and Behavior*, 23(4), 324–337. https://doi.org/10.1080/10508422.2013.787359;
Heersmink, R. (2016). The internet, cognitive enhancement, and the values of cognition. *Minds and;*
Machines, 26(4), 389–407. https://doi.org/10.1007/s11023-016-9404-3;
Hertwig, R. & Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, Vol. 12(6) 973– 986. https://doi.org/10.1177%2F1745691617702496;
Jeffares, B. (2010). The co-evolution of tools and minds: cognition and material culture in the hominin lineage, *Phenomenology and Cognitive Science*, 9,503–520. https://doi.org/10.1007/s11097-010-9176-9;
Malafouris, L. (2008a). Between brains, bodies and things: Tectonoetic awareness and the extended self. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499), 1993–2002. https://doi.org/10.1098/rstb.2008.0014;
Malafouris, L. (2008b). Beads for a Plastic Mind: the 'Blind Man's Stick' (BMS) Hypothesis and the Active Nature of Material Culture Lambros. *Cambridge Archaeological Journal*, 18(3), 401–414. https://doi.org/10.1016/j.cub.2015.10.018;
Malafouris, L. (2008c). Is it 'me' or is it 'mine'? The Mycenaean sword as a bodypart. In J. Robb & D. Boric (eds.) *Past Bodies*. Oxford: Oxbow Books, 115-23.;
Malafouris, L. (2013). *How things shape the mind*. A theory of Material Engagement. Massachusetts Institute of Technology;
Montag, C., Duke, É., & Reuter, M. (2015). A short summary of neuroscientific findings on internet addiction. In Montag, C. & Reuter, M. (eds.) *Internet Addiction* (pp. 131–139). Springer International Publishing. https://doi.org/10.1007/978-3-319-46276-9_12;
Munson, S. (2012). Mindfulness, Reflection and Persuasion in Personal Informatics. *Workshop on Personal Informatics*, 1–4.;

Nagenborg, M. (2014). Surveillance and persuasion. *Ethics and Information Technology*, 16(1), 43–49. https://doi.org/10.1007/s10676-014-9339-4;
Nakajima, T., Lehdonvirta, V., Tokunaga, E., & Kimura, H. (2008). Reflecting human behavior to motivate desirable lifestyle. *Proceedings of the 7th ACM Conference on Designing Interactive Systems - DIS '08*, 405–414. https://doi.org/10.1145/1394445.1394489;
Nakajima, T., & Lehdonvirta, V. (2013). Designing motivation using persuasive ambient mirrors. *Personal and Ubiquitous Computing*, 17(1), 107–126. https://doi.org/10.1007/s00779-011-0469-y;
Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, 16(1), 105–114. https://doi.org/10.1007/s00779-011-0412-2;
Smart, P. R. (2017). Situating machine intelligence within the cognitive ecology of the internet. *Minds and Machines*, 27(2), 357–380. https://doi.org/10.1007/s11023-016-9416-z;
Smart, P. R. (2018). Emerging digital technologies: Implications for extended conceptions of cognition and knowledge. In J. A. Carter, A. Clark, J. Kallestrup, S. O. Palermos, & D. Pritchard (Eds.), *Extended Epistemology* Oxford: Oxford University Press;
Spahn, A. (2012). And Lead Us (Not) into Persuasion...? Persuasive Technology and the Ethics of Communication. *Science and Engineering Ethics*, 18(4), 633–650. https://doi.org/10.1007/s11948-011-9278-y;
Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology* 53, 1–25.;
Verbeek, P.P. (2006). Persuasive Technology and Moral Responsibility Toward an ethical framework for persuasive technologies. *Persuasive*, 1–15. https://doi.org/10.1111/j.1600-0668.2005.00303.x;
Verbeek, P.P. (2009). Ambient intelligence and persuasive technology: The blurring boundaries between human and technology. *NanoEthics*, 3(3), 231–242. https://doi.org/10.1007/s11569-009-0077-8;
Walsh, P. J. (2017). Cognitive extension, enhancement, and the phenomenology of thinking, *Phenomenology and the Cognitive Science*, 16, 33–51. https://doi.org/10.1007/s11097-016-9461-;
Zahavi, D. (2007). Self and Other: The Limits of Narrative Understanding. *Royal Institute of Philosophy Supplement* 60: 179–202.

SECTION 2

SECTION 2

LANGUAGE AND MIND – SOCIAL MEDIA AND IDENTITY CONSTRUCTION

Andrea Addis, Olimpia G. Loddo, and Massimiliano Saba
Social Acts in Digital Environments

Silvia Donzelli
Countering harmful speech online. (In)effective strategies and the duty to counterspeak

Henk Jasper van Gils-Schmidt
Hilde Lindemann's Counterstories: A Framework for Understanding The #MeToo Social Resistance Movement on Twitter

ANDREA ADDIS
Infora S.c.a.r.l.
andrea.addis@gmail.com

OLIMPIA G. LODDO
University of Cagliari
olimpia.loddo@gmail.com

MASSIMILANO SABA
University of Cagliari
massimiliano.saba@unica.it

SOCIAL ACTS IN DIGITAL ENVIRONMENTS[1]

abstract

Adolf Reinach's theory of social acts and Czesław Znamierowski theory of the environment can show a new perspective of analysis in the fields of computer science and digital communication. This paper will begin analysing the performance of social acts in two categories of digital environments: (i) fictional digital environment and (ii) real digital environment. The analysis will be supported by examples from the history of computer science. In both kinds of digital environments, organigrams play a significant role and depend on the users' digital power to perform a real or fictional social act. Finally, the paper will analyse one of the possible roles that AI plays in performing social acts in digital environments. It will show how AI could affect the perception of social acts.

keywords

Adolf Reinach, Czesław Znamierowski, social act, digital environments

1 Acknowledgments. We are thankful to the two referees and to prof. Giuseppe Lorini for their thoughtful and constructive suggestions.

Phenomenology and Mind, n. 20 - 2021, pp. 64-75
DOI: https://doi.org/10.17454/pam-2006
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

The grammar of the word "knows" is evidently
closely related to that of "can", "is able to".
But also closely related to that of "understands".
('Mastery' of a technique)
(Wittgenstein, 1958, § 150).

1. Introduction

In this paper, the authors will try to apply Adolf Reinach's theory of social acts and its ideal development, Czesław Znamierowski theory of the environment, to the field of informatics and digital communication. The first paragraph gives a short account of the two approaches. In the second paragraph, the authors distinguish and analyze two categories of digital environments, (i) fictional digital environment and (ii) real digital environment. The research will take into account examples from the history of computer science. The third paragraph points out that, in a digital environment, there can be organigrams. These organigrams depend on the user's power to perform different forms of social acts. More specifically, the hierarchical status concerns both the extension and the intention of the social act: in fact, the hierarchical status of the user could be affected both by the typology of the performable social acts (commercials, invitations to treat, posts) and the (number and the kind of) potential addressees that are reachable in the digital environment. The fourth paragraph focuses on the role of AI in the performance of social acts in digital environments and how could AI affect their perception.

Reinach's and Znamieroski's historical-cultural context was very different from the one in which digital environments achieved social significance. However, both Reinach and Znamierowski provide a useful conceptual apparatus that can clarify two aspects of the social dynamics: the structure of the social acts and the relevance of the environment for their performance. These aspects are particularly relevant in a digital environment since the user can perform a specific act if the digital environment embodies its structure (type). More precisely, the conceptual apparatus proposed by Reinach and Znamierowski can clarify both how the users perform social acts and the developers' way of acting (how developers can create new forms of acts performable in a digital environment, and how the users can perform those acts).

One can argue that other more recent philosophical theories could better clarify the digital environments' social dynamics. One of these philosophical theories could be, for instance, the theory of speech acts, developed more recently, mainly in the framework of analytic

philosophy. Contemporary authors, who were more aware of the dynamics that characterise digital environments since they operate in a more technologically advanced context supported the theory of speech acts. However, Reinach's phenomenological approach allows us to build a conception of social acts under which the distinction between the social act structure and its manifestation is more explicit.
Interestingly, Reinach (1911) clearly distinguishes three moments of the social act: (i) the internal experience, that is, the presupposition of the social act; (ii) the social act that is rooted in the inner experience; (iii) the external form of expression of the social act. A question (that is an example of social act) presupposes an internal experience (a doubt) and is manifestable, for example, with a verbal expression.
Moreover, Znamerowski develops the analysis of the manifestation of the social act focusing on the environment where the expression of the act takes place. According to Znamierowski, on the one hand, the environment makes possible the social act and, on the other hand, affects the form of its manifestation.
The relevance of the environment in the form of manifestation of social acts emerges in computer sciences. Curiously, computer scientists distinguish mark-up language from programming languages; both languages are fundamental for creating the same digital environment, where, as we will see, it is possible the performance of social acts. The use of these two different languages can often mirror the distinction between the social act and the form of manifestation. Presentational markup languages (such as HTML) often are used for defining how to present (make visible) the structured data. For instance, through HTML, the developer can create a button that the user can click to accept a specific proposal. Acceptance and proposal are typical examples of social acts and, in this case, to click on the button 'accept' counts as the manifestation of a social act. Indeed, through the expression of the social act, the social act becomes visible to the addressee.

2. The Social Act and the Environment

In 1911, during a lecture, the young phenomenologist Adolf Reinach first enunciated his social acts theory. He further analysed the concept of "social act" in his most famous book, *The A Priori Foundations of the Civil Law* (1913).
According to Reinach, social acts do not rest on themselves but need another subject to be performed. Unlike other acts (such as deciding or walking), social acts are not individually performable. These acts are necessarily addressed to another person, and their eidetic aim is to be perceived (*vernommen*) by that person. The main characteristic of social acts is the necessity to be perceived (*Vernehmungsbedürftigkeit*). Promises, commands, communications, acceptances are all examples of social acts.
Since social acts' aim is their perception by an addressee, when they are performed between human beings (who do not possess telepathic powers), the agent shall manifest them to the addressee. In this sense, Reinach talks about the social act's function of manifestation (*Kundgabefunktion*).[1] Thanks to its manifestation, the social act projects itself into the external environment. From a merely material point of view (at the pre-semiotic level. See Lorini 2008, p. 116), a social act occurs in the external physical environment. In this respect, the social act's manifestation is a modification of the environment: the agent modifies the environment to make the perception of the act by the recipient possible. The Polish philosopher of Law, Czesław Znamierowski, proposes a reinterpretation of the concept of social act highlighting the role of the environment (*środowisko*) in its performance and the role of a specific kind of norms "construction norms" (*norma konstrukcyjna*). To explain this peculiar category of

1 See Loddo 2020.

norms, Znamierowski (1924) hypothesizes that two children in a play make an agreement: if one of them raises a little flag, the other should respond by raising a little flag himself. The connection between the two raisings of the flag is instituted by a norm that gave them a specific significance. In virtue of this constructed connection, the first physical movement becomes a greeting, while the other becomes a response to that greeting.
Znamierowski calls "construction norms" the norms which bestow new conventional meaning on objects and acts.
We can find a similar schema in digital environments. When a new type of action is embodied in a digital environment, the developer attributes an institutional meaning to a phenomenon that in itself it is not institutional, and this operation reflects the logical structure of constructive rules. Indeed, some constructive rules can pre-exist the construction of the virtual environment. The developer can import in the digital environment institutional concepts from the non-digital social world. For instance, the constructive rules that create and regulate bitcoin as a virtual institutional fact presuppose the constitutive rules that build a non-virtual institutional fact, such as money. Some other constructive rules can be developed by the creator of the virtual environment, for instance, when the developer elaborates a hierarchy of users for an online platform. The developer incorporates in the digital environment both the constitutive rules from non-digital social reality and the ones created to define the different roles of the users that will operate in the digital environment. These rules are a condition of possibility for the performance of a social act by the user.
Starting from the analysis of the social act proposed by Reinach, Znamierowski identified three fundamental moments in the social act: (i) the experience, intentional and active, (ii) the change of the environment, necessary in the realization of the intention, (iii) the change in the social structure. Only the first moment, according to Znamierowski, was analysed by Reinach. Znamierowski thinks that a change in the external (physical) environment plays a significant role in fulfilling the social act. Any change in social reality is a consequence of the recipient's perception of the social act (Lorini 2008, pp. 23-41). The environment is the *medium* that makes possible social interaction between humans. There is always an impersonal "environment" between agent and recipient (Znamierowski 1921, p. 13). The modification of the physical environment is necessary for the performance of social acts to the extent that conscious changes in one person's mind cannot directly cause psychic changes in another person's mind.
Moreover, given different types of social acts (such as communication, promise, offer), the environment determines the diversification of the acts' manifestation.
Besides, the environment "makes long-term activity possible" (Znamierowski 1921, pp. 13-15). Changes in the environment (produced by an agent who intends to manifest a social act) do not necessarily have to reach the recipient immediately. It is sufficient that they produce a structural change in the environment and, even after some time, this change may become the cause of the psychic processes that will lead the recipient to the perception of the social act. For example, a note, addressed to a friend (that is not in the room) and left on his table, represents a lasting change in the environment that will allow us to postpone the moment of perception.
To sum it up, the environment is a condition for the existence of social relationships [*stosunki społeczne*] (Lorini and Żełaniec 2016).

3. Digital Environments

Digital communication systems and virtual platforms are examples of how the environment can play a role in social acts' performance. As mentioned before, the architecture of the digital environment embodies rules and institutional concepts. Therefore, it could be a condition of possibility for the performance of a set of social acts. These environments are fundamental in our daily life, and no company can survive without undergoing a digital transformation (Weber 2020; Pal 2019).

3.0. Fictional Digital Environment vs Real Digital Environment

We can observe two macro-categories of digital environments: *fictional* digital environments and *real* digital environments.
Fictional digital environments are structured digital environments designed for carrying out fictitious social acts. These environments are typical of the videogame world (the virtual setting *par excellence*). They not only transposes real settings with the possibility of meeting, socializing, interacting and communicating (indeed useful even to promote these activities in daily life (Peng et al. 2012)), but they add additional powers and tools as, e.g., the possibility of teleportation, flying, planning to build with means and materials unavailable in reality, pushing the potential of the users to the extreme. For this reason, scholars analysed their impact on real social acts (Greitemeyer and Mügge 2014). However, these environments do not coincide entirely with the videogame field. Sites that allow playing board games (such as online chess) are excluded. The competitive dimension of the game, which is a real dimension, must be kept distinct from the role-playing game, which involves introducing the user into a fictitious social reality. The two dimensions can be present in the same game, but they do not coincide.
Real digital environments are digital environments that allow the fulfilment of real social acts. These environments include both the world of multimedia platforms and digital tools aimed at conducting video conferences. The world of multimedia platforms, which is generally more constricting and binding in terms of relationships, however, is hugely more pervasive: online newspapers, service sites, e-commerce, but above all social networks, can host a massive number of users (hundreds of millions and, sometimes, billions).
Video conferencing tools must also be mentioned, having undergone a vast diffusion and popularity during the lockdown due to the coronavirus outbreak.
For decades, these worlds have been considered parallel to conventional reality (Damer 1998). Their future implications are being more and more investigated (Mazuryk and Gervautz 2012). Reasoning on the possibility not only that they will eventually become indistinguishable on a sensorial level (Tallinn 2012), but dusting off philosophical questions raised already by Plato and revisited with theories like *the simulation hypothesis* (Bostrom 2003), which advocates the non-excludability of our existence in an already fictitious reality.

3.1. Fictional Digital Environment

Before the advent of social networks, video games were the most common digital environment. Since then, the term '*creator*' indicates the person who created the game, while the players are the *actors* because they play roles planned by the *creator-director*, who puts several limits to the actors' free will. The actors cannot perform actions that are not foreseen by their hierarchical role built by the platform's creator.
Moreover, the creator can encourage certain behaviours awarding the actors, therefore forcing them to follow the rules and influencing them through behavioural psychology paradigms such as positive reinforcement.
Some of the most common examples of fictional digital environments naturally involve the videogame field. For example, the online virtual world *Second Life*, developed by *Linden Lab* around 2003, has experienced a rapid expansion reaching approximately one million regular users in a few years.[2] The *residents* live the virtual reality through social and economic relationships also dictated by the trade of goods and lands and institutional activities, which is still a pioneering feature.
Sandbox Video Games as Minecraft are another example.
Minecraft was officially released in 2011 and, since then, has not seen a significant decline in playing, becoming a best seller with 200 million copies sold on all gaming platforms and

2 Linden Lab 2014.

126 million monthly active users (Warren 2020). It is playable in a direct multiplayer fashion, for example, on LAN (local area networks). Still, the worlds that host entire communities of players are possible thanks to the *Minecraft Servers*[3]. The servers allow players to interact, communicate, sometimes fight (through the Player versus player combat, PvP (Bartle 2003)), plan the use of resources and actions as a real community, with all the advantages of the virtual setting. The operators cover the vital role of guaranteeing and maintaining order, the playability of the worlds, avoid or contain the actions of *cyber vandals* or *griefers*[4], often drawing up intricate lines of conduct that must be learned and followed by the players before being allowed to join the community (Garrelts 2014), to not having some of their social skills (e.g., communication, teleportation, interaction with the world's objects) blocked, or, for more serious matters, their temporary or permanent exclusion (banning).
To get an idea of these worlds' size, one of the most populous servers, *Hypixel*, has been visited by about 14 million *unique players*[5] (Fogel 2018).
As the last example, the rapid expansion of the latest edition of the *Animal Crossing* video game series, belonging to the class of *social simulation video games*, developed and published by Nintendo. In this case, the players are anthropomorphic creatures that can perform various activities that include social interactions and non-social acts (such as fishing, building, carrying out archaeological excavations).
Since its release, the game, and its spin-offs, sold tens of millions of copies, the only *New Horizon* has sold almost 14 million copies[6].
In all these fictional digital environments, even within a single game, hierarchical organizational structures, spontaneously defined during the game, emerge. These structures are often strongly influenced by leading figures who, by charisma or ability, take over the team's management.

3.2. Real Digital Environment

Some authors stress that the interaction in a digital environment is as real as face-to-face interaction. The mental reactions to a non-digital social event, and digital social event can be identical (Chayko 2018, p. 55). Despite that digital environments often find their expression in ludic environments, the "virtual world" does not necessarily consist of fictitious transpositions of real environments. Indeed, digital multimedia environments can represent the substratum of real institutional facts (Mathiason 2009). A digital platform could allow interactions between different users who intend to perform real social acts (purchases, questions, requests, communications). Within the platform, users will operate within different roles or "privileges" that enable them to perform a subset of social acts. Some users will need, e.g., to acquire the platform administrator role to complete administrative and organizational tasks. Furthermore, the developer will set up some constraints outlining the platform architecture based on the roles and features being made available to the users (Scuderi 2015). Creating groups, pages, specific issues, and managing their availability to other users are examples of administrative-social acts that can be precluded to low privileged users.
While selecting the pool of users to whom attribute social acts, the programmer will be able to create a hierarchy of privileges inside the platform. In order to do that, s/he will have to model the platform's structure by changing the virtual environment in which social acts are performable. Within this paradigm, the social platform will inherit all the characteristics that

3 Minecraft Multiplayer Server - https://www.minecraft.net/en-us/download/server/.
4 Griefing is an act explicitly intended to irritate a person or a community in a virtual world through vandalism, destruction, theft of resources, or in the most subtle cases of using social hacking.
5 One player can play with different avatars.
6 Minotti 2020.

Znamierowski attributes to the physical environment for carrying out social acts: it transmits the activity of one person to another, models and diversifies its activities, making them irreducible to mere manifestations of the intent of the individual agent, thus allowing long-term activity, keeping track of the acts performed within it (Krämer and Conrad 2017).
Digital platforms are now innumerable. They are differently classifiable, e.g., by type of service (social, e-commerce, institutional) or by the interaction style (unidirectional, collective).
Among the real digital environments, we can list the cloud-based peer-to-peer software platforms used for teleconferencing, telecommunications, distance education, and social relationships.
In the near future, the possibility of consolidating these features to virtual environments through Virtual Reality (VR) and Augmented Reality (AR) technologies (Schroeder 1996), for example, VRChat (over 50,000 communities) offers an endless collection of social VR experiences by giving the power of creation to the users.

4. Digital Organigrams

A hierarchy of roles and competencies characterizes both fictional and real digital environments. A set of commonly accepted rules make social interaction in the digital environment accessible and not frustrating.
Some of these rules create institutional figures that can perform administrative acts and sanction users for creating a disturbance in the digital environment (Bakioglu 2009) (ultimately to prevent cybercrimes).
Both fictional and real digital environments are therefore generally populated (but not exclusively) by these types of figures: (i) the *administrators* who are generally responsible for the management and configuration of the available resources, either because they are the owners or because they have acquired "jurisdiction" on the platform, (ii) the *operators* who generally assist the administrators in the operations of placing the resources on-site or moderating them through various forms of control made available by the same platform, and obviously (iii) the *users*, the agents that actively participate in community activities (e.g. the players of a video game, the readers of a blog or newspaper, the participants in a videoconference) (Bertino and Matei 2014).

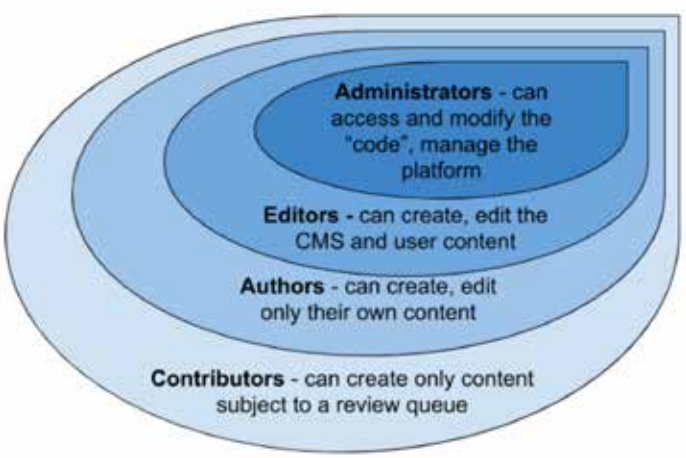


Figure 1.
The organization chart of the most popular Content Management System (CMS), WordPress. Above the platform users, there are operative and managerial roles. In WordPress, roles overlap, inheriting or restricting the capabilities.

These organizational hierarchies develop in three steps: (i) in *the design phase*, in which, following the analysis of the requirements of the platform, are established roles and operating procedures of the administrative sector; (ii) *during the platform life*, where inadequacies or physiological imperfections of the preliminarily established hierarchies emerge, and therefore the need to modify the availability of features or roles and management modes according to the new experiences of use; (iii) *following the spontaneous emergence of intrinsic hierarchies*, dictated by the physiological, communicative, social and attitudinal characteristics of the actors, which stabilize without particular structural impositions on the platform.

5. Other-Direction and Perceptibility of the Social Act: the Impact of an AI

Social acts, according to Reinach, are "other-directed" (*fremdpersonal*) acts that need to be perceived (*vernehmungsbedürftig*) by an addressee.
AI is an instrument that could help the addressee's perception. For instance, when the AI makes more visible certain information than others in accordance with the addressee profile, it is facilitating the social act of communication.
AI can affect the performance of social acts under these two different profiles; in particular, a well-structured AI could:

i. Determine the recipient of a particular social act, or automatically direct a social act towards the most suitable recipients for its acquisition
ii. facilitate the perception of a social act by being able to discern a speech with imperative language by one with a more proactive or descriptive language, e.g., thanks, to the identification of a different vocabulary.

5.1. The Impact of AI in Determining the Recipients of Social Acts

In an environment with defined hierarchies, AI plays a crucial role. AI "delivers" contents and resources through planning, labeling, categorizing. These operations have direct consequences on the perception of social acts and their direction between the nodes of the hierarchy or of the organizational structure of the social group that operates in a digital environment.
The benefits of a supportive approach are obvious. Without a search engine, even if trivial, it would be impossible to allow users of a platform to select the topics of their interest, thus directing the social act (in this case of the communication (*Mitteilung*)) in a precise manner. This possibility would allow not only a new paradigm in the manifestation of social acts but also their classification.

5.2. AI and the Perception of Social Acts

Perception (*Vernehmung*) in the Reinachian lexicon is not reducible to the merely sensory sphere but implies a full understanding of the social meaning of that act. The recipient of the social act must understand the nature of the social act addressed to him (a promise, a command, a communication).
Social dynamics require extremely heterogeneous skills from a communicative, relational, cognitive point of view and are still too vast and complex for AI to simply transpose a digital representation of it (McCarthy 2007).
However, AI is proving itself capable of managing increasingly superhuman abilities, increasingly complex tasks linked to action and strategies in controlled contexts, which sometimes even require great abstraction skills (Baum, Goertzel, Goertzel 2011).
In particular, nowadays, artificial intelligence is focused on narrow tasks (Artificial Narrow Intelligence, ANI)[7]. According to Searle, ANI "would be useful for testing hypotheses about minds, but would not actually be minds" (Searle 1980). On the contrary, strong AI is defined as

7 Io9.com mentions narrow AI. See Dvorsky 2014.

a machine with the ability to apply intelligence to any problem, rather than just one specific problem, considered to require consciousness, sentience, and mind.
AI is a valid and essential tool in the field of the categorization of texts and the understanding of natural language. This is possible through algorithms that in general, perform the following operations:

i. translation of documents into entities that can be represented within a multidimensional vector space. This operation is possible by simplifying the document representation in order to reduce the computational burden for their treatment through a preprocessing phase (Uysal and Gunal 2014) thus, e.g., eliminating poorly informative terms, speech particles, prepositions and pronouns (stop wording) and eliminating endings and declination, reducing the terms to their root (stemming) (Riloff 2014).
ii. building a representation of a set of documents belonging to the same class, then reducing the categories of interest to models of meta documents (Yang and Pedersen 1997).
iii. classification of new incoming documents in one or more of the selected/trained categories, possibly with confidence scores (Sebastiani 2002).

Generally, these algorithms are based on statistical analysis (Yang 1999) of the occurrence characteristics of terms within documents and classes. However, these techniques are also expanded in a semantic sense by using tools and networks that relate lemmas with the abstract concept to which they refer (Miller, Beckwith, Fellbaum, Gross and Miller 1991), or with additional algorithms that automatically process these relationships. These *semantic* engines, more sophisticated (Cai and Hofmann 2003), are used by the largest stakeholders in the digital world. The most advanced search engines can interpret concepts expressed with syntactically unrelated and sometimes semantically ambiguous terms, decoupling the search tasks both from the very presence of the words searched and subsequently even extrapolating the concepts hidden in the search query, not only by related terms (for example, "*movie about a kid being left at home*").
With these assumptions, the analysis in determining the form and quality of a social act seems to be plausible and will be investigated in future research.

6. Conclusions

From the discussion above, it is clear that, in today's society, social acts must necessarily deal with digital environments.
The relationship between the social act and the virtual environment is not one-sided but bi-univocal. The environment contributes to defining the social act that takes place in it and, symmetrically the social acts (fictitious or real) contribute to defining the nature of the digital environment where they are typically performed.
Moreover, the developer often intervenes by registering *ex-post* new rules in the digital environment to regulate acts that have become customary.
The hierarchical division of users is fundamental to define the role of social acts in virtual environments. Organization charts tend to be a common trait shared more or less explicitly in different digital environments. The presence of a hierarchical structure has a mainly technical-functional purpose, granting controlled interaction between users who share the same interests.
These hierarchies impose specific possibilities of action to the users based on privileges and flexibility of the digital environment, which differs from the traditional social environment (Letaifa, Edvardsson and Tronvoll 2016).
Digital environments are channels of online social interaction that could contribute to identity-based oppression and hierarchy. Compared to traditional social environments, this

conditioning is more easily found (thanks to the user experience and affordance techniques), more restrictive (imposed by structural constraints, not by violable rules) (Lorini and Moroni 2020), and almost immediately modifiable by the platform operator (given the very nature of the virtual environments). These structural constraints can be understood as the product of the "onticisation" of originary deontic entities: in other words, rules that in other conditions could be violable (and possibly sanctioned) are transformed into patterns of action that the user cannot bypass, except by hacking the system or changing its structure.

It is not difficult to imagine that political control could take over the technical-functional purpose. This aim cannot be reduced to the set of rules that determine the design of the digital architecture; it has a meta-institutional nature (Lorini 2014).

In other words, to understand the nature of the social acts and that of the digital environment within which they take place, it is necessary to put oneself in the perspective of the person who *designs* the system of rules in terms of a game, an instrument of social interaction or an instrument of social control.

In this sense, the distinction proposed in this essay between fictitious digital environment, real digital environment and extra-digital social reality represents a theoretical distinction that does not preclude the transformation of an originally fictitious digital environment into a real digital environment capable of producing consequences in the extra-digital reality.

REFERENCES

Adrian, A. (2010). Beyond Griefing: Virtual Crime. *Computer Law & Security Review*, 26(6), 640-648. doi: 10.1016/j.clsr.2010.09.003;

Bakioglu, B. S. (2009). Spectacular Interventions of Second Life: Goon Culture, Griefing, and Disruption in Virtual Spaces. *Journal of Virtual Worlds Research*, 1(3), 4-21. doi: 10.4101/jvwr.v1i3.348;

Bartle, R. (2004). *Designing Virtual Worlds*. Indianapolis: New Riders;

Baum, S. D., Goertzel, B., & Goertzel, T. G., (2011). How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change*, 78(1), 185-195. doi: 10.1016/j.techfore.2010.09.006;

Bertino, E., & Matei, S. A. (Eds.). (2014). *Roles, Trust, and Reputation in Social Media Knowledge Markets: Theory and Methods*. Berlin: Springer;

Bostrom, N. (2003). Are You Living in a Computer Simulation? *Philosophical Quarterly*, 53(211), 243-255. doi: 1467-9213.00309;

Cai, L., & Hofmann, T. (2003). Text Categorization by Boosting Automatically Extracted Concepts. In *SIGIR: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval* (pp. 959-970). Toronto. doi: 10.1145/860435.860470;

Cantú-Paz, E. (2004). Feature Subset Selection, Class Separability, and Genetic Algorithms. In Deb K. (Eds.) *Genetic and Evolutionary Computation – GECCO 2004. Lecture Notes in Computer Science,* vol. 3102, Berlin-Heidelberg: Springer. doi: 10.1007/978-3-540-24854-5_96;

Chayko, M. (2018). *Superconnected: The Internet, Digital Media, and Techno-Social Life*. Los Angeles, CA: SAGE;

Damer, B. (1998). *Avatars! Exploring and Building Virtual Worlds on the Internet*. Berkeley: Peachpit Press;

Dvorsky G. (2014). How Much Longer Before our First AI Catastrophe? *Io9 Gizmodo*. Retrieved from http://io9.com/how-much-longer-before-our-first-ai-catastrophe-464043243;

Fogel, S. (2018). Creators of Popular 'Minecraft' Server Announce New Game' Hytale'. Retrieved from https://variety.com/2018/gaming/uncategorized/hypixel-studios-reveals-hytale-1203089201/;

Garrelts, N. (Ed.). (2014). *Understanding Minecraft: Essays on Play, Community and Possibilities*. Jefferson: McFarland;

Greitemeyer, T., & Mügge, D. O. (2014). Video Games Do Affect Social Outcomes: A Meta-Analytic Review of the Effects of Violent and Prosocial Video Game Play. *Personality and Social Psychology Bulletin*, 40(5), 578-589;
Krämer, B., & Conrad, J. (2017). Social Ontologies Online: The Representation of Social Structures on the Internet. *Social Media + Society*, 3(1), 1-11. doi: https://doi.org/10.1177/2056305117693648;
Letaifa, S. B., Edvardsson, B., & Tronvoll, B. (2016). The Role of Social Platforms in Transforming Service Ecosystems. *Journal of Business Research,* 69(5), 1933-1938. doi: https://doi.org/10.1016/j.jbusres.2015.10.083;
Linden Lab (2014). "Infographic: 10 Years of Second Life", https://www.lindenlab.com/releases/infographic-10-years-of-second-life;
Loddo, O.G. (2020). *Manifestare gli atti sociali. I canali della giuridicità dopo Reinach*. Milano: FrancoAngeli;
Lorini, G. (2014). Meta-institutional Concepts: A new Category for Social Ontology. *Rivista di estetica*, 56, 127-139;
Lorini, G., & Moroni, S. (2020). Ruling without Rules: Not Only Nudges. Regulation beyond Normativity. *Global Jurist*, 20. doi: 10.1515/gj-2019-0051;
Lorini, G., & Żełaniec, W. (2016). Czesław Znamierowski's Social Ontology and Its Phenomenological Roots. In A. Salice & H. B. Schmid (Eds.), *The Phenomenological Approach to Social Reality* (pp. 75-90). Dordrecht: Springer;
Mathiason, J. (2009). *Internet Governance: The New Frontier of Global Institutions*. London: Routledge;
Mazuryk, T., & Gervautz, M. (1996). Virtual Reality - History, Applications, Technology and Future. Institute of Computer Graphics and Algorithms, Vienna, University of Technology;
McCarthy, J. (2007). From Here to Human-Level AI. *Artificial Intelligence*, 171(18), 1174-1182. doi: 10.1016/j.artint.2007.10.009;
Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1991). Introduction to WordNet: An Online Lexical Database. *International Journal of Lexicography* 3(4), 235-244. doi:1093/ijl/3.4.235;
Minotti, M. (2020). Animal Crossing: New Horizons has sold over 13.41 million copies. *VentureBeat*. Retrieved from https://venturebeat.com/2020/05/07/animal-crossing-new-horizons-has-sold-11-77-million-copies/;
Pal, D. (2019). Future of Workplace: Collaboration of Human Resource with Artificial Intelligence. *Journal of HR Organizational Behavior Entrepreneurship Development*, 3(3), 14-19. Retrieved from http://management.nrjp.co.in/index.php/JHROBED/article/view/429/590;
Peng, W., Crouse, J. C., & Lin, J.-H. (2013). Using Active Video Games for Physical Activity Promotion: A Systematic Review of the Current State of Research. *Health Education & Behavior*, 40(2), 171-192. doi: 10.1177/1090198112444956;
Reinach, A. (1911). Nichtsoziale und soziale Akte. In K. Schuhmann & B. Smith (Eds.), *Sämtliche Werke. Kritische Ausgabe und Kommentar* (1989 ed., pp. 95-139). München: Philosophia Verlag;
Reinach, A. (1913). Die apriorischen Grundlagen des bürgerlichen Rechtes. In K. Schuhmann & B. Smith (Eds.), *Sämtliche Werke. Kritische Ausgabe und Kommentar* (pp. 141-278). München: Philosophia Verlag;
Riloff, E. (1995). Little Words Can Make a Big Difference for Text Classification. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 130-136). New York: Association for Computing Machinery.
Schroeder, R. (1996). *Possible Worlds: the Social Dynamic of Virtual Reality Technology*. Boulder, CO: Westview Press.

Scuderi, R. (2015). Identità virtuali. Un fenomeno giuridico postmoderno. In P. Moro (Ed.), *Etica, Informatica, Diritto* (pp. 193-121). Milano: FrancoAngeli;
Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417-424.
Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1-47;
Tallinn, J. (2012). *Why Now? A Quest in Metaphysics*. Paper Presented at the Singularity Summit. Retrieved from https://docs.google.com/document/d/1-yQGSPjXPB64rey7nQ3YNvWTf4ND05_3Pkqud80O_uM/pub;
Uysal, A. K., & Gunal, S. (2014). The Impact of Preprocessing on Text Classification. *Information Processing & Management*, 50(1), 104-112. doi: 10.1016/j.ipm.2013.08.006;
Warren, T. (2020). Minecraft is Still Incredibly Popular as Sales Top 200 Million and 126 Million Play Monthly. *The Verge*, Retrieved from https://24newsorder.com/minecraft-still-incredibly-popular-as-sales-top-200-million-and-126-million-play-monthly/;
Weber, A. (2020). *Digitalization for Value Creation: Corporate Culture for a Digital World*. Dordrecht: Springer;
Wittgenstein, L. (1958). *Philosophical Investigations*, Oxford: Blackwell;
Znamierowski, Cz. (1921). O przedmiocie i fakcie społecznym. *Przegląd filozoficzny*, 24, 1-33.
Znamierowski, Cz. (1924). *Podstawowe pojęcia teorji prawa. Układ prawny i norma prawna*. Poznań: Fiszer i Majewski.

SILVIA DONZELLI
Bielefeld University
silvia_donzelli@yahoo.com

COUNTERING HARMFUL SPEECH ONLINE. (IN)EFFECTIVE STRATEGIES AND THE DUTY TO COUNTERSPEAK

abstract

The concept of counterspeech denotes a non-coercive and non-censoring method for reacting to harmful speech, with the aim of impeding or at least diminishing its damaging effects. Remarkable work is being done by researchers and activist groups on elaborating practical strategies of countering hate speech online. Though, research in moral and political philosophy exploring the effectivity of counterspeech and grounding the reasons for engaging in it still remains in its early stages. In the following paragraphs I will address recent contributions which elaborate on the viability and normative aspects of counterspeech. Outlining their valuable insights, but also their failure to give due importance to the peculiarities of online speech dynamics, I will highlight relevant features on which future research about online harmful speech and counterspeech can build.

keywords

counterspeech, digital harmful speech, social media, bystander, complicity

Phenomenology and Mind, n. 20 - 2021, pp. 76-87
DOI: https://doi.org/10.17454/pam-2007
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Counterspeech goes digital

Recent decades have seen a growing body of theoretical and empirical research exploring the relationship between speech and harm. Along with the undoubted conceptual interest, the topic of harmful speech[1] raises egregious moral and political issues, especially when it comes to elaborating viable strategies of harm prevention.

Strategies to address harmful speech can be broadly classified into three main approaches: legal sanctions, content takedown regulations, and counterspeech. The first two strategies have a censoring nature. For this reason, they can come into conflict with the core democratic value of freedom of speech, raising challenging issues of political and economic coercion, definition and evidence of speech-related harm, proportionality requirements in balancing competing values, among others.

The third strategy, counterspeech, lacking a censoring approach, does not seem to pose similar normative issues and can therefore appear as an appealing and drawback-free solution. In spite of this, counterspeech is by far not unproblematic, as researchers have already pointed out, though mainly in the context of offline harmful speech.

The rise of digital communication to an overall pervasive phenomenon is bringing new challenges to harmful speech theory and policy. This does not mean that approaches from law, philosophy and social psychology, which have been designed with offline speech dynamics in mind, should now be dismissed. Rather, they should be up-graded, taking into account the particular ways in which digital speech is impacting on individuals, groups and socio-cultural structures.

In the following, I will deal with recent contributions on harmful speech and counterspeech, highlighting their valuable insights, but also their failure to give due importance to the distinguishing features of digital speech. I will start by outlining feminist philosophy of language insights on harmful speech and counterspeech strategies (2), as well as Lepoutre alternative proposal (3). Then, I will argue that both approaches fail to consider some specific features of digital speech dynamics profoundly affecting harmful speech force, and I offer one example of counterspeech strategy tailored for digital speech challenges (4). Finally, I will deal with the normative question of how to ground a possible individual moral duty to engage in

1 For present purposes, I define harmful speech broadly, comprising legally defined hate speech and incitement to hatred as well as forms of discriminatory speech which do not fit legal standards. Moreover, part of the paper deals with theories about speech constituting, as opposed to causing, harm.

counterspeech. Both Howard's proposal drawing on positive duties and the idea of complicity by failure to counterspeak will be considered, as well as their implications for offline and online speech (5-6).

2. Insights from philosophy of language

The concept of counterspeech denotes a non-coercive and non-censoring method for reacting to harmful speech, with the aim of impeding or at least diminishing its damaging effects. Usually, advocates of counterspeech do not exclude legal bans or content takedown as complementary strategies in order to achieve such goals. Yet, they trust in the power of speech as an effective and democratic valuable remedy to speech-related harms. Speech can be harmful in various ways – it can directly harm the feelings of the audience (both targeted and not), induce to harmful actions, contribute to the establishment and functioning of structural injustice. According to philosophers of language, speech can also constitute, as opposed to cause, harm, if certain conditions are given. I will start by considering this approach, since it offers valuable insights both on the ways in which speech can be harmful and on related counterspeech strategies – though, focusing on offline speech, it is not entirely able to match digital speech challenges, as should become clear in section 4.

Integrating philosophy of language conceptual resources with feminist critical thinking about language and systemic oppression, Rae Langton's seminal work elaborates on the topic of speech acts, specifically on the ways in which speech can *constitute* harm. Langton adopts John Langshaw Austin's distinction between locution (the content of an utterance), illocution (what the speaker does with the utterance, like ordering or promising) and perlocution (the multifarious causal effects of the utterance), in order to show that pornography can not only have harmful perlocutionary effects, but can indeed have normative illocutionary force, constituting a speech act subordinating and silencing women (Langton, 1993).

An utterance can have illocutionary force and thus constitute harm if some conditions are met (felicity conditions). Langton focuses mainly on authority felicity condition, particularly on speaker authority. To grant illocutionary force to an utterance, authority does not need to be institutional, nor formal. It suffices that the speaker has authority in the specific contextual domain in which speech is uttered.

There are various sources of authority, but I will focus here on one particular source, namely authority gained by default. As Ishani Maitra (2009) shows, a speaker can obtain authority in a given activity by simply assuming to have it: if the other participants and bystanders do not raise objections, the speaker's presupposed authority can become factual by default. Bystanders' silence in the face of harmful speech can have a licensing effect, i. e. the effect of conveying acceptance of speaker's authority. Failing to object can legitimize authority even if such an effect is unintended and even in case of dissent, if dissent remains unexpressed.

The crucial role of silence, understood as the omission of objection, in granting force to harmful speech has been explored also by Mary Kate McGowan, yet from a different perspective. McGowan focuses on one particular way in which speech can constitute harm, namely the dynamic process of adjustment and accommodation occurring within conversations (McGowan 2018; 2019). Accordingly, every contribution to a conversation can modify the conversational score, which is defined as the set of all elements counting as appropriate for that conversation. This includes, but is not limited to, the psychological and epistemic common ground of participants. In order for a contribution to change the conversational score, no authority is required. Indeed, changes in the conversational score occur routinely and mostly in a covert, "sneaky" way, remaining unnoticed by speakers and hearers. If none of the participants raises an objection, score changes are automatically accommodated in the conversation. This applies also to harmful presuppositions and associations, like sexist remarks, which, if not challenged, are automatically accommodated.

Utterances triggering implicit bias like "C'mon, even a woman could do it!" (Ayala & Vasileva, 2016, p. 206) are a case in point.

It should be noted that both authority felicity condition and conversational accommodation appear to hinge, at least in part, on the failure of hearers to object. This can have notable normative consequences: as Maitra notes, if remaining silent in the face of harmful speech can contribute to it, a moral obligation to speak up could follow (Maitra 2009, p. 20). This idea, as well as its implications for online speech, will be sketched below in section 6.

Philosophy of language insights downright suggest a particular counterspeech strategy. Since speech can constitute harm only in case certain felicity conditions are met, making such conditions unavailable would hinder speech-related harm. Langton talks of "blocking" harmful speech felicity conditions (Langton, 2018a). Hence challenging the speaker's epistemic and practical authority, or drawing attention on presuppositions covertly creeping into a conversation in order to undermine their accommodation appear viable ways of counterspeech. Though, counterspeech has shortcomings[2] and in particular the strategy of drawing attention to presuppositions.

3. Lepoutre's alternative approach: *ex ante* positive counterspeech

In a recent paper, Maxime Lepoutre addresses precisely this issue (Lepoutre, 2019). Explicitly focusing on speech spreading misinformation, Lepoutre outlines two kinds of it: speech diffusing falsehoods about policies (a case in point: "vaccine causes autism" rumors) and about persons, based on their group membership. The latter is a form of hate speech, delivering a distorted representation of out-group members.[3]

Lepoutre is concerned with the question of how to draw speech strategies which can effectively counter misinformative speech. Even if they cause different kinds of damages, the two prongs of "ignorant speech" (Lepoutre, 2019, p. 155) share a crucial feature: they can profoundly affect the beliefs of the audience, making it particularly difficult to eradicate the false beliefs – and thus to counter the related harms.

In order to account for this phenomenon, Lepoutre draws on the works of McGowan and Robert Simpson, among others. As Simpson (2013) explains by expanding McGowan's theory of presupposition accommodation, introducing an association (say, associating out-group members with dangerous attitudes or demeaning attributes) in a conversation or in a public debate makes such association *salient* in that context: this means, that the attention of the audience is drawn to such an association, increasing the possibility that someone would believe it.[4] This is likely to be the case especially if the association corroborates already preexistent bias or if it prospects a potential danger. Research on stereotyping and cognitive bias confirm this mainly unconscious process (Blum 2004). Associations are sticky: they are difficult to reverse. Hateful utterances and negative information are especially sticky (Simpson 2013). This explains why counterspeech attempting to challenge such associative claims can possibly backfire: since negating an association requires naming it, every counter-utterance

2 For a critical assessment of counterspeech flaws, see Mc Gowan 2018; Langton 2018a, pp. 16-18.

3 Note that Lepoutre does not focuses on hate speech, but on speech spreading falsehood, though maintaining that falsehood about groups can be crucial in hate speech. I endorse Lepoutre's view that discriminatory speech and hate speech are often corroborated by bias and falsehood. I should add that the epistemic component indeed plays a huge role in the debate about harmful speech and counterspeech.

4 Simpson, as well as Lepoutre, appear to use the adjectives *salient* and *relevant* interchangeably, endorsing an understanding of salience which is commonly used in psychology: salient is something that stands out to our perceptive and cognitive attention. For present purposes, I adopt this understanding of salience, for reasons that should become clear in section 4 in relation to Facebook's definition of relevance. McGowan uses salience in a more narrow and technical way, meaning picking up one particular noun or pronoun from a broader category within a conversation (see also Simpson 2013, footnote 17).

can make the association even more salient. "X are not lazy parasites" (Lepoutre 2019, p. 161) can be an instance.
Hence, as McGowan notes (2018, p. 191), in the attempt of designing an effective counterspeech strategy, we face an uncomfortable challenge: on the one hand, not speaking up could possibly back speech-related harms, favoring felicity conditions such as authority by default and presupposition accommodation. On the other hand, by replying to harmful speech one risks to strengthen its impact.
At this point, Lepoutre introduces his own suggestion for a counterspeech approach addressing falsehood which can avoid the "stickiness" impasse. Firstly, instead of negating false information and associations, counterspeech should be of a positive type, conveying facts and aspirational values. Secondly, Lepoutre recommends diachronic counterspeech: instead of reacting *ex post* to speech already uttered - which must deal with the impossibility of making at least some perlocutionary harmful effects retroactively undone - the diachronic approach aims at building an epistemic and moral common ground which is hostile to political and hateful misinformation. This should be achieved "in at least two ways: by entrenching important facts in the conversation's common ground, and by eroding ignorance-promoting speakers' status, so that they no longer count as authorities" (Lepoutre 2019, p. 182). Once such a background is created, misinformative utterances and hateful associations should hardly be possible, or, more accurately, they would not find the felicity conditions that can make them harmful, such as authority and the covert accommodation of harmful presuppositions. Lepoutre illustrates this with the example of a democratic elected head of state explicitly affirming that the government categorically rejects racism: "The hate speaker who *then* pronounces deeply racist views thereby marks himself as a minority voice, who cannot speak for the majority" (Lepoutre 2019, p. 182).
Some critical remarks are in order. Lepoutre appears to propose a form of citizens inoculation, which should be carried out by the state, against hateful ideologies. I do not dispute the legitimacy and possibly also the necessity of such enterprises. Yet when one considers that similar efforts were made in European Countries after the Holocaust, especially and very systematically in Germany, and one looks at the current rise of far right-wing discriminatory ideologies, including in that Country, then some doubt arises as to the effectiveness of the inoculation approach.
It should be noted that on the contemporary political scene worldwide racism, as well as right-wing extremism, cannot be properly defined as marginal phenomena. Currently, it appears not easy to reverse perceived authority, or factual acceptance, of racist and discriminatory speech in the way Lepoutre suggests. Possibly, it is already time to react, rather than to prevent. Moreover, even if inoculation strategies can do their part – and in some specific historical and political contexts, a huge part – in making the terrain unsuitable for speech to be harmful, the concept of counterspeech explicitly refers to a response, and thus to an *ex post* reaction to speech. Thus, Lepoutre partly bypasses the question of how to design viable counterspeech strategies.
Undoubtedly, since digital communication has been expanding as one of the most powerful means for conveying speech, research on counterspeech is presently urged to take online speech specificities into account. Since the philosophical positions outlined above focus on traditional forms of speech and counterspeech offline, the question arises as whether the tools they provide can be capable of matching the specific challenges of digital communication. This is what I would now like to explore.

4. Digital conversational dynamics

The recent proliferation of ideologies of discrimination and of far-right groups and political parties is accompanied by the rise of social media, which in many relevant respects differ from one-way communication. Importantly, social media are designed for interaction, in part following traditional non-digital conversational patterns, in part radically diverging from them. Physical unavailability, anonymity, spatial reach and temporal pervasiveness are just the most evident features of digital communication. Moreover, the peculiar interaction patterns of social media, made of an interplay of posts, re-posts, likes and comments require a rethinking of the relationship between speech and salience, confirmation bias and authority.
Let start by considering the problem of speaker's authority. As discussed above, authority can be granted by default, i. e. by the audience's failure to question the presupposed speaker's authority. The examples provided by Maitra, Langton and McGowan depict almost exclusively situations of small-scale personal interaction, mainly in public spaces: that of a man targeting an Arab woman with racist speech in a crowded subway car (Maitra 2009, p. 19) being a case in point.[5]
Now consider a racist post by an unknown private user on a public online platform.[6] Is the "authority by default" model, transposed in the context of digital communication, still convincing? I think the answer is no. Other users' failure to react to the racist post would be unlikely to have the effect of legitimizing it. Maintaining that the post can still have harmful perlocutionary effects on readers, both targeted and not, if nobody takes it up it can hardly exercise normative force, nor can its author be granted epistemic and practical authority in that digital conversation.
This does not mean that in digital communication other users cannot grant authority to a speaker, to a speech act or to a particular conversational tone. On the contrary, in the digital context speech impact - and the phenomenon of granting authority in particular - appear to be dependent on what other users do in a fundamental way. Though, it is not other users' omissions, but rather their active participation, which can substantially determine processes of legitimation and normalization; active participation meaning answering, sharing, rating or linking other users' contributions.
Importantly, the structure of harmful speech on social media is not, or by far not only, that of an authoritative speaker and an audience, according to a "top down" model of communication. More aptly, it can be described as a choral process of mutual confirmation and validation. Political leaders who do not endorse social equality values have quickly learned how to use digital communication participative patterns to strengthen their legitimacy.[7] A look at the online communication of right-wing political leaders reveals accurately calculated propaganda strategies: carefully avoiding explicit racist speech, which would probably backfire, though sometimes intentionally slipping into it, winking to supremacist or fascist followers, far-right leaders leave, more often than not, the "dirty job" of the most explicit hateful speech to their followers.
Beyond the use made by political leaders and on a more general level, other users' interaction by replying, linking and liking is necessary for posts and comments to get attention, fundamentally determining digital processes amplifying the impact of that speech. Such processes follow partly social patterns of communication and interactional dynamics

5 Langton briefly touches upon digital communication, for instance in 2018a, p. 2, though not elaborating on it.
6 I focus here on public communication platforms, since they play a crucial role in shaping public discourse about online harmful speech. In small private digital conversations among participants knowing each other, things may look different.
7 Thanks to an anonymous reviewer for bringing the work of Jennifer Saul (2017) and Tail Mandelberg (2001) to my attention.

characterizing offline speech: the attention, and the interest, of an audience is most likely to be attracted by speech already getting a considerable uptake.
But importantly, digital communication is also driven by automatic digital processes, which do not depend from patterns of human interaction. Besides human users, digital conversations are crucially shaped by non-human "participants", namely algorithms, which are responsible (in a technical sense) of some decisive, routinely activated online conversational dynamics. Facebook, for instance, automatically ranks posts on public pages according to the number of interactions. Closely connected with Facebook's dynamic ranking process is the visibility of posts and comments: they are shifted between top and bottom of the comment thread, which has a substantial impact on how many users will see and read them. In this way, contributions to a digital conversation obtain what Facebook calls *relevance*.
In light of this, we can reconsider McGowan's thoughtful insights on conversational kinematic, specifically on presupposition accommodation dynamics. It should be noted that McGowan is explicitly concerned with small-scale conversations occurring face to face. As explained above, her theory has already been expanded by others, notably Simpson and Lepoutre, in order to cover broader speech contexts and with a particular focus on the phenomenon of the "stickiness" of presuppositions and expectations introduced in a conversation. Now, it could be useful to consider the implications of the abovementioned digital speech dynamics for McGowan's theory and especially the role played by algorithms in determining speech relevance. We can hardly maintain the view that every contribution to a public platform digital conversation automatically changes the score, i. e. what counts as appropriate in that conversation, until someone tries to block this accommodation process. On the contrary, in absence of other users' active interaction, digital conversational contributions do not have much chance of affecting the score of that conversation.
Or more carefully, we could indeed maintain McGowan's view, if we integrate it with the consideration of the cumulative effect of interactions on further development of the conversation. Not the single contribution, but the aggregative effect of more interactions with that contribution can change the permissibility rules of the public digital conversation. In this process, human active interactions (rather than omissions) and algorithmic dynamics are closely interconnected. They determine which content stands out to attract users' attention and thus its salience (using Facebook's terminology: its relevance) which is crucially dictated by social media platform standards.
Beyond the level of the changes produced in a particular conversational score, digital speech relevance can have a huge impact on users' beliefs and behavior. The concept of norm perception, which has been developed in psychology, can be useful for understanding the relationship between online ranking, beliefs, values and behavioral choices. While individuals tend to perceive as norms the standards of persons and groups with which they identify (Tanckard & Paluck, 2015), they can also be conditioned by the value judgement of the majority, as Genocide Studies show: "sufficiently suffuse an ideological environment, so that a belief becomes something like 'everybody says' and that is liable to receive wide endorsement even if never properly substantiated" (Maynard 2014, pp. 10-11).
Though empirical work on this topic is surely needed, the impact of social media ranking on the perceptions and decisions of users is already well known - specifically, in the form of the relationship between accounts ranking and influencer economy. Further research on the phenomenon of online influencers' epistemic and practical power could surely shed light on the topic of speaker's authority in the digital domain. Needless to say, digital conversational dynamics are fundamentally driven by and exploited for economic interests.
However, the phenomenon of ranking can be turned to the advantage of counterspeech efforts as well. Activist counterspeech groups around the world are already practicing such approach,

showing a fresh way of engaging in effective action beyond the dichotomy of *ex ante* positive speech and *ex post* critical counterspeech.
The Swedish group #jagährär[8] have implemented an organized system of specifically online-tailored, collective counterspeech. #jagährär is not bound to a political party, and is framed by a set of rules, based both on democratic values like freedom of speech and diversity of opinion and on considerations of effectiveness. In order to avoid the drawback of making the posts they want to counter more relevant through their interactions (the problem MCGowan, Simpsons and Lepoutre are concerned with, though relating to offline speech), #jagährär activists rather focus on interaction with positive comments and counterspeech posts, in order to shift them higher in the comment thread. Researcher Susan Benesch explains: "Most of the news outlets have their comments ranked by what Facebook calls 'relevance'. Relevance is, in part, determined by how much interaction (likes and replies) a comment receives. Liking the counterspeech posts, therefore, drives them up in relevance ranking, moving them to the top and ideally drowning out the hateful comments." (Benesch 2019).[9]

5. Counterspeech as a positive duty to rescue: Howard's proposal

While the project of organizing counterspeech groups of private citizens appears praiseworthy, a related question emerges, which, as Jeffrey Howard rightly recognizes, has been unjustly neglected in the literature (Howard, 2018). Do individuals have a moral duty to engage in counterspeech, and if they do, which is the moral source of this duty?[10]
Touching upon Brettschneider's important work on state counterspeech (2012), Howard argues that individual action and state engagement in speaking back are not mutually exclusive, rather, they are complementary. Focusing on the question of how to justify the moral duty to counterspeak of private citizens, Howards concentrates on a specific kind of harmful speech, namely "speech that implicitly or explicitly encourages wrongful criminal violence, such as speech that advocates terrorism and incites racial hatred" (Howard, 2018, p. 2).
According to Howard, the normative source of the duty to counterspeak against incitement to violence is "the Samaritan obligation to rescue others from risks of harm" (Howard 2018, p. 1). Howard considers the Samaritan obligation a natural duty, pertaining to every moral agent, under the condition that the obligation is not too demanding (Howard 2018, p. 6). Since the counterspeech duty requires just to speak, and not to engage in dangerous rescue operations, the obligation is, *prima facie*, not too demanding.
Despite valuable insights on the counterspeech topic, Howard's route appears unconvincing in many respects, both on a general level and specifically when related to digital speech.
Firstly, Howard postulates the Samaritan duty to rescue others from risks of harm as universally accepted: "The general idea that we have natural duties to defuse unjustified threats posed by others is not controversial" (Howard 2018, p. 6). However, this issue is by far more controversial, both in philosophy and law, than Howard admits.[11]
The duty to rescue is *prima facie* imposed to private citizens under further specific conditions, in addition to the requirement that the intervention is not too demanding. There are various approaches to this. Generally, the duty relates to life-or-death-cases, in which the rescuer,

8 https://www.jagarhar.se/. Presently, #jagährär groups are present in 12 different countries. See also Cathy Buerger 2020.
9 On using "like" interactions as a strategy for countering ISIS propaganda, see Yadron 2016.
10 Though in philosophy the concepts of moral duty and moral obligation are sometimes differentiated, for present scope they are used interchangeably.
11 See for instance Schiff 2005, Dressler 2000.

standing close to the perilous event, is capable of effectively intervening. The example of a good swimmer's duty to rescue a drowning person is a case in point.
Incitement cases are structured differently, involving a speaker (the inciter), the possible perpetrator of violence and a bystander (the possible bearer of the duty to counterspeak).
In Howard's example (2018, p. 6), an inciter is intentionally striving, with high chances of succeed, to induce someone to kill. In such clear structured cases, there can well be evidence of serious peril, and trying to dissuade the speaker or the perpetrator can be plausibly framed as a way to fulfill the moral duty to rescue; However, I would argue, in this case the duty also entails the moral (and possibly, legal) obligation to report to authorities and under certain conditions even to physically impede the violence, if the counterspeech strategy does not prove effective.
So, if Howard has just this type of cases in mind, then his argument of extending the duty to rescue to cases of incitement to violence can be plausible. But this does not seem to be the case. Indeed, Howards is also concerned with "speech that risks inspiring agents to engage in criminal violence" (Howard 2028, p. 5), which makes the analogy with life-or-death rescue cases less convincing. The category of speech that *risks* inspiring to criminal violence is very broad, surely including other kinds of speech rather than just incitement (whose legal definitions, incidentally, usually include the *purpose* to induce someone to violence).
It is extremely difficult to establish if and in which way a specific utterance would lead the audience to violence, since the effects of speech are highly context-dependent, hinging on a broad range of variable factors. Moreover, it should be noted that the variety of speech which is apt to inspire to violence is more often than not harmful even in the absence of such an effect: speech unjustly discriminating, subordinating, propagating distorted information such as the depiction of the target as dangerous can lead to grave harms, including the establishment and legitimation of social injustice and rights deprivation. Grounding a duty to counterspeak on the moral obligation to rescue from criminal violence leaves these harms unaddressed.
Thus, the duty-to-rescue route raises the following quandary: either the counterspeech obligation addresses just inciting speech which would bring about criminal violence with a high degree of probability, in which case other kinds of regulations could be better suited. Or the duty has a wider scope, including speech that risks inspiring to physical violence: then we face both the huge empirical difficulty of singling out which speech should be countered, and the problem of leaving other harmful effects unaddressed. Alternatively, we further relax the requirements, abandon the focus on physical harm in order to address other harmful effects and establish a general individual duty to counter potentially harmful speech – which would have troublesome implications, even if only for the vagueness of its scope.
Moreover, Howard does not seem to consider the specificities of digital speech. In the contemporary world, the main (though not exclusive) vehicle of incitement to terrorism and racial hatred is online communication. Words of hatred, refusal and dehumanization, scapegoating and misrepresentation of the other as a threat are spreading from myriads of digital sources, amplified by interactive processes and echo chambers among global users. Incidentally, it should be noted that the practice of linking and liking is not always driven by hateful purposes, nor accompanied by awareness of related harms: a subtle dynamic of reward and gamification-based incentives binds the users to social media participation and interaction. Facebook first president Sean Parker's comments are illuminating, explaining how the "like" button was specifically designed to induce in the users "a little dopamine hit" (Deibert, 2019, p. 30).
However playful the participation to hateful speech online may be: the harms it contributes to are multifarious and steered by specific digital conversational dynamics. Howards suggested

counterspeech strategies - among others, to change speaker's and audience's emotional and moral attitudes "introducing her to a new person, recommending a particular film or simply telling her a provocative story that powerfully conveys an alternative narrative" (Howard 2018, p. 14) - do not appear to be best suited for meeting the specific challenges of digital harmful speech.

6. The negative duty approach: complicity by failure to counterspeak?

Beside the proposal of grounding counterspeech duty in the moral obligation to rescue others from risks of harm, and thus in the positive duty to make others better off, it is worth considering an alternative strategy, drawing on the negative duty not to harm others. In the literature, this route has been repeatedly touched upon, though not systematically elaborated. As Maitra (2009) notes, failure to raise objection can contribute to speech-related harm, grounding the responsibility of a silent bystander as an accomplice in that harm; complicity being a form of responsibility for (reasonably foreseeable) contribution to harm committed by someone else.[12] Of course, in order to make someone accountable as an accomplice, some further conditions must be met, in addition to a possible causal contribution. Defining the appropriate epistemic, intentional and causal requirements for complicity is a challenging task, and the more so in cases of complicity by omissions.

Though, the phenomenon of contribution-by-silence and the concept of bystander complicity reveal important insights relating to moral choice. Notably, in the face of harm a bystander's options for action do not always include a threefold range of possibilities: actively supporting harm, objecting, or remaining neutral. The idea that a bystander's inactivity does not affect the setting, leaving things happening in the very same way, as if the bystander was not there, is, more often than not, an illusion. In case of small-scale offline context, the presence of a bystander can have remarkable causal and normative implications. Bystanders are in a privileged position, witnessing what is going on and having the chance to directly intervene to counter harm, which could make them the natural bearer of the duty to engage in counterspeech in that specific situation. Moreover, other actors being aware of bystander's presence can enact expectations, leading to record the omission of intervention and generating a variety of psychological and social harmful effects.

Though, these insights are not transferable to the context of digital communication without more work. Digital conversations on public platforms indeed seem to concede the possibility of remaining neutral by failing to interact: since what other users - and algorithms - track is active participation, not omission, it seems possible to look away from harmful speech without impacting on it. Thus, a user running into a potentially harmful digital conversation - becoming a virtual bystander in the privileged position of epistemic awareness and intervention capability - can simply decide to leave the platform and navigate into more peaceful digital waters. Nobody would notice, no expectation would arise, nothing seems to qualify this particular user as bearer of counterspeech duty.[13] The analogy between digital users and offline bystanders seems of little use to an account of digital counterspeech obligation based on the negative duty not to contribute to harm.[14]

However, digital harmful speech could be framed from a broader perspective, as a large-scale process systematically improving the erosion of aspirational values such as social justice and human rights. Refraining from dealing with this, shirking one's own political responsibility,

12 This definition is widely shared in law and philosophy, diverging both from Brown's (2019) and from definitions requiring participatory intention.

13 Note that the name of the Swedish digital counterspeech group #jagährär means "I am here".

14 See also Brown 2019, p. 217.

means leaving the development of unjust social patterns and institutional orders in the hands of others – which can still be a way of reinforcing it.
Perhaps, framing the question of why we should engage in counterspeech in terms of a moral duty, attempting to ground this duty in a single general moral obligation – either to benefit, or not to harm others – is neither adequate, nor necessary. We indeed have a variety of good moral and political reasons for taking an active part in countering both small-scale and systemic harm, and speech can be a powerful means at our disposal. This implies becoming aware of how counterspeech can be effective, and of possible shortcomings. Learning new ways in which digital technology can promote harm – but also how it can be effectively used against it – are mandatory first steps.

REFERENCES

Ayala, S. & Vasilyeva, N. (2016) Responsibility for Silence. *Journal of Social Philosophy* 47(3), 256-272;
Benesch S. & Jones D. (2019, August 13) Combating Hate Speech through Counterspeech. Retrieved from: https://cyber.harvard.edu/story/2019-08/combating-hate-speech-through-counterspeech;
Blum, L. (2004) Stereotypes and Stereotyping: A Moral Analysis. *Philosophical Papers* 33(3), 251-289;
Brettschneider, C. (2012) *When the State Speaks, What Should I Say?* Princeton: Princeton University Press;
Brown, A. (2019) *The Meaning of Silence in Cyberspace: The Authority Problem and Online Hate Speech*, in: S. Brison, K. Gelber, Free Speech in the Digital Age. Oxford: Oxford University Press;
Buerger, C. (2020, December 14) The Anti-Hate Brigade. Retrived from: https://dangerousspeech.org/anti-hate-brigade/;
Citron, D. & Richards N. (2018). Four Principles of Digital Expression (you won't believe #3!). *Washington University Law Review,* 95, 1353-1387;
Deibert, R. (2019) The Road to Digital Unfreedom: Three Painful Truths About Social Media. *Journal of Democracy*, 30(1), 25-39;
Dressler, J. (2000) Some Brief Thoughts (Mostly Negative) About "Bad Samaritan" Laws. *Santa Clara Law Review* 40, 971-989;
Howard, J. (2019) Terror, Hate and the Demands of Counter-Speech. *British Journal of Political Science* 1-6. doi:10.1017/S000712341900053X;
Klonick, K. (2018) The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review* 131, 1598-1670;
Langton, R. (1993), Speech Acts and Unspeakable Acts. *Philosophy and Public Affairs*, 22(4), 292-330;
Langton, R. (2018a), Blocking as Counter-speech, in: *New York on Speech Acts*, Oxford University Press;
Langton, R. (2018b), *The Authority of Hate Speech*, in: J. Gardner, L. Green, B. Leiter (eds.), Oxford Studies in Philosophy of Law, vol. 3, Oxford: Oxford University press 123-152;
Lepoutre, M. (2019). Can "More Speech" Counter Ignorant Speech? *Journal of Ethics and Social Philosophy*, 16(3), 155-191;
Maitra, I. (2012) *Subordinating Speech*, in: I. Maitra, M.K. McGowan (eds.), Speech and Harm: Controversies Over Free Speech. Oxford: Oxford University Press;
McGowan, M.K. (2009) Oppressive Speech. *Australasian Journal of Philosophy* 87(3), 398-407;
McGowan, M.K. (2018) *Responding to Harmful Speech. The More Speech Response, Counter Speech, and the Complexity of Language Use*, in: Johnson, C.R. (ed.), Voicing Dissent. The Ethics and Epistemology of Making Disagreement Public. London, New York: Routledge;

McGowan, M.K. (2019) *Just Words: On Speech and Hidden Harm*. Oxford: Oxford University Press, 2019;
Schiff, D. (2005) Samaritans: Good, Bad and Ugly: A Comparative Law Analysis. *Roger Williams University Law Review,* 11(1) 77-141;
Simpson, R. (2013) Un-Ringing the Bell: McGowan on Oppressive Speech and the Asymmetric Pliability of Conversations. *Australasian Journal of Philosophy* 91(3), 555-575;
Tankard, M., Levy Paluck E. (2015) Norm Perception as a Vehicle for Social Change. *Social Issues and Policy Review,* 10(1) 181-211;
Yadron, D. (2016, Januar 21) Facebook's Sheryl Sandberg: "likes" can help stop Isis recruiters. The Guardian. Retrieved from: https://www.theguardian.com/technology/2016/jan/20/facebook-davos-isis-sheryl-sandberg.

HENK JASPER VAN GILS-SCHMIDT
Independent scholar
henkjaspervangils@gmail.com

HILDE LINDEMANN'S COUNTERSTORIES: A FRAMEWORK FOR UNDERSTANDING THE #METOO SOCIAL RESISTANCE MOVEMENT ON TWITTER[1]

abstract

This paper proposes a framework for understanding and analysing online social resistance movements based on Hilde Lindemann's concept of counterstories (Damaged Identities, Narrative Repair, *2003). This framework is based on the premise that we shape our identities in shared social spaces, and that such shared spaces are structured according to so-called 'master narratives'. Master narratives define the 'realm of possible identities' that we can assume, and form the basis for either recognizing or denying recognition to various social groups in specific roles that they might occupy. Social oppression occurs when master narratives preclude or forbid a certain form of self-expression, or alternatively force members of a specific social group into a determinate societal role (say, women who receive recognition only in the roles of mother or housewife). Counterstories serve as a corrective to these aspects of oppression by challenging the oppressive facets of master narratives. Based on this framework, I propose an interpretation of the #MeToo movement as a counterstory that aims to change the oppressive aspects of the patriarchal master narrative that (partially) structures many shared social spaces in the modern Western world. I end this paper by applying the framework to consider potential obstacles #MeToo may encounter as a distinctively online movement.*

keywords

social resistance, counterstory, master narrative, Twitter, #MeToo

1 I would like to thank participants of the workshop 'Sisterhood on Twitter' in Tübingen and of the PF Kolloquium at Greifswald for the insightful discussion. I also would like to thank two anonymous reviewers for their extensive and invaluable remarks. This work was supported by the *Evangelisches Studienwerk Villigst* (PhD Fellowship) and the University of Antwerp (DOC-PRO1 Fellowship).

Phenomenology and Mind, n. 20 - 2021, pp. 88-99
DOI: https://doi.org/10.17454/pam-2008
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Social media platforms constitute online technologies for social interaction and community building. Since the establishment of social media platforms such as Facebook, Twitter, and Snapchat, users have been able to express themselves online, often to a global audience. Such platforms are used to tell stories, to discuss problems and to share information. Although they are often used to share positive experiences, many members of marginalized groups have come to appreciate these platforms for the role they can play in resisting social structures (as indicated by the use of hashtags such as #MeToo, #BlackLivesMatter, #Occupy, or #IfTheyGunnedMeDown). The role of social media platforms in providing a shared social space for social resistance movements is still under-investigated. As a basis for future inquiries aimed at understanding and evaluating online social resistance movements, I propose an overarching interpretive framework in which Hilde Lindemann's concept of counterstories (2003) takes central stage. After introducing this conceptual framework in Section 2, I demonstrate in Section 3 how it can be applied to the Twitter-based #MeToo movement. In Section 4, I use the framework to highlight potential obstacles to the online #MeToo movement posed by Twitter's status as an online public platform.[1]

2. Identity, Master Narrative, and Counterstory

To understand and evaluate online social resistance movements, I propose a conceptual framework based on Lindemann's concepts of identity, master narrative, and counterstory. For Lindemann, our identities are narrative constructs created in an intersubjective, shared space: "identities are collaborative ventures requiring a number of people to bring them into being" (p. xiii).[2] Our identity – our autobiographical narrative – weaves together "acts, experiences, and characteristics we care most about, and the role, relationships, and values to which we are most deeply committed" into a story of our life (p. 71). By embedding our past and future actions within the broader story of our lives, our identity functions both as an interpretative framework to make our past actions intelligible to ourselves and as a decision-making framework to guide our deliberations (p. 70). As our identities guide our decisions, we reveal our identities – who we are – through our actions (p. 25).
However, because our identities are collaborative ventures – taking shape in intersubjective

1 Related hashtags are #NiUnaMenos in Latin America (which preceded #MeToo), #QuellaVoltaChe in Italy, and #TimesUp. #YesAllWomen is a powerful example similar to #MeToo (see Rodino-Colocino, 2014 for discussion).
2 Unless otherwise indicated, page numbers in parentheses refer to Lindemann's *Damaged Identities, Narrative Repair*.

spaces – we are constrained by the social contexts in which we live and act. This means that we are not entirely free in deciding how we express ourselves through our choices and actions, and are thus also restricted in what life-story we can create for ourselves: "freedom of agency requires not only certain capacities, competencies, and intentions that lie within the individual, but also recognition on the part of others of *who one is*" (p. 24).[3] For example, if someone values 'being employed' as a way to express herself, she will depend on others to see her as a competent employee – an ascription historically denied to many women. We are only able to express ourselves by taking up a specific role (and thereby integrating this role into our identity) if others grant us the necessary recognition.

Lindemann identifies widely shared master narratives as important grounds on which recognition is granted. Master narratives are "stories found lying about in our culture that serve as summaries of socially shared understandings" about the available roles we (can) take up (p. 6). At the same time, master narratives function as the "repositories of common norms", providing an interpretative framework for assessing the self-expressions of others (pp. 6-11). In this way, master narratives to a large extent govern what we perceive as justified self-expressions by the people with whom we share our social space (p. 6). For instance, the master narrative 'all human beings are born free and equal' undergirds (in the eyes of ourselves and others) our common possession of the right to determine our own lives.

As we need recognition from others, the construction of our identities depends on more than merely making our actions intelligible to ourselves by fitting them into our personal life-stories. We also need to make ourselves intelligible to others by fitting our actions into the dominant master narratives that structure the social spaces we share.

2.1. Master Narratives and Social Oppression

Within this framework, a social group may be oppressed in a given society if prevalent norms deprive this group of "valuable roles, relationships, and goods" (p. 7). For example, in modern Western societies, women have long been pushed into taking up the roles of housewives and mothers. According to norms implicit in the master narrative 'family is the cornerstone of society', women are not typically recognized as 'autonomous' or 'able to be employed'. This prevailing narrative may limit the extent to which women can pursue independent lives. The influence of this master narrative in our Western societies continues today (see, for example, Gotlib, 2016), as the expectation that mothers should function as the primary caretakers of children is still widespread. This makes women less free than men to pursue the expression of other aspects of their identities, such as a career, fulfilling friendships, and hobbies.

Dominant master narratives are reproduced in societies in many different places and ways, which extend beyond the control of any one individual. As individuals do not control the master narratives that govern much of their lives, this lack of control exacerbates the oppressive force of master narratives. Lindemann articulates this in a recent paper: "fragments of history, biography, film, fables, jokes, and similar narrative forms ring changes on the theme [of dominant master narratives], as do proverbs, music, advertising slogans, and other cultural artifacts" such as fairy tales or children's books like Harry Potter (2020, pp. 288-289). The fact that master narratives are reproduced in so many different places and forms makes it difficult for individuals to challenge them solely by themselves. To counteract the oppressive effects of a master narrative, a more united effort involving a large group of people that start to tell and reproduce a 'counterstory' seems necessary.

3 The idea of relational identity and autonomy is explored in feminist philosophy and recognition theory. For a useful introduction and overview, see Anderson (2014).

2.2. Counterstories – Aims

Lindemann introduces the concept of a counterstory as a way to challenge the oppressive relations created and reproduced by master narratives. A counterstory confronts "an oppressive but shared moral understanding", and attempts to shift it by "rejecting its assumption that people with a particular group identity are to be subordinated to others or denied access to personal and social goods" (p. 8). A counterstory is itself a narrative, one that reidentifies the oppressed group under different, non-oppressive norms. The counterstory introduces a new understanding of the oppressed group into the socially shared space, an understanding which gains greater authority over the norms by which we grant recognition to others as this story is shared and recounted. If successful, the counterstory starts to structure the socially shared spaces, reconfiguring the possibilities open to the oppressed and providing the social recognition necessary for taking on new roles in society at large.

A counterstory, by telling and reproducing an alternative story about an oppressed group, changes the social imaginary in relation to which roles are considered appropriate for the members of this oppressed group. For a counterstory to succeed, therefore, it must be able to change the interpretative framework of the dominant, oppressive group: "the [oppressors] would not only have to hear [the counterstory] but accept it and alter [their] behaviour accordingly" (p. 6). Since recognition is crucial to one's expressive ability, the behaviour and beliefs of the oppressor must be changed so that they can recognize these new identities.

Henrik Ibsen's 1879 play *A Doll's House* (Ibsen, 2008) offers a microcosm that illustrates the process of telling a counterstory. In the play, Nora attempts to convey to her husband Torvald that she is not merely a tool to assist him in the pursuit of his aims. At a pivotal moment in the play, Nora tries to explain to Torvald that she has duties to herself as well. Not understanding this, Torvald replies: "Before all else, you are a wife and a mother" (Ibsen, 2008, Act III). In an attempt to challenge the master narrative by which Torvald lives, Nora responds by telling a counterstory that presents both herself and Torvald as beings of equal moral worth, saying, "I don't believe that any longer. I believe that before all else I am a reasonable human being, just as you are – [...] a being of equal moral standing" (ibid.). Torvald, however, is unable to understand Nora because he is incapable of questioning the expectations and shibboleths of the 'traditional family' master narrative. This narrative ascribes a subservient role to women within the family in relation to men, whether fathers or husbands. Nora's attempts to emancipate herself within her marriage fail, and it comes as no surprise when she leaves Torvald at the end of the play.

2.3. Counterstories – Necessary Conditions

A counterstory thus aims to free both the oppressed and the oppressor from the shared norms that maintain the oppressive relation. What are the conditions under which a counterstory can be developed and told? The aim of a counterstory is to tell a positive story that reidentifies the oppressed group as equal to their oppressors. In describing this process, Lindemann stresses the importance of a 'chosen community'. A chosen community is one that individuals join voluntarily, for example a trade union, political action network, or support group (cf. Friedman, 1992).

The chosen community is important in that it provides a safe and private space for the construction of a counterstory. Chosen communities assist in the development of counterstories in three ways. First, the chosen community makes the oppressive nature of the master narrative (that the oppressor lives by) intelligible to the oppressed person (p. 31). In other words, a chosen community creates a basis for consciousness-raising. Second, the chosen community provides the recognition required to allow the oppressed person to imagine and build new, non-oppressive modes of self-identification: "the moral space created by the chosen community" provides "a space for reflecting on who the [oppressed] are and want to be" (p. 11). This allows the oppressed person to try on new perspectives, distinct from

those offered by the oppressive master narrative (p. 5). Third, chosen communities allow for the safe expression of developed counterstories, by supplying alternative evaluative standards (p. 10). These alternative standards are important because under the oppressive norms the oppressed person takes herself as well to occupy a submissive role. This makes it impossible to confidently imagine and communicate a counterstory. The alternative perspectives provided by a chosen community do more than merely position the oppressed on an equal moral standing equal with her oppressors; they also make the act of telling the counterstory in society at large conceivable to the oppressed herself (p. 10).
To summarize, a counterstory is developed within a chosen community that provides a safe and private social space. In this space, mutual recognition makes the act of telling a counterstory conceivable to the oppressed person, who can then retell the counterstory with the aim of changing widely accepted, oppressive norms. The counterstory proposes a more positive (self)understanding of the oppressed group as being of equal moral standing. If successful, the counterstory can by its own manifold reproduction impact society at large, eventually replacing the oppressive aspects of master narratives with alternative (counter) stories. This process opens up the eyes of both the oppressed and oppressor – changing the behaviour of both by offering alternative norms that grant recognition to the oppressed.

3. Social Resistance on Twitter: #MeToo as a Counterstory

Tarana Burke initiated the Me Too movement in 2006 to highlight the prevalence of sexual abuse of young women of colour and create "empowerment through empathy" (Just be Inc, 2013).[4] The movement took on a new global form after actress Alyssa Milano tweeted on October 15, 2017, asking women who had been sexually harassed or assaulted to "write 'me too' as a reply to this tweet" in order to draw attention to the problem of sexual harassment and assault.[5] In doing so, she aimed to place the focus on survivors of sexual assault, rather than their assailants (Sayej, 2017). The response to the tweet was overwhelming: within 48 hours, the hashtag #MeToo had appeared in over a million Tweets and more than 12 million Facebook posts worldwide (Park, 2017). Since this event, #MeToo has been identified as an online social resistance movement against the sexual harassment and abuse of women that continues to impact our lives (Jackson *et al.*, 2019; Xiong, 2019).
Although #MeToo had clear (and ongoing) offline consequences,[6] it predominantly developed online through shared testimonies and expressions of support. This makes #MeToo an interesting case study to evaluate the usefulness of the counterstory framework for understanding and evaluating online social resistance movements. In the following, I develop this line of enquiry through two questions. First, can we understand #MeToo as a collective act of telling a counterstory? Second, can the counterstory framework serve as a tool to critically reflect on #MeToo as an online social resistance movement capable of changing the relation between oppressed and oppressor? Section 3 discusses the first question, while Section 4 takes up the second.

4 Some commentators have argued that we need to distinguish between the Me Too movement started by Burke and the #MeToo moment instigated by Milano's tweet, see, e.g., Karen Boyle (2019, pp. 3-8).
5 Statistics show that one out of every two women in the EU have experienced sexual harassment or assault (European Union Agency for Fundamental Rights, 2014).
6 Virtually every industry has seen change as a result of the #MeToo movement. Examples include mandatory annual anti-harassment training for members of the US Congress, panic buttons for hotel workers in Chicago, and some luxury fashion brands introducing a Model Charter (involving, among other requirements, private spaces to be provided for models to change of clothes during events).

3.1. #MeToo as Counterstory: Shared Community and Challenging Master Narratives

To assess whether the #MeToo movement can be understood within the counterstory framework, we need to determine whether we can locate the central features of a counterstory within the #MeToo movement itself. Through the use of hashtags, Twitter has enabled survivors of sexual assault and harassment to create an extensive community. Within this community, survivors can connect and share their stories of sexual assault and harassment in a space where these stories are recognized and accepted by others. In an interview with *The Guardian*, Milano affirms that #MeToo movement has been effective as "a standing in solidarity to all those who have been hurt" (Sayej, 2017). This claim finds further support in a recent empirical study by Suk et al. (2019), which shows how the use of #MeToo created a network of acknowledgement "through testimony and witnessing about harassment and assault" (see also Gallagher *et al.*, 2019, p. 1).[7] An additional advantage of social media is its international reach, allowing women to build extensive communities across geographical locations to fight against their oppression.

Within the counterstory framework, #MeToo can be understood as a movement that fights social oppression by challenging widely accepted master narratives. These master narratives give our modern Western societies a patriarchal orientation that places women in oppressed and submissive relations (see Jackson, 2019; Hu *et al.*, 2020). #MeToo challenges many aspects of patriarchal master narratives; in what follows, I highlight some aspects that have been addressed in empirically-oriented literature.

Hu *et al.* argue that "[w]omen are supposed to be subservient in many aspects in order to become intelligible under the aggressive male gaze" (2020, p. 2). This enforced subservience undermines women's opportunity to build their own identities (Hu *et al.*, 2020, p. 1). This point is reflected in the above example of *A Doll's House*: within the patriarchal master narrative of our society, women's actions only make sense to members of the oppressing class if they conform to this expectation of subservience. It is this structure that makes the male 'gaze', or the dominant interpretative framework of patriarchy, so oppressive of women and their actions.

A further element of the patriarchal master narrative is 'the myth of chivalry', discussed in a recent analysis by Ilinskaya and Robinson (2018). This myth involves men telling themselves they 'would do anything for her': "[t]hey are being nice to women; telling them how beautiful and wonderful they are; making them feel loved and appreciated" (Ilinskaya & Robinson, 2018, p. 384). But underlying this myth is the oppressive idea that the pure and perfect woman needs to be taken care of. By implication, she cannot do this herself. According to this narrative, if a woman denies the advances of a man she is both casting doubt on his good intentions and character as well as acting contrary to her own interests. In other words, "this privileging narrative confers" to men a position of "dominance" over women, giving them "permission to control" women's actions (ibid., p. 386).[8] By posting '#MeToo', women assert that the fantasy of chivalry frequently underlies acts of sexual harassment and abuse, and that it is actually oppressive, domineering, and controlling. Moreover, the act of posting #MeToo is already a challenge to the subservient role women are supposed to accept under the patriarchal master narrative, as women claim control over their own narrative through the act of posting itself.

In her 1978 essay 'The Rape Culture', Herman argues that one oppressive myth of the patriarchal master narrative we live by is "that victims are usually the ones to be responsible for their victimization" (as cited by Hu *et al.*, 2020, p. 2). According to this myth, victims invite

7 Their study used computational approaches, such as part-of-speech tagging, dependency analysis, hashtags extraction, and retweet network analysis.

8 McAdams (2020) gives a personal account of these and other relevant aspects of her own experience.

abuse by behaving in seductive ways or wearing particular types of clothing. This myth creates the false idea that women have the power to stop men abusing them, because they are the ones "in charge of men's sexual interest" (McAdams, 2020, p. 26). This is illustrated by the following tweet posted in response to Milano: "I never told anyone because I thought it was my fault for sending the wrong signals" (Jackie, 2017). Research shows that this perspective is reproduced in the way police treat women who report sexual assault, asking questions such as what they were wearing at the time of the assault (see, e.g., see Maier, 2008; see also Hockett *et al.*, 2016 on rape myths) even though such questions are irrelevant to building a criminal case or identifying an offender. #MeToo provides an opportunity for survivors to counter this narrative of assault, and instead articulate the counternarrative that clothing is irrelevant. Here again, we see how #MeToo builds a community that can form a basis for consciousness-raising.

#MeToo also aims to raise awareness and change the behaviour of the oppressors. This is necessary as male offenders are often unaware of harm they have caused, particularly because the patriarchal master narrative upholds the idea that only 'deviant' men with low moral standards commit acts of sexual harassment and assault. This leads to the common response 'not all men!' from those who "feel that feminist critiques unfairly tarnish all men" when in fact there are only a few bad apples (Flood, 2019, p. 288). Evidence of this reasoning can be found in a study by Flood, who found that men tend to understand sexual harassment to mean actual sexual coercion, failing to recognize that sexist jokes and offensive gendered commentary are also examples of harassment. Instead, they tend to view such comments as natural expressions of "male sexuality" (Flood, 2019, p. 288). The #MeToo movement created an increased awareness that sexual harassment includes unwanted sexual attention, sexist jokes, and other gendered commentary, and at the same time revealed how prevalent this type of behaviour is among many men. Flood suggests that #MeToo served in this way as a "call to action among men" comprising "three key tasks" (2019, p. 285): first, to listen to the stories of women and take them seriously; second, to change their own behaviour and thinking; and third, to speak up if they see other men harassing women. This shows how the #MeToo movement can be understood as a counterstory aimed in part at opening up the eyes of the aggressor and demanding a changes in their behaviour.

By sharing their stories of sexual harassment and abuse, women created a counterstory that revealed the scale of the issue of unwelcome sexual advances and harassment from men. This counterstory was amplified through the power of Twitter, creating a community across time and geographical location. Moreover, #MeToo's counterstory is clearly aimed at changing the norms embedded in patriarchal master narratives. Used to assess the behaviour of both men and women, these norms are shared by oppressor and oppressed alike. I thus conclude that the #MeToo movement can be understood as telling a powerful counterstory that contradicts the prevailing patriarchal master narrative, which has tended to minimize or even justify the abusive and oppressive acts of men.

3.2. #MeToo as Counterstory: Changing the Norms of the Oppressor

To close this section, I want to make a qualification. On the surface, we might think that #MeToo has opened the eyes of the oppressors and impacted their perceptions and behaviour. The following tweet appears to provide an example of this: "Guys, it's our turn. After yesterday's endless #MeToo stories of women being abused, assaulted and harassed, today we say #HowIWillChange" (Law, 2017). But beneath the surface, the reality is more complex. In his article, Flood uses survey data to show that we "should not overestimate #MeToo's reach among men" (2019, p. 286). Although around 50% of men in two 2018 surveys from the US and UK reported changes in how they perceived (in)appropriate behaviour, around 40% of men in the US and 60% in the UK claimed they had not heard of the #MeToo movement (Flood,

2019, p. 287). Only about one third of the men interviewed in the UK had had a conversation about #MeToo. Perhaps even more tellingly, only 30% of the men interviewed in the US said that they had changed their behaviour towards women in response to the #MeToo movement (Flood, 2019, p. 292). These results suggest that #MeToo still has the potential to effect further changes in the attitudes and behaviour of men.

4. #MeToo on Twitter: Drawbacks of the Online Public Sphere

Despite the ways in which Twitter has empowered women (making it possible for them to connect, speak up, and support each other), this platform also comes with its drawbacks that may undermine the force of #MeToo as a counterstory. In this section, I use the counterstory framework as a tool to critically reflect on the hurdles faced by the online #MeToo movement. Due to space limitations, I will focus exclusively on Twitter as a public platform.

4.1. Diverging Voices within the Oppressed Group

One hurdle that in which the public nature of Twitter poses to the creation of a coherent counterstory is that communication regarding the new narrative takes place in public. This undermines the process by which new identities, roles and ways of acting are imagined and negotiated. This can be observed, for example, in the open letter written by a group of prominent French women led by Catherine Deneuve. The letter expressed concern that the #MeToo movement "includes a moralist backlash, [...] [in which] women's bodies and sex become again this forbidden territory and that a new moral order introduces a new censorship against the free movement of desire" (Poirier, 2018). Further, it claimed that men's "freedom to pester" is "indispensable to sexual freedom" (ibid.). As a consequence of such challenges, the message of the #MeToo counterstory becomes more diffuse. On this point, it is helpful to revisit Lindemann's emphasis on the importance of initially creating a counterstory within the safe and private confines of a chosen community. Creation of the counterstory in private enables the oppressed to negotiate their new identities together first, so that after developing an effective counterstory they can speak in public with a unified voice, thereby increasing the potential of the counterstory to replace the oppressive master narrative.

4.2. Who Controls the Act of Telling?

The public nature of Twitter also has the idiosyncratic consequence that the #MeToo narrative is, as Paul Dawson puts it, "not told by anyone" (2020, p. 979). The #MeToo movement mainly consists of individual posts and responses, collected together by use of a hashtag and regulated by an algorithm in the background. No one person or group is in charge of creating the counterstory. This has at least two consequences.

First, the counterstory itself may fall short of the aims imagined by those who started it. In a speech shortly after #MeToo took off, Burke described herself as "desperate to change the narrative about the #MeToo movement before it's too late". She said that she regretted the movement's focus on prosecuting powerful men, and wished instead to see a more fundamental change in society's (sceptical) attitudes towards women who speak out about sexual assault and harassment (Bitran, 2018). Furthermore, some have maintained that the emergent #MeToo narrative has solely encouraged survivors "to publicly reflect on their experiences without explicitly pushing perpetrators to examine their own behaviour" (Clark-Parsons, 2019, p. 11)[9] – something that should be a central purpose of a counterstory, as we have seen above. But because no one person or unified group is in control of the narrative as it emerges from individual posts, it can become difficult to develep one consistent and coherent counterstory.

A second negative consequence of Twitter's public design is that uncovering a counterstory from a set of individual posts becomes an interpretative act, leaving ample space for

9 This is in line with Flood's discussion (2019), as detailed above.

interpreters to construct their own versions of the narrative. As there is no clearly defined group in control of the counterstory, and a plurality of diverging voices can be heard, one person's understanding of the narrative may be very different from someone else's. As such, when the emergent counterstory is reproduced in mainstream media it may be framed in ways that have potentially harmful effects, as Starkey et al. (2019) show in their qualitative media framing analysis. For instance, some mainstream media outlets have portrayed the women posting under #MeToo as 'hysterical sluts' or 'uptight feminists'. A concrete example can be found in a *Washington Post* op-ed by Garrison Keillor, a former Minnesota Public Radio host, who "used his position to marginalize alternative narratives of assault while promoting a masterplot of women as hypersensitive" (Johnson, 2018, p. 52).

4.3. Inclusion of the Oppressed, Problematized

Clark-Parsons' (2019) analysis of tweets using the #MeToo hastag emphasizes a third potential drawback associated with the public nature of Twitter: many women require anonymity to be able to give voice to the oppressive relations they find themselves in. These women may be unable to use the platform to find the support and recognition they need to change their circumstances. Clark-Parsons points out that making oneself visible "as a survivor on a globally networked platform" can involve significant personal risk (2019, p. 10). As one online participant in the #MeToo movement has "put it bluntly, 'Twitter is NOT a safe space'" (2019, p. 10). Of course, Twitter allows for the possibility of posting under a pseudonym. However, as each Twitter account is associated with an email address and often a mobile phone number, true anonymity cannot be guaranteed. Regardless, the fact that some women may feel the need to post anonymously highlights the fact that many do not feel empowered to address sexual harassment and abuse in the contexts where they take place, whether online or offline.[10]

One takeaway from the study by Clark-Parsons is the need to create private and safe social spaces where women can find the support and recognition needed to cope with oppressive relations without immediately revealing themselves as engaged in this effort. In this paper, I can only touch upon possible avenues for further exploration. One possibility is the creation of private social spaces on social media platforms such as Facebook or WhatsApp. However, like Twitter, these social media platforms also request an email address or mobile phone number, undermining the possibility of true anonymity. Another possibility is to base initiatives in offline contexts, creating in-person communities that provide safe and private spaces for women, as consciousness-raising groups have already done for decades (Suk *et al.*, 2019). Another option is to engage men in offline groups. Corbin describes an inspiring example of a program consisting of five two-hour discussion sessions with migrant Hispanic men in the US around sexual and intimate partner violence. Corbin's study found that the program's focus on "reflection on past behaviours and current decision-making supported self-initiated behaviour change" (Corbin, 2018, p. 923). The success of examples like these may inspire us to combine the online activism of the #MeToo movement with offline initiatives. These initiatives could aim either at the creation of safe spaces to empower women to fight and escape the oppressive relations they find themselves in, or at behavioural change in men.[11]

10 A related, and equally important, issue is that public social media platforms do not provide equal representation for those whose identities lie at the intersections of multiple axes of oppression. Empirical research shows that it is (white) women in already privileged positions who have benefitted most from the #MeToo movement (Hu *et al.*, 2020; Onwuachi-Willig, 2018-2019). A telling example is the R. Kelly scandal, which received broad media coverage only after prominent white women shared their stories under #MeToo (for discussion, see Leung & Williams, 2019).

11 The Black Lives Matter movement (#BLM) can be understood as a fairly successful attempt to combine online and offline action (for an excellent overview and discussion, see Nunmings *et al.*, 2019).

Although Twitter is a powerful tool for connecting people and creating communities around urgent social problems and injustices, the #MeToo movement provides an opportunity to reflect on how spaces both online and offline can provide safe environments for women. A mix of online and offline environments may empower women, providing them with opportunities to receive the recognition they need to express themselves.

5. Conclusion

Lindemann points out that we depend on our communities to provide legitimacy and intelligibility to our identities and actions. Moreover, we are assessed by widely shared standards and norms within the social spaces we occupy. If our social spaces are structured according to master narratives that uphold oppressive norms, one powerful way to counteract this oppression is to create a counterstory that highlights the equal moral standing of the oppressed. However, the act of creating and telling this counterstory itself depends on recognition from others. Here Lindemann stresses the importance of chosen communities, where shared norms help facilitate the development of such counterstories among the oppressed themselves.

Using Lindemann's framework as a basis, I have argued that the #MeToo movement on Twitter can be understood as a counterstory. This counterstory centres the freedom and moral authority of women, and, in doing so, challenges the oppressive patriarchal master narrative that tends to dominate in how shared social spaces are structured in our Western modern societies. I have discussed how one significant drawback of Twitter as a platform for social resistance movements is its public character, which can allow dissenting voices to weaken a unified message, and may also ultimately fail to empower the women who feel the need to remain anonymous. Despite these potential shortcomings, it is clear nonetheless that social media-based movements like #MeToo can and do have a powerful influence in driving societal change and overcoming oppression and injustice.

REFERENCES

Anderson, J. (2014). Autonomy and Vulnerability entwined. In C. Mackenzie, W. Rogers, & S. Dodds (Eds.), *Vulnerability – New Essays in Ethics and Feminist Philosophy* (pp. 134-161). Oxford: Oxford UP;

Bitran T. (2018, April 13). Tarana Burke on MeToo: 'It's a mistake to think of this as a moment'. *Variety*. Retrieved from https://variety.com/2018/biz/news/viola-davis-tarana-burke-power-ofwomen-1202751993/;

Boyle, K. (2019). *#MeToo, Weinstein and Feminism*. Cham: Palgrave macmillan;

Corbin, J.H. (2018). Editorial - Health promotion and #MeToo: meeting men where they are. *Health Promotion International*, 33, 921–924. doi: 10.1093/heapro/day097;

Dawson, P. (2020). Hashtag narrative: Emergent storytelling and affective publics in the digital age. *International Journal of Cultural Studies*, 23, 968-983. doi: 10.1177/1367877920921417;

European Union Agency for Fundamental Rights. (2014). Violence against women: an EU-wide survey. Luxembourg: Publications Office of the European Union. Retrieved from https://fra.europa.eu/en/publication/2014/violence-against-women-eu-wide-survey-main-results-report;

Flood, M. (2019). Men and #MeToo: Mapping Men's Responses to Anti-violence Advocacy. In B. Fileborn & R. Loney-Howes (Eds.), *#MeToo and the Politics of Social Change* (pp. 285-300). Cham: Palgrave macmillan;

Friedman, M. (1992). Feminism and modern friendship: Dislocating the community. In S. Avineri & A. de-Shalit (Eds.), *Communitarianism and Individualism* (pp. 101-119). Toronto: Oxford UP;

Gallagher, R.J., Stowell E., Parker A.G., Foucault Welles B. (2019). Reclaiming Stigmatized

Narratives: The Networked Disclosure Landscape of #MeToo. *Proc. ACM Hum.-Comput. Interact.*, 3. doi: 10.1145/3359198;
Gotlib, A. (2016). "But You Would Be the Best Mother": Unwomen, Counterstories, and the Motherhood Mandate. *Bioethical Inquiry*, 13, 327-347. doi: 10.1007/s11673-016-9699-z;
Herman, D.F. (1978). The rape culture. In F. Jo (Ed.), *Women: A feminist perspective* (pp. 41–63). Mountain View, CA: Mayfield;
Hockett, J.M., Saucier D.A., & Badke C. (2016). Rape Myths, Rape Scripts, and Common Rape Experiences of College Women: Differences in Perceptions of Women Who Have Been Raped. *Violence Against Women*, 22, 307-323. doi: 10.1177/1077801215599844;
Hu, Y., Mu Y., Huang Y. (2020). The #MeToo narrative: Reconstructing the cultural intelligibility of female subjects. *Women's Studies International Forum*, 80. doi: 10.1016/j.wsif.2020.102365;
Ibsen, H. (2008). *A Doll's House*. Produced by Martin Adamson & David Widger. The Project Gutenberg Ebook. Retrieved from https://www.gutenberg.org/files/2542/2542-h/2542-h.htm;
Ilinskaya S., & Robinson D. (2018). #MeToo and the Estrangement of Beauty-and-thee-Beast Narratives. *Social Research: An International Quarterly*, 85, 375-405;
Jackie (@samandea). '#metto and I never told anyone because I thought it was my fault for sending the wrong signals...' October 21st 2017, 01:14. Tweet;
Jackson, S., Bailey M., & Foucault Welles B. (2019). Women Tweet on Violence: From #YesAllWomen to #MeToo. *Ada: A Journal of Gender, New Media, and Technology*, 15. doi: 10.5399/uo/ada.2019.15.6;
Johnson, M.E. (2018). Feminist judgments & #metoo. *Notre Dame Law Review Online*, 94, 51-54. Retrieved from http://scholarship.law.nd.edu/ndlr_online/vol94/iss1/10;
Just Be Inc. (2013). Retrieved December 15, 2020, from Just Be Inc website, http://justbeinc.wixsite.com/justbeinc/the-me-too-movement-c7cf;
Law, B. (@mrbenjaminlaw). 'After yesterday's endless #MeToo stories of women being abused, assaulted and harassed, today we say #HowIWillChange.' October 16th 2017, 17:15. Tweet;
Leung, R., & Williams R. (2019). #MeToo and Intersectionalitiy: An Examination of the #MeToo Movement Through the R. Kelly Scandal. *Journal of Communication Inquiry*, 43, 349-271. doi: 10.1177/0196859919874138;
Lindemann, H. (2001). *Damaged Identities, Narrative Repair*. Ithaca: Cornell UP;
Lindemann, H. (2020). Counter the Counterstory: Narrative Approaches to Narratives. *Journal of Ethics and Social Philosophy*, 17, 286-298. doi: l0.26556/jesp.v17i3.1172;
Maier, S.L. (2008). "I Have Heard Horrible Stories...": Rape Victim Advocates' Perceptions of the Revictimization of Rape Victims by the Police and Medical System. *Violence Against Women*, 12, 30-45. doi: 10.1177/1077801205277539;
McAdams, C. (2019). Perspectives on Sexual Power, #MeToo. *Academic Psychiatry*, 44, 26-28. doi: 10.1007/s40596-019-01146-3;
Milano, A. (@Alyssa_Milano). "If you've been sexually harassed or assaulted write 'me too' as a reply to this tweet." October 15th 2017, 13:21. Tweet;
Onwuachi-Willig, A. (2018-2019). What About #UsToo: The Invisibility of Race in the #MeToo Movement. *Yale Law Journal Forum*, 128, 105-120. Retrieved from https://scholarship.law.bu.edu/faculty_scholarship/331;
Park, A. (2017, October 24). #Metoo Reaches 85 Countries with 1.7M Tweets. *CBS News*. Retrieved from https://www.cbsnews.com/news/metoo-reaches-85-countries-with-1-7-million-tweets/;
Poirier, A (2018, January 14). After the #MeToo backlash, an insider's guide to French feminism. *The Guardian*. Retrieved from https://www.theguardian.com/world/2018/jan/14/french-feminists-catherine-deneuve-metoo-letter-sexual-harassment;
Rodino-Colocino, M. (2014). #YesAllWomen: Intersectional Mobilization Against Sexual Assault

is Radical (Again). *Feminist Media Studies*, 14, 1113-1115. doi: 10.1080/14680777.2014.975475;
Clark-Parsons R. (2019). "I SEE YOU, I BELIEVE YOU, I STAND WITH YOU": #MeToo and the performance of networked feminist visibility. *Feminist Media Studies*. doi: 10.1080/14680777.2019.1628797;
Sayej, N. (2017, December 1). Alyssa Milano on the #MeToo movement: 'We're not going to stand for it any more'. *The Guardian*. Retrieved from https://www.theguardian.com/culture/2017/dec/01/alyssa-milano-mee-too-sexual-harassment-abuse;
Starkey J.C., Koerber A., Sternadori M., & Pitchford B. (2019). #MeToo Goes Global: Media Framing of Silence Breakers in Four National Settings. *Journal of Communication Inquiry*, 43, 436-461. doi: 10.1177/0196859919865254;
Suk J., Abhishek A., Zhang Y., Ahn S.Y., Correa T., Garlough C., & Shah D.V. (2019). #MeToo, Networked Acknowledgment, and Connective Action: How "Empowerment Through Empathy" Launched a Social Movement. *Social Science Computer Review*, 39(2), 276–294. doi: 10.1177/0894439319864882;
Xiong Y., Cho M., & Boatwright B. (2019). Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Publix Relations Review*, 45, 10-23. doi: 10.1016/j.pubrev.2018.10.014

SECTION 3

SECTION 3

ETHICAL AND POLITICAL IMPLICATIONS OF DIGITAL TECHNOLOGIES

Helena de Preester
Life is what you fill your attention with – the war for attention and the role of digital technology in the work of Bernard Stiegler

Francesco Tava
Solidarity and Data Access: Challenges and Potentialities

Natalia Satokhina, and Yulia Razmetaeva
“The Loss of Experience” in Digital Age: Legal Implications

HELENA DE PREESTER
The Royal Academy of Fine Arts and the Royal Conservatory, School of Arts, University of Applied Sciences and Arts (BE) and Department of Philosophy and Moral Sciences, Ghent University (BE)
helena.depreester@hogent.be

LIFE IS WHAT YOU FILL YOUR ATTENTION WITH – THE WAR FOR ATTENTION AND THE ROLE OF DIGITAL TECHNOLOGY IN THE WORK OF BERNARD STIEGLER[1]

abstract

This contribution focuses on the topic of attention and sets forth the main points of Bernard Stiegler's analysis of the interplay between capitalist consumer society, the destruction of attention and the consequences for individual and collective life. We look at how current digital technologies in service of the needs of the market are a major factor in the destruction of attention and discuss two counterforces that do not destroy but form attention: education and meditation. If life is what you fill your attention with, then focusing or directing attention is one of the most valuable abilities for knowing how to live. Instead of letting our attention be hijacked by the market and the economic needs of neoliberal capitalism, being in charge of what happens to our attention may be a basic right that needs protection given the current conditions of the attention economy.

keywords

attention, Bernard Stiegler, (digital) technology, education, meditation, neoliberal capitalism

1 Research for this paper was funded by a research grant from KASK & Conservatory, the school of arts of HOGENT and howest (BE).

Phenomenology and Mind, n. 20 - 2021, pp. 102-116
DOI: https://doi.org/10.17454/pam-2009
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind
© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In today's *attention economy* individuals' attention is the rarest and most crucial resource that companies chase to run successful business. In this type of economy, the supply of capital, labor and information is abundant whereas human attention is in shortage. Even though it is easy to start a business and to reach consumers, it is a challenge to get and hold consumers' attention – attention necessary to devote to the information available even before buying or consuming anything.

Apart from the physical and analogue world of ads, brands and commercials, the digital online world is the place per excellence where attention is competed for. However, and notwithstanding the idea of multitasking (which is a matter of switching attention not of increasing it), attention has its limits. "Telecommunications bandwidth is not a problem, but human bandwidth is" (Davenport & Beck, 2001, p. 2). Or in the famous words of Herbert Simon: "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention" (Simon, 1971, pp. 40-41). For the organization of businesses and companies, the attention problem is one of efficiency, i.e. a matter of the efficient allocation of attention among the overabundance of information resources that might consume it.

For the mass of consumers, the problem is a different one: it is not just a matter of the efficient allocation of attention but primarily a matter of unhooking attention from everything that wants (or is designed) to capture attention, especially in the online world. For if attention is not unhooked, conscious time is consumed without us being able to determine how to allocate our attention resources. Before we know it, our scarce resource of attention is depleted and our behavior is modified in accordance with the goals of those who capture our attention. This modification of our behavior is not necessarily in line with decisions that we would want to make.

What we experience as "our life" is mostly the time we experience consciously. What we experience thus depends on what enters consciousness. Crucially, what enters consciousness is a function of attention processes. Therefore, who or what is in control of our attention is of utmost importance for how we experience our lives and what we make out of it.

This contribution focuses on the topic of attention and sets forth the main points of the analysis by Bernard Stiegler, a major contemporary philosopher of technology who analyses in several of his works the interplay between capitalist consumer society, the destruction of attention and the consequences for individual and collective life. The next section (section 2) describes what attention is from a philosophical point of view. In the third section, we look at

how current digital technologies in service of the needs of the market are a major factor in the destruction of attention, to such an extent that we may say that there is a war for attention going on. Next, two counterforces are discussed that do not destroy but form attention: education (section 4) and meditation (section 5). The aim is to formulate insights into both the destruction and the formation of attention under early 21st century conditions.

2. What is attention? Primary, secondary and tertiary retentions

Bernard Stiegler's philosophical examination of attention (2010; 2014) clarifies that attention is much more than concentration or vigilance.[1] Attention is about desire, waiting, attending and caring. Attention is the result of education, of the formation of the individual as such. It is also an intergenerational relationship. In Stiegler's philosophy, attention ties together all these phenomena. This section sets out how it is possible that attention is so encompassing.
Stiegler's view on attention is rooted in Edmund Husserl's account of the (temporal) structure of consciousness. In his phenomenological analysis of inner time-consciousness, Husserl (1991 [1966]) traces the origin of the temporal way in which objects appear to us. His phenomenological account is not just based on our subjective *experience* of time, but on the laws governing the experience of an object appearing in time. According to Husserl, each momentary phase of perceptual consciousness is a continuum made up of a now-consciousness (present impression) and a number of points representing what has just passed (retentions). The succession of these phases is in turn a continuum.
Husserl distinguishes primary retentions (impressions) and secondary retentions (memories). A primary retention (or impression) is the presentation in consciousness of what is just past (e.g. the just passed notes when you are listening to a melody). A secondary retention (or memory) does not *present* but it *re*presents the past (e.g. when you think back to that melody the next day). According to Stiegler, secondary retentions function as a filter for primary retentions because the secondary retentions that constitute one's memory select the primary retentions. As such, secondary retentions determine what appears in the flux of consciousness. Vice versa, primary retentions influence secondary retentions because they determine what constitutes a secondary retention. Linking primary with secondary retentions, consciousness is also able to anticipate or to project into the future. That forward stretch of consciousness is made up of so-called protentions.
Stiegler (2010; 2014) emphasizes that attention is the result of the accumulation of primary and secondary retentions and the anticipation of protentions. Protentions draw out attention, but are themselves the result of retentions. "It is as an accumulation of experiences in what I have previously called secondary retentions that the horizons of anticipation are formed" (Stiegler, 2014, p. 65). Put in a nutshell: consciousness is a flux of backward (memory) and forward stretches (anticipation) and what we pay attention to is rooted in our past experience. Primary and secondary retentions are of the level of the individual and appear and disappear with the individual subject. In the wake of Jacques Derrida's analysis of Husserl's *On the Origin of Geometry* (Husserl, appendix to 1970 [1954]; Derrida, 1989 [1962]) and in addition to the notions of primary retention and secondary retention (or memory), Stiegler (1998) elaborates the notion of tertiary retention. As humans are mortal, the retentional finiteness of memory implies that the ability to remember is limited, finite and temporary. Tertiary retentions break through this limitation of memory because they are *external* memories. Tertiary retentions are externalized memory traces and they range from stone tools to contemporary technology.

1 See Lee (2019) for a short introduction to how Stiegler's account differs from Katherine Hayles' distinction between hyper and deep attention (Hayles, 2007). See also Stiegler's discussion of Hayles' distinction (Stiegler, 2010, especially pp. 77-83).

They are artificial and technical and through a process of accumulation over time they form the technical environment of man. This technical environment consisting of the accumulated tertiary retentions precedes the individual.
According to Stiegler (1998; 2010), tertiary retentions are constitutive and conditioning for primary and secondary retentions. Primary and secondary retentions, impressions and memories, belong to the order of the individual. Tertiary retentions not only transcend the individual but they also constitute and condition the human subject. They are of the order of the tradition or of the collective memory. On the individual level, the selectivity of perception shows that perception happens in function of what has already been, in function of memories or secondary retentions. Secondary retentions are in turn overdetermined by the system of tertiary retentions: our memories depend on the external memories in the technical environment. For Stiegler it means that the human condition is fundamentally a technological condition.
Tertiary retentions transcend the individual but in contrast to genetic information transmitted by the genes (phylogenetic) tertiary retentions are *epi*-phylogenetic. Tertiary retentions are of the order of tradition or the collective memory and are transmitted not by the genes but by the generations (Stiegler, 2014, p. 7). They consist of all that belongs to the human world and carries knowledge and skills. They are the materialization or the concretion of secondary retentions. However, and in contrast to secondary retentions, they do not disappear when the individual disappears. Tertiary retentions are the collective memory of generations. They form an intergenerational relationship and they are the source for new secondary retentions in the next generation's individual consciousness.
The transmission of memory is formative of the next generation's attention, constituting the retentions which then create protentions (Stiegler, 2010, p. 8). That is the reason why attention is the result of education or of the formation of the individual as such. That is also why attention is an intergenerational relationship.
Typical of Stiegler's philosophy is that tertiary retentions are *artificial* memories or memory *technics* or *mnemotechnics.* They range from the invention of writing to the printing press and contemporary information and communication technologies. Stiegler pleas for a recognition of the profound role of mnemotechnics (also called more broadly psychotechnics) in the constitution of human consciousness and subjectivity. His whole philosophy is an elaboration of the thesis that human consciousness is essentially *technically* constituted and conditioned because tertiary retentions are of a technical nature. That implies that attention is also constructed *technically* because secondary retentions are built on tertiary retentions which are the materializations of memories based on the use of technics (with writing as a prime example). The life of the mind has always been technically determined. Until recently, writing was central to the technical environment that defines spiritual life (spiritual in the sense of the French *esprit*). Since the Enlightenment, writing has been the technical condition for the emancipation of the citizen and has created a public space (Stiegler, 2010). Similarly, humans have always used attention technics to focus attention. Strategies for concentrating attention are thus not unique to our times, because the formation of attention requires that attention is concentrated or captured. All attention formation requires a technics for capturing attention (e.g. a book, but also education, as every teacher knows). Attention is something that originally can only exist because it is captured, directed and modulated – and that happens on the basis of technics (Lemmens, 2009, p. 90). Without technics, there is no human mind. Both the individual and the social formation of the subject happen on the basis of technics and attention is therefore always at once psychic and collective (Stiegler, 2010).
In today's technological environment digital technology and the internet have a central position. Individual life and society are more and more determined by digital information and communication technologies. Accordingly, the life of the mind is more and more determined

by what Stiegler calls *psychotechnologies,* i.e. technologies that condition and even constitute psychic life. Stiegler, however, points out that all technics can either foster or destroy the human mind.

3. The destruction of attention

The previous section presented the intimate relationship between consciousness, the interweaving of primary, secondary and tertiary retentions and attention. It was also pointed out that attention is intergenerationally formed because tertiary retentions are formative of secondary retentions or individual memories. Next, the technical nature of tertiary retentions – and thus of human consciousness and attention - was emphasized. Finally, it was mentioned that all technics can be either beneficial or detrimental to the human psyche. That remark is the point of departure of this section. More in particular, this section focuses on the detrimental effects that current-day digital information and communication technologies have on attention.

The broader context in which current-day digital technologies have detrimental effects on the human psyche is the consumer society's reduction of citizens to consumers. The core process in the consumer society is that all energy is mobilized for production and consumption whereas the higher plane of ideals and spiritual culture ("spiritual" again in the sense of the French *esprit*) is destroyed. Life is reduced to the level of immediate satisfaction of drives, a situation typical of consumerism. In *cultural capitalism* even the higher plane of ideals and culture is annexed by neoliberal capitalism. In other words, in cultural capitalism, culture is put in service of the economy. What is new in *cultural* capitalism is that commodities are acquired because they offer a certain experience (of pleasure or meaning). This differs from the earlier version of capitalism in which commodities are acquired because they have intrinsic value or because they are useful or give the owner a certain status.

Stiegler (2010; 2011) observes that we live in societies in which marketing has become the central function of psychic and social development and even replaces traditional social regulation (which used to be intergenerational). The control of *culture* is at the heart of current consumerist capitalism, controlling the behavior of the masses (mass production of behavior). The computational nature of capitalism tends to eliminate all values not calculable on the market of economic exchange. Consciousness itself, in terms of attention, has become a commodity. Libidinal energy, i.e. the energy of desire, being exhausted both the formation of the individual and the society are under threat. Whereas drives are related to basic needs such as food and shelter, desire is the spiritual transformation of drives and is related to goals and ideals in the context of social life. The systematic exploitation of desire in the consumer society, however, tends to destroy desire. The result is that the subject is thrown back to the level of subsistence.

Stiegler calls the kind of society in which desire is exploited and exhausted 'irrational' because it precisely demotivates those who constitute the society. "(...) this control is an exploitation of libidinal energy that *exhausts* this energy, and it is in this way that the industrial model emerging from twentieth-century modernity reaches its limit (...)" (Stiegler, 2011, p. 26).

The cause of the destruction of both the individual and the society lies in the industrial exploitation of libidinal energy in an unlimited way. Put more simply, it is the reduction of the citizen (both psychically and collectively) to the status of pure consumer. People merely *subsist* in this kind of society, they do no longer *exist* in the sense of believing in a future, supported by symbolic activities (the higher plane of ideals and spiritual culture). In the consumer society, consumer behavior has been standardized and *culture* has become the means of *controlling* consumer behavior. According to Stiegler (2011), the consumer is the new proletarian figure, submitted to standardized behavioral models of consumption.

Neoliberal capitalism implies the subjugation of the whole of society to the imperatives of

capital. The neoliberal society is a consumerist or even hyper-consumerist society in service of economic growth in which citizens are addressed exclusively as consumers (also when voting). Stiegler insists that this happens on the basis of technical apparatuses, more in particular on the basis of the *psychotechnologies* of the mass media, exploited in a systematic way by marketing and advertising (cf. also Lemmens, 2012a). Neoliberal capitalism deploys the possibility of information and communication technologies to promote consumption and to control the behavior of people. This control determines the behavior and life style of citizens in a way more profound than biopower as analyzed by Michel Foucault (and as something in hands of the state) (Foucault, 2008; 2009). "The primary role of bio-power in our time is not so much to turn people into obedient and efficient production machines, but to produce consumers. And in fact it is no longer about biopower, but about psychopower. Or more precisely, it is no longer so much about a power over the body and life, but about a power over the 'soul' (psyche), about an acquisition of consciousness or desire for the sake of consumption" (Lemmens, 2009, p. 88, our translation). In the above terms of technically based retentions: there is a systematic exploitation of the mnemotechnical system that is the basis of human consciousness, attention and culture. As explained above, this system of artificial memories or *mnemotechnics* is the condition of possibility of the formation of attention and human spiritual life. However, marketing deploys a technologically organized attention *control apparatus* that destroys attention itself.

Radio, film and television and later digital network technologies are deployed to capture the consciousness (and the desires) of people, both individually and collectively. The crucial factor in controlling behavior and in channeling or nudging the people towards (more) consumption is individuals' attention. "Today, attention control via cultural and cognitive technologies ('technologies of the *spirit [esprit]*' those malignant spirits haunting the adult minor as apparatuses for capturing, forming, and deforming attention), has become the very heart of hyperindustrial society; however, it no longer relies on psychotechnics but on psychotechnological apparatuses whose devastation we see on TF1, Channel Y, and so on" (Stiegler, 2010, p. 22). Stiegler only observes what others (Rouvroy & Berns, 2013; Yeung, 2017; Zuboff, 2019) also observe: the capturing of the attention of the public is motivated by the business plans of companies. At the same time, Stiegler notices a shift from (psycho-)technics to (psycho-)technologies, the latter enabling the exploitation of human attention and desire on an unprecedented industrial scale.

The aim of this use of technologies is to adapt the behaviors and ways of living to what the market offers, i.e. to the output of capitalist industry (Lemmens, 2012b; Stiegler, 2009, p. 60). The control over attention by current digital technologies implies the control of our behavior. The notion of "algorithmic governmentality" (Rouvroy & Berns, 2013) is an extension of Foucault's conception of power in terms of the ability to direct people's behavior (*conduire des conduites*) without direct coercion or direct prohibition. Algorithmic governmentality is a form of *soft* power, an unobtrusive yet powerful way to direct behavior. Its current algorithmic nature refers to the phenomenon of big data and the use of statistical knowledge to anticipate the behavior of individuals and to relate it to a wealth of correlations (so-called profiles) obtained from data mining. That statistical knowledge is not accessible to the individual but it is nevertheless applied to the individual to infer knowledge or predictions about his or her preferences, intentions and propensities. On that basis, behavior is gently steered in the desired direction (read: in the direction of more consumption or in the direction of certain beliefs and opinions).

Yeung (2017) discusses the phenomenon in terms of "hypernudge" or the influencing of choice behavior in the context of big data. Hypernudge techniques push us "in the right direction" and are used to shape individual decision-making in the interests of commercial

big data tycoons. While governments and citizens have so far mainly been concerned about privacy, Yeung also points to the erosion of our capacity for democratic participation and the misleading and non-transparent nature of hypernudge techniques. Both algorithmic governmentality and hypernudge are forms of psychopower and presuppose that the attention of people is captured by digital technologies.

In summary, psychopower conditions the mind and desires of people for the purpose of consumption and adaptation to the capitalist production apparatus. The true market is therefore consciousness itself because the access to the consciousness and attention of the people determines all other markets. Consciousness (or available conscious time and attention) is the most important commodity as it is the *metamarket* that gives access to all other markets (Lemmens, 2009, p. 89). Attention is therefore the commodity most searched for and marketing is the most important instrument of the hyperindustrial producer (Stiegler, 2015; Rossouw, 2015, p. 187). That is the reason why digital technologies are massively designed and employed: to capture consciousness channeling it towards more consumption. As such, psychopower technologies take massive control of our behavior (see also the notion of hypernudge, Yeung, 2017; algorithmic governmentality, Rouvroy & Berns, 2013). Psychopower technologies control the mental activities of individuals by capturing consumer attention and causing consumers to adopt new psychomotor behaviors that help form the markets of current neoliberal capitalism. Consciousness, or attention, is reduced to commodity, to "available brain time" which is maximally exploited, leading to systemic stupidity in the information age (cf. also Fitzpatrick, 2020). Stiegler adopts the phrase "available brain time" from the CEO of the French television channel TF1 Patrick Le Lay, who announced the following from the business perspective of TF1: " [...] in the end, TF1's job is helping Coca-Cola, for example, sell its product. What *we* sell to Coca-Cola is *available human brain time.* This is where permanent change is located. Nothing is more difficult than getting access to it: we must always be on the lookout for popular programs, follow trends, surf on tendencies, all in a context in which information is speeding up, getting diversified and trivialized" (cited in Stiegler, 2010, p. 196, fn. 34 by the translator, emphasis in original). Because attention is captured by psychotechnologies exploited on an industrial scale, attention can no longer be cultivated and focused on consistence. In Stiegler's work, the term "consistence" refers to the order of ideals and the symbolic but is – in line with his philosophy – itself technically conditioned (cf. the notion of tertiary retention). Without consistence human existence shrinks back to the mere condition of subsistence, a condition of symbolic misery (Stiegler, 2014; 2015). Symbolic misery is the loss of participation in symbolic production, especially the formation of ideas and ideals, which are the basis of spiritual culture (cf. Lemmens, 2014a, p. 24). What distinguishes human existence from mere subsistence is precisely the capability of having goals or ideals, reasons and motives to live for that are beyond competition and consumption and that belong to the order of consistence. Humans cannot live as humans without investing in a future they can desire, without ideals or "objects of consistence" (Lemmens, 2012a, p. 6). When human existence shrinks back to subsistence, all that is left is consumer behavior. The shrinking back of human existence to subsistence is also noticeable in the fact that the time of leisure is now also controlled by culture and programming industries (roughly, the products of cultural capitalism and the media). Section 4 explores this into greater detail.

The reduction of human existence to subsistence leads to what Stiegler calls the proletarianization of citizens, not only as worker-employee but also as consumer. The lives of people are controlled and modulated by management and marketing techniques. Those management and marketing techniques determine the lives of people to such an extent that they are increasingly relieved of inventing and shaping their own ways of life and tend to

lose the ability to shape existence themselves in a free and autonomous way (Stiegler, 2011). The result is a reduction of existence to labor and consumption (cf. supra the reduction of existence to subsistence; see also Lemmens, 2012a). In a nutshell, people lose their knowledge of how to exist or how to live (*savoir-vivre*). Consciousness, attention, desire, time – all dimensions of existence are annexed by production or consumption for the purpose of capital accumulation. Unfortunately, what is taken is at the same time destroyed and the spiritual world is annexed by the world of economy.

The concomitant shift from psychotechnics to psychotechnologies is an important one. "It is this industrial exploitation of the (now predominantly digital) 'technologies of the mind' by the psychopower of capital that is chiefly responsible for the deep social, spiritual and political crisis in which we find ourselves embroiled at the moment (and which is the actual root cause of the current financial and economic crisis). In fact, it destroys the libidinal energy and therefore our 'spiritual life' and the critical attention for ourselves, for others, and for the world as the expressions of this energy" (Lemmens, 2012b, p. 51). The whole of our technical milieu is annexed by the psychotechnologies of marketing, management and entertainment industry. Stiegler says our technical milieu is *intoxicated* by commercial and marketing psychotechnologies. Whereas psychotechnics traditionally serves *care* (of the individual and the collective) the transformation into psychotechnologies has implemented an attitude of "I don't give a damn" (Stiegler, 2010, p. 41). The previously unknown technology of power that marketing is, leads to subjects reduced to consumers and not in charge of their very life. Marketing psychopower creates a colossal regression, according to Stiegler, and leads to a massive irresponsibility because it infantilizes adults and cuts the intergenerational bond.

On the first pages of *Taking Care of Youth and the Generations* (2010), Stiegler presents responsibility as a learnt social competency. Society is responsible for transmitting responsibility to the next generation. The name for transmitting responsibility is *education* (Stiegler, 2010, pp. 1-2, pp. 42-43; see also section 4 for the issue of education). Irresponsibility is the reduction of critical consciousness to a mere brain deprived of consciousness. "In order to be made available to marketing imperatives, the brain must early on be literally deprived of consciousness in the sense that the creation of synaptic circuits responsible for the attention formation resulting in 'consciousness' is blocked by the channeling of attention toward the programming industry's objects" (Stiegler, 2010, p. 72). That people – consumers – are stripped of critical consciousness amounts to saying that they are stripped of consciousness itself. From Stiegler's perspective on what consciousness is (cf. the interweaving of primary, secondary and tertiary retentions) this regression implies a colossal deficit of attention, " [...] a *global attention deficit disorder*, stemming directly from the proliferation of psychotechnologies that no political power can now control" (Stiegler, 2010, p. 57).

Stiegler contrasts this situation with the Enlightenment thought that promotes maturity and critical consciousness. Consumption is intensified by capturing attention, and at the cost of widespread irresponsibility, resulting into the opposite of what we inherited from the Enlightenment. This goes hand in hand with the destruction of the social or the intergenerational bond. "The cognitive faculty – what we call reason – is the only solid link between the psychic and the social, in that it is passed through the succession of generations transformed and sublimated by disciplinary learning; this process constitutes knowledge. Informational saturation, on the other hand, desocializes the consumer of that information" (Stiegler, 2010, p. 184). Knowledge and understanding must be psychically assimilated and made one's own (one's own *self*) while information is merchandise made to be consumed – and therefore disposable.

Of course, it is possible for neoliberal capitalism to control (and proletarianize) the human mind and human desire because the psyche and attention, or human culture, have always

had (and also originally have) a technical basis. The spirit has always been controlled on the basis of psychotechnics. However, in contemporary hyperindustrial societies, the control over attention by current digital technologies happens on a unseen scale and with disastrous consequences. There is no public authority nowadays that can "arbitrate the conflict between psychopower and the attention *diversion,* on the one hand, and attention *formation* as the psychic and social faculty of responsibility, on the other" (Stiegler, 2010, p. 53, emphasis original). The result is a battle or even a war for attention. The education of young people stands at the heart of this battle.

4. Remedies for cure: education and a new *otium* of the people

According to Stiegler, the mass media and so-called programming industries (e.g. television in the second half of the 20th century) have ruined public education systems and training programs instituted in France in the 1880s, leading to psychological and social disaster. The programming industries replace the programming institutes of public education. Their goal is "complete control of the behavior-formation programs regulating social groups, indeed their removal from the public education system and their adaptation to immediate market needs" (Stiegler, 2010, p. 58). The programming industries destroy the attention needed to obtain knowledge and they deform or destruct the individual that education has constructed. "The work of forming attention undertaken by the family, the school, the totality of teaching and cultural institutions, and all the apparatuses of 'spiritual value' (beginning with academic apparatuses) is systematically undone in the effort to produce a consumer stripped of the ability to be autonomous either morally or cognitively [...]" (Stiegler, 2010, p. 184). Even worse, the goal of the programming industries would be to destroy the programming institutes themselves or any public structure that obstructs the dominance of the market (Lemmens, 2009, p. 93). The war between programming industries and programming institutions is ongoing. Whereas educational institutes have always been the centers where culture is transmitted to the next generation, digital media are almost exclusively dominated by marketing. In sum, there is a clash between the programming institutes and the programming industries or between the school and the market, the latter trying to capture and channel attention towards consumption.

Stiegler, however, seems to have a strong belief in the possibilities and the role of the state, the public education system (the programming institutions) and a new politics as counterforces to the programming industries and the psychopolitics that emerges from these industries. Nonetheless, the situation is critical because the reconstruction of attention has to happen in a time in which attention is not only exploited but also being destroyed (cf. Fitzpatrick, 2020, p. 349). That means that before we can fight back, attention should be cured or healed. Stiegler's remedy lies in (a restoration of) education and a new form of *otium* of the people. This section therefore frames Stiegler's belief in education or the school as emancipatory discipline in the broader context of a new *otium* of the people. When discussing the spiritual world (still in the sense of the French *esprit*) Stiegler also uses the (German) term *Bildung* or the (Latin) term *otium. Otium* can be translated as "leisure" but has a meaning that predates the industrialization of the time not devoted to its counterpart *negotium* or the world of economy and its calculus. The modern form of leisure starts in the beginning of the 20th century with the program and culture industries and the invention of the consumer. This modern form of leisure effaces the difference between *otium* and *negotium* – *otium* now being the extension of *negotium* and *negotium* absorbing the spiritual or consistence (Rossouw, 2015, p. 189; Stiegler, 2011, chapter 3).

Stiegler analyses the school as the traditional apparatus for the formation of critical

attention.[2] As already mentioned, the different forms of attention that oppose each other in this battle have always been constituted by psychotechnics. That means that the history of humanity has always been a permanent battle for the psyche and that this war has always been waged on the basis of psychotechnics and attention technics (from rituals over printed books to digital technologies) (Lemmens, 2009, p. 90). The current *psychotechnologies* that control attention to the benefit of the market have to be transformed into *nootechnologies,*[3] i.e. technologies that form a new kind of critical attention, both individually and socially, and that foster responsibility (Stiegler, 2010, p. 51). *Nootechnologies* are emancipatory technics of the psyche that foster autonomous and critical thinking. Stiegler also pleas for the translation of *psychopolitics*, i.e. politics that regulates psychotechnologies, into a *noopolitics*, i.e. a politics that not only limits and regulates psychotechnologies but also transforms the current technologies with their poisonous effects into a remedy. In other words, a *noopolitics* transforms psychotechnologies producing stupidity into nootechnologies producing a new, collective intelligence. All technics, and mnemo- or psychotechnics in particular, have the character of a *pharmakon*, i.e. they can either support or undermine the human soul.[4]

Stiegler does not call for an immediate struggle "against the disastrous effects of the savage use of psychotechnologies by the programming industries as they destroy attention and consciousness" (Stiegler, 2010, p. 179). What is needed first is *meta*care, i.e. a psychopolitics that shapes care, or a politics of techniques of the psyche. Education is a good example of metacare. "Truthfully, education is [...] in fact a *metacare,* not care of the body nor even of numbers of bodies but of what have for centuries been called 'souls,' whose collectivity constitutes *a* spirit" (Stiegler, 2010, p. 177). The programming institutions in general are for Stiegler "the sole guarantors of a system of care worthy of the name and the supporters of the battle for intelligence" (Stiegler, 2010, p. 191). Stiegler thus seems to put his hope in the state or institutions financed by the state and the reformings that can be effectuated by (state) institutions. Those reformings will be based on reclaiming technologies and using them to the benefit of attention and care. Thus, instead of political resistance and action, the politics needed should have the character of a therapy for society (cf. Lemmens, 2014b). However, as the ownership or the control of digital network technologies seems to be a crucial factor in this, it is not clear how Stiegler could effectuate a therapy for society without a battle for the ownership or at least a political appropriation of psycho- (or noo-)technologies.

What is clear, however, is that Stiegler's sociotherapy has to detoxify the human psyche and to transform the poison of psychotechnologies into the medicine of nootechnologies. Sociotherapy is the conception of "a new age of the formation of care and attention for facing the care-lessness of a global consumer" (Stiegler, 2010, p. 180). The formation of care and attention would lead to the development of a planetary consciousness because "[T]o take care also means to pay attention, first paying attention to taking and maintaining care of oneself, then of those close to us, then of their friends – and thus, by projection, of *everyone* (...)" (Stiegler, 2010, p. 179).

Even though a detailed picture of the relationship between technologies, care and attention is not offered by Stiegler, it is telling that he wrote the preface for the French translation (2015) of a book on peer-to-peer economy by Michel Bauwens and Jean Lievens. The book pleas for a new distributed and decentralized economic model and highlights the role of technology in the making of a new society that fits well with Stiegler's notion of a sociotherapy.

2 For an analysis of how this compares to Foucault's analysis of the school, cf. Lemmens (2009).

3 "Noo-" relates to noetic, which refers in Stiegler philosophy and following Aristotle's tripartition of the soul, to the spiritual over and above the sensory and the nutritive.

4 For the notion of *pharmakon,* cf. Derrida's (1972) analysis of Plato's condemnation of writing as a pharmakon in the *Phaedrus*. Stiegler was a pupil of Derrida.

The importance of education as a counterforce to the destruction of attention is better understood in the broader frame of *otium*. As mentioned above, *otium* is free time or time not devoted to activities related to business and subsistence. That free time is time for activities related to education, intellectual work and spiritual contemplation (e.g. reading, writing, meditation, prayer, ...). For Stiegler, it has to be cultivated through disciplines or practices that cultivate attention, that facilitate community, commemorate tradition and thus open up (hope for) a future (Rossouw, 2015, p. 195). In brief, culture is the development of intellectual faculties and it is the content of the *otium.* In principle, it is open to all and intimately related to consistence or the formation of ideas and ideals. As such, it counteracts the shriveling of existence into subsistence. That Stiegler understands *otium* in terms of discipline is important. A discipline is the repetition of exercises that forms the discipline (cf. also Rossouw, 2015). Stiegler reverts to the old Greek term *melete* for this. *Melete* refers to care and attention, practice and exercise. We come back to *melete* in the next section.

Intellectuals, scientists, artists, philosophers and other "spiritual" people are the elite troops in the constitution of a new *otium* of the people (Stiegler, 2015) but it also is a state affair, i.e. of a state characterized by participatory democracy and the political appropriation of psychotechnologies (transformed into nootechnologies). Those nootechnologies would support the creative and intellectual activities of the elite troops. How exactly technology plays a role in the spiritual upliftment and the emancipation or the socialization of the people is not elaborated by Stiegler (and is not the subject of this paper) but Stiegler mentions peer-2-peer examples such as Wikipedia and opensource software (for more on this, cf. Bauwens and Lievens, 2015).

Conceived as part of *otium,* education is a combination of culture and *Bildung* (in the Enlightenment sense). Education is attention formation. However, the renewal of the educational system requires that the symbolic milieu in which we all live – children and parents – is no longer a systematic obstacle to this (Stiegler, 2010, p. 54). Unfortunately, the destruction of attention has resulted in the impossibility of education because *everyone* (including artists, professors, writers, scientists, ...) who should deliver the elements of a new *otium* of the people is affected by the destruction of attention or consciousness (Stiegler, 2010, p. 55). The socially configured system of care as a kind of attention that was developed in the educational system in France in the 19th century can no longer be adopted. That can also be formulated in terms of tertiary retentions. Tertiary retentions are the external inscription of secondary retentions, which hold the primary retentions. "It is tertiary retention which is, arguably, the most relevant for the question of education today. The question of education, therefore, is posed within Bernard Stiegler's work through the disruption to attention and attentional processes which are caused by digital tertiary retentions. [...] The process of exploiting our attentional being in the world has led to the destruction of the very forms of attention that are being harvested" (Fitzpatrick, 2020, p. 349). User profiles are, after all, the result of an individual's attention and are subsequently exploited to nudge the behavior of the people (cf. Rouvroy & Berns, 2013; Yeung, 2017).

Stiegler seems to set his hopes on state institutions or institutes supported by the state, such as public education. This may sound as if there is little that individuals can do on their own to restore attention formation. Today, however, many people turn to meditation as another discipline of attention based on repeated practice. The last section explores how meditation could contribute to Stiegler's new *otium* of the people. Or rather, it shows how meditation practices, traditionally part of *otium,* also offer a contemporary answer to the destruction of attention.

5. Remedies for cure: meditation and restoring the awareness of consciousness

Meditation is a practice that Stiegler does not explore as an answer to the destruction of attention. Yet meditation is part of the practices that (traditionally) belong to *otium*. Meditation is discussed by Michel Foucault in the context of technologies of the self (Foucault, 1986; 1988) and Stiegler discusses Foucault's technologies of the self at length in *Taking Care of Youth and the Generations* (2010). According to Foucault, technologies of the self "permit individuals to effect by their own means or with the help of others a certain number of operations on their bodies and souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality" (Foucault, 1988, p. 18). Technologies of the self are about taking care of one's self and encompass a broad range of practices and disciplines (praying, reading, meditation, writing, ..). Stiegler focuses on writing as technique of the self and thus also as care of the self. He focuses on Foucault's *écriture de soi* (Foucault, 1983) or the writing down of one's actions and thoughts as part of the art of governing oneself. His preference for writing probably stems from the concept of tertiary retentions with writing as a prime example. Of course, meditation traditions have their own written sources and philosophies, which are externalized, but the core of meditation is the practice that one discovers by oneself, often under guidance of a master with whom one enters into dialogue. Since meditation is primarily an (embodied) practice it does not sit very well with Stiegler's notion of tertiary retention, which is an externalization based on the model of writing. For Stiegler, attention that is focused is based on writing and text whereas a distributed kind of attention occurs, e.g., during the motor activity of walking as a technology of the self. In contrast to "literate" and focused attention, this kind of attention can lead to daydreaming and accidental thoughts or remembrances (cf. Stiegler, 2010, pp. 80-81). Stiegler does not explicitly comment on the position or role of meditation practices in this matter of attention. Nonetheless, we can infer its position and role from a number of elements that are present in his writings.

Stiegler connects paying attention with waiting (*attendre*). The translator, Stephen Barker, specifies that waiting is precisely the act of *in*action. "(...) as the act of inaction *and* as anticipation *attente* [waiting] is itself infinite, in that as a condition it is not reliant on any *end to waiting*: waiting *is*" (Stiegler, 2010, p. 213, footnote 5 by the translator). Inaction is an essential part of meditation which is explicitly not goal-directed (i.e. the goal is not to attain a state of relaxation or any alternative state of consciousness). Moreover, there is a close link between meditation and discipline. Stiegler discusses this close link in the context of Foucault's techniques of the self. The Greek *melete* is the muse of meditation and the term is derived from the verb *meletao,* which means "to take care of something" but also "to exercise", "to prepare oneself for something", and thus it refers to a kind of training. Stiegler, in the wake of Foucault, considers this as the ancient meaning of the term *discipline,* such that *melete*, meditation and discipline are intimately interconnected.

The essence of proletarianization (cf. section 3) is the loss of knowledge (savoir). Next to *savoir faire* or knowing how to do something and *savoir théorique* or theoretical knowledge, the lost knowledge is also knowing how to live (*savoir-vivre*). But as Lemmens (2012a) correctly remarks, it is also the loss of the sense or the awareness that one is a knowing and thinking being that relates to the world and gives meaning to it. As this sections aims to explain, meditation plays a crucial role precisely in training this awareness. Moreover, meditation is a discipline that trains the practitioner in *not reacting*, i.e. in interrupting the chain between stimulus and response. It does so by restoring awareness of subjectivity itself such that reactivity is counteracted and consciously responding becomes possible. The ability to counteract reactivity is important from the point of view where attention, care and responsibility are interwoven. The ability to *respond* in a conscious way and to pay attention to our own reactivity (instead of to *react*) is exactly what psychotechnics in service of the market

tries to circumvent, stimulating preconscious reactions and canceling out the subject between stimulus (marketed commodities) and response (consumption) (cf. also Rouvroy & Berns, 2013; Yeung, 2017).

What happens in many forms of meditation is that one is no longer primarily occupied with the objects toward which consciousness is directed but becomes aware of consciousness itself.[5] What happens in the process of proletarianization is that consciousness is constantly bombarded with objects (often digital ones) that call for attention, diverting consciousness from being aware of itself. The result is that the constant occupation with contents or "desired" objects weakens the ability to become aware of consciousness or subjectivity itself. This weakening of awareness is necessary in order to be a present-day (hyper-)consumer. To step back from the objects or contents that are presented to us is a necessary condition to come to critical reflection because this step back opens up the space to transform reactions into responses. What is exercised in meditation is non-reactivity as "a sustained refusal to actively pursue [whatever may occur within or outside oneself], i.e. to react to it cognitively or practically" (Fasching, 2008, p. 464). All intentional activity comes to a halt, and "one is simply conscious without doing much else" (Fasching, 2008, p. 465). The key point of being "simply conscious without doing much else" is that "one does not allow anything to distract consciousness from itself" (Fasching, 2008, p. 465), exactly the opposite of the unrelenting distractions set up by present-day psychotechnologies.

During meditation, the mind time and again begins to occupy itself with something that affects it. The exercise is to notice this affection and the distraction of the mind and to return to the awareness of consciousness itself "until the affective power of the stimuli gradually diminishes" (Fasching, 2008, p. 469). Non-reactivity thus means that we consciously notice that we are affected and that we have drifted away with what affects us (a thought, an emotion, a desire, ...). When we notice that, we can return to a state of awareness in which we are no longer carried away by that which affects us but stay present and aware of this presence.

To be affected by something and subsequently to be occupied by it (i.e. to be distracted by something, to desire it, to be curious about it, to evaluate it, to make plans, etc.) belongs to the nature of the human mind. To make full circle, that is precisely also the reason why psychotechnologies can so easily capture our attention.

What consequences could the practice of meditation have for daily life? And more in particular, how could it counteract the poisonous effects of psychotechnologies? Even though the practice of meditation itself is not goal-directed, it does not imply that there is no ultimate aim in meditation. Persistent practice transforms the everyday way of experiencing the world. "The daily object-experience is re-structured. [...] I am there not only as one who is active (who perceives, thinks, desires) but also as the very being of activity (of perceiving, thinking, desiring) itself, which is essentially non-activity in each and every moment of acting" (Fasching, 2008, p. 480). This "very being of activity" is the same as being present to the subjective presence we are in the world. It is awareness itself, a presence to oneself as a conscious, knowing and acting being in the world. As such, presence is a necessary element in the training of non-reactivity or in the ability to interrupt the chain between stimulus and response by restoring awareness of subjectivity itself. It is exactly this ability that is under treat by the exploitation of our attention processes by current-day psychotechnologies that poison the technical milieu. Training this ability is training attention and is taking care

5 We refer here to non-ideational forms of meditation. In ideational meditation, the meditator focuses on an idea and pursues a certain intellectual activity (see also Fontana, 2007). In non-ideational meditation, one does not have an object-directed state of mind.

of oneself. Meditation is therefore an act of de-proletarianization, it is the cultivation of attention itself, in an act of self-care but also in an act of responsibility for what is non-self.

6. Conclusion

Our point of departure was Stiegler's description of attention rooted in the interweaving of retentions and protentions and the role of psychotechnologies in the destruction of attention. Stiegler, not unlike many others, highlights how current digital technologies in service of the needs of the market are a major factor in the destruction of attention to such an extent that we may say that there is a war for attention going on. That war for attention is part of the battle between the programming institutions and the programming industries. Whereas the latter destroy attention, the programming institutions form attention. Educational institutes are part of this counterforce to attention destruction, but await a context of sociotherapy in which the psychotechnologies with their poisonous effects are turned into emancipatory, politically and democratically appropriated nootechnologies. Stiegler seems to put his hopes on the state and institutions financed by the state. At the same time, the elite troops that should provide the necessary elements for a new *otium* of the people are also affected by the destruction of attention. The last section discussed meditation as a second kind of counterforce. Even though Stiegler does not explicitly discuss the position and role of meditation, there are a number of elements in Stiegler's exposition that lead to a view on meditation as an important means in the formation of attention and thus the process of de-proletarianization.

If life is what you fill your attention with, then focusing or directing attention is one of the most valuable abilities for knowing how to live. Instead of letting our attention be hijacked by the market and the economic needs of neoliberal capitalism, being in charge of what happens to our attention is not only a basic need for a human existence that transcends subsistence but maybe also a basic right that needs protection given the current conditions of the attention economy.

REFERENCES

Bauwens, M. & Lievens, J. (2015). *Sauver le monde. Vers une société post-capitaliste avec le peer-to-peer* (with a preface by B. Stiegler). Paris: Editions Les Liens Qui Libèrent. French translation of *De Wereld Redden. Met Peer-to-Peer naar een Post-Kapitalistische Samenleving.* Antwerpen: Hautekiet;

Davenport, T.H. & Beck, J.C. (2001). *The Attention Economy. Understanding the New Currency of Business.* Boston, Massachusetts: Harvard Business School Press;

Derrida, J. (1972). *La Dissémination.* Paris: Editions de Seuil. English translation (1981) *Dissemination* (transl. B. Johnson). London: The Athlone Press;

Derrida, J. (1989 [1962]). *Edmund Husserl's Origin of Geometry: An Introduction* (trans. J.P. Leavey). Lincoln/London: University of Nebraska Press;

Fasching, W. (2008). Consciousness, self-consciousness, and meditation. *Phenomenology and the Cognitive Sciences,* 7, 463-483;

Fontana, D. (2007). Meditation. In M. Velmans and S. Schneider (Eds.), *The Blackwell Companion to Consciousness* (p. 164-162). Oxford: Blackwell;

Fitzpatrick, N. (2020). Questions concerning attention and Stiegler's therapeutics. *Educational Philosophy and Theory,* 52(4),348-360;

Foucault, M. (1983). L'écriture de soi. *Corps écrit, 5*(1), 3-2;

Foucault, M. (1986). *The History of Sexuality, Vol. III The Care of the Self.* New York: Pantheon Books;

Foucault, M. (1988). Technologies of the self. In L.H. Martin, H. Guman, & P.H. Hutton (Eds.), *Technologies of the Self. A Seminar with Michel Foucault* (pp. 16-49). Massachusetts: The University of Massachusetts Press;

Foucault, M. (2008). *The Birth of Biopolitics. Lectures at the Collège de France, 1978-1979* (Ed. M. Senellart, transl. G. Burchell). New York: Palgrave MacMillan;
Foucault, M. (2009). *Security, Territory, Population. Lectures at the College de France 1977-1978* (Ed. M. Senellart, transl. G. Burchell). New York: Palgrave MacMillan;
Hayles, N.K. (2007). Hyper and deep attention: The generational divide in cognitive modes. *Profession,* 13, 187-199;
Husserl, E. (1970 [1954]). *The Crisis of the European Sciences and Transcendental Phenomenology* (Transl. D. Carr). Evanston, Ill.: Northwestern University Press;
Husserl, E. (1991 [1966]). *On the Phenomenology of the Consciousness of Internal Time (1893–1917)* (Trans. J. B. Brough). Dordrecht: Kluwer Academic Publishers;
Lee, S. (2019). Educational methods and congitive modes: Focusing on the difference between Bernard Stiegler and N. Katherine Hayles. *Educational Philosophy and Theory,* (52)4, 376-383;
Lemmens, P. (2009). Van de biomacht van de staat naar de psychomacht van de markt. De receptie van Foucault in het werk van Bernard Stiegler. *Krisis, Tijdschrift voor Actuele Filosofie,* 3, 86-99;
Lemmens, P. (2012a). *Reclaiming the mind.* Het neoliberalisme als voltooid nihilisme en de noodzaak een nieuwe cultuur van de geest uit te vinden. In T. Bakker and R. Brouwer (Eds.), *Vrijheid. Maar voor wie?* (pp. 1-11). Utrecht: Uitgeverij IJzer;
Lemmens, P. (2012b). De strijd om de geest in het huidige kapitalisme. Inleiding bij Bernard Stiegler. *Open,* 24, 50-52;
Lemmens, P. (2014a). Eros en techniek. De annexatie van het verlangen door het kapitaal bij Marcuse en Stiegler. *Wijsgerig Perspectief op Maatschappij en Wetenschap,* 54(3), 22-31.
Lemmens, P. (2014b). Digital network technologies and the (dis)empowerment of the general intellect. A Stieglerian pharmacological view. Paper for the symposium *The Art of Living with ICT*, Brussels, April 25th 2014. Retrieved from https://www.academia.edu/8144444/Digital_network_technologies_and_the_dis_empowerment_of_the_general_intellect_in_cognitive_capitalism_A_Stieglerian_pharmacological_view;
Rossouw, J. (2015). Bernard Stiegler se esteteologie en die nuwe *otium* van die volk. *LitNet Akademies, 12*(3), 176-2010;
Rouvroy, A. & Berns, Th. (2013). Algorithmic governmentality and prospects of emancipation. *La Découverte, 1*(177), 163-196;
Simon, H. (1971). Designing Organizations for an Information-Rich World. In M. Greenberger (Ed.), *Computers, Communications, and the Public Interest* (pp. 38-52). Baltimore: The Johns Hopkins University Press;
Stiegler, B. (1998). *Technics and Time, 1. The Fault of Epimetheus* (Transl. R. Beardsworth & G. Collins). Stanford: Stanford University Press;
Stiegler, B. (2009). *Acting Out* (Transl. D. Barison). Stanford: Stanford University Press;
Stiegler, B. (2010). *Taking Care of Youth and the Generations* (Transl. S. Barker). Stanford: Standford University Press;
Stiegler, B. (2011). *The decadence of industrial democracies. Disbelief and discredit* (Vol. 1) (Transl. D. Ross and S. Arnold). Cambridge: Polity Press;
Stiegler, B. (2014). *Symbolic Misery 1: The Hyperindustrial Epoch* (Transl. B. Norman). Cambridge: Polity Press;
Stiegler, B. (2015). *Symbolic Misery 2: The* katastrophè *of the sensible* (Transl. B. Norman). Cambridge: Polity Press;
Yeung, K. (2017). "Hypernudge": Big Data as a mode of regulation by design. *Information, Communication & Society, 20*(1), 118-136;
Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

FRANCESCO TAVA
University of the West of England, Bristol (UK)
francesco.tava@uwe.ac.uk

SOLIDARITY AND DATA ACCESS: CHALLENGES AND POTENTIALITIES

abstract

This paper provides an account of the challenges and potentialities of a solidarity-based approach to data access and governance. To do that, it offers an infraethical understanding of solidarity that describes it as a structural moral enabler that can sustain collective action and risk taking. The paper ends with a brief discussion of health data access as a possible case study to test this approach.

keywords

solidarity, data access, distributed morality, infraethics

Phenomenology and Mind, n. 20 - 2021, pp. 118-126
DOI: https://doi.org/10.17454/pam-2010
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

The emergence of big-data technologies and the consequent increasingly networked character of society caused a paradigm shift in contemporary ethics. Traditional ethical theories, which assume the relevance of individual responsibility and individual decision making, needed to be conceptually reframed insofar as new moral actors—for example, artificial multiagent systems—entered the moral domain and transformed it deeply. From this reconceptualisation, new, big-data-based ethical models emerged, which soon took a central position in philosophical debates. These include models of distributed morality (Floridi, 2013a; Heersmink, 2017), which allow ethicists to address the fact that, because of complex interactions among multiagent systems, moral actions and responsibilities are no longer centred on individuals, but are also distributed across society.

In this paper, I will first provide a brief analysis of these new models, highlighting both their main characteristics and their potential societal effects (section 2). In doing this, I will focus on the notion of infraethics (or infrastructure ethics), which refers to the "first-order framework of implicit expectations, attitudes, and practices that can facilitate and promote morally good decisions and actions" (Floridi, 2013a, p. 738; see also Floridi, 2017). Among these practices are trust, respect, transparency, and reliability. I will then argue in favour of including solidarity among such infraethical practices (section 3). To do that, I will provide a brief account of solidarity, which I contend is particularly relevant because it is conducive to sustained collective action and risk taking. More precisely, solidarity can be understood as a moral and political desideratum insofar as it stimulates and supports longer-term and risk-laden collective action aimed at addressing perceived injustices (Meacham and Tava, 2021; Prainsack and Buyx, 2017; Scholz, 2008)—all characteristics that clarify how solidarity is not merely a descriptive notion that indicates a certain form of human togetherness, but a fundamental, infrastructural dimension of democratic life.

In section 4, I will then contend that an infraethical model of solidarity might offer a powerful tool for tackling one of the most challenging issues that underpins our digital way of life—namely, how we own and use digital data. Looking at one of the most advanced models of data access and governance (the evidence-based, default-open, risk-managed, user-centred [EDRU] model—see Ritchie, 2014 and Ritchie and Green, 2016), one can easily detect the growing importance that collective and societal aspects have in this domain. Whilst traditional models are defensive in nature as they are essentially anchored in the costs and risks to the data owner, this more advanced model relies on the principle that society is the relevant locus of costs and benefits. I will argue that adding an infraethical narrative to this approach, which

would pinpoint solidarity-based practices' impact on data sharing, would permit data-access researchers and stakeholders to better understand the inner mechanisms of the community of interest that data owners and users constitute. Introducing the concept of solidarity into data-access practices would therefore highlight the growing relevance of collective interests and aims to the determination of how personal information is owned and used.
I will conclude by showing why developing an infraethical, solidarity-based approach to data access is particularly urgent now that humanity is facing the Covid-19 pandemic (section 5). Today more than ever, access or lack of access to data (specifically, health data) might have major consequences for public security and public health. I will show how an approach based on solidarity can help implement models of data access and sharing that place societal needs at their centre.

2. Distributed Morality and Infraethics

Distributed morality (Floridi and Sanders, 2004; Floridi, 2013a; Floridi, 2013b, chap. 13) is a phenomenon with ever-increasing impact on society due to the emergence and dissemination of new information and communication technologies that range from digitisation techniques to algorithmic decision making. The idea that morality is not exclusively centred on the individual and on her capacity to act autonomously and responsibly is not new. For instance, phenomena such as collective moral responsibility and obligation, which highlight the significance of non-individual decision making and accountability, have often been analysed by ethicists, jurists, and political scientists (Isaacs, 2011; Isaacs and Vernon, 2011; Hess et al., 2018). However, this research is gaining momentum in light of the central role that artificial agents and hybrid multiagent systems are playing in the infosphere in which we live. Our everyday life (on both the personal and societal levels) is increasingly influenced by operations undertaken by artificial agents—that is, by "sufficiently informed, 'smart', autonomous artefacts, able to perform morally relevant actions, independently of the humans who engineered them, causing 'artificial good' and 'artificial evil'" (Floridi, 2013a, p. 728). Interlinked technical innovations such as, for instance, big-data technologies, deep learning, the semantic web, and the Internet of Things have given rise to a variety of new agents that are progressively less dependent on their creators and therefore play an independent role in the public sphere. This phenomenon has a major impact on certain pivotal moral concepts, such as responsibility and power, whose relational (rather than individual) facet is progressively coming to light (Zwitter, 2014).
This scenario is made even more complex by the formation of multiagent systems, which can be human, artificial, or hybrid. The fundamental assumption of distributed morality is that a series of small actions, which a number of agents (human, artificial, or mixed) perform and which taken individually are morally neutral or morally negligible, can generate morally charged (either good or evil) big actions as soon as they are embedded in a powerful multiagent system. Floridi (2013a) gives several examples of this phenomenon, focusing on the good side of the coin. Take, for instance, a company that reinvests part of the profits deriving from its customers' purchases to sustain humanitarian projects. Such an operation involves a series of actions (performed by the company itself, its customers, and all the artificial agents that constitute the company's technological platform) that taken individually can be described as morally neutral. Only the big action of this multiagent system, and not the actions of its components, seems to be morally loaded.
Assuming that the majority of individual actions that human or artificial agents perform are morally neutral or morally negligible, the central challenge in a distributed-morality framework is how to ensure that the sum of these individual actions will generate a good rather than an evil outcome. This can be ensured by strengthening the resilience and fault tolerance of our ethical environment whilst weakening its inherent inertia (Floridi, 2013a,

p. 736). In other words, we have to learn how to identify potentially good actions and facilitate their aggregation into virtuous multiagent systems, while at the same time we have to isolate and neutralise possibly evil actions. To do that, we need to establish what Floridi (2013a) calls "infraethics"—that is, a "first-order framework of implicit expectations, attitudes, and practices that can facilitate and promote morally good decisions and actions" (p. 738). These expectations, attitudes, and practices operate as moral enablers insofar as they help aggregate potentially good actions and disperse potentially negative actions, although they do not need to be morally characterised either positively or negatively. Examples of such enablers are information transparency (Turilli and Floridi, 2009) and privacy (Floridi, 2005; 2006), trust online (Taddeo, 2009; 2010), and openness (Chopra and Dexter, 2008). In what follows, I will contend that solidarity (as I define it) should be understood as an infraethical phenomenon that can facilitate the emergence of positive moral behaviours in an environment characterised by distributed morality.

3. Why Solidarity?

The concept of solidarity has a relatively short history compared with other pivotal moral and political concepts such as democracy and freedom (Metz, 1999; Stjernø, 2004). Various accounts of solidarity have identified several traits that are deemed as essential to understanding its meaning and function. Solidarity relations are often described as historically grounded on a principle of equal and freely given support among peers (Metz, 1999). People who establish a bond of solidarity must recognise their similarity in a relevant respect (Prainsack and Buyx, 2017), whether it be similarity of features (for example, belonging to a certain social group or sharing the same language, nationality, or religion) or to similarity of motivation and agency (for example, sharing the same goals). Other commonly identified traits include mutual responsibility (Bayertz, 1999), the recognition of individual freedom (Honneth, 1995) and the aim of democratically establishing it (Brunkhorst, 2005), and the entailment of positive duties such as cooperation and reciprocity (Scholz, 2008, p. 58).

I think that at least two features stem from these characterisations, which I contend are essential to defining solidarity relations. First, in order for people or institutions to set solidarity in motion, they must share certain goals and ideals of justice. In this sense, although acknowledging their shared status (for example, qua compatriots or fellow workers) might help create or strengthen solidarity bonds, the emergence of solidarity also requires collectively aiming to overcome perceived inequalities or injustices that a certain social condition might involve.[1] Take, for instance, Thelma and Louise. They might enjoy knowing that they are both American citizens, and the knowledge of this similarity may be the origin of a number of sentiments such as sympathy, trust, or even camaraderie. Nonetheless, this would not be a sufficient condition for the emergence of a solidarity relation. To be in solidarity, Thelma and Louise must also act in unison in response to (for instance) a threat or act of violence and to re-establish what they believe is just.

Second, the sharing of goals depends on the willingness of individuals and groups to also assume the costs and burdens that these goals might involve with at least the expectation that their action will be reciprocated. This element helps differentiate solidarity from other intersubjective relationships such as benevolence and charity. For instance, when a group of workers decide to establish a mutual relationship of solidarity, they know that this decision has consequences that might be detrimental (for example, job loss, discomfort due to a

1 By justice and injustice, I do not mean here any specific theory of justice or juridical system but both the perception of such phenomena that individuals and groups may have as part of their lifeworld experience and the meanings and values that they form from this perception.

prolonged strike) and nonetheless decide to act anyway and to equally share the burden of these potential consequences.
We can also argue that the first element justifies the second in that sharing goals or ideals incites willingness to also share costs and risks with an eye towards longer-term gain and an increased probability of achieving the shared goals. The willingness to shoulder burdens and take risks with the aim of pursuing potentially longer-term political and socioeconomic goals is furthermore supported, in a virtuous circle, by the sharing of the risk, which can mitigate individual risk. Hence solidarity facilitates collective decision making, collective action, and longer-term political action and planning—all characteristics that can be described as desiderata of a democratic society. Solidarity is thus not just a descriptive notion that indicates a certain form of human togetherness, but a fundamental, infrastructural dimension of democratic life.
This infrastructural dimension is what suggests that solidarity might be employed within an infraethical framework as a powerful moral enabler. Like other enablers (for example, trust or privacy), solidarity as such is morally neutral insofar as it is not inherently good or evil (even a group of criminals can share goals, a perception of justice, and risks and burdens). Despite this neutrality, however, solidarity has the power to foster collective action and collective risk taking and may or may not (depending on the use that we make of it) therefore be conducive to morally good outcomes and consequently to a general reinforcement and amelioration of the infraethical environment in which it operates. In the next section I will portray a concrete scenario in which solidarity can be employed in this way.

4. Solidarity and Data Access

How we collect, own, and employ data is one of the oldest questions that humankind has had to address. The capacity to retain information from our own experience and to make good use of this information over time without having to relearn notions and practices that (as a species) we have already learned is a fundamental human skill whose evolution and perfection enabled the progress of humanity. In the present age, new problems have arisen concerning how to analyse and organise the increasing amount of information that new technological devices allow us to obtain. The ontological and ethical consequences of big data are tightly connected to the topic of distributed morality insofar as they require us to move away from traditional moral concepts, such as individual responsibility and decision making, causality, and culpability, and to envision forms of relational, systemic morality. The development of network and information ethics (Bynum, 2011) has allowed researchers to address this modified scenario and to tackle new ethical dilemmas stemming from it. The ethics of big data shows how moral responsibility is no longer entirely ascribable to individual agents, but is rather spread throughout a network of data generators, collectors, and users. Consequently, whilst concepts that were once pivotal, such as individual agency, lose their centrality, other phenomena, such as network knock-on effects—that is, the unintended consequences of and collateral damage caused by actions within a network—become of primary importance (Zwitter, 2014).
One of the sectors that the current setup has most profoundly altered is data access and data security (Micheli et al., 2018). How do we grant access to data in a responsible and productive way—that is, without compromising (for instance) the property, security, and privacy of data owners and users? What are the ethical guidelines that citizens and institutions should follow to make good use of their data? An example of how new digital, data-driven technologies are altering established notions in this area is group privacy. It is well known how public and private companies make strategic use of data analysis in order to mine data about habits and customs of citizens and customers. These datasets constitute raw material that, once thoroughly refined, can generate valuable outcomes such as higher revenues or more efficient

policies. The collection and analysis of personal data involves the risk of privacy breach, which is traditionally prevented through de-individualisation. In other words, according to traditional privacy protocols, personal data can be collected as long as they are disjointed from the specific person who generated them. Although de-individualisation makes *personal* anonymisation possible, it cannot guarantee *group* anonymisation (Dwork, 2006; Zwitter, 2014). This means that de-individualised data still provide information regarding the opinions, habits, and tastes of social groups and population strata, and those data can be employed in a targeted way for specific purposes (from marketing to political campaigning). Big-data technologies have the power to facilitate this process and to provide more and more fine-grained pictures of group characteristics thanks to their ability to enhance hyperconnectivity and identify hidden correlations among data. This aspect raises huge ethical issues and must be considered in the development of innovative and data-informed privacy policies.

These and other aspects have substantially informed the design of new models of data access and governance. Traditional models of data access are defensive in nature and anchored in the costs and risks to the data owner. This perspective implies that the primary aim of any data-access strategy is to prevent malicious misuse, which results in the extensive use of worst-case scenarios and protection against hypothetical possibilities (Ritchie and Green, 2016, p. vi). A default-closed strategy underpins these models. Their developers start from the fundamental question "Are we allowed to do this?", which is defensive insofar as it interprets regulations as a shield. The aforementioned technological advancements in the field of digital, data-driven technologies, and the subsequent emergence of hypernetworked societies widely characterised by distributed moral frameworks, help reveal the inadequacy of default-closed traditional models. Recent analytic reports such as the Data Access Project[2] exemplify an alternative approach: "An alternative is to consider the law as one of the tools to be used in designing data strategies; the appropriate question is 'how do I lawfully achieve what I want?'. This alternative approach, of deciding objectives and studying the legal framework to see how an objective can be achieved, is a key part of the EDRU ethos" (Ritchie and Green, 2016, p. 33). According to its developers, an EDRU approach has the power to substantially modify and improve the way in which data are owned and used (Ritchie, 2014; 2016; Ritchie and Green, 2016).

The aim of this paper is not to provide an exhaustive analysis of the EDRU approach or of any other innovative model of data access. Of all the traits that are central in the EDRU approach, the one that is most relevant here is its focus on the collective and societal aspects and constraints of data access. According to this model, the relevant locus of costs and benefits of data access is no longer the data owner but society, which corresponds to a community of interest between data owners and users. Therefore, the decision making is not grounded on individual agency and responsibility, but (in line with the fundamental assumptions of distributed morality) rather corresponds to a balance of subjective probabilities. On the basis of these premises, this model establishes that data should be made available for research purposes if "the expected benefit to society outweighs the potential loss of privacy for the individual" (Ritchie and Green, 2016, p. v). The most challenging question is how to calculate this benefit. I contend that a solidarity-based approach would offer a decisive contribution to overcoming this challenge. In section 3, I characterised solidarity as a principle that has the power to foster collective action and risk taking and that can therefore be conducive to morally good outcomes. This makes solidarity a moral enabler in an infraethical environment.

2 This report was prepared for the Australian Department of Social Services by Elizabeth Green and Felix Ritchie of Bristol Economic Analysis at the University of the West of England, Bristol.

For the same reason, solidarity can also be seen as a valuable tool to identify what societal benefits justify individual risk taking. In other words, analysing the solidarity relations among the members of a society might help determine their willingness to shoulder burdens and take risks (for example, to partially renounce their privacy by granting broader data access) in order to benefit their community of interest. In the next section, I will briefly point to a potential case study that might highlight this mechanism.

5. Data Access and the Public Good

In the middle of a global pandemic, the major impact of data access on public health has become one of the most discussed topics not just among experts but more generally in the public discourse. This discussion essentially concerns the potentialities and boundaries that accessing and sharing health data in order to track the confirmed cases of Covid-19 infections and prevent further epidemic outbreaks would imply. This ongoing crisis is a perfect example of how solidarity practices might successfully be implemented in the governance of data access. Appeals to solidarity have emerged in various scientific reports on the pandemic. The European Group on Ethics in Science and New Technologies (2020) issued a statement on 2 April 2020 in which they claimed that "this pandemic should be seized, not as an opportunity but as a call, to foster solidarity at the European and global level. This must manifest itself in concrete actions such as the honest sharing and pooling of information, experiences, innovations and resources". Barbara Prainsack (one of the authors of the aforementioned statement), discussing the difficulties of raising solidarity in times of pandemic, has recently claimed that "rather than only celebrating solidarity where we see it happen, we need to build institutions and circumstances that can make solidarity stable and lasting" (Prainsack, 2020; see on this also Wagenaar & Prainsack, 2020). These institutions include public infrastructures, solidaristic healthcare, and fair taxation systems. Implementing a solidarity-based healthcare system depends on the willingness of data owners to share data in order to help trace the contagion and allow public health institutions to take countermeasures that might lead to clear public goods and societal benefits. Several analyses of the importance of a solidarity-based approach to biomedicine and public health precede the pandemic and highlight new areas of application for this concept (for example Prainsack and Buyx, 2011; 2017). What is still lacking is a more general account of the potentialities of this solidarity-based approach—whereby solidarity is understood as an infrastructural moral enabler in today's society—in the broader field of data access. This paper provides a first step in this direction.

REFERENCES

Arbuckle, L. & Ritchie, F. (2019). The Five Safes of Risk-Based Anonymization. *IEEE Security and Privacy*, 17, 5, 84–89. doi:10.1109/MSEC.2019.2929282;

Bayertz, K. (1999). Four Uses of 'Solidarity'. In K. Bayertz (Ed.), *Solidarity*. Dondrecht: Kluwer;

Brunkhorst, H. (2005). *Solidarity: From Civic Friendship to a Global Legal Community*. Cambridge, MA: MIT Press;

Bynum, T. (2011). Computer and Information Ethics. In *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2011/entries/ethics-computer/;

Chopra, S. & Dexter, S. (2008). *Decoding Liberation: The Promise of free and Open Source Software*. New York: Routledge;

Dwork, C. (2006). Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, I. Wegener I. (Eds.), *Automata, Languages and Programming*. ICALP 2006. Lecture Notes in Computer Science, vol. 4052. Berlin-Heidelberg: Springer;

Floridi, L. (2005). The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology*, 7, 4, 185–200. doi:10.1007/s10676-006-0001-7;

Floridi, L. (2006). Four Challenges for a Theory of Informational Privacy. *Ethics and Information Technology*, 8, 3, 109–119. doi:10.1007/s10676-006-9121-3;
Floridi, L. (2013a). Distributed Morality in an Information Society. *Science and Engineering Ethics*, 19, 3, 727–742. doi:10.1007/s11948-012-9413-4;
Floridi, L. (2013b). *The Ethics of Information*. Oxford: Oxford University Press;
Floridi, L. (2017). Infraethics—on the Conditions of Possibility of Morality. *Philosophy & Technology*, 30, 4, 391–394. doi:10.1007/s13347-017-0291-1;
Floridi, L. & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, 14, 3, 349–379. doi:10.1023/B:MIND.0000035461.63578.9d;
Heersmink, R. (2017). Distributed Cognition and Distributed Morality: Agency, Artifacts and Systems. *Science and Engineering Ethics,* 23, 2, 431–448. doi:10.1007/s11948-016-9802-1;
Hess, K., Igneski V. & Isaacs T. (2018). Collectivity: Ontology, Ethics, and Social Justice. London: Rowman and Littlefield International;
Honneth, A. (1995). *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Cambridge: Polity Press;
Isaacs, T. (2011). *Moral Responsibility in Collective Contexts*. Oxford: Oxford University Press;
Isaacs, T. & Vernon R. (2011). *Accountability for Collective Wrongdoing*. Cambridge: Cambridge University Press;
Meacham, D. & Tava F. (2021). The Algorithmic Disruption of Workplace Solidarity: Phenomenology and the Future of Work Question. *Philosophy Today*, 65, 3, 571–598. doi:10.5840/philtoday2021519408;
Metz, K.H. (1999). Solidarity and History: Institutions and Social Concepts of Solidarity in 19th Century Western Europe. In K. Bayertz (Ed.), *Solidarity*, pp. 191–207. Dondrecht: Kluwer;
Micheli, M., Blakemore M., Ponti M. & Craglia, M. (2018). The Governance of Data in a Digitally Transformed European Society. Second Workshop of the DigiTranScope Project, European Commission, JRC114711. Retrieved from https://ec.europa.eu/jrc/communities/sites/jrccties/files/jrc_digitranscope_report_-_oct_2018_data_governance_workshop_1.pdf;
Prainsack, B. (2020). Solidarity in Times of Pandemic. Democratic Theory, 7, 2, 124-133. doi:10.3167/dt.2020.070215;
Prainsack, B. & Buyx, A. (2011). *Solidarity as an Emerging Concept in Bioethics*. London: Nuffield Council on Bioethics;
Prainsack, B. & Buyx, A. (2017). *Solidarity in Biomedicine and Beyond*. Cambridge: Cambridge University Press;
Ritchie, F. (2014). Access to Sensitive Data: Satisfying Objectives Rather than Constraints. *Journal of Official Statistics*, 30, 3, 533–545. doi:10.2478/JOS-2014-0033;
Ritchie F. (2016). Can a Change in Attitudes Improve Effective Access to Administrative Data for Research? Working Papers in Economics No. 1607, University of the West of England, Bristol. Retrieved from http://eprints.uwe.ac.uk/29646/1/1607.pdf;
Ritchie, F. & Green, E. (2016). *Data Access Project Final Report*. Australian Department of Social Services. Retrieved from http://eprints.uwe.ac.uk/31874/1/Data%20Access%20Project%20Final%20Report%20v2.00%20Final%20DSS.pdf;
Scholz, S. (2008). *Political Solidarity*. University Park, PA: Penn State University Press;
European Group on Ethics in Science and New Technologies (2020). *Statement on European Solidarity and the Protection of Fundamental Rights in the COVID-19 Pandemic*, 2 April 2020. Retrieved from https://ec.europa.eu/info/sites/info/files/research_and_innovation/ege/ec_rtd_ege-statement-covid-19.pdf;
Stjernø, S. (2004). *Solidarity in Europe: The History of an Idea*. Cambridge: Cambridge University Press;

Taddeo, M. (2009). Defining Trust and e-Trust: Old Theories and New Problems. *Journal of Technology and Human Interaction*, 5, 2, 23–35. doi:10.4018/jthi.2009040102;
Taddeo, M. (2010). Modelling Trust in Artificial Agents, a First Step toward the Analysis of e-Trust. *Minds and Machines*, 20, 2, 243–257. doi:10.1007/s11023-010-9201-3;
Turilli, M., Floridi, L. (2009). The Ethics of Information Transparency. *Ethics and Information Technology*, 11, 2, 105–112. doi: 10.1007/s10676-009-9187-9;
Wagenaar, H. & Prainsack B. 2020. The New Normal: The World after COVID- 19. A Blog in Four Parts. Retrieved form https://medium.com/@hendrik.wagenaar/the-new- normal-the-world-after-covid-19-201189e22545.

NATALIA SATOKHINA
Yaroslav Mudryi National Law University
nataliasatokhina@gmail.com

YULIA RAZMETAEVA
Yaroslav Mudryi National Law University
yulia.razmetaeva@gmail.com

"THE LOSS OF EXPERIENCE" IN DIGITAL AGE: LEGAL IMPLICATIONS

abstract

Exploring the history of our experience, Hannah Arendt reveals not only a radical transformation of its structure, but also the loss of experience as such and its replacement with technology. In order to identify the place of law in this process, we are trying to clarify the legal aspect of experience in terms of phenomenological hermeneutics and to trace its transformation in the digital age. The experience of law is thought of as one of the aspects of our mode of being-in-the-world, which is based on openness to the world and consists in the mutual recognition of people in their dignity. Digital technologies, in turn, contribute to replacing fundamental openness with illusory freedom in cyberspace. The latter, unlike the public realm as a realm of action of many, and in this sense legal realm, is based primarily on productive activities of one and no longer requires law.

keywords

phenomenological hermeneutics, experience of law, the loss of experience, digital technologies, social media

Phenomenology and Mind, n. 20 - 2021, pp. 128-136
DOI: https://doi.org/10.17454/pam-2011
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

The idea of experience as such embodies one of the key motifs of phenomenology – the desire to think of the world, first of all, not as the object of our cognition or technical domination, but as what happens to us. With regard to the philosophy of law, it is about comprehending those aspects of our being-in-the-world that make up the experience of law and are a kind of DNA of law, the only resource that makes it possible to revive law even in the ashes of wars and genocides, when norms and institutions are completely destroyed.

However, experience is obviously not constant. Hannah Arendt in *The Human Condition* explores the history of our experience, showing not only a radical transformation of its structure, but also the loss of experience as such and its replacement with technology. Today, digital technologies have greatly changed the experience of many. Does this leave room for law in the structure of human experience?

In this article we are trying to clarify the legal aspect of experience in terms of phenomenological hermeneutics and to trace its transformation in the digital age. Thinking law in terms of human experience, we move from eidetics (identifying the essential contours of the idea of law, extracted from its real-word instantiations) to constitution analysis (describing how law registers within human experience).

2. Experience of Law in the Structure of Fundamental Experience

Phenomenology addresses a basic level of experience as a person's fundamental engagement with the world. Moreover, phenomenologists refuse to search for a "true" reality outside the reality that is given to us through our experience. In turn, phenomenological hermeneutics regards any form of experience as an experience of understanding, or a meaningful relation. From birth to death, we understand the world and understand ourselves in it, and in this way give rise to meaning. There is no meaning as such, either in the world or in a man himself; meaning is always a meeting.

2.1. The Essence of the Hermeneutic Experience

The ontological structure of the experience of meaning is a circle: being-in-the-world is that whole, on the understanding of which depends the understanding of its parts – the world and the man. In turn, the understanding of being-in-the-world is determined by the understanding of the world and the self-understanding of the man (Heidegger, 1996, pp. 142-143). The ontological hermeneutic circle assumes that understanding, on the one hand, has always prerequisites (what Martin Heidegger defines as "thrownness"), and on the other, is always incomplete, that is, it is not something present, but its own possibility ("project") (Heidegger, 1996, pp. 134-139). It is not about the dictatorship of the world or the free-floating creativity of a man, but rather about the dialogue, as a result of which meaning is born.

According to Hans-Georg Gadamer, dialogue is a universal model of all experience. The situation in which we are in relation to the world is like a conversation, the participants of which try not to defend their position and not to accept the position of the interlocutor, but to understand the essence of what is being said (Gadamer, 2004, p. 387). Accordingly, the logical structure of any experience is the question, that presupposes the principled openness to the various answers (Gadamer, 2004, pp. 356-357). Thus, a condition for the possibility of understanding is a recognition of the interlocutor in his claim to be heard not in the sense of simply acknowledging of otherness, "but in such a way that it has something to say to me" (Gadamer, 2004, p. 355). It is noteworthy that as a paradigmatic example of the hermeneutic experience Gadamer considers the experience of law as a *phronetic*[1] experience, which, unlike the experience of the theoretical (cognition of the unchanging truth) and the experience of art (free creativity), is an experience of meaning as a meeting, which always presupposes the experience of the Other as a co-author of the common meaning and involves constant work of recognition (2004, pp. 321-336). Also Paul Ricoeur, following Gadamer, views *phronesis* as the essence of the experience of law, and the experience of law as an illustrative example of *phronesis* as a responsible judgment in a particular situation that always contains an element of risk (2000, pp. XXI-XXII).

2.2. Experience of Law

The concept of experience of law is not a well-established concept either in phenomenology or in the philosophy of law. In its most general form in phenomenology, the experience of law is considered as a part of being-in-the-world, the experiencing of law, during which the ontological region of law is constituted (see for instance Reinach, 1953; Husserl, 1955; Alekseev, 1999). In this, despite the variety of answers to the question of how are we experiencing law, they are all somehow connected with overcoming the initial asymmetry between the I and the Other and the concept of mutual recognition. This is probably why, in the theory of hermeneutic experience as the experience of the Other, the experience of law becomes a paradigmatic example. Indeed, it is the refusal to recognize someone or the threat of such a refusal that gives rise to the claim to justice, and with it the experience that we usually consider to be the experience of law.

In this article, we regard the experience of law as an aspect of fundamental experience, which is localized in the constant tension between mutual recognition, that enables the experience of meaning, and the unavoidable risk of people not recognizing each other. In this sense, Ricoeur stresses that although it is possible to talk about mutual recognition only in the mode of desirability, which is neither descriptive nor normative, it, at the same time, is rooted in the essence of law and forms its basis (1996, p. 33). One can assume that it is this shaky status of mutual recognition between fact and norm that caused the existence of institutions designed to increase the weight of the latter – legal institutions in which we confirm each other's recognition. Whereas in traditional societies, mutual recognition is confirmed by rituals (such as gift exchanges), today, the direct expression of it is human rights, which guarantee everyone the minimum public recognition of his dignity (Hénaff, 2019, part 2). Thus, recognizing the norm, we simultaneously recognize the generalized Other, and vice versa. Thereby our hypothesis is that law is brought about by the fact that, being immanent to experience, the possibility of recognition is always accompanied, like a shadow, by the

1 Gadamer appeals to the Aristotelian idea of *phronesis* as the capacity of acting in an unforeseen situation, the willingness to meet with the unexpected, in contrast to cognition based on axioms. Aristotle distinguishes between three forms of experience: 1) theoretical experience based on scientific knowledge (*episteme*), that concerns unchanging things; 2) art based on know how (*techne*) as a kind of technology; 3) practical experience based on *phronesis* – practical wisdom regarding correct actions in a particular situation (See Aristotle, 2009, pp. 102-117).

risk of non-recognition. In this risky zone, halfway from non-recognition to recognition, we are dealing with law. To put it differently, the experience of law arises from the claim to recognition as a claim to justice. In this, a theory of the hermeneutic experience shows that the meaning of justice, like any meaning, is not contained in us or outside us; it appears in a meeting again and again, and in this sense, "the just in the final analysis qualifies a unique decision made within a climate of conflict and incertitude" (Ricoeur, 2000, p. XXI). In other words, it is about a *phronetic* experience.

It is in *phronesis* that Heidegger sees not only a model of his own idea of philosophy, but also the most authentic way of dealing with the world in general, that is, the original basis of experience (See Heidegger, 2001). We find a similar opinion in Gadamer: "this is perhaps the fundamental form of experience (Erfahrung), compared with which all other experience represents an alienation, not to say a denaturing" (2004, p. 319). Being an integral (and the most authentic) part of human experience, *phronesis* at the same time is an aspect of our fundamental experience that underlies the experience of law.

However, the structure of human experience changes under the influence of technology, which forces us to rethink also the experience of law.

2.3. "The Loss of Experience"

Tracing the dramatic history of our experience, Hannah Arendt reveals a trend, which she describes as the loss of experience. It is about the gradual disappearance from the structure of human experience of the specifically human capacity for generating meaning in the process of thinking and action. According to Arendt, thinking and action, impossible in isolation and occurring in being with others, are supplanted first by work, where a person is connected at least with the world of things, and then by labor – a process aimed at satisfying biological needs, when people are thrown back to their own bodies. In turn, she says, technological progress turns the laboring society into a society of jobholders, that demands of its members nothing but automatic functioning, and all human activities "appear not as activities of any kind but as processes" (1998, p. 322). As a result, we are observing not just a radical transformation of our experience, but "the loss of experience" as such as an experience of *worldliness*, which makes meaningful human existence possible (Arendt, 1998, pp. 320-325). Such world-oriented forms of experience are thinking and action – properly human meaning-generating experience, which, unlike work and labor, "needs the surrounding presence of others" (Arendt, 1998, p. 188).

Drawing on Kant's distinction between intellect and reason, Arendt distinguishes the monological faculty of cognition and the dialogical faculty of thinking, aiming respectively at truth and meaning. Thinking, she says, "does not ask what something is or whether it exist at all – its existence is always taken for granted – *but what it means for it to be*" (1978, p. 57). Thinking, understood as a craving for meaning, like action, has the structure of an open question described above, and therefore assumes a readiness for an unexpected answer. It is in this sense that "there are no dangerous thoughts; thinking itself is dangerous" (Arendt, 1971, p. 435).

Thinking in turn has a liberating effect on another faculty, that realizes thinking and makes it manifest in the world of appearances. This "manifestation of the wind of thought" is judging – "the ability to tell right from wrong" (Arendt, 1978, p. 193),

> a human faculty which enables us to judge rationally without being carried away by either emotion or self-interest, and which at the same time functions spontaneously, that is to say, is not bound by standards and rules under which particular cases are simply subsumed, but on the contrary, produces its own principles by virtue of the judging activity itself (Arendt, 2003, p. 27).

And only if we assume the existence of such a human ability – an analogue of the Aristotelian *phronesis* – is morality possible. "For behind the unwillingness to judge lurks the suspicion that no one is a free agent, and hence the doubt that anyone is responsible or could be expected to answer for what he has done" (Arendt, 2003, p. 19). It is judging, along with thinking, that Arendt considers the most human experience, but at the same time the most vulnerable.
In her opinion, the modern age in its concern with maximum profits, as well as technical progress that provides the means to achieve this goal, only contributed to the eternal temptation of a person to get rid of the unpredictability of action's outcome, the irreversibility of the process, and the anonymity of its authors. The experiments proposed to resolve these calamities always end the same way. Generally speaking, Arendt writes, it's about replacing the acts of many in their being with each other by the productive activity of one person, be it the outright tyranny, benevolent despotism or those forms of democracy in which the many form a collective body, excluding any pluralism (1998, pp. 220-221).
Thus, despite the seeming intensification of communication, in fact, we are witnessing the unprecedented disappearance of public realm as a common world that gathers us together and separates us, like a table between those sitting at it. However, the weirdness of contemporary situation

> resembles a spiritualistic seance where a number of people gathered around a table might suddenly, through some magic trick, see the table vanish from their midst, so that two persons sitting opposite each other were no longer separated but also would be entirely unrelated to each other by anything tangible (Arendt, 1998, p. 53).

In turn, the disappearance of public realm means the disappearance of thinking and judging as a meaning-generating experience, which is intersubjective in nature (Arendt, 1978, p. 266). What is it about here is more than loss of common sense; it is "the loss of the quest for meaning and need for understanding" and in this sense the loss of experience, for understanding "is the specifically human way of being alive" (Arendt, 1994, pp. 308, 317).
Thus, the lack of public realm makes an action, and therefore law, impossible. In the absence of a common world, the experience of law as an experience of mutual recognition is fundamentally impossible, and even unnecessary. Driven into our subjective experience, we can no longer hear the other, nor be heard by them.
It is no coincidence that Ricoeur considers the public realm, as Arendt thought of it, as a realm of responsibility and, in this sense, a legal realm (1996, p. 30), and the erosion of the concept of responsibility (and, therefore, law) in the today's world associates with the gradual elimination of the acting from the structure of our experience and the transition from individual fault management to a socialized management of risk (2000, pp. 24-26).
Is it not the disappearance of the public realm that we are observing in the contemporary world of digital technology, where we "are all imprisoned in the subjectivity of our own singular experience, which does not cease to be singular if the same experience is multiplied innumerable times" (Arendt, 1998, p. 58)? As a result, the ability to make judgment has fallen under a hail of blows. It is a constant exhausting presence in the overwhelming flow of information, where we are also being manipulated with the help of the digital tools that we are exposed to. This is a general simplification of complex issues, wrapped in an attractive and bright wrapper, which attracts us to take someone's position – because we are ready to rely on someone's experience (perhaps studying the issue on our own seems too complicated or we are under pressure from society, or we are in the bubble of opinions).

3. Legal Implications of Digital Technologies

It is the prospect of a radical transformation of our moral experience that motivates the concerns of Jürgen Habermas about the rapidly developing biotechnologies, which can lead to the fact that we will no longer be able to understand ourselves as ethically free and responsible creatures (Habermas, 2003, pp. 16-74). However, is it really necessary to intervene in the genome for this? Digitalization seems to have a similar effect.

3.1. Digital Identity in Public and Private Life

Digital world and the dynamics of its growth influence every single individual despite the fact that the degree of its influence differs. It goes without saying that the world is extremely heterogeneous, but in the end the ongoing and massive digitalization reaches even individuals with radically different cultural, economic and social experiences. Legal implications of digital technologies are linked to fundamental things such as autonomy, human rights, rule of law, justice and democracy. Even the least evident manifestations of the digital world still bear subtle influence on everyone and even transform the experience of many. In particular, digital identity has born an unreasonably strong influence on human identity as such. Online activities and social media gradually and imperceptibly change our ideas of ourselves. The images that one share in cyberspace are cemented by an incredibly long digital footprint. The images others share can be far from the truth or intentionally fragmented.

Every day one comes across the visible part of other people's experience in cyberspace. To what extent the images of others could be one-sided? At the same time, one is constantly compelled to compare oneself with others. Disconnecting is becoming increasingly impossible in the digital age.

The fundamental openness which presupposes the initial immersion in the world, certain limits of self-understanding and self-presentation, honesty and, therefore, vulnerability before the Other making a meaning-generating dialogue possible, is replaced by a seemingly safe and therefore attractive false openness of the digital person capable of infinite variability, but incapable of genuine dialogue, and therefore also of thinking and acting.

Accordingly, yet another sphere of life, associated not so much with personal as with public experience, is also undergoing changes. This may be the result of direct impact, in particular the activities of authoritarian and semi-authoritarian governments. Digital technologies allow us to express opinions, mobilize protests, and expand the horizons of freedom, but autocratic governments are also learning to master these technologies (Diamond, 2010). Optimistic opinions regarding the impact of digital tools have overestimated the role of access to alternative and independent sources of information and unfiltered access to the Internet, so more attention should be paid to overcoming the difficulties of online organization in the face of authoritarian governments in an increasingly digital geopolitical environment (Etling, *et al*., 2010). This may be the result of indirect exposure. The repeatedly discussed Cambridge Analytica case is just one striking example of the gradual indirect influence. Influences like this might have led to election results that were unexpected for many and to some extent jeopardized the values of democracy. Ultimately, this casts doubt on democratic procedures as the most preferred and legitimate. According to Federica Liveriero such procedures "allow solving conflicts and avoiding indeterminacy, while respecting every agent that takes part in the deliberation" (2019, p. 97). The digital environment, on the contrary, is multiplying uncertainty and disrespect.

3.2. False Openness and Illusory Control

We seem to have new patterns of interaction in the digital age. The phenomenon of false openness often arises on social media – users want to reveal a little more about themselves. What people visualise on the other side of the screen is a friendly community or group. Users tend to often perceive social media as a private diary, a safe harbor for self-expression and a public tribune at the same time.

Social media stimulate subscribers to reveal data with rewards-likes, frequent offers to share

thoughts, assurances to serve to better match our preferences. Online entities are extremely interested in using the trust resource. According to Christopher W. Savage, they "want people to trust them so they will keep coming online, sharing information, being surveilled, viewing ads, and buying things" (2019, p. 98).
The feeling that cyberspace is a free and self-regulatory space, on the one hand, and the ability to manage our devices, applications and data, on the other, give the illusion of control over what is happening. It is the lack of Arendtian public realm that we can see firsthand when dealing with cyberspace.

3.3. The Asymmetries of Power, Corporate Norms and Policies

In today's cyberspace some feel that there exists the asymmetry of power and the increasing power of corporations reflected in the asymmetry of power in democratic communities. It is believed that cyberspace was created as free, decentralized, self-regulatory and transnational (See: Gilden, 2000; Svantesson, 2007; La Chapelle & Fehlinger, 2016). Initially a free platform for opinions and information circulation, the cyberspace still allows certain people, or virtual, anonymous projects, to become opinion leaders without institutional support. But just as the asymmetry of power transforms the public space of democratic communities from a neutral to a dictate of the will of the established majority (See: Liveriero, 2019, p. 95), giant corporations reign in cyberspace and the digital world today.
Who determines which words or images will be deleted as hateful? Where to complain if your social media account has been deleted? However unfair it may seem, you will not be able to seek justice in a traditional way since your claims are not directly supported by any law, but are instead regulated by corporate standards and policies. It's peculiar how one can get a decision in their favor much faster than in court – or find oneself in the void, without tools to appeal that decision.
Thus, law based on the universal structures of our fundamental experience and guaranteeing everyone a minimum public recognition of dignity is replaced by conflict resolution mechanisms focused on corporations' interests.

3.4. The Digital Divide, Human Rights, Equality and Justice

In the digital age, people run the risk of being in a situation where they refuse others in claiming the truth, while remaining in their own filter bubbles. They gradually lose the opportunity for dialogue, since the radicalization of views leads to poorer interaction. At the same time, there is an ever-increasing transition of activity to cyberspace and the widespread use of digital tools which eventually deepens the digital divide that is inequality in access and use of digital technologies, which together with gender, economic, political and other dimensions of inequality deepens general inequality and injustice.
It entails the emergence of new types of gaps, as gender digital divide, which creates further problems. Although digital technologies can improve everyday life and be tools to achieve equality through their educational potential, they can still reinforce injustice and give false hopes, while excluding important issues from political agenda. Therefore, feminist approaches to gender digital divide suggest that we should analyze in more depth present sociotechnical transformations and emphasize that "the political activity involved in the recognition of, reflection on, and action regarding the techno-gendered codes requires a technical configuration based on openness and articulation" (Pujol & Montenegro, 2015, p. 183).
New waves or levels of digital divide are being considered today: 1) to access 2) to use and 3) to benefit from digital technologies. According to Massimo Ragnedda the third level of digital divide describes benefits from the previous two and the ability to exploit these benefits in a digital-driving market to improve one's life chances (2017, p. 5). Duncan Campbell's optimistic forecast that digital technologies can drive progress and smooth out the development levels of countries "if the digital divide is not to worsen existing patterns of inequality" (2001,

p. 136) has not come true. On the contrary, some abysses have deepened so much that it becomes difficult to believe that such dramatically different practices could even exist. Some of us seem to be so immersed in filter bubbles, that we tend to ignore significant differences and inconsistencies, which leads to an even greater isolation from other opinions and the inability to recognize the Other. This, in turn, jeopardizes the central thesis of the concept of human rights – their recognition to all humans being, regardless of differences, by virtue of recognition of human dignity.

3.5. Uncertainty and Its Legal Implications

We never know what delayed or indirect negative effects various actions in cyberspace will have. We are yet to learn how the almost invisible radicalization of opinions undermines democracy, in what way big data affect vulnerable groups and how algorithmic discrimination reinforces inequality. We strongly feel that we must act immediately since the potential legal response might cease to matter.

The potential consequences of many democratic procedures going online are: (1) the habit of joining due to the lack of interaction (we join already formed communities on social media, often clicking "I agree" without reading); (2) silencing of voices, since memorable content is valued higher than the ability to make judgments (demand for more emotional and sensational content); (3) breached trust, because we cannot be guaranteed privacy and security; (4) non-participation in the discourse for those who do not use digital tools.

4. Conclusion

Paradoxically, the development of technology, which initially forms an integral part of our mode of being-in-the-world, leads at the same time to the disappearance of the most authentic structures of fundamental experience, which make such a component of being-in-the-world as experience of law necessary and possible.

Digital technologies take this process to a new level, contributing to the replacement of the acting with simple functioning, and fundamental openness with illusory freedom in cyberspace. The latter, unlike the public realm as a realm of action and responsibility of many, and in this sense the legal realm, is more and more a realm of the productive activity of one and no longer requires law.

Thus, we are facing not only with the transformation of the structure of experience, but with the loss of experience as such, and the unprecedented crisis of law in today's world is perhaps the most alarming symptom of this process.

REFERENCES

Alekseev, N. (1999). *Osnovyi filosofii prava* [Foundations of the philosophy of law]. Saint-Petersburg: Lan;

Arendt, H. (2003). Personal responsibility under dictatorship. In *Responsibility and judgment* (pp. 17-48). New York: Schoken Books;

Arendt, H. (1998). *The human condition* (2nd ed.). Chicago & London: The University of Chicago Press;

Arendt, H. (1978). *The life of the mind*. San Diego, New York, London: Harcourt, Inc.;

Arendt, H. (1971). Thinking and moral considerations. *Social Research*, 38:3, 417-446. Retrieved from https://jonudell.net/h/arendt.pdf;

Arendt, H. (1994). Understanding and Politics (The Difficulties of Understanding). In *Hannah Arendt: Essays in Understanding, 1930–1954* (pp. 307-327). New York: Harcourt, Brace;

Aristotle (2009). *The Nicomachean ethics* (D. Ross, Trans.). New York: Oxford University Press;

Campbell, D. (2001). Can the digital divide be contained? *International Labour Review*, 140(2), 119-141;

Diamond, L. (2010). Liberation technology. *Journal of Democracy*, 21(3), 69-83. doi:10.1353/jod.0.0190;

Etling, B., Faris, R., & Palfrey, J. (2010). Political change in the digital age: The fragility and promise of online organizing. *SAIS Review of International Affairs*, 30(2), 37-49. Retrieved from https://www.muse.jhu.edu/article/403437;
Gadamer, H.-G. (2004). *Truth and method* (2nd ed.) (J. Weinsheimer, & D. G. Marshall, Trans.). London & New York: Continuum;
Gilden, M. (2000). Jurisdiction and the Internet: the "real world" meets cyberspace. *ILSA Journal of International & Comparative Law*, 7, 149-160;
Habermas, J. (2003). *The future of human nature* (H. Beister, M. Pensky, & W. Rehg, Trans.). Cambridge: Polity Press;
Heidegger, M. (1996). *Being and time* (J. Stambaugh, Trans.). New York: State University of New York Press;
Heidegger, M. (2001). *Phenomenological interpretations of Aristotle (Initiation into phenomenological research)* (R. Rojcewicz, Trans.). Bloomington: Indiana University Press;
Hénaff, M. (2019). *The philosophers' gift: Reexamining reciprocity* (J.-L. Morhange, Trans.). New York: Fordham University Press;
Husserl, G. (1955). *Recht und Zeit*. Frankfurt/Main: Vittorio Klostermann;
La Chapelle, B., & Fehlinger, P. (2016). Jurisdiction on the Internet: How to move beyond the legal arms race. Observer Research Foundation and Global Policy Journal series. *Digital Debates. CyFy Journal*, 3, 8-14;
Liveriero, F. (2019). The social bases of self-respect. Political equality and epistemic injustice. *Phenomenology and Mind*, (16), 90-101. Retrieved from https://doi.org/10.13128/Phe_Mi-26076;
Pujol, J., & Montenegro, M. (2015). Technology and feminism: A strange couple. *Revista de Estudios Sociales*, (51), 173-185. Retrieved from https://doi.org/10.7440/res51.2015.13;
Ragnedda, M. (2017). *The third digital divide: A Weberian approach to digital inequalities*. London; New York: Routledge, Taylor & Francis Group;
Reinach, A. (1953). *Zur Phänomenologie des Rechts: die apriorischen Grundlagen der bürgerlichen Rechts*. München: Kösel-Verlag;
Ricoeur, P. (2000). *The just* (D. Pellauer, Trans.). Chicago: The University of Chicago Press;
Ricoeur, P. (1996). Torzhestvo yazy`ka nad nasiliem. Germenevticheskij podkhod k filosofii prava [The victory language gains over violence. The hermeneutic approach to the philosophy of law]. *Voprosy`filosofii*, 4, 27-36;
Pujol, J., & Montenegro, M. (2015). Technology and feminism: A strange couple. *Revista de Estudios Sociales*, (51), 173-185. https://doi.org/10.7440/res51.2015.13;
Savage, C. W. (2019). Managing the ambient trust commons: The economics of online consumer information privacy, *Stanford Technology Law Review*. 22 (1), 95-162;
Svantesson, D. J. B. (2007). *Private international law and the Internet*. Alphen aan den Rijn: Kluwer Law International.

SYMPOSIUM

SYMPOSIUM

PENSARE L'INFOSFERA. LA FILOSOFIA COME DESIGN CONCETTUALE

Carlo Crosato
Introduzione: se la rivoluzione informazionale si rivela troppo conservativa

Maurizio Ferraris
Web tolemaico e web copernicano

Leonardo Manna
Il pensiero come relazione o intero semantico? Intorno alla filosofia di Luciano Floridi

Roberto Mordacci
Sapere, progetto, azione. Sul primato del pratico

Luigi Vero Tarca
Infosfera, apertura e verità

CARLO CROSATO
Cultore della materia presso l'Università degli studi di Bergamo
car.crosato@gmail.com

INTRODUZIONE: SE LA RIVOLUZIONE INFORMAZIONALE SI RIVELA TROPPO CONSERVATIVA

abstract

Luciano Floridi proposes to reconceive philosophy as a constructive practice committed to facing the issues of the present and to dismantle and assemble the categories of contemporary thought. Introducing this symposium, I aim to explore what I consider to be the inherent risks of a philosophy conceived in this manner. To put it briefly, such a philosophy appears to be both consumed in the plots of contemporary times and over-committed towards formulating evermore effective and useful answers, each time aiming to somehow transcend itself and to critically portray the specific context in which it arises. Can informational philosophy criticize the present epoch and open new ways of emancipation, or is it too functionally dependent on the actual processes it examines hindering a critical point of view?

keywords

infosphere, information and communications technology, present, actuality

Phenomenology and Mind, n. 20 - 2021, pp. 140-145
DOI: https://doi.org/10.17454/pam-2012
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1 L'interrogazione intorno a cosa sia la filosofia non è certo nuova e, anzi, ricorre frequente in molti pensatori: ci si chiede quale sia il suo statuto e che cosa significhi praticarla, si pone la questione della natura troppo accademica di alcune questioni o dell'opposto eccessivo annacquamento divulgativo, ci si interroga sulla funzione degli intellettuali, a volte chiacchieroni altre volte rinserrati nella loro torre d'avorio. Le recenti riflessioni di Luciano Floridi sulla "quarta rivoluzione" e sull'"infosfera" offrono una nuova occasione di sondare la natura della filosofia, impegnando con considerevole potenziale provocatorio in merito a questioni attualissime, come le tecnologie computazionali e informazionali e le numerose implicazioni morali, politiche, legali che esse hanno importato nella contemporaneità. Cogliendo questa occasione, il presente simposio intende avviare un confronto schietto, che sfrutti come punto di partenza alcune intuizioni di Floridi, per criticarle o evidenziarne gli elementi progressivi e, soprattutto, per giungere a sviluppare riflessioni autonome.
Il libro *Pensare l'infosfera. La filosofia come design concettuale* (2020) è la traduzione di parte di un'opera più voluminosa, pubblicata in inglese nel 2019 e pensata come il terzo volume di una tetralogia sui fondamenti della filosofia dell'informazione. La sua pubblicazione in Italia segue la traduzione, sempre a cura di Raffaello Cortina, del volume *La quarta rivoluzione*. Copernico ha tolto la Terra dal centro dell'universo; Darwin ha rivelato le umili origini di *homo sapiens* privandolo della sua posizione privilegiata nel mondo dei viventi; Freud e, più di recente, le neuroscienze hanno messo in crisi la presunta sovranità del soggetto sulla propria vita interiore. Oggi le tecnologie dell'informazione e della comunicazione hanno investito in poco tempo ogni dimensione della nostra vita, dalla quotidianità alle istituzioni, su scala microscopica o su scala macroscopica, colmando ogni possibile anfratto in maniera così efficiente che ormai ci siamo abituati a considerarle come i normali strumenti di interazione con il mondo e con gli altri; ma proprio questa abitudine e la funzionalità con cui mediano il nostro mondo relazionale ci rende tali tecnologie difficilmente apprezzabili nella loro essenza di vere e proprie forze ambientali, antropologiche, sociali e interpretative. Esse rappresentano un fattore di mutamento nelle prerogative dell'uomo sulle realtà naturali e artificiali che lo circondano, impossibile da slegare dagli effetti inerenti al rapporto dell'uomo con se stesso, con la comprensione che ha di sé e della propria storia. Uno snodo rivoluzionario, che conforma le relazioni umane su un registro non più – o non solo – politico o economico, ma anche e prima di tutto informazionale e comunicativo; e questo orizzonte necessita di una filosofia all'altezza per poter essere compreso e abitato, se non si vuole piombare in una inversione di ruolo fra servo e padrone in senso hegeliano.

Non siamo enti isolati e unici, quanto piuttosto degli organismi, il cui sostrato è informazionale (*inforgs*), reciprocamente connessi e parte di un ambiente costituito da informazioni (*infosfera*), che condividiamo con agenti naturali o artificiali simili a noi sotto più profili (Floridi, 2017, p. 106). La quarta rivoluzione, sulla cui soglia ci troviamo, contribuisce a sua volta a togliere all'uomo la presunta centralità che un ingenuo umanismo gli vorrebbe spacciare; ma questa perdita di unicità e di centralità è, insieme, un impoverimento e un arricchimento, costringendoci ad adottare una prospettiva consapevole, attiva, sollecita nella relazione con l'altro. Ci troviamo in un momento di fondamentale mutamento storico, afferma Floridi, e la filosofia deve sincronizzarsi con gli eventi e forgiare gli strumenti semantici adeguati a descriverli con consapevolezza. *Pensare l'infosfera*, prendendo molto sul serio questi elementi trasformativi, ambisce a ridestare l'attività filosofica, i suoi strumenti e i suoi obiettivi, gettando le basi per un rinnovamento intellettuale dell'etica, dell'estetica, dell'epistemologia.

La filosofia, paragonata a un computer piantato, secondo Floridi va riavviata e rimessa al lavoro su simili questioni, se si vuole che essa conservi un senso e un ruolo. E un senso e un ruolo irriducibili la filosofia ce li ha da sempre, con la sua capacità di relazionarsi con l'attuale in maniera problematica, o, come dice Floridi, ponendo domande aperte, impossibili da affrontare in chiave esclusivamente empirica o logico-matematica. La filosofia si è sempre proposta e dovrà continuare a proporsi come una postura problematica: anche a seguito di osservazioni o calcoli, anche dopo che ogni formulazione sarà stata ben strutturata, e dopo che ogni tema e concetto saranno stati chiariti, gli interlocutori che praticano la filosofia dovranno accettare la possibilità di ritrovarsi in una condizione di perdurante disaccordo. Non un ottuso muro contro muro, bensì un disaccordo informato, razionale e onesto, ma potenzialmente non ricucibile, se non in quel dato minimale consistente nella necessità di continuare a sostare in quell'apertura discorde. E tuttavia, perché il domandare filosofico non si riduca a mero manierismo fra filosofi, pur concentrandosi su questioni fondamentali esso non potrà pretendere di abbracciare il reale in senso assoluto: il debito che, secondo Floridi, la filosofia contemporanea e del futuro dovrà riconoscere a Alan Turing deriva soprattutto dall'aver compreso l'urgenza di attribuire legittimità al disaccordo – altrimenti derivante dall'incomprensione reciproca e fonte di confusione – attraverso l'identificazione e il chiarimento di livelli di astrazione, capaci di offrire il corretto punto di vista a partire dal quale affrontare la domanda e attendersi risposte ragionevoli e produttive. **2**

L'idea di Floridi è che, assunti questi presupposti tutt'altro che semplicemente metodologici, la filosofia possa ripensarsi come un'impresa costruttiva, coordinata al presente in cui viene praticata grazie all'inquadramento delle proprie domande entro livelli astrattivi concordati. Non si tratta semplicemente di un appello ai filosofi a una maggiore responsabilità nei confronti della loro contemporaneità; passando attraverso lo stesso contesto informazionale e tecnologico che è chiamata a indagare, la filosofia riscopre se stessa come una pratica relazionale e come studio della relazione, in cui metodo formale e oggetto indagato coincidono in un curioso isomorfismo. **3**

La rivoluzione dell'infosfera è un'occasione: essa contribuisce al decentramento dell'uomo e dell'*ego*, collocandoci costantemente in una periferia da cui poter meglio osservare la galassia relazionale che ci coinvolge; ma il digitale è, secondo Floridi, anche un'occasione per superare alcune narrazioni politiche legate a ideologie o vecchie strutture, per individuare strumenti economici ed ecologici adeguati; il digitale aiuta l'analogico non rimpiazzandolo, ma permettendo di ottimizzare le risorse; perfino in campo etico, l'allargamento degli strumenti dovuto al digitale permette di concentrare l'attenzione sul destinatario, adeguando a esso l'azione, la cura, la parola. D'altra parte, la rivoluzione dell'infosfera implica anche una sfida,

consistente nell'abbandono di ogni velleità metafisica assolutistica, nell'assunzione delle più promettenti novità metodologiche e tecniche proposte dal digitale, e nella definizione di soluzioni efficaci nella complessità caratteristica dell'apertura filosofica.
"Costruttivismo" e "*design* concettuale" sono le parole che Floridi usa per descrivere questo rapporto, al contempo di avvicinamento e di gestione, tra la filosofia a venire e le tecnologie informazionali e comunicative. Per agganciare questa novità, Floridi propone di risalire al punto in cui, con Platone e la sua separazione del sapere e della tecnica, il pensiero occidentale avrebbe preso a interpretare la propria conoscenza del mondo privilegiando il punto di vista dell'utente, di chi usa l'oggetto, di chi lo scopre e ne disvela la verità intrinseca, relegando in secondo piano la conoscenza di chi quell'oggetto, quella verità, quel concetto ha forgiato. Invertendo la tendenza, Floridi pone l'accento sul valore irriducibile del sapere come costruzione, e non come semplice fruizione mimetica: la conoscenza, lungi dall'essere adeguamento del soggetto all'oggetto, va ripensata come vera e propria tecnica della formulazione di domande e della ricerca di risposte efficaci. Qui sta la pretesa, secondo Floridi assente in gran parte della filosofia da Platone in poi, di vivere attivamente la filosofia, come una vera ingegneria concettuale, che monta e smonta i problemi, sempre con l'urgenza di collocarsi dentro un contesto reale.

4 Ciò che sembra legittimo chiedersi è se una filosofia come quella suggerita da Floridi, così presa dalla propria operatività, dall'efficacia e dall'utilità delle risposte che è chiamata a dare, sia in grado di dar conto di se stessa e del contesto in cui opera. L'isomorfismo tra l'ambiente informativo e la prassi filosofica che Floridi consiglia è senza dubbio il motivo per cui si può ben sperare che la filosofia dell'informazione saprà aderire all'impresa digitale e consigliare il migliore uso degli strumenti in vista dei fini da raggiungere; ma tale isomorfismo rischia di trasformarsi nell'incapacità di mantenere aperte domande non immediatamente rilevanti, e di trascendere in maniera davvero filosofica il "tutto pieno", la "gabbia d'acciaio" che la quarta rivoluzione edifica attorno a noi. Il rischio è quello di trasformare la filosofia in consulenza tecnico-ingegneristica, in predisposizione di mezzi in vista di fini la cui elezione e i cui significati rimangono difficili da problematizzare. Insomma, se il metodo filosofico, più che trascendere, è in funzione di una saggia transizione avviata già altrove, se l'armamentario che la filosofia utilizza è mutuato dall'oggetto che essa è chiamata a pensare, se lo stesso lessico – *design*, capitale semantico coerente, efficienza, prestazione – è affratellato alle dinamiche contemporanee, il pericolo che la filosofia divenga conferma di un esistente appena ritoccato rimane un problema che la filosofia di Floridi dovrà dimostrare di saper affrontare.
In un passaggio del suo libro, sostenendo la sua prospettiva del costruttore, Floridi parafrasa Austin, scrivendo che «facciamo cose con le informazioni»; e proprio perché *sappiamo come* fare possiamo ambire a *sapere che*. A proposito, tornano alla mente i lavori di Michel Foucault sulla *parrhesia*, la pratica del parlare francamente, attività di vera critica filosofica e rottura con il presente: Foucault presenta la *parrhesia* come un'attitudine speculare all'atto discorsivo di matrice austiniana, essendo quest'ultimo conservazione delle funzioni che a ciascuno vengono imposte nelle varie situazioni, laddove il parresiasta, specie il cinico, sa tagliare il presente in maniera obliqua trascendendo le condizioni fattuali e materiali, e sa perturbare tali condizioni irrompendovi con un atto di incoerenza, mostrando così la contingenza degli armamentari semantici in uso e la fragilità delle condizioni pragmatiche in funzione (Foucault 2015, pp. 66 ss).
Fra le molte trattazioni su che cosa sia la filosofia di cui si è detto all'inizio, si può scegliere di rileggere quella di Deleuze e Guattari, in cui, proprio in relazione a Foucault, viene proposta la distinzione tra presente e attuale: «L'attuale non è ciò che noi siamo, ma piuttosto ciò che diventiamo, ciò che stiamo divenendo, ossia l'Altro, il nostro divenir-altro. Il presente, al

contrario, è ciò che siamo e proprio per questo, ciò che già non siamo più» (Deleuze-Guattari, 1996, p. 106). La filosofia di Floridi, così avvertita di quanto le avviene attorno, sa calarsi nel presente e ancorarsi nel reale, immedesimarsi con esso e, appunto in termini immanenti, riordinarne gli elementi. Saprà essa rivelarsi anche "attuale", e perciò intempestiva non solo rispetto alla vecchia filosofia ma anche alle nuove dinamiche tecniche? Secondo un adagio aristotelico ormai frusto, la filosofia non serve perché non è serva: saprà la filosofia dell'informazione contraddire il presente e aprire spazi di possibilità e critica, o si rivelerà uno strumento troppo coinvolto nelle dinamiche presenti per poterle riconsiderare dalla giusta distanza?

5

Nel *simposio* che qui si viene presentando, queste e altre osservazioni critiche vengono sollevate nei confronti della proposta di Luciano Floridi: sono intuizioni importanti, stimoli di riflessione che, pur nello spazio di una conversazione colta, possono offrire alla filosofia un'occasione di progresso prima di tutto metodologico.
Maurizio Ferraris si occupa delle implicazioni teoriche e pratiche del *web* nella nostra epoca, interrogando l'approccio di Floridi nella sua ambizione di sostenere e meglio funzionalizzare quella che egli chiama "Quarta rivoluzione". Con l'espressione "*web* copernicano", Ferraris si distingue dal metodo di Floridi, sottolineando l'ampiezza irriflessa del *web* di cui l'Infosfera rischia di cogliere solo la parte esplicita e consapevole. Ferraris pone l'accento sulla natura di registrazione e non solo di comunicazione del *web*, sulla sua performatività mobilitante irriducibile alla sola informazione, sulla sua materialità all'interno della realtà sociale e sull'urgenza di penetrarlo con un'operazione di trasparenza. Di qui, la necessità politica di un "*Webfare*", una sorta di *Welfare* riattualizzato nell'ambiente del *web*.
Leonardo Manna interroga Floridi in merito alla possibilità di superare i livelli di astrazione della sua filosofia, per avvicinare domande ultime che, altrimenti, rimarrebbero inaccessibili da una prospettiva destinata al relativismo costruzionista.
Roberto Mordacci si concentra sull'urgenza di conservare il primato della dimensione pratica della filosofia, ciò che invece la filosofia del *design* tende a trascurare. Floridi, argomenta Mordacci, non vuol disfarsi della filosofia pratica, ma, anziché considerarla come il quadro critico capace di offrire un fine e un movente alla meditazione filosofica, la colloca in subordine rispetto alla preminenza del *design* concettuale. Un'inversione nuova e originale, che tuttavia rischia, da un lato, di perdere la presa sul vivere da cui solo possono sgorgare i motivi della filosofia, e, dall'altro, di privarsi della sufficiente mobilità critica rispetto agli esiti tecnici che pure il *design* concorre a raggiungere.
Luigi Vero Tarca invita al simposio tre personaggi fittizi, che dialogano in merito alla logica profonda delle domande filosofiche proposte e in particolare al concetto di "chiusura"; in tal modo vengono sollevate alcune questioni di fondo relative alle conseguenze che l'avvento dell'infosfera ha sulla natura dei nuovi soggetti storici che si impongono nello scenario dominato dalla tecnologia. Proprio attraverso le riflessioni relative alla coppia concettuale chiusura/apertura si arriva pure a intuire il motivo del particolare stile compositivo adottato, dovuto appunto all'intenzione di avviare una modalità di dialogo nella quale l'apertura giunga a mettere in questione persino la nozione di negazione, che di solito pare invece definire e quindi *rinchiudere* a priori il senso di tutti i discorsi filosofici.

Post scriptum

L'idea originaria era che la pubblicazione comprendesse anche una risposta dell'autore (Luciano Floridi) del libro (*Pensare l'infosfera*) al quale sono rivolti gli scritti qui presentati. Purtroppo egli non si è trovato nelle condizioni di replicare ai vari interventi; sicché non è stato possibile portare a termine l'impresa nella modalità inizialmente prevista. Considerata però la qualità dei discorsi qui proposti, che riguardano temi la cui importanza va ben al di là

di qualsiasi motivazione occasionale, mi è parso non solo giusto ma anche opportuno proporli comunque per la pubblicazione per offrire un'occasione di dialogo tra gli stessi autori e tra costoro e il pubblico. È chiaro che i testi palesano, soprattutto in qualche caso e in alcuni passaggi, la peculiarità del fine per il quale erano stati originariamente pensati, ma, data la natura generale dei temi proposti, mi pare che essi conservino intatto il loro valore filosofico. Naturalmente lo stesso Floridi potrà in seguito intervenire a sua volta se ne avrà l'interesse e l'occasione.

RIFERIMENTI

Deleuze, G., & Guattari, F. (1996). *Qu'est-ce que la philosophie?* (1991), tr. it. a cura di A. De Lorenzis, *Che cos'è la filosofia?*, Torino: Einaudi;
Floridi, L. (2017). *The Forth Revolution. How the Infosphere is Reshaping Human Reality* (2014), tr. it. a cura di M. Durante, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano: Raffaello Cortina;
Floridi, L. (2020). *The Logic of Information. A Theory of Philosophy as Conceptual Design* (2019), tradotto parzialmente da M. Durante, *Pensare l'infosfera. La filosofia come design concettuale*, Milano: Raffaello Cortina;
Foucault, M. (2015). *Le gouvernement de soi et des autres. Cours au Collège de France (1982-1983)* (2008), tr. it. a cura di M. Galzigna, *Il governo di sé e degli altri. Corso al Collège de France (1982-1983)*, Milano: Feltrinelli.

MAURIZIO FERRARIS
Università degli Studi di Torino
maurizio.ferraris@unito.it

WEB TOLEMAICO E WEB COPERNICANO

abstract

There are two opposite ways of interpreting the web: one Ptolemaic and one Copernican. The first analyzes the web starting from the concept of information, as in the case of Luciano Floridi. The latter recognizes the importance of information, and as a consequence Floridi's work, but it takes a step forward. Indeed, it highlights the importance of documents and the relationship between information and human life. In this way, it makes it possible to comprehend the human way of being and the revolution we are witnessing.

keywords

web, documents, infosphere, digital revolution, digital welfare

Phenomenology and Mind, n. 20 - 2021, pp. 146-162
DOI: https://doi.org/10.17454/pam-2013
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. La quarta di che?

Sono sin troppo ovvi i motivi che mi spingono ad ammirare il lavoro di Floridi: anzitutto la sua capacità di avere imposto all'attenzione del dibattito filosofico internazionale il problema del *web*. Così come sono completamente d'accordo sul fatto che l'idea di una filosofia chiamata non a risolvere i problemi, ma a porli, è futile e stucchevole.

Davvero, i momenti alti della filosofia sono stati quelli in cui i filosofi non interrogavano l'essere o il linguaggio, ma intervenivano nel mondo risolvendo problemi sollevati non da loro colleghi in tanto tempo libero, ma dall'umanità, sotto l'urgenza di tempi stretti e difficili. È così che ai tempi di Marx, di Leibniz, o di Platone, nessuno si sarebbe chiesto se la filosofia servisse a qualcosa o se fosse viva o morta, mentre nel secolo scorso era impossibile leggere un testo con un minimo di ambizioni speculative che non si aprisse con la dichiarazione di morte della filosofia. E ho detto tutto. La rivoluzione in corso che sintetizziamo con il nome di "*Web*" è una occasione unica per ridare un senso e una utilità alla filosofia, non ce la dobbiamo lasciare sfuggire, e se questo tentativo potrà andare a buon fine, come credo fortemente, ciò dipende in un modo significativo dall'azione di Luciano.

Sono anche d'accordo sul fatto che la filosofia deve essere concepita come una costruzione di concetti, e non come una futile analisi o una invenzione di parole d'ordine parapolitiche. Negli anni Sessanta del secolo scorso, Bertrand Russell descrisse la filosofia analitica dei suoi tempi, ridottasi a scolastica e manierismo, raccontando di quando si era perso in un giro in bicicletta e chiese a un passante se poteva indicargli la via per Oxford. L'uomo gli chiese "Intende la via?" "Sì"; "Per Oxford?" "Sì". Al che l'uomo gli rispose "No". E tutto sommato a Russell era andata bene, perché, se invece che un analitico avesse incontrato un continentale, questi lo avrebbe intrattenuto con una lunga concione sull'impossibilità di trovare la via giusta e sulla necessità di perdersi, con testi all'appoggio sull'abilità necessaria per perdersi secondo Benjamin, e sull'elogio heideggeriano dei sentieri interrotti.

Le mie differenze incominciano quando si tratta di definire questa costruzione di concetti. Non sono troppo convinto sulle virtù del *design* concettuale: sebbene oggi sia molto di moda (il che non è un male), mi ricorda troppo la dialettica hegeliana, o peggio ancora (visto che Hegel era comunque un grande) il costruzionismo postmoderno, che tratta il mondo come una pasta per biscotti cui i concetti danno la forma desiderata dal filosofo (Putnam 1991). Passando dalla pasticceria alla macelleria, preferirei seguire la via della dialettica platonica, quella che disseziona il reale seguendone le linee naturali[1] che – aggiungo io – anche nel caso di artefatti

1 Platone, *Fedro* 265e.

tecnologici come il *web* sono molto meno il frutto di una costruzione deliberata che non di una emergenza, frutto di tante circostanze in cui intervengono il *medium*, il mondo e i bisogni degli umani. Il che sembra più facile e meno creativo, ma impone l'obbligo di capire che cosa sia il reale di cui si sta trattando, il che non sempre è facile e non sempre è ovvio. Beninteso, uno si può sempre sbagliare, sia nell'identificare il reale sia nel sezionarlo, ma questo è un rischio che impende su ogni attività umana, ed è con questo spirito che vorrei proporre una idea alternativa al rilancio della nozione di "Infosfera" su cui Luciano ha costruito la sua interpretazione del *web*.

Per essere chiari sin dall'inizio. Il *web* è *anche* una Infosfera, ma questa non ne è che la minima parte; l'Infosfera poggia su una *Docusfera* (ossia su documenti che registrano le azioni umane senza necessariamente portare informazioni) e quest'ultima poggia su una *Biosfera*, ossia appunto sul mondo della vita che, questa l'autentica rivoluzione che il *web* ha portato nel mondo, oggi è in linea di principio sempre registrabile, dunque documentabile. Il che però non significa che sia di per sé informazione, perché in quel caso confonderemmo il *web* con Wikipedia.

Luciano ci parla di una "rivoluzione copernicana" apportata dal *web*, e sono perfettamente d'accordo con lui. Solo che ho l'impressione che la sua rivoluzione copernicana sia solo apparente, così come lo era quella di Kant, che a tutti gli effetti era una controrivoluzione tolemaica, che riportava l'uomo (in veste di Io penso) al centro dell'universo. Altro che Copernico. Lo stesso vale per il rilancio della nozione di "Infosfera"[2] che sta al centro delle analisi del *web* offerte da Luciano. Si badi bene: io non nego che nel *web* ci sia anche una Infosfera. Ma, ripeto, questa non è la stoffa fondamentale di cui è intessuto, né il valore principale di cui è portatore. Il seguito del mio discorso consisterà dunque nella contrapposizione tra un *web* tolemaico (quello di cui parla Luciano) e un *web* copernicano (quello di cui mi piacerebbe parlare con lui).

Che il *web* costituisca una *rivoluzione* è una delle poche cose su cui specialisti e profani si trovano completamente d'accordo. Il disaccordo, però, incomincia subito dopo, quando si tratta di definire la natura di questa rivoluzione. Non tanto per ciò che riguarda l'estensione della trasformazione in corso, che (su questo siamo tutti d'accordo) non è solo tecnologica, ma politica, economica e sociale (Bunz, 2012; Lanier, 2013). E nemmeno per la collocazione nella serie delle rivoluzioni, visto che un largo e miracoloso consenso vuole che sia la *quarta*; il punto è però che non c'è accordo su quali siano le tre che la precedono, e questo ovviamente non è un problema da poco, che va risolto preliminarmente per capire di cosa stiamo parlando.

Per taluni, è la quarta rivoluzione industriale (Zuazua, 2019). In questo quadro la prima rivoluzione sarebbe basata sull'acqua, la seconda sull'elettricità, la terza sulla tecnologia dell'informazione e l'elettronica, e l'attuale su una fusione di tecnologie che fa saltare la differenza tra fisico, digitale e biologico. Il problema, però, è che questa analisi presenta un doppio limite. Il primo consiste nel considerare la trasformazione in atto come una variante della rivoluzione industriale, mentre ne scuote le fondamenta, e in particolare i rapporti tra produzione e consumo. Il secondo è che circoscrive l'evento con cui abbiamo a che fare a una cronologia esigua, mentre ha potenzialmente la stessa portata del passaggio dell'umanità dallo stato di cacciatori e raccoglitori a quella di allevatori e agricoltori.

Per altri, e Luciano rientra in questo novero, è una rivoluzione radicale, che riguarda anzitutto il modo di pensare e di vedere il mondo, ed è paragonabile a quelle promosse da Copernico, Darwin, e Freud (Floridi, 2017), e consisterebbe in una ferita narcisistica: Copernico aveva

2 Il cui conio, come è noto, si deve a Toffler (1987) e che si ritrova, nove anni dopo, nel romanzo di fantascienza *Hyperion* di Dan Simmons.

tolto la Terra dal centro dell'universo, Darwin aveva dimostrato la nostra discendenza dalle scimmie, Freud aveva rivelato che la coscienza non è che la punta emersa di un continente inconscio, e Turing ci ha mostrato che delle macchine possono pensare molto meglio di noi. Questa genealogia mi sembra fabbricata da un araldico confuso e ambizioso, per due motivi. Primo, deriva da una affermazione, quella sì narcisistica, di Freud (Freud, 1989, pp. 657-664), che non solo aveva considerato l'inconscio una scoperta, e per di più sua, il che è ovviamente falso; ma soprattutto l'aveva presentata come una ferita narcisistica, laddove, come ognuno ha potuto constatare, si è trattato esattamente del contrario: un narcisismo in libera uscita, un egotismo senza limiti, raccontare i propri sogni a destra e a manca, considerare rilevanti e piene di significato le proprie *gaffes*, *lapsus*, azioni ed omissioni.
Questo però è un problema di Freud e della sua terza rivoluzione. Il fatto è invece che, quanto alla quarta rivoluzione, quella di Turing, è semplicemente falso che si sia dovuto aspettare così tanti millenni per capire che le macchine possono pensare meglio degli umani. Il nostro remoto antenato che aveva iniziato a intagliare un osso tenendo traccia delle fasi lunari aveva compreso, ben prima di Turing, che la mente umana ha prestazioni molto più deboli e inaffidabili di una memoria esterna. Ed è a queste medesime considerazioni che erano giunti coloro che, con un processo collettivo difficile da situare e con datazioni e ubicazioni differenti hanno dato vita alla scrittura, ai dispositivi di calcolo, ai calendari. L'abaco non è certo uno strumento sofisticato, eppure permette calcoli aritmetici molto superiori alla portata di una normale mente umana. E lo stesso si può dire di carta e penna: Eulero soleva dire che tutta la sua matematica si concentrava nella matita con cui calcolava. A chi poi obiettasse che se si dà una matita a una scimmia questa non ne cava granché, sarebbe facile rispondere che lo stesso avverrebbe se alla scimmia si desse una macchina di Turing. E qui si viene a un secondo problema. Le macchine non solo pensano meglio dell'umanità da molto tempo prima di Turing, ma in compenso, in assenza di umanità, sia un abaco sia una macchina di Turing non pensano. Anche qui, non si vede la ferita narcisistica, anzi, è piuttosto l'inverso.
Tra il minimalismo della rivoluzione industriale e il massimalismo della ferita narcisistica credo che la più corretta genealogia della quarta rivoluzione sia quella che la concepisce non come un rivolgimento del modo di pensare, bensì di annotare e trasmettere i pensieri[3]. Invece che un cambiamento ideologico delle visioni del mondo, abbiamo una quarta trasformazione tecnica dei supporti della registrazione e della comunicazione, dopo il passaggio dall'oralità alla scrittura, quello dai rotoli ai libri, e l'invenzione dei caratteri mobili. In prima approssimazione, è proprio di qui che conviene prendere l'avvio, ossia dalla trasformazione delle tecnologie della registrazione, senza restringere la scrittura alla semplice sfera dell'alfabeto[4]. Questa opzione ha, ai miei occhi, il vantaggio di indicare in un ambito specifico di tecnologie la quarta rivoluzione di cui ci stiamo occupando, che viene collocata in una successione empirica circoscritta e verificabile. Al tempo stesso, però, tocca un punto sensibile tanto dal punto di vista empirico quanto da quello trascendentale, ossia precisamente le tecniche della registrazione, che non sono tecniche fra le altre, bensì il fondamento fisico e metafisico della realtà in cui viviamo[5].
Ancora un punto. Se la rivoluzione fosse consistita davvero nell'Infosfera, saremmo nei guai, perché non sarebbe una rivoluzione. Per caratterizzare la rivoluzione in corso, Luciano sostiene che consisterebbe nel fatto che abbiamo scoperto – e poco alla volta accettato – di

3 D'accordo con la visione di Roncaglia (2010).
4 Incorrendo in un errore comune, una delle cui manifestazioni più pregevoli e recenti si trova in Havelock (2005).
5 Argomenterò estesamente questo punto in *Documanità. La filosofia del mondo nuovo*, Laterza, Roma-Bari 2021.

essere animali sociali e dotati di linguaggio[6]. Bisognava aspettare Turing? A me pare che già Aristotele (seguito da una pletora di altri, filosofi e non filosofi) lo avesse detto, per giunta richiamandosi a Omero[7]. Direi che, semmai, il presente conferma, con una evidenza senza precedenti (e questo indubbiamente è un vantaggio), ciò che sapevamo da tempo, un po' come un medico viennese ha ritrovato nei comportamenti di certi suoi compassati pazienti fenomeni tragici e deplorevoli già ben noti ai Greci antichi. Quel che vale per Edipo, vale per il *web*: i filosofi, antichi e moderni, insegnano che l'uomo è un animale dotato di linguaggio e che è un animale sociale, e i telefonini e i *social* lo dimostrano come meglio non si potrebbe. Ma la predisposizione era in atto sin da principio, la tecnica si è limitata a rivelarla, non ci ha portato lontani da noi, ma ci ha fatti diventare quelli che siamo.

2. *Web* tolemaico

L'Infosfera, muovendo da un presupposto tolemaico, sostiene che ciò che abbiamo intorno a noi, grazie al *web*, sono informazioni, cioè pensieri. Ma sono *davveroe anzitutto* pensieri e informazioni quelli che ci scambiamo? E soprattutto, la rivoluzione è consistita nell'accrescere la registrazione di pensieri e di informazioni, o non ha rappresentato piuttosto una moltiplicazione del numero di atti e di comportamenti (significativi o meno, informativi o meno, coscienti o meno) che possono venire registrati? Se ammettiamo, come sembra ragionevole, che un pensiero è tale in senso filosofico e non psicologico solo se è vero (Frege, 1918, pp. 58-77), allora anche una informazione è una informazione solo se è vera. Se una informazione è falsa non è una informazione più di quanto un diamante falso sia un diamante. Ora, il parente più prossimo dell'Infosfera non è l'Intelletto Unico degli averroisti (che conteneva solo nozioni vere), ma la Biblioteca di Babele: «È ormai risaputo: per una riga ragionevole, per una notizia corretta, vi sono leghe di insensate cacofonie, di farragini verbali e di incoerenze» (Borges, 1961).
È chiaro che nella teoria della informazione di Shannon, "informazione" è solo contare il numero di stati possibili di qualcosa, cioè si avvicina molto alla registrazione (contare presuppone che si registri), ma purtroppo si crea una confusione tra l'informazione nel senso ordinario del termine e in quello tecnico-informatico, e sicuramente i teorici cui Floridi si riferisce, così come i suoi lettori, pensano prima di tutto al termine nel suo senso ordinario. Immagino che la scelta di Luciano dipenda dalla necessità di trovare un termine forte e

6 «Nel presente, stiamo lentamente accettando l'idea che non siamo enti isolati e unici, quanto piuttosto organismi, il cui sostrato è informazionale (*inforgs*), reciprocamente connessi e parte di un ambiente costituito da informazioni (*infosfera*), che condividiamo con agenti naturali o artificiali simili a noi sotto più profili. Turing ha cambiato la nostra antropologia filosofica tanto quanto Cartesio, Darwin o Freud» (Floridi, 2020, p. 130).
7 «L'uomo è l'animale politico per natura [...] colui che è senza città per via della <sua> natura e non per qualche accidente è un meschino oppure è superiore all'uomo; egli è come colui che è stato ingiuriato da Omero <con le seguenti parole>: "senza famiglia, senza legge, senza focolare". Chi è di tal natura è immediatamente desideroso di guerra, essendo privo di legame alla maniera di <un pezzo isolato> nel gioco della *petteia*. Pertanto l'uomo è un animale politico ben altrimenti che ogni ape e ogni animale gregario» (Aristotele, *Politica*, 1253a, 2-8).

chiaro per tutti, ma questa chiarezza ha un costo molto elevato. Perché un conto è parlare di informazioni disponibili a tutti, un altro dire che il libro del *web* è scritto in caratteri accessibili solo a pochi umani aiutati da automi potentissimi. In questo senso, le informazioni ci sono anche nella docusfera: qualsiasi comportamento futile o insignificante io abbia *online* (aprire video a caso per noia, scrivere cose insensate su Facebook, o mettere "mi piace" di qui e di là) è un'informazione per quei pochi che sanno come utilizzarla per profilarmi. Il che è vero, ma genera una grandissima differenza tra coloro che sanno leggere solo le informazioni in chiaro e quelli che sanno interpretare i documenti: una differenza non meno grande che quella che intercorre tra l'analfabetismo e la cultura, o tra l'alchimia e la chimica.
Luciano, non ne dubito, ne è perfettamente consapevole, ma a questo punto i casi sono due: o troviamo un altro nome, oppure ci impegniamo in una lotta perdente in partenza, cioè moralizzare l'Infosfera, insegnarci a diventare tutti veritieri, e, quel che è ancor più improbo e mefistofelico, a conoscere tutti la verità: *eritis sicut deus scientes bonum et malum*. Vasto disegno, non nuovo, e mai realizzato in milioni di anni. Forse è opportuno orientarsi verso obiettivi più a portata di mano, e invece che applicarsi esclusivamente a raddrizzare il legno storto dell'umanità, opera doverosa ma in linea di principio inattuabile, comprendere e valorizzare quello che sappiamo fare meglio, cioè agire e interagire, magari fregandoci a vicenda. Nel riesumare un nome così inadatto, Luciano è stato fuorviato, a mio parere, da quattro ordini di circostanze che rendono l'Infosfera subito comprensibile (e questo è il vantaggio), ma alla fine insensata (e questo è lo svantaggio). Se vogliamo, possiamo classificarli secondo gli *idòla* baconiani.
Gli *idòla tribus* sono quelli che abbiamo come specie. E che si manifestano nell'idea che la conoscenza sia il massimo desiderio degli umani. Conviene invece, con un rapido esame di coscienza, considerare la descrizione degli umani come naturalmente vocati alla conoscenza una nobile fanfaronata, basti pensare a quanto tardiva sia l'apparizione della conoscenza nella storia dell'umanità, quanto pocola conoscenza (non la credenza, non l'abitudine) operi nella vita di ognuno di noi, professori compresi, e quanto poche volte il nostro accesso al *web* sia dettato da bisogni conoscitivi.
Gli *idòla specus*, cioè i pregiudizi che ognuno di noi ha, e in particolare quelli, così influenti all'inizio, così ininfluenti alla fine, che animano gli inventori di un apparato tecnico. Questi avevano in mente una accademia di dotti, mentre il *web*è diventato un bazar, un'arena, un cinema a luci rosse. Ora, come sappiamo (e Luciano è il primo a ricordarcelo), raramente le invenzioni obbediscono agli intenti degli inventori, e il *web* non deroga da questa legge. Contrariamente a quanto per lo più si crede, il *web* non è dunque principalmente una sfera in cui si raccolgono informazioni[8], ma un archivio in cui si registrano azioni, ben più che idee o opinioni. Più precisamente il *web* è un archivio che registra (trasformandoli in documenti) gli atti degli umani per addestrare le macchine, perseguendo il mandato di una automazione perfetta, che sollevi gli umani dalla fatica (e ovviamente li licenzi tutti, ma per questo abbiamo soluzioni, come vedremo). È in questo senso che mi pare più appropriato parlarne come di un apparato di registrazione invece che di comunicazione, e di una Docusfera in cui si depositano gli atti che gli umani compiono nel mondo della vita, cioè nella Biosfera.
Gli *idòla fori* derivano, come sappiamo, dalle imprecisioni del linguaggio. Nel caso del *web* queste imprecisioni sono errori belli e buoni, e si manifestano nella forma della paleonimia, ossia, in parole povere, del conservare nomi che non si adattano più alle cose cui si riferiscono. L'acronimo ICT, che designa le tecnologie del *web*, è eloquente sotto questo profilo: "Information and Communications Technologies". Ma le tecnologie a cui si riferisce,

8 Una critica precoce dei limiti dell'infosfera si può trovare in Day (2001).

oggi, hanno a che fare con l'informazione e la comunicazione solo in forma derivativa, e si rischia di generare lo stesso equivoco per cui la rete telefonica italiana, quando ero bambino, si chiamava "SIP", acronimo che nelle intenzioni designava, e non si capisce come, la Società Italiana per l'Esercizio Telefonico (di qui il mio stupore infantile: perché non la chiamano SIET?), ma era semplicemente il risultato della riconversione, nel 1964, della "Società Idroelettrica Piemontese", attiva dal 1918 e poi specializzatasi in telefonia.
Gli *idòla theatri* derivano dalle dottrine filosofiche del passato, e, per la loro autorevolezza, credo che abbiano giocato un ruolo predominante nella scelta di un termine tanto problematico. L'Infosfera ha un *pedigree* di tutto rispetto[9], visto che i filosofi, da Platone in avanti, hanno amato concepire un mondo delle idee, una noosfera o un Intelletto Generale da intendersi vuoi come una regione dell'essere spirituale, vuoi come un esito possibile della evoluzione sociale.
In questo quadro, l'Infosfera risulta concettualmente imparentata con gli ideali enciclopedici dell'Illuminismo, con il progetto idealistico di un sapere assoluto, e con la successiva teorizzazione di una Noosfera, ossia di una regione del conoscere e della interazione tecnologica in quanto contrapposta alla Biosfera, la regione della evoluzione e della vita. Il prevalere della connessione e della registrazione di atti, però, suggerisce che tra l'empireo della Infosfera e la fertile bassura della Biosfera si intromette una Docusfera, con un ambito di iscrizione di fatti, registrazione di atti e connessione di documenti che abbisognano di interpretazione molto più di quanto non apportino conoscenza e magari *General Intellect*. Il *General Intellect* è a sua volta il patriarca di una famiglia di concetti che ha invaso l'Infosfera (qui la parola non è inappropriata): economia della conoscenza, capitalismo cognitivo, intelligenza collettiva, intellettualità di massa, lavoro immateriale, società dell'informazione, classe creativa, condivisione dei saperi, postfordismo. A sua volta, l'intelligenza collettiva è stata considerata un sistema emergente, simile a quello che ha luogo nei superorganismi (Johnson, 2004) Il che è vero, ma ciò che emerge non è intelligenza, ma solo le condizioni per la formazione di una intelligenza.
Non che Luciano sia responsabile di tutto questo, ma diciamo che il richiamo all'Infosfera contribuisce ad accrescere la confusione, e funge da distrattore rispetto a quello che, a mio parere, è il problema fondamentale. Nel *web* gli umani, molto prima che scambiarsi informazioni e conoscenze, registrano azioni fissate in documenti, che proprio grazie al *web* divengono ubiqui e automatici; queste azioni possono, qualora se ne dispongano gli strumenti, diventare informazioni, ma in quanto tali sono un libro chiuso di cui ignoriamo il codice: diciamo, sono come Enigma prima che Turing lo decifrasse. Insomma, parlare di infosfera e di informazione genera controsensi, secondo cui per agire abbiamo bisogno di sapere[10], e che

9 La *noesis noeseos* di Aristotele, l'intelletto unico degli averroisti, gli ideali enciclopedici di Wolff, Diderot e d'Alembert, il Sapere Assoluto di Hegel e, più prossimamente, il General Intellect di Marx, la noosfera di Pierre Teilhard de Chardin o il "terzo regno", quello del pensiero, di Frege o di Popper. Ma con una differenza essenziale, e cioè che (tranne nel caso del *General Intellect* di Marx) in tutti gli altri esempi abbiamo a che fare con delle ipotesi teleologiche o con delle proposte tecnologiche. Per quasi tutti i filosofi che ho menzionato, l'Infosfera è un ideale teleologico, una ipotesi speculativa per cui i singoli e limitati pensieri che ognuno di noi ha dipendono da un dominio generale universalmente valido (2 + 2 fa 4 non perché lo penso io, ma perché è così in generale). A loro volta, gli enciclopedisti cercano di fornire una versione mondana di questo intelletto, in forma di sapere positivo, ma sono ben consapevoli del fatto che, come disse Borges, è vero che anche la più modesta delle enciclopedie è più colta del più dotto tra i suoi compilatori, ma che si tratta solo di un modo di dire: il compilatore sa qualcosa, poco o tanto che sia, l'enciclopedia non sa niente.
10 «Facciamo cose con le informazioni, per parafrasare la celebre espressione di Austin (1962). Pertanto, è perché *sappiamo come* fare questo o quello creativamente, interattivamente e collettivamente che possiamo correttamente asserire che *sappiamo che* questo o quello è il caso, passivamente e individualmente. La conoscenza proposizionale è la gloriosa conclusione del processo informativo, non il suo umile principio» (Floridi, 2020, p. 99).

conosciamo davvero solo ciò che noi abbiamo fatto, dunque la società e non la natura[11]. Anche qui, si può avere più di una riserva, per ragioni che partono dall'osservazione spassionata della nostra vita e ascendono all'ermeneutica della tradizione filosofica sino alle questioni più strettamente politiche.

Quanto all'esperienza di vita. Facciamo la maggior parte delle cose senza sapere. Non ho bisogno di una teoria economica per comprare un caffè, e posso comprarlo anche se dispongo di una teoria completamente sbagliata. Sostenere che noi conosciamo solo quello che abbiamo creato poggia su un equivoco linguistico, e cioè che l'uomo che vive nella società sia anche l'uomo che l'ha creata. Ora, se c'è qualcosa che non mi sento di aver creato è proprio il mondo sociale: mi son trovato delle regole, alcune le ho imparate, altre no, alcune le ho seguite, altre no. Mentre con le leggi naturali mi son trovato benissimo anche senza saperle, per lo stesso motivo per cui l'asino va dritto al fieno senza aver letto una riga di Euclide.

Quanto all'ermeneutica filosofica, Vico era convinto di qualcosa di tutt'altro che certo, e cioè che nell'ambito della storia, diversamente che nello studio della natura, che è opera di Dio, avessimo la possibilità di conoscenze dotate di peculiare certezza, ciò che viceversa è smentito dalle controversie tra gli storici e in generale dalla costitutiva opacità del mondo sociale[12]. Soprattutto, e in parziale contraddizione con la su prima tesi, Vico, riferendosi all'origine del mondo sociale, diceva – e non vedo come si possa dargli torto, con un minimo di esperienza del mondo – che *Homo non intelligendo fit omnia*[13] (ossia la tecnologia umana procede come la tecnologia naturale): gli umani agiscono prima di comprendere, generano gesti e processi[14], e la comprensione, se e quando arriva, non è la premessa (come pensano i cartesiani) ma il risultato. Si tratta di un qualcosa che è anteriore ed esteriore rispetto al pensiero[15], ossia di un *know how* e non di un *know that*, di un *können* e non di un *kennen*. Questa competenza potrà, ma non necessariamente dovrà, diventare comprensione, il che significa che la tecnologia potrà dar vita a isole di epistemologia, ossia di sapere concettuale, ma la grande forza fondamentale è l'azione (che nel *web* è potenziata in modo inaudito dalla registrazione), non l'informazione.

Quanto al problema politico. È molto concreto il rischio di trasformare l'appello all'Infosfera e alla intelligenza collettiva nel compenso immaginario di un lavoro reale e non riconosciuto (Lanier, 2011): la produzione di documenti che sono informazioni sfruttabili per alcuni (le piattaforme) e non per altri (gli utenti), che divengono così lavoratori non retribuiti di un lavoro di cui non hanno neppure contezza. Trascurare l'aspetto politico ha un costo, che non

11 *Ivi*, p. 118. «Una *white box* è un sistema di cui l'osservatore conosce tutto, poiché l'osservatore è in realtà colui che l'ha costruita. Questa prospettiva, ben conosciuta in informatica e nell'intelligenza artificiale, si iscrive nell'alveo della cosiddetta tradizione della conoscenza del costruttore, secondo la quale: 1) si può conoscere solo ciò che si costruisce, e pertanto 2) non si può conoscere l'autentica natura della realtà in sé. Come Vico e Hobbes, i filosofi che sostengono (2) argomentano che, dal momento che qualsiasi tentativo di conoscere la natura intrinseca della realtà è inevitabilmente votato alla sconfitta, è preferibile concentrare l'attenzione su quelle scienze il cui oggetto è creato da noi stessi, come la politica e le scienze».

12 Diversamente da ciò che pensava Vico, fenomeni come le recessioni, i cicli economici e i sistemi di potere dimostrano che la realtà sociale, sebbene sia dipendente dalle nostre credenze e azioni, cioè sia «fatta da noi», ci sia tutt'altro che trasparente. Su questa intrasparenza si vedano, per esempio: Ruben (1985); Goldman (1970); Gilbert (1996).

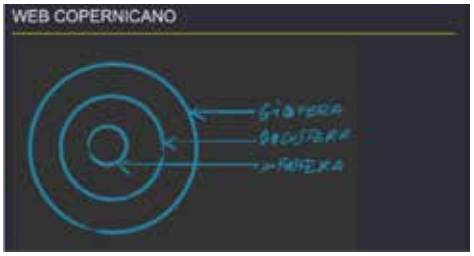
13 G. Vico, *Scienza nuova*: «Perché come la Metafisica ragionata insegna che "*homo intelligendo fit omnia*", cosi questa Metafisica fantasticata dimostra che "*homo non intelligendo fit omnia*"».

14 Gilbert Simondon definisce l'operazione tecnica non come una messa in forma secondo uno scopo prestabilito ma come "un movimento, una transizione", cfr. Simondon, 2017, pp. 101-142. In un ambito solo apparentemente diverso, Luigi Pareyson ha definito la creazione artistica come un fare che nel processo del proprio farsi inventa il proprio modo di fare. Cfr. Pareyson (1988). Cfr. anche Dorfles (2003): ciò che caratterizza ciascuna arte è esattamente non il risultato che ottiene l'artista, ma come lo può ottenere attraverso il medium che usa e le tecniche di cui si serve, che permettono di differenziare le arti le une dalle altre.

15 Secondo la definizione dello scambio mercantile in Sohn Rethel. Cfr. Sohn Rethel 1977; 1971; 1994.

è solo conoscitivo, perché sostituisce il problema politico ed economico con una deriva etica (ricordate i socialisti umanitari dell'Ottocento?) che non so quanto porti lontano, giacché il *web*, come del resto la strada per l'inferno, è lastricata di buone intenzioni, e i primi a enunciarle sono i proprietari di piattaforme, che ci avvisano (mentre stiamo cercando un ristorante) che la nostra *privacy* è la cosa più importante per loro. Ma visto che per noi il ristorante è più importante che la nostra *privacy*, accettiamo i *cookie* e tiriamo diritto. Parlando dei problemi dell'Infosfera Floridi menziona la fiducia *online*, l'hackeraggio, il *digital divide*, la società della sorveglianza, la *privacy*[16], la libertà di espressione, Wikileaks, i partner artificiali e la cyberguerra (Floridi, 2020, 128-129). Tutti problemi reali, ma con ogni probabilità insolubili in modo definitivo, mentre un problema più reale e radicale, quello del lavoro implicito e gratuito degli umani nel *web* non viene messo a fuoco, e nemmeno menzionato. Ed è così che questi problemi tendono ad assolvere il ruolo di distrattori rispetto al problema fondamentale, ossia, *in primis*, il radicamento strutturale della Infosfera trasparente in una Docusfera opaca e, *in secundis*, la dipendenza della Docusfera dalla Biosfera, e dunque la centralità dell'apporto umano, da riconcettualizzarsi come lavoro e da ridistribuirsi come *Welfare*, anzi, come *Webfare*, se vogliamo ribattezzare così il *Welfare* del XXI secolo reso possibile da un *design* concettuale che permetta davvero di comprendere che cosa è il *web* e cosa possiamo farcene.

3. *Web* copernicano



Diversamente dal *web* tolemaico dell'Infosfera e della comunicazione, propongo un *web* copernicano della Docusfera e della registrazione. In sintesi: 1. il *web* è anzitutto registrazione, e non solo comunicazione; funziona non come una televisione, ma come un archivio; 2. è azione e performatività prima che informazione, non si limita ad accumulare conoscenza, ma definisce uno spazio in cui hanno luogo atti sociali come promesse, impegni, ordini; 3. è reale prima che virtuale, ossia non è una semplice estensione immateriale della realtà sociale, ma si definisce come lo spazio elettivo per la costruzione della realtà sociale[17]; 4. è mobilitazione prima che emancipazione, ossia non fornisce immediatamente liberazione (come si credeva quando il *web* mosse i suoi primi passi) né semplicemente si configura come uno strumento di dominio, ma è piuttosto un apparato che mobilita, ossia fa compiere delle azioni; 5. è emergenza molto più che costruzione, nel senso che non è il progetto deliberato di qualcuno, ma piuttosto il risultato di molte componenti che sono venute convergendo in forma non programmatica; 6. infine, è opacità e non trasparenza[18], ossia non si chiarisce da solo (gli umani agiscono e interagiscono, però nella stragrande maggioranza dei casi non conoscono i princìpi che li guidano) ma, al contrario, chiede di essere chiarito, anche in questo caso

16 Nozione del resto non meno ambigua e problematica di "informazione", cfr. Thomson, 1975, pp. 295-314.
17 Latour et al., 2012; Marres, 2017, pp. 106-115.
18 Cfr. Van Den Eede, 2011, pp. 139-159.

rivelando uno stretto isomorfismo con la realtà sociale, e in particolare con quella sua punta emersa che è il capitale.
È questo il punto cruciale e dolente (perché non compreso, eppure molesto come un arto fantasma) della rivoluzione in corso. In realtà, il valore sul *web* non è prodotto dall'intelligenza e dall'informazione, ma da una mobilitazione che ha gli scopi più diversi (socialità, sessualità, consumo, noia...) e che se opportunamente documentata può trasformarsi in automazione, distribuzione, accumulo primario di dati che possono essere comprati e venduti. Siamo passati dalla strutturale penuria di documenti a una sovrabbondanza che non ha precedenti nella storia, e questo nuovo assetto si è manifestato come la vera essenza del capitale, che ora è capace di operare senza residui, *capitalizzando l'interazione sociale, gli atti registrati che hanno luogo sul web*, e che divengono valore potenziale per il Capitale Documediale (Ferraris, 2018, pp. 31-50), che nasce dalle nozze tra Mercurio e Filologia o, se preferiamo, tra denaro, *media* e documenti. I documenti vengono infatti capitalizzati dalle piattaforme a costo zero, attraverso un accumulo primario (i nostri dati divengono di proprietà delle piattaforme) che offre possibilità di automazione dei processi (si pensi alle macchine per traduzione che evolvono grazie agli utenti), di profilazione dei comportamenti, e ovviamente di monetizzazione (le piattaforme vendono i loro archivi, mentre nessuno di noi potrebbe pagare un caffè con i dati che ricava dal suo telefonino: al massimo potrebbe essere dissuaso a farlo dal contatore di battiti cardiaci).
Il capitale documediale non è più un capitale ristretto, riferito a un ambito particolare, quale era il caso del capitale commerciale, industriale e finanziario (nei quali peraltro era osservabile la forma di una progressivo ampliamento), bensì un capitale generale di cui si tratta di comprendere i caratteri e le potenzialità emancipative. Se l'ipotesi della Infosfera presenta il capitale semantico come una entità distinta dal capitale economico e da quello sociale (Floridi, 2018, pp. 481-497), io propongo invece di vedere nel capitale documediale la forma attuale di un processo di capitalizzazione coestensivo con l'ominizzazione, che ha avuto inizio con la produzione dei primi utensili, si è sviluppato attraverso la scrittura (e quella forma specifica di scrittura che è il denaro) e che oggi si manifesta non in una Infosfera, bensì in una Docusfera che a sua volta rimanda a una Biosfera.
In effetti, è indiscutibile, nel *web* si producono, raccolgono e distribuiscono delle informazioni (Floridi, 1996). Tuttavia, se non parlo l'ungherese, di un libro in ungherese posso solo dire che è un libro. Non molto, come informazione. Per la maggior parte degli utenti, il *web* è – nella stragrande maggioranza delle sue funzioni davvero decisive, quelle legate alla raccolta e al calcolo dei comportamenti – un libro in ungherese, ed è una autentica Infosfera solo per chi parla l'ungherese, ossia, fuor di metafora, per chi dispone degli strumenti di archiviazione, calcolo e informazione. I documenti non parlano a chiunque, non più di quanto la sintomatologia di una malattia possa suggerire la diagnosi a un profano. Questa circostanza è importante non solo dal punto di vista concettuale, della definizione di che cosa è il *web*, ma anche dal punto di vista politico ed economico, perché spiega che cosa il *web* può fare, e motiva le trasformazioni che introduce nella vita delle persone e delle società. Il *web* registra azioni e produce documenti, che il più delle volte gli utenti non sono consapevoli di produrre (e che in effetti sino a non molto tempo fa non sarebbero stati prodotti). Documenti che, in quanto tali, non sono più chiari di un codice a barre o di una lingua sconosciuta. Limitandosi all'Infosfera, si esamina solo la punta dell'iceberg, ossia il fatto che indubbiamente nel *web* ci sono delle informazioni, e si ritiene che questa costituisca la totalità del *web*.
Questo limite vale anche per i teorici del capitalismo cognitivo[19]. Non è vero che il capitale raccoglie informazioni e rilancia conoscenza. Il capitale raccoglie mobilitazione, l'indaffararsi

19 Cfr. Vercellone, 2006; Marazzi, 2002.

degli umani nel mondo, e genera documenti, grazie al suo potere di registrazione; documenti che solo se raccolti e interpretati possono generare conoscenza. Ecco il limite di tutti i discorsi sugli *open data*, vecchi del resto quanto la riflessione sul postmoderno[20]. C'è indubbiamente un senso per cui siamo arrabbiati per un motivo socratico, cioè perché siamo ignoranti. Ma la nostra ignoranza non consiste tanto nelle lacune formative quanto alle dinastie egizie o al PIL del Belgio, bensì piuttosto perché ignoriamo una cosa molto importante per la nostra vita, ossia che agendo produciamo valore, e che se il lavoro è produzione di valore, allora lavoriamo gratis. Alla faccia dell'Infosfera. Siamo talmente informati che non conosciamo nemmeno il nostro statuto sociale e lavorativo. Alla faccia dell'economia della conoscenza, se la maggior parte della conoscenza la produciamo noi, ma non ci torna indietro. Alla faccia della intelligenza collettiva, se non capiamo neppure che produciamo gratuitamente valore sul *web*, e ce la prendiamo invece con le banche, la borsa, i migranti. Come distinguere questa intelligenza collettiva dalla più banale stupidità? E qui, certo, diviene necessaria la filosofia come *design* concettuale, come invenzione (ripeto, su questo sono totalmente d'accordo con Luciano) a patto però che si sia capita la realtà con cui si ha a che fare, ossia appunto si sia passati dal *web* tolemaico al *web* copernicano.Il punto, insomma, non è adattare i vecchi parametri del lavoro alla nuova realtà, che probabilmente è nel suo insieme destinata a divenire, proprio a causa dell'automazione favorita dalla registrazione delle forme di vita umana attraverso il *web*, postlavorativa. Ed è per questo che trovo inappropriato parlare di fabbrica-rete o di postfordismo, se non come degli ultimi vestigi di un mondo destinato a tramontare. A chi dicesse che oggi c'è un nuovo schiavismo, chiunque obietterebbe che il lavoro fordista era ripetitivo, obbligatorio, alienante e retribuito, e che la mobilitazione non è né ripetitiva né obbligatoria né alienante né, soprattutto, retribuita. È appunto, con un gesto di *design* concettuale, che si può comprendere la nuova forma che ha preso il lavoro e fare i conti con essa.

Perché limitarsi alla Infosfera, ossia alla punta emersa dell'iceberg, quando possiamo disegnare un concetto più comprensivo che descriva un sistema di inclusioni, dove l'Infosfera è il minuscolo sottoinsieme della Docusfera, e quest'ultima a sua volta è il risultato di una Biosfera che non pensa ma agisce? Tutto sommato, è una ferita narcisistica anche questa, ma non consiste nel dire che le macchine pensano meglio di noi (questo è vero da sempre, ma senza di noi le macchine non vanno da nessuna parte), bensì nel riconoscere che noi pensiamo molto meno di quanto credessimo, e che *ego cogito, ego sum* è la descrizione non di una condizione normale dell'umanità, ma dell'esperimento mentale che un uomo di ingegno ha fatto una volta nella vita, dopo aver fatto il soldato, squartato gatti in cantina e pianto una figlia morta bambina. Ridurre tutto all'Infosfera è un platonismo, che coglie un punto (indubbiamente ci sono informazioni) ma perde il tutto, ossia l'obiettivo della filosofia.

4. *Design* alternativo

Ecco dunque il *design* alternativo che vorrei proporre. Il *web* tolemaico è un eidocentrismo, quello copernicano un etocentrismo. L'eidocentrismo ritiene che la funzione prioritaria del *web*consista nel raccogliere, diffondere e capitalizzare l'idea (εἶδος) e le sue varianti, da intendersi non letteralmente come forme, bensì come pensieri, convinzioni, credenze, informazioni e modi di sentire. L'etocentrismo muove invece dall'ipotesi che l'obiettivo del *web* consista nel raccogliere, classificare e capitalizzare delle forme di comportamento (ἦθος): le macchine automatizzano qualunque processo perché registrano e classificano i nostri atti,

20 È proprio sul libero accesso ai dati che si concludeva J.F. Lyotard, *La condizione postmoderna* (1979), tr. it. di C. Formenti, Feltrinelli, Milano 1981.

abitudini e desideri, sicché ciò che chiamiamo "intelligenza artificiale" non è che il grande archivio della commedia umana.
Gli umani sono mobilitati dai meccanismi e insieme conferiscono loro uno scopo; l'Archivio (il *web*) tiene traccia di questa mobilitazione e la elabora non tanto per scopi inquisitivi e ideologici, quanto piuttosto per migliorare l'automazione della produzione e l'efficacia della distribuzione, ossia per scopi commerciali. I dati raccolti, ovviamente, potranno essere riutilizzati anche per obiettivi di inquisizione ideologica, ma questa non è né la loro funzione primaria né, soprattutto, corrisponde alla natura dei dati raccolti, che sono per l'appunto ben più la manifestazione di comportamenti che non l'espressione di idee, la quale certo ha luogo, ma in forma minoritaria. Questo errore prospettico si propaga e condiziona una grande varietà di atteggiamenti nei confronti del *web*, che risultano affetti (in forma consapevole o meno) da eidocentrismo, con ciò pregiudicando l'efficacia della analisi. L'eidocentrismo poggia su più equivoci.
Il primo è l'antropomorfismo per cui quando si parla di "intelligenza artificiale" ci si rappresenta una sorta di riproduzione dell'intelletto umano che include caratteristiche come ad esempio la finalità, la volontà di potenza, il desiderio e la speranza. Nasce da questo equivoco il timore, così ricorrente nel discorso pubblico, di una presa del potere da parte dei computer. Quando Heidegger parla di intelligenza artificiale e di "cibernetica" immagina qualcosa come Hal di *2001 Odissea nello spazio*. Un cervellone che pensa meglio degli umani e che a un certo punto decide di comandarli. Non può pensare che il cervello è diffuso nell'ambiente e che forma lo stesso ambiente, pensa solo a una funzione di imposizione (Heidegger, 2002, pp. 35–57) che vede nella tecnica uno sgravio di responsabilità umane, perché è lei che dà gli ordini e siamo noi che li eseguiamo[21].
A ben vedere, siamo ancora sotto lo scacco delle leggi di Asimov (Asimov, 1991), che suppongono concordemente, e fallacemente, che gli automi possano avere intenzioni umane. E questo è un antropomorfismo comprensibile ai suoi tempi, non ai nostri. Un gorilla, magari, potrà prendere il potere. Un automa non potrà mai farlo. La documedialità non prende decisioni, ma esegue decisioni già prese. È una ingenuità fiabesca quella che si figura la crescita dell'automazione come un potenziamento della volontà di potenza delle macchine, che prenderanno il posto degli umani. Ciò di cui le macchine hanno bisogno (ed è la loro unica parentela con il vampirismo) non è l'intelligenza umana, bensì la vita organica.
Il secondo equivoco, che in qualche misura è complementare, nasce dal fatto che l'intelligenza artificiale, assimilata all'intelligenza umana, genererebbe come ho detto una forma di intelletto generale e condiviso, un sapere assoluto nato dalla interazione uomo-macchina[22]. Questa visione può declinarsi sia in forma euforica, come avveniva quando il *web* mosse i suoi primi passi, sia nella forma disforica, oggi prevalente, per cui il *web*, in quanto intelligenza collettiva, diventerebbe una grande macchina di controllo e di spionaggio. Entrambe le visioni sono fuorvianti. L'intelligenza artificiale, infatti, non mira alla comprensione e a ciò che ne

21 Argomento che – e non posso credere sia un caso – Heidegger affaccia solo dopo il 1946, cioè dopo che era stato usato dai nazisti a Norimberga, che si difendevano dicendo che eseguivano gli ordini trasmessi da un apparato tecnico perfetto diretto da Hitler, che dunque era l'unico responsabile. Nell'autobiografia di Speer tutto questo è esposto molto chiaramente. Togli Hitler e metti al suo posto il *Gestell* e hai la visione della tecnica proposta da Heidegger. E non finisce lì: Kittler, 1999; Kelly, 2011. Occorre superare questa forma di auto-assoluzione. Eppure il 29 ottobre 2019 è uscita su *La Repubblica* una intervista a Jaron Lanier che prima dice (ed è sacrosanto) che produciamo valore con la nostra mobilitazione sul *web*, dunque dobbiamo essere pagati, e poi che ha luogo una "manipolazione di massa delle coscienze", come se esistessero delle coscienze in sé, non manipolate dalla cultura, dalla religione, dalle ideologie politiche. E come se le nostre coscienze potessero interessare a qualcuno.
22 Tra le varie teorizzazioni dell'intelletto generale ricordo: Lévy, 1996; Vattimo, 1989; Hardt & Negri, 2017; Virno, 2002; Hardt & Negri, 2001; Marazzi, 2002.

deriva, finalità e volontà di potenza comprese, bensì alla automazione di processi. Un grande archivio e una enorme capacità di calcolo divengono capaci di surrogare tutto ciò – ed è moltissimo – che nell'umano è meccanico, trasformando tutto il lavoro vivo in un lavoro morto incorporato nella macchina.

In terzo luogo, la visione eidocentrica e tolemaica rimuove una circostanza essenziale: ciò che offre l'intelligenza artificiale non è una intelligenza ma una capitalizzazione che di per sé non ha nulla di cognitivo. La capitalizzazione è l'accumulo di risorse, non necessariamente finanziarie, che si oppone alla ridistribuzione, cioè all'uso periodico delle risorse in questione. Ciò che è realmente interessante, nel caso del sapere, così come in quello della competenza tecnologica, è che ridistribuzione e uso non si contrappongono. Tutto ciò che l'umanità ha appreso in termini di comprensione e di competenza può essere adoperato *senza che il capitale venga intaccato, e anzi incrementandolo ulteriormente*. La sola condizione perché ciò avvenga è il possesso di buoni strumenti di registrazione che consentono insieme capitalizzazione e ridistribuzione.

Comprendere il *web* copernicano significa dunque capovolgere l'eidocentrismo a vantaggio dell'etocentrismo. Questo capovolgimento non ha solo una portata teorica, giacché indirizza la comprensione degli obiettivi politici del *web*, che per l'eidocentrismo è il controllo politico, mentre per l'etocentrismo è il profitto economico, e comporta due diverse visioni del *web* nel suo insieme. L'eidocentrismo adotta infatti il modello Informazione → Documento → Azione: delle informazioni si fissano in documenti e producono delle azioni.

L'etocentrismo segue invece il modello Azione → Documento → Informazione. Quanto dire che all'inizio ci sono dei comportamenti che vengono fissati in documenti (i dati) e che possono, in determinate circostanze e disponendo di opportuni strumenti di calcolo e di interpretazione, trasformarsi in informazioni, che generalmente non sono trasparenti né di pubblico dominio[23].

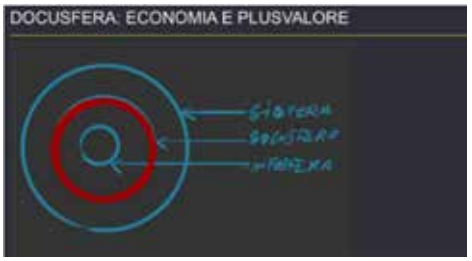
Trattandosi di una produzione di oggetti sociali resa possibile dalla esplosione della registrazione, il *web* è anzitutto una Docusfera, un ambito in cui si producono documenti, che non necessariamente veicolano in chiaro delle informazioni, non più di quanto possa farlo una traccia di DNA che diviene informativa solo quando viene sottoposta a esame scientifico. Definendo "Infosfera" ciò che è a tutti gli effetti una Docusfera, ossia un ambito di documenti, ci lasciamo sfuggire il carattere distintivo della società prodotta dal *web*, una società della registrazione prima che della comunicazione, per il semplice motivo che, come ho detto, il tratto saliente del digitale rispetto all'analogico sta nel fatto che in quest'ultimo la registrazione può (anche se non necessariamente deve) seguire la comunicazione, mentre nel digitale la registrazione è logicamente, tecnologicamente e ontologicamente preliminare alla comunicazione.

Ma, di là dall'aspetto puramente teorico, si apre un problema di gestione politica e di responsabilità critica. Gli ideali che solitamente animano la riflessione politica e sociale sul *web* sono quelli di una *tutela* (anche se per l'appunto non sembra che gli umani vogliano essere tutelati più di tanto); e della *trasparenza* di un ambito che un tempo veniva considerato come lo spazio della trasparenza per eccellenza (l'Infosfera, appunto) e che oggi si scopre essere opaco. Entrambi gli obiettivi non possono venire conseguiti in modo semplice – per esempio, con la richiesta del libero e universale accesso ai dati – sia perché in questo caso si aprirebbe un conflitto tra tutela e trasparenza, sia perché la Docusfera è un ambito intelligibile ma non intelligente. Il primo passo per una diversa consapevolezza deve dunque consistere proprio nel comprendere questa circostanza, ossia nel passaggio dall'eidocentrismo all'etocentrismo.

23 Cfr. il notevole libro di Badia (2019), *The Information Manifold. Why Computers Can't Solve Algorithmic Bias and Fake News*, che ha il merito di sottolineare come l'informazione ha non solo un valore semantico, ma anche sintattico e pragmatico, ossia di mostrare come l'infosfera sia radicata nella Docusfera (sintassi) e nella Biosfera (pragmatica).

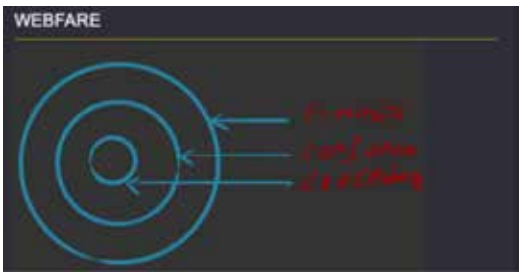
Mi avvio alla conclusione. Con tutto questo, non intendo minimamente negare l'esistenza dell'Infosfera, ma solo far presente che, a seconda del peso che diamo a queste sfere, abbiamo soluzioni diverse ai problemi del *web*, e che si tratta piuttosto di avere una diversa scala di rilevanze. Propongo pertanto un *design* concettuale alternativo a quello di Luciano che riassumo in quattro figure.

Concentrarsi sull'Infosfera significa focalizzare l'attenzione non sulla totalità del *web*, ma su una sua parte minoritaria, quella delle comunicazioni esplicite e riconosciute come tali. A questo punto, il problema si riduce a quello di un'etica della comunicazione, estendendo delle tendenze già in corso nel secolo passato e in riferimento alla comunicazione in generale, senza riferimento al *web*, di cui vien meno la specificità.



Portare l'attenzione sulla Docusfera significa invece mettere in chiaro la dimensione economica del *web* e la produzione di plusvalore che ha luogo nello scambio, che solo apparentemente è equo, tra utente e piattaforma. Riconoscendo così che gli utenti del *web*, nel momento in cui consumano informazioni, producono documenti a loro non necessariamente accessibili, e molto più pregiati delle informazioni e dei servizi ricevuti dagli utenti. Nello scambio, dunque, ha luogo una produzione di valore a esclusivo vantaggio delle piattaforme che propongo di chiamare "plusvalore documediale". Un plusvalore che, se tassato in misura ragionevole, potrebbe costituire la base per un *Welfare* digitale.

Ovviamente ci si chiede se davvero lo scambio sia iniquo e se non si possa parlare piuttosto di mutua convenienza, specie se ci si riferisce all'Infosfera (dopotutto, chi avrebbe pensato di avere delle biblioteche e discoteche gratuite?). Così come è legittimo domandarsi se l'abbassamento dei prezzi consentito dall'automazione e dal perfezionamento della distribuzione che le piattaforme ricavano dalla profilazione della Docusfera non costituisca già un risarcimento per la produzione di valore generata dagli utenti. Ma, anche ammettendo queste osservazioni, resta che Infosfera e Docusfera perdono qualunque significato se non sono animate dalla Biosfera, dai bisogni che mobilitano gli umani da sempre, e che ora determinano la necessità e la finalità del *web*. In poche parole, possiamo benissimo immaginare una umanità senza *web* (si tratterebbe della storia dell'umanità che va dal primo ominide a trent'anni fa), ma non possiamo immaginare un *web* senza umanità, perché l'automa non va da nessuna parte se non è spinto e motivato dai bisogni delle anime.



Una volta riconosciuto tutto questo, diviene non solo concepibile e legittimo un *Webfare*, ma troviamo la giusta collocazione dell'Infosfera: la Biosfera costituisce il mondo della vita fondamentale che conferisce finalità al *web* così come a qualunque altro apparato tecnico; la Docusfera genera valore attraverso il consumo, ed è l'ambito in cui le anime umane incontrano gli automi; e l'Infosfera costituisce una zona di comunicazioni responsabili e di informazioni corrette che, ben lungi dal costituire l'essere del *web* (ne è, come abbiamo visto, una parte assai minoritaria) ne rappresenta il dover essere, lo sviluppo che l'umanità deve seguire sotto forma di educazione e di progresso razionale.

REFERENCES

Asimov, I. (1991/1950), *Io, Robot* . Tr. it. di R. Rombelli. Milano: Bompiani;
Badia, A. (2019). *The Information Manifold. Why Computers Can't Solve Algorithmic Bias and Fake News*. Cambridge: The MIT Press;
Borges, J.L. (1961/1944). *Finzioni*. Tr. it. di A. Melis. Torino: Einaudi;
Bunz, M. (2012). *Die stille Revolution: Wie Algorithmen Wissen, Arbeit, Öffentlichkeit und Politikverändern, ohne dabei viel Lärmzumachen*. Berlin: Suhrkamp;
Day, R. (2001). *The Modern Invention of Information: Discourse, History, and Power*. Edwardsville: Southern Illinois University Press;
Dorfles, G. (2003/1952). *Discorso tecnico delle arti*. Milano: Christian Marinotti;
Ferraris, M. (2021). *Documanità. La filosofia del mondo nuovo*. Roma-Bari: Laterza;
Ferraris, M. (2018). *From Capital to Documediality*. In A. Romele, E. Terrone (Eds.), *Towards a philosophy of digital media*. (pp. 31-50). Basingstoke: Palgrave MacMillan;
Floridi, L. (2017/2014). *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Tr. it. di M. Durante. Milano: Raffaello Cortina Editore;
Floridi, L. (2020/2019). *Pensare l'infosfera*. Tr. it. di M. Durante. Milano: Raffaello Cortina Editore;

Floridi, L. (2018). *Semantic Capital: Its Nature, Value and Curation*, in «Philosophy and Technology», 31, 481-497;
Floridi, L. (2015). *The Onlife Manifesto. Being Human in a Hyperconnected Era*. London: Springer;
Formenti, C. (2011). F*elici e sfruttati. Capitalismo digitale ed eclissi del lavoro*. Milano: Egea;
Frege, G. (1918), *Der Gedanke. Eine Logische Untersuchung* (1918-19), in Id., *Beiträgezur Philosophie des deutschen Idealismus*, I, (pp. 58-77). Erfurt Verlag der Keyfer'fchen Buchhandlung, Erfurt;
Freud, S. (1989/1917). *Una difficoltà della psicoanalisi*. Tr. it. di C.L. Musatti. In Id., *Opere*, vol. 8 (1915-1917), (pp. 657-664). Torino: Bollati Boringhieri;
Gilbert, M. (1996). *Living Together. Rationality, Sociality, and Obligation*. Lanham: Rowman & Litterfield;
Goldman, A. (1970). *A Theory of Human Action*. Princeton: Princeton University Press;
Hardt, M., Negri, A. (2001/2000). *Impero*. Tr. it. di A. Pandolfi e D. Didero. Rizzoli, Milano: Rizzoli;
Hardt, M., Negri, A. (2017). *Assembly*. New York: Oxford University Press;
Havelock, E. (2005/1986). *La musa impara a scrivere: riflessioni sull'oralità e l'alfabetismo dall'antichità al giorno d'oggi*. Tr. it. di M. Carpitella. Roma-Bari: Laterza;
Heidegger, M. (2002/1949). *L'impianto*. In Id., *Conferenze di Brema e Friburgo*, tr. it. di G. Gurisatti (pp. 35-57). Milano: Adelphi;
Johnson, S. (2001/2004). *La nuova scienza dei sistemi emergenti. Dalle colonie di insetti al cervello umano, dalle città ai videogame e all'economia, dai movimenti di protesta ai network*. Tr. it. di A. Antonini. Milano: Garzanti;
Kelly, K. (2010/2011). *Quello che vuole la tecnologia*. Tr. it. di G. Olivero. Torino: Codice;
Kittler, F. (1986/1999). *Gramophone Film Typewriter*, Stanford: Stanford University Press;
Lanier, J. (2010). *Tu non sei un gadget.* Tr. it. di M. Bertoli. Milano: Mondadori;
Lanier, J. (2013). *Who Owns The Future?*. New York: Simon & Schuster;
Latour, B. et al. (2012). *Enquête sur les modes d'existence: une anthropologie des modernes*. Paris: La Decouverte;
Lévy, P. (1994/1996), *L'intelligenza collettiva. Per una antropologia del cyberspazio*, tr. it. di D. Feroli e M. Colò. Milano: Feltrinelli;
Lyotard, J.F. (1979/1981). *La condizione postmoderna*, tr. it. di C. Formenti. Milano: Feltrinelli;
Marazzi, C. (2002). *Capitale e linguaggio. Dalla New Economy all'economia di guerra*. Roma: DeriveApprodi;
Marres, N. (2017). *Digital Sociology*. Cambridge: Polity Press;
Pareyson, L. (1954/1988). *Estetica. Teoria della formatività*. Milano: Bompiani;
Putnam, H. (1987/1991). *La sfida del realismo*, tr. it. di N. Guicciardini. Milano: Garzanti;
Roncaglia, G. (2010). *La quarta rivoluzione. Sei lezioni sul futuro del libro*. Roma-Bari: Laterza;
Ruben, D.H. (1985). *The Metaphysics of Social Reality*. New York: The Free Press;
Simondon, G. (1970/2017). Nascita della tecnologia, in *Sulla tecnica*, tr. it. di A.S. Caridi (pp. 101-142). Napoli-Salerno: Orthotes;
Sohn Rethel, A. (1976/1994). *Il denaro. L'a priori in contanti*, tr. it. di F. Coppellotti. Roma: Editori Riuniti;
Sohn Rethel, A. (1970/1977). *Lavoro intellettuale e lavoro manuale: per la teoria della sintesi sociale*, tr. it. di V. Bertolini e F. Coppellotti. Milano: Feltrinelli;
Sohn Rethel, A. (1971). *Warenform und Denkform*. Frankfurt/M: Suhrkamp;
Thomson, J.J. (1975). The Right to Privacy. *Philosophy and Public Affairs*, 4, 295-314;
Toffler, A. (1980/1987). *La terza ondata*, tr. it. di L. Berti. Milano: Sperling & Kupfer;
Van Den Eede, Y. (2011). In Between Us: On the Transparency and Opacity of Technological Mediation. *Foundations of Science*, XVI, 2, 139-159;
Vattimo, G. (1989). *La società trasparente*. Milano: Garzanti;

Vercellone, C. (ed.) (2006), *Capitalismo cognitivo*. Roma: Manifestolibri;
Virno, P. (2002). General Intellect, Exodus, Multitude, *Archipélago*. Retrieved from https://www.generation-online.org/p/fpvirno2.htm
Zuazua, M. (2019). The Fourth Industrial Revolution Will Change Production Forever. Here's How. *World Economic Forum*, 18 Jan 2019.

LEONARDO MANNA
Facoltà di Teologia di Lugano, Università Vita-Salute San Raffaele
mannal@teologialugano.ch

IL PENSIERO COME RELAZIONE O INTERO SEMANTICO? INTORNO ALLA FILOSOFIA DI LUCIANO FLORIDI

abstract

In Pensare l'infosfera La filosofia come design concettuale, *Luciano Floridi states the value of a correct analysis and philosophical responses within the information age we are living in. The aim of this paper is to suggest several ideas for reflecting on some of Floridi's topics such as the analysis of presence, relationship and semantization. In the second paragraph, after a brief introduction about the author, I will analyse philosophical insights on the relationship between being and determination. In the third paragraph, I am going to suggest the importance of a correct analysis of presence. In the fourth paragraph, a possible analogy between the philosophy of information and a certain form of neo-idealism will be shown. At last, I try to show how the philosophy of information can play a key role in the way we think about reality, but also its possible conceptual risks.*

keywords

information, semantization, thought, presence

Phenomenology and Mind, n. 20 - 2021, pp. 164-174
DOI: https://doi.org/10.17454/pam-2014
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. La filosofia come primarietà dell'informazione

Luciano Floridi da anni delinea con crescente precisione, competenza e acume speculativo il ruolo fondamentale della filosofia dell'informazione all'interno della contemporaneità, con tutti gli elementi che da questa dipendono: etica, società, *privacy*, politica, *governance*, per citare solo alcuni ambiti di applicazione. L'informazione nel nostro tempo ha un ruolo di primaria importanza e, di conseguenza, è imprescindibile la comprensione di questa nel suo statuto ontologico e quindi filosofico. Negli ultimi anni, Floridi sta ampiamente portando avanti il discorso sul valore rivoluzionario dell'informazione[1] e su come l'uomo contemporaneo in quest'era si stia riconfigurando in rapporto ad essa come *inforg*.
In merito a queste pregevoli e imprescindibili questioni, di cui la filosofia ormai non può assolutamente fare a meno, il testo di Floridi[2] andrebbe considerato in modo ambivalente come indice riassuntivo rispetto alle pubblicazioni degli ultimi anni dello stesso Floridi e, dall'altro lato, come passo aggiuntivo rispetto a queste. Questo saggio rappresenta un invito ulteriore a prender sul serio la filosofia dell'informazione, vista come conoscenza mediante la quale "negoziamo la corretta tipologia di comunicazione che intratteniamo con il mondo" (p. 100) per comprendere la realtà di esso. Floridi suggerisce che si può divenir "amici della verità, senza essere nemici di Platone" (p. 101), in forza del principio secondo il quale non bisogna identificare colui che conosce con colui che utilizza l'informazione. Da questo presupposto è possibile intravedere tutta l'impostazione costruzionista di Floridi, da cui si apprende che il mondo si configura a partire da chi costruisce l'informazione e non attraverso artefatti eidetici già predisposti. Il filosofo nell'era dell'informazione è un costruttore, un agente poietico oltre che un utente, è sia un produttore che un consumatore. Floridi dimostra che le domande filosofiche sono sempre aperte anche se non ultime e così la filosofia si rivela essere una evoluzione continua di artefatti semantici, di domande aperte. La filosofia è un continuo disegno di concetti, che chiarisce domande, che son sempre aperte, disegnando risposte valide e convincenti.
La filosofia, nel suo esser dinamica, è sistematicamente in evoluzione:

1 Riporto solo alcuni dei lavori apparsi recentemente in Italia come: Floridi, L. (2012). *La rivoluzione dell'informazione*. Torino: Codice; Floridi, L. (2017). *La quarta rivoluzione: Come l'infosfera sta trasformando il mondo*. Milano: Cortina; Floridi, L. (2009). *Infosfera: Etica e filosofia nell'età dell'informazione*. Torino: G. Giappichelli; e il recente Floridi, L. (2020). *Il verde e il blu: Idee ingenue per migliorare la politica*. Milano: Raffaello Cortina Editore.
2 Floridi, L. (2020). *Pensare l'infosfera: La filosofia come design concettuale*.

> È uno spazio aperto: chiunque può avervi accesso, non importa quale sia il punto di partenza, e il disaccordo vi resta sempre possibile. È anche uno spazio dinamico, poiché quando l'ambiente culturale cambia, la filosofia fa lo stesso ed evolve (Floridi, 2020, p. 54).

L'approccio filosofico è indispensabile per l'accrescimento del capitale semantico legato alla produzione, relazione e comprensione delle informazioni. Il minimalismo, il metodo dei livelli di astrazione e il costruzionismo, nel modo in cui li concepisce Floridi in *Pensare l'Infosfera*, costituiscono gli strumenti centrali per una corretta filosofia dell'informazione, in cui l'ultimo stadio della riflessione filosofica, su cui va mantenuta aperta la riflessione, deve essere una continua semantizzazione dell'essere.
Floridi si dichiara "costruzionista", e così deve essere la stessa filosofia: costantemente in fase di progettazione in atto di *design* concettuale, in un contesto in cui il concetto stesso di *poiesis* viene lentamente identificandosi a quello di noesi:

> Se l'autentica conoscenza è conoscenza della natura intrinseca dell'oggetto conosciuto (conoscenza dell'ontologia del conosciuto) e se non sussiste alcuna innata acquisizione di tale prototipo, allora conoscere un fenomeno, un artefatto o, nel nostro caso, conseguire un'informazione ed essere in grado di renderne conto, significa essere capaci di produrre e riprodurre, comporre e scomporre, costruire, montare e smontare tale fenomeno, artefatto o informazione, ovvero di migliorarli o rispondere a domande al loro riguardo, e in ciascun caso per le ragioni corrette. Perché la conoscenza sia possibile, la *mimesis* deve essere sostituita dalla *poiesis* (Floridi, 2020, p. 109).

Lo spazio all'interno del quale ogni possibile informazione rimodella la natura dell'uomo come agente informazionale e poietico va compreso superando la dicotomia platonica fra *techné* ed *epistème*, ovvero, come ribadisce l'autore, tra sapere pratico e sapere teoretico. Non si tratta quindi solo di comprendere il potenziale fine a sé stesso dell'informazione come ancella della tecnologia, ma al contrario di rendersi conto del valore dell'informazione in rapporto a molteplici campi scientifici nei quali il valore aggiunto sta nella creazione di nuovi ambienti e significati. L'informazione e la comprensione di essa nelle sue dinamiche ha qui il suo ruolo primario.
Per la formulazione di una corretta comprensione della filosofia dell'informazione, il testo di Floridi offre questo tipo di struttura: 1) la rilevanza di una domanda filosofica; 2) l'importanza della conseguente risposta filosofica; 3) la filosofia come disegno concettuale; 4) cinque lezioni tratte dalle conclusioni raggiunte nelle parti precedenti: fissare il corretto livello di astrazione, quali sono le giuste domande filosofiche da porre, riformulare una nuova antropologia filosofica, utilizzare la filosofia dell'informazione per comprendere il mondo contemporaneo e come formare il capitale semantico futuro. Nel nostro intervento ci soffermeremo sui punti 2) e 3), corrispondenti al secondo e terzo capitolo dell'opera in questione, ossia sulle tesi poste da Floridi nel libro in oggetto riguardanti il concetto di presenza e il metodo dei livelli di astrazione[3].

3 Per un approfondimento del metodo dei livelli di astrazione proposto da Floridi in modo decisamente più analitico si guardi al suo Floridi, L. (2008). *The method of levels of abstraction. Minds and machines*, 18(3), 303-329. Lo stesso articolo è stato poi rimaneggiato dallo stesso Floridi ed arricchito con notevoli altri approfondimenti ed inserito nel terzo capitolo della sua opera: Floridi, L. (2013). *The philosophy of information*. Oxford: Oxford University Press. Ed ancora il metodo dei livelli di astrazione viene sistematizzato nella sua versione originale in inglese, di cui la versione italiana è contenuta nell'opera in questione del presente saggio, ovvero: Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.

2. Semantizzazione e determinazione

Floridi è perentorio, sin dall'inizio dell'opera, nel dichiarare l'importanza di mantenere sempre aperta la "semantizzazione dell'essere" (Floridi, 2020, p. 53), non dando troppo peso al punto di partenza, ma stabilendo la rilevanza del disaccordo informato, poiché tale disaccordo è sempre una possibilità. Viene spontaneo, quindi, considerare la semantizzazione, come Floridi converrebbe, in quanto produzione e disegno costante di ontologie, una modellazione sistematica di enti possibili. La produzione, ontologicamente parlando, è sempre un qualcosa di cui possiamo parlare, è sempre legata al soggetto presente, un ente che opera direttamente sulla realtà. Nel saggio di Floridi in questione, non è stata data una definizione di ente, benché questo venga spesso associato ad altri termini come: agente, utente o ente informazionale. Potremmo definire l'ente come l'*origo* concettuale dal quale ogni domanda e relativa risposta viene formulata e originata. Per una corretta definizione di ente si potrebbe dire *simpliciter*: quell'elemento minimo di significato da cui ha origine qualsivoglia postulazione. Il primo capitolo si occupa dell'origine di una postulazione e di come questa parta da una corretta formulazione di domande, le quali, assieme alle relative risposte, non sono mai sospese, ma "incorporate in una rete di ulteriori domande" e risposte. Le domande sorgono sempre all'interno di un contesto e, nello specifico, sono sempre lette a un livello di astrazione (LdA) più o meno complesso.

Ebbene, l'ente, in quanto portatore di significato e punto di partenza dal quale è possibile costruire ogni discorso, secondo Floridi non è scisso dal domandare filosofico, non può mai oltrepassare il suo orizzonte. Si noti una analogia con il principio heideggeriano, che lo stesso Floridi descrive come "semantica della presenza o del *Dasein*" (p. 65). Essere situato sempre all'interno di un determinato ambiente, rimarca a grandi linee l'esser-gettato di cui parla lo stesso Heidegger in *Sein und Zeit*. Questo trovarsi non è pre-strutturale, non è trascendentale, ma è *Geworfenheit,* il qualificarsi da parte dell'ente nel suo essere situato, collocato. In questo modo l'esserci di Heidegger, il *Dasein*, rivoluziona l'approccio alla comprensione fenomenologica di un concetto originario e alla cognizione del mondo inteso nel suo essere semantizzato, sempre a una certa comprensione. Dice Heidegger che "l'esserci è innanzitutto e per lo più presso il "mondo" di cui si prende cura" (Heidegger, 1927, p. 279).

Nella filosofia dell'informazione viene proposta una semantizzazione dell'ente e, così, viene dichiaratamente abbandonata una indagine delle cose ultime o delle domande ultime. Ciò accade perché, per Floridi, la natura di ogni domanda/problema va investigata dal punto di vista della sua complessità e le azioni necessarie per risolverlo/comprenderlo sono sempre determinate dalla complessità dell'analisi che guarda al problema stesso e alle sue risorse: l'orizzonte del problema.

Le domande ultime, quindi domande assolute, per l'autore, sono domande che possono generare solo una quantità enorme di ulteriori domande e risposte. Kantianamente, Floridi porta avanti il principio secondo il quale il mondo noumenico delle cose in sé è inconoscibile, poiché è impossibile da comprendere nella sua interezza, ma in modo molto diverso. La stessa domanda qui non ha senso di porsi. Interrogarsi sull'Essere come concetto di Intero o come termine primo/ultimo è un procedimento che non deve essere metafisicamente compreso o meglio de-finito (*ti estì* socratico), ma va semantizzato, parafrasando Gustavo Bontadini[4]. Su questa linea possiamo porre ora una netta distinzione fra semantizzare e definire. La definizione dell'Essere è a posteriori e compete all'ente in quanto orizzonte già semantizzato, già dotato di identità e significato. Se l'ente è il punto nodale dal quale partiamo per costituire una

4 Un breve testo, fra i vari, che posso citare di Gustavo Bontadini, che offre importanti spunti per dialogare sui temi trattati è sicuramente: *Per una teoria del fondamento*, in Bontadini, G. (2009). *Metafisica e deellenizzazione*. Milano: Vita e pensiero.

domanda – poi si dirà a un determinato LdA – allora ogni risposta/domanda è costituita da un modello di essa, da una formalizzazione mediante la quale possiamo dire qualcosa.
Secondo questa prospettiva non è possibile dire, filosoficamente, cosa sia una cosa in sé, ma solo se questa cosa che è (ente) funziona in base a uno scopo determinato. Il modello funziona se questo produce uno scopo funzionale all'accrescimento o decremento di significato. Una prima conseguenza che viene spontanea da queste brevi note è che i modelli, di cui possiamo fare esperienza e di cui possiamo parlare, sono finiti. Nello specifico, il modello è in funzione degli osservabili che sono disponibili, lo stesso Floridi definisce il LdA come "un insieme finito, ma non vuoto, di osservabili" (Floridi, 2020, p. 48). Dal modello possiamo ricavare una definizione primitiva di mondo. Più precisamente: ciò che sappiamo, lo sappiamo mediante l'ente (ciò che è), che dà notizia del modello e permette che possa esserci una somma determinata di modelli finiti a cui fa riferimento. Floridi mostra che questa definizione è pluralista e non relativista, in forza del fatto che i modelli possono render conto solo della realtà conoscibile, relativa all'ente e poi a un dato LdA e mai di quella ultima, escatologica o teleologica.
Andando più a fondo, però, questo è possibile solo per mezzo dell'informazione, che consiste noeticamente in dati ben formati e dotati di significato. L'informazione va definendosi come qualcosa che è, di cui possiamo parlare e che è già dotato di significato e sempre inteso in modo determinato, poiché analizzato a un dato LdA.
Possiamo parlare sempre di qualcosa che è in quanto ente, preso nel suo momento speculativo a un determinato LdA, presente e in perenne discussione rispetto al suo esser presente.
Uno stato di cose, in quanto è sempre posto in base alle sue risorse e quindi orientato semanticamente, è anche sempre già determinato. Conosciamo i suoi termini, secondo Floridi, in quanto sappiamo di essi mediante quel LdA determinato, che ricordiamo essere dinamico: la complessità è analizzata a seconda del "diversificarsi" della sua posizione.
Citando Deleuze, potremmo dire che "la differenza è lo stato in cui si può parlare della determinazione" (Deleuze, 1968, p. 53) o, meglio, che possiamo definire la presenza dell'ente, come un qualcosa dotato sempre di significato, ma anche sempre soggetto a una modellazione della sua ontologia, della differenziazione concettuale. "D'altro canto, quando determiniamo la differenza come differenza concettuale, crediamo di aver fatto abbastanza per la determinazione del concetto di differenza in quanto tale" (p. 51); questo perché il concetto per Deleuze segue:

> la determinazione da un capo all'altro, in tutte le sue metamorfosi, e la rappresenta come pura differenza consegnandola a un fondamento, in rapporto al quale non importa più di sapere se si è davanti a un minimo o a un massimo relativi, davanti a un grande o a un piccolo, davanti a un principio o a una fine, poiché entrambi coincidono nel fondamento come un solo e stesso momento "totale", che è anche quello dello svanire e del prodursi della differenza, quello della sparizione e dell'apparizione (Deleuze, 1968, p. 76).

L'"impegno ontologico" di cui parla Floridi ha questo obiettivo. Formalmente, l'impegno ontologico che l'ente assume porta questo stesso a esser il soggetto del suo diversificarsi: considerando sé e modificando il suo LdA, l'ente, per mantenere il suo statuto di esistente e variabile, dovrà sempre negare la sua posizione ontologica, ma porsi anche a fondamento di essa. Impegnarsi ontologicamente rispetto a uno specifico sistema significa che tale sistema viene modellato, sostituito in favore di un sistema diverso, non per forza qualitativamente migliore o peggiore. La scelta di un determinato LdA permetterà che vi siano quelle informazioni necessarie o osservabili, che utilizzeremo per elaborare il modello. L'osservazione è – secondo

questa posizione – soggetta alla sua negazione in favore di una osservazione diversa. Ed è possibile affermare ciò a partire dalla affermazione che Floridi fa in merito al sistema, per cui ogni mutazione del sistema corrisponde a una mutazione di stati e viceversa; una continua transizione fra stati. Questi sistemi si determinano dal valore delle proprietà poste in ogni momento di questa loro mutazione.

Risulta indispensabile la tematizzazione relativa a un altro termine adoperato da Floridi: la presenza. L'autore definisce la presenza mostrando i limiti legati a una analisi meramente epistemologica: il fallimento epistemico (FE) è insufficiente per definire una descrizione ben formata delle cose e non permette che si possa costituire la sua negazione o l'assenza di essa. La presenza siffatta può inglobare semanticamente altri spazi come la co-telepresenza e l'assenza – ciò che non è presente – e può assumere solo una connotazione metafisica, aperta a speculazioni su termini e problemi astratti. Se il FE non può definire precisamente l'assenza, Floridi assume come metodo rilevante, approfondendo l'uso degli LdA, l'osservazione di successo (OS).

Floridi parafrasa Quine nella definizione delle modalità di esistenza dei livelli di astrazione e formula lo stesso LdA come il valore di una variabile tipizzata, in cui la necessità di un LdA è a sua volta dipendente da un ulteriore LdA scelto. La presenza, se è dipendente dall'ontologia che scegliamo per analizzare qualsiasi cosa, si pone in ogni analisi possibile come precomprensione oggettiva, ancora, sempre determinata ed immediata. Un ulteriore elemento in accordo con la filosofia heideggeriana, dove l'esserci, il *Dasein*, è sempre situato, finito e trasmesso.

3. La filosofia come interezza o mediazione della presenza

È in corso, secondo il filosofo dell'informazione, una vera e propria rivoluzione dell'informazione. L'ontologia che nasce a partire da questa rivoluzione, come detto, predispone la filosofia al *design* come progetto futuro perenne e mediante l'uso dei LdA, Floridi tematizza la collocazione spaziale e temporale della presenza, dimostrando il modo di intendere correttamente la sua localizzazione.

Se la presenza è tutto ciò che è possibile dire del mondo e mediante cui possiamo superare la dicotomia fra presenza epistemologica e presenza dislocata, allora è necessario mostrare in che modo l'analisi della presenza siffatta sia anch'essa soggetta a questa stessa analisi. *In primis*, Floridi tiene a distinguere l'analisi della presenza (ente) dall'analisi dall'Essere inteso metafisicamente, come qualcosa di scarso interesse per una filosofia forte. Già l'ente, quindi, che si costituisce come fondamento e, come visto sopra, non viene affermato in base alla sua presenza, bensì come mediazione razionale. L'ente informazionale, di cui parla Floridi, viene esibito come "presente", presupponendo che la determinabilità dell'osservabile sia un elemento dialetticamente scisso dalla presenza stessa. L'ente ed il suo contenuto semantico emergono posteriormente alla loro affermazione e perciò dopo la valutazione funzionale della loro esistenza.

Le questioni poste fino ad ora sollevano nuovi quesiti: chi è il *subjectum*, ovvero il soggetto, che vede questa presenza? A sua volta è una presenza? In che modo si possono postulare concetti al di fuori di questo? Il problema sollevato apre a due vie ulteriori: 1) la presenza è luogo dell'intero semantico oltre il quale non possiamo andare e, pertanto, bisogna ammettere la possibilità che il modello della presenza sia fallibile, in quanto la stessa analisi, poiché è un modello possibile, è soggetta a modificazione; 2) oppure l'analisi della presenza è esterna a essa, tale da supporre che vi sia una presenza, eccedente l'ente analizzato, un ente maggiore che vede l'ente minore, dunque, una analisi dell'oggetto che dice di più dell'oggetto analizzato. Il *definiens* è maggiore del *definiendum*.

Questa analisi se posta a un determinato LdA non elimina il problema, ma lo complica. Nel primo caso non è mai possibile uscire dall'analisi parziale del fenomeno, mentre nell'altro caso

dobbiamo ammettere che l'analisi del fenomeno non è esterna, ma mediata da colui che pone l'analisi. Nel secondo caso, avendo una scissione concettual-oggettuale, l'oggetto analizzato non compete *in toto* alla presenza, incorrendo in un ingenuo realismo interpretativo.
La presenza, essendo sempre in relazione a sé stessa, si pro-getta sempre mediante il trascendimento del suo contenuto. Inoltre, il contenuto è l'oggetto intenzionale e immediato dell'esperienza.
Se la risposta alla prima critica conferma che il problema non abbia senso di porsi, in quanto la risposta è contemplata a un LdA diversificato, si dovrà, in ogni caso, giustificare perché tale diverso LdA può avere valore ontologico *super partes*, considerato che l'osservato deve esserci oggettivamente a priori dell'analisi, altrimenti non parleremmo di nulla. Ossia, in quanto c'è qualcosa e questo qualcosa è analizzabile se e solo se utilizziamo un determinato LdA con le sue relative proprietà, questo qualcosa è sempre determinato, mediato razionalmente.
Possiamo quindi porre questa obiezione: perché questo "qualcosa" non viene definito meglio nella sua essenza, ma continua ad essere proposto a partire dalla sua "presupposizione" e quindi letto già a un determinato LdA?
Se il capitale informazionale semantizza qualsiasi cosa, il discorso di Floridi rischia di diventare un "relativismo costruzionista", in cui ciò che si va a costruire è una catena relazionale di concetti che non riescono mai a de-finirsi, ma solo a dichiararsi estensionalmente. Se si dichiara la non assolutezza del modello (possibilità del modello), allora questo stesso potrà essere aperto, sì, alla sua modificazione ontologica perpetua, alla costante re-ontologizzazione, ma dovrà anche essere disposto ad accettare che la sua stessa valenza può essere soggetta a modificazione, sostituzione o addirittura annullamento, non per forza mediante un artefatto logico similare, bensì attraverso qualcosa di totalmente nuovo, non importa se migliore o peggiore.
Concludendo, se ogni analisi è posta a un determinato LdA e in forza di ciò si afferma la validità di tale analisi, anch'essa a un determinato LdA, allora l'analisi deve ammettere di essere una parziale. Il discorso sulla "osservazione di successo" (OS), posta nel secondo capitolo del saggio in questione, esaspera questa dinamica, poiché ipotizza si possa vedere ciò che è altro dall'analizzante non superando la semplice aporia sopra descritta. Per Floridi il flusso delle informazioni ad un dato LdA permette che un ente sia presente in uno "spazio remoto di osservazione" (che chiama SRO) soltanto se è anche osservato con "successo" ad un determinato LdA in quanto portatore di proprietà (sistema informativo). Tale applicazione degli LdA permette che vi sia una valutazione della presenza e il suo relativo impegno ontologico eludendo così errori di valutazione di portata soggettivistica, cioè: fenomenologici o psicologistici.
Se la filosofia dell'informazione, in quanto *design*, desidera propendere per questa via, si metterà sicuramente sempre in gioco e modificherà il proprio "capitale semantico" Una filosofia radicale dovrebbe riflettere in modo critico il valore della "presenza", così come si è provato molto brevemente a delineare.

4. La filosofia come pensiero

"L'unico vero metodo in filosofia è pensare", afferma Floridi (2020, p. 43): questo è il "compito" filosofico, in cui il pensiero viene assimilandosi al concetto di *design*, pensiero poietico. Ebbene, tornando al concetto di "pensiero", per la tradizione idealista e neoidealista l'esercizio del pensare è stato sempre sinonimo di pensare la Verità oltre che identificato al concetto di metodo e sistema. Il pensiero nella tradizione neoidealista è un pensiero mai scisso dal suo essere attuale, in atto. È possibile trovare svariate consonanze sia teoretiche che metodologiche tra il discorso di Floridi, che segue il concetto di produzione come *poiesis* e il discorso attualista di Giovanni Gentile secondo il quale c'è analogia tra il l'azione, quel fare che è indispensabile per la conoscenza e il pensiero stesso.

Secondo il filosofo Gentile, mediante il pensiero, che è sempre dinamico e localizzato, è possibile risolvere l'antinomia kantiana; è interessante l'analogia tra l'attualismo gentiliano e l'esercizio filosofico suggerito in *Pensare l'infosfera*, in cui vi è un obiettivo similare.
Floridi afferma, come già detto, che l'errore epistemico sta nella riduzione di ogni possibile analisi del reale alla soggettività. Gentile propone in svariate opere una indagine speculativa su come si possa evitare la fallacia oggettivo-interpretativa e il limite dialettico nell'indagine del reale.
Il filosofo italiano suggerisce che la realizzazione oggettiva di un qualcosa sta nel pensiero di quel qualcosa, vista come operazione consequenziale, ma soggetta alla convinzione che l'alterità rispetto al pensiero stia fuori dalla struttura di questo:

> Un pensiero altrui, pur volendolo pensare come altrui, non possiamo pensarlo se non pensandolo come pensiero [...] facendolo nostro. [...] pensare un pensiero (o porre il pensiero oggettivamente) è realizzarlo; ossia negarlo nella sua astratta oggettività per affermarlo in una oggettività concreta (Gentile, 1913, p. 184).

Questo vizio speculativo costituisce una catena di errori basati sulla convinzione che il pensiero possa essere il *subjectum* per esperire qualcos'altro da sé di qualsivoglia forma. Nel secondo momento speculativo della dialettica gentiliana, e consequenzialmente al primo, la determinazione dell'alterità come comprensione delle proprietà di questa e come ogni possibile altro dal pensiero permette che vi sia una lenta negazione dell'astratto. Il pensiero si riempie della consapevolezza crescente che l'oggetto assume in esso (modellazione):

> Questo secondo momento, reso possibile dal primo, se annulla l'attualità del pensiero altrui, o nostro e non più nostro, l'annulla in un nuovo atto di pensiero; per cui l'oggettività nuova [...] conferita a cotesto pensiero [...] è realizzata in funzione del nuovo pensiero, nostro attuale; ed è membro organico dell'unità immanente di questo (Gentile, 1913, p. 185).

Si veda che per Gentile non vi è solo una descrizione logica del processo epistemico, ma vi è anche un processo prasseologico[5]: è parte della consapevolezza logica nei momenti del pensiero che si realizza la realtà di questo stesso pensiero. Anche per Floridi il raggiungimento/consapevolezza di ogni oggetto è all'interno della prassi speculativa ed è quindi genetico, si evolve con la consapevolezza del capitale semantico che si porta dietro.
Oltretutto, non avrebbe senso il tentativo di fuoriuscire da questo orizzonte, poiché se il dicibile è tale all'interno di un orizzonte semantico già definito, allora il pensiero (ente determinato) si configura come *designer*. Non si tratta, quindi, di voler ridurre tutta la realtà all'Uno del pensiero, ma al contrario mostrare che la totalità del dicibile è tale perché è esperito secondo un modello ontologicamente inscindibile dal pensiero in atto[6].

5 Per l'importanza di questi concetti legati alla filosofia della prassi è d'obbligo citare sia: *Teoria generale dello spirito come atto puro* e *Genesi e struttura della società*, entrambi contenuti in Gentile, G., Severino, E., & Cicero, V. (2015). *L'attualismo*. Milano: Bompiani.

6 Oltre alla filosofia gentiliana vi sono stati svariati altri autori nostrani a partire dai quali si son sviluppate riflessioni sulle innumerevoli implicazioni problematiche del concetto di presenza, del rapporto fra pensiero e realtà o del pensiero ed i suoi limiti. Ne cito solo alcuni relativi al panorama italiano che ritengo rilevanti: Calogero, G. (2015). *Filosofia del dialogo*. Brescia: Morcelliana; Carabellese, P. (1948). *Critica del concreto*. Firenze: Sansoni; Spirito, U. (1948). *Il problematicismo*. Firenze: Sansoni; Bacchin, G. R., & In Castegnaro, G. (2017). *Theorein*. Roma: Aracne; Severino, E. (1981). *La Struttura originaria*. Milano: Adelphi.

Floridi sostiene che "il capitale semantico digitale inizia a fare la differenza nel nostro stesso processo di semantizzazione" (Floridi, 2020, p. 117), il capitale semantico è quindi in uno stato procedurale di affermazione, in una *dynamis* come atto, come movimento. Anche per Gentile non si può mai trascendere l'atto, non si può guardare come un "fatto" qualcosa di compiuto, visto dal di fuori del pensiero. Un'altra analogia gentiliana assimilabile con l'analisi che Floridi propone del processo di semantizzazione riguarda la dichiarazione delle risorse disponibili per il compimento delle operazioni utili alla risoluzione di problemi. Se questo processo viene visto come *energheia* illimitata, per la produzione di domanda aperte, al di fuori delle quali non è possibile uscire se non come accordo informato, allora il pensiero come interfaccia andrebbe inteso, come sosteneva anche Gentile, in quanto fare, come atto creativo. L'attualista italiano, contestando l'impostazione metafisica, propose di evitare l'errore trascendente dell'assoluto come un qualcosa di già "dato". Ancora, Floridi, così come Gentile, condivide l'errore della tesi platonica secondo la quale un approccio valido alla conoscenza non può essere *mimesis*, ma deve essere *poiesis*, quindi non imitazione di un già dato, ma produzione di artefatti. Inoltre, per Gentile l'essere non può negarsi dialetticamente in favore di un altro essere perfettibile, l'essere e la presenza non possono migliorarsi col fine di corrispondere ad un vero già dato. Gentile contesta della dialettica platonica esattamente questa ammissione, in cui la realtà che si vuol conoscere è presupposta al pensiero e con esso la concezione della natura e dell'eternità delle idee. Per Gentile quella positività delle idee, che Platone ambiziosamente cercava, sta nella realizzazione della realtà di cui l'idea stessa è principio. Chiarificando, se l'idea è comprensione della "cosa" allora deve essere prodotta (*poiesis*) dall'idea stessa. Direbbe Gentile che "il pensiero che è vero pensiero, deve generare l'essere di cui è pensiero" (Gentile, 1916, p. 98), in questo modo per l'attualista si può scorgere il vero significato del cogito cartesiano. In Gentile emerge con forza l'importanza di una costruzione genetica e poietica della storia.
Una ulteriore analogia può esser trovata tra l'accumulo del cosiddetto "capitale semantico" di cui parla Floridi e nel come Gentile intende l'analisi della realtà mediante ciò che chiama "autocoscienza". Il sapere genetico della storia della filosofia, che Floridi definisce "*progetto umano* [corsivo suo]", rasenta quel "momento dialettico" come "incremento della nostra intelligenza" (p. 253) di cui Gentile parla in *Teoria generale dello spirito come atto puro*. L'intelligenza come capitale semantico per l'attualista non è già acquisita, tale per cui non necessita di attività ulteriori, ma abbisogna di costante incremento (consapevolezza) per diventare "nuovo sapere" (p. 253). Passato e futuro divengono reali e concreti se visti da quell'intelligenza che li rende attuali, dal pensiero in atto[7].
L'esercizio del pensare l'atto del domandare come esercizio rigoroso e come realtà integrale può essere utilizzato in forza del superamento del pensiero come delimitazione ad un esclusivo contesto di riferimento o ad una interpretazione della domanda, ad una interfaccia possibile. Andrebbe considerata la forza speculativa in seno all'ontologia del domandare, come autocoscienza, come indispensabilità del dato, quindi come posizione irriducibile ed imprescindibile.

5. La filosofia come rischio: domanda e risposta

Quella di Floridi è senza dubbio una filosofia democratica e pluralista, che riesce a creare un ponte fra innumerevoli discipline, riuscendo ad evitare un certo vizio logico verificazionista, un approccio realista o empirista delle diverse filosofie contemporanee, che si interrogano sul corretto modo di orientarsi all'oggetto speculativo.

7 Riguardo all'Atto gentiliano come metodo e come problema, oltre che un breve approfondimento riguardante il neoidealismo gentiliano e gli sviluppi successivi nella filosofia italiana si potrebbe consultare il recente volume Manna, L., & Iosi, M. (2018). *L'eredità di Giovanni Gentile nella filosofia italiana.*Roma: Aracne.

Il suo scopo sembra esser quello di riportare sotto un'unica bandiera la filosofia nel suo statuto primario, come conoscenza e come primato del pensiero. Ciò che emerge dal discorso di Floridi non è il *design* come fine ultimo, ma come nuovo fine da cui iniziare a comprendere i limiti che la filosofia odierna ha assimilato e mantenuto dopo vizi postmoderni, d o altre forme di pensiero riduzionista.
La filosofia dell'informazione offre una valida prospettiva rispetto a un modo di porre domande filosofiche, esente da errori empirico-logici, ma mantiene altresì il rischio di fermarsi solo all'inizio della sua formulazione principale: nella comprensione della domanda e della risposta. Nonostante lo stesso Floridi a più riprese dichiari il suo come progetto iniziale di filosofia dell'informazione (p. 141). Poiché l'errore è insito nel cattivo domandare, nella formulazione del principio del rispondere c'è un importante bivio da tenere in considerazione. La filosofia della tradizione, da Aristotele a Kant, passando per Hegel o dalla filosofia italiana neoidealista o neotomista, ha mantenuto comunque alto il valore della Verità come fine primo ed ultimo. Inoltre, il bivio che si presenta nel discorso di Floridi può essere pericoloso, perché può portare la filosofia a riscoprirsi forte a partire da ciò che già possiede e a ricoprire il buono della tradizione del pensiero aletico. L'alternativa è che questo sistema possa divenire pura modellazione di concetti, col rischio di trasformarsi in pura relazionalità e portare l'oggetto stesso dell'indagine a costituirsi come catena concettuale indefinita. Una filosofia forte deve dare risposte assiologiche per mantenere il suo statuto filosofico, non deve temere, al contrario di quanto Floridi sostiene, di parlare delle Verità ultime.
La Verità nella tradizione filosofica dai Greci al nostro tempo dovrebbe ancora essere tematizzata intorno al concetto di intero. Proporre una visione dell'intero semantico democraticamente valido, ma mai de-terminato, conduce il pensiero all'autoreferenzialità del suo pro-getto, senza mai costituirsi fattualmente. Non è sbagliato partire dal valore che ogni risultato è progressione. Anche Vico e Croce condividerebbero la virtù genetica e dinamica del pensiero, vedrebbero la progressione storica, orientata al futuro dell'accrescimento come punto di partenza continuo per una costruzione più forte del pensiero.
Ciò che potrebbe apparire come errato dell'immagine che Floridi dà del pensiero è però questo assunto recuperato dalla logica e dalle teorie computazionali: la presupposizione che ogni risultato sia un risultato aperto al dubbio e alla sua riformulazione/costruzione/orientamento. In questo la sua filosofia incorre nel pericolo di diventare l'ultimo stato di una filosofia del possibile, una filosofia aperta solo alle domande, ma mai alle risposte.

RIFERIMENTI BIBLIOGRAFICI

Bacchin, G. R., & In Castegnaro, G. (2017). *Theorein*. Roma: Aracne;
Bontadini, G. (2009). *Metafisica e deellenizzazione*. Milano: Vita e pensiero;
Calogero, G. (2015). *Filosofia del dialogo*. Brescia: Morcelliana;
Carabellese, P. (1948). *Critica del concreto*. Firenze: Sansoni;
Deleuze, G., & Guglielmi, G. (ed.) (1971). *Differenza e ripetizione*. Collezione di testi e di studi. Bologna: Il mulino;
Floridi, L. (2008). *The method of levels of abstraction*. *Minds and machines*, 18(3), 303-329. DOI:10.1007/s11023-008-9113-7
Floridi, L. (2009). *Infosfera: Etica e filosofia nell'età dell'informazione*. Torino: G. Giappichelli;
Floridi, L. (2012). *La rivoluzione dell'informazione*. Torino: Codice;
Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.
Floridi, L. (2013). *The philosophy of information*. Oxford: Oxford University Press.
Floridi, L. (2017). *La quarta rivoluzione: Come l'infosfera sta trasformando il mondo*. Milano: Cortina;
Floridi, L. (2020). *Il verde e il blu: Idee ingenue per migliorare la politica*. Milano: Raffaello Cortina Editore;

Floridi, L. (2020). *Pensare l'infosfera: La filosofia come design concettuale*.
Gentile, G., Severino, E., & Cicero, V (eds.). (2015). *L'attualismo*. Milano: Bompiani;
Gentile, G. (1954). *La riforma della dialettica Hegeliana*. Firenze: Sansoni;
Gentile, G. (1987). *Teoria generale dello spirito come atto puro*. Firenze: Le Lettere;
Heidegger, M., & In Chiodi, P (ed.). (1969). *Essere e tempo: L'essenza del fondamento*. Torino: Unione Tip. Editrice Torinese;
Manna, L. (2018). *L'eredità di Giovanni Gentile nella filosofia italiana.* Roma: Aracne;
Severino, E. (1981). *La Struttura originaria*. Milano: Adelphi;
Spirito, U. (1948). *Il problematicismo*. Firenze: Sansoni.

ROBERTO MORDACCI
International Research Centre for European Culture and Politics (IRCECP), Vita-Salute San Raffaele University
mordacci.roberto@unisr.it

SAPERE, PROGETTO, AZIONE. SUL PRIMATO DEL PRATICO

abstract

Recent proposals to think about philosophy as conceptual design (Floridi, 2019) fail to understand the scope and method of philosophical thinking. The goal of philosophy is truth, but truth has a double direction: towards theory and towards practice. The design is in the middle and it lacks the criteria in order to evaluate itself, since its truth depends on the theory on which it is based and on the practical use of its functions. Furthermore, design lacks critical ability, since only the horizon of existence, in social and political terms, offers the possibility to judge about the validity of a project.

keywords

design, philosophy of design, ethics, critical theory, praxis

Phenomenology and Mind, n. 20 - 2021, pp. 176-182
DOI: https://doi.org/10.17454/pam-2015
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Filosofia come prassi

L'antica suddivisione stoica della filosofia in Logica, Fisica ed Etica conserva la propria pertinenza. Si può intendere la logica come l'ambito teoretico – che lo si affronti per via metafisica o di analisi del linguaggio; si può riconoscere nella fisica l'orizzonte della filosofia della natura o dell'ontologia, analitica o fenomenologica; e si può intendere l'etica come l'ampio spettro della filosofia pratica, dalla filosofia della tecnica alla morale e alla politica. L'ambito che sembra meno facile ricondurre a questo schema è l'estetica, che tuttavia costituisce una sorta di mediazione fra natura e prassi, fra rappresentazione e azione. La tripartizione disegna dunque il territorio ove si pongono le "domande fondamentali" che, parafrasando Kant, chiedono: che cosa significa pensare? Che cosa esiste? Che cosa fare?
Si pone tuttavia una questione architettonica decisiva: qual è il fine – e a un tempo il movente – che orienta l'indagine filosofica? La guida del pensiero spetta alla formalizzazione inattaccabile, alla comprensione ordinata dell'intero, o alla pratica riflessiva?
Potrà sorprendere constatare che tutti i grandi sistemi, anche quelli che assegnano alla speculazione il grado più alto, concepiscono come *scopo* della filosofia la pratica del pensiero e, più specificamente, la vita filosofica come esercizio costante della riflessione critica.
Per Socrate, è la vita senza ricerca che non merita di essere vissuta; per Platone – su cui torneremo – la contemplazione è dialettica e la verità appare nella relazione; per Aristotele, l'orizzonte in cui si inscrive la stessa vita contemplativa è la vita buona; e, per tacere d'altri, non è un caso che il sistema di Spinoza sia un'etica, che Kant assegni il primato alla ragion pratica e che persino l'idealismo hegeliano veda nel compimento dello spirito la realizzazione della libertà. La filosofia non ha mai considerato la mera dottrina o il solo metodo lo scopo *finale* del pensiero. Non si spiegherebbe, altrimenti, l'inveterato anti-dogmatismo dell'interrogare filosofico, né l'incessante rinnovarsi dei metodi come vie per una destinazione che ha sempre i contorni di una città, terrena o celeste. Piuttosto, la filosofia non ha mai inteso distaccarsi del tutto dal movimento dell'esistenza, persino quando si è data come compito la semplice chiarificazione concettuale: quale scopo può avere la terapia linguistica di Wittgenstein se non quello di ridurre l'incomprensione, la confusione mentale e l'alienazione degli esseri umani?
In tal senso va interpretato il primato che Platone assegna all'uso, su cui si sofferma criticamente Luciano Floridi in un suo recente libro (Floridi 2019). Riassumendo le argomentazioni del *Cratilo* e dell'*Eutidemo*, Floridi riconosce in Platone la tesi della priorità della prassi: non chi crea i manufatti, bensì chi agisce conosce a fondo le cose, perché ne valuta

l'adeguatezza nell'uso. A questa prospettiva, che Floridi chiama «conoscenza dell'utente»[1], egli oppone la «conoscenza del costruttore», ovvero di chi progetta e realizza un artefatto, che si tratti di uno strumento tecnico o di un sistema di informazioni, richiamandosi variamente al punto di vista del *Novum Organum* baconiano, al *verum factum* di Vico, alla filosofia critica kantiana e al costruzionismo derivato dalla teoria dell'informazione. La conoscenza corrisponde alla capacità di progettare apparati informativo-concettuali che rendono possibili risposte alle domande fondamentali. Tali risposte risultano dalla suddivisione dei grandi problemi in unità più facilmente affrontabili (minimalismo), da operazioni di astrazione controllata (metodo dei livelli di astrazione) e dalla "costruzione" di artefatti concettuali (modelli) internamente coerenti (Floridi 2019, pp. 113-121).
In questo consiste l'idea della filosofia come "design concettuale", di cui Floridi si fa portatore, ovvero «l'arte di identificare e chiarire le domande aperte e di disegnare, proporre e valutare risposte convincenti e chiarificatrici» (p. 53). Non sembra che Floridi voglia sostituire il disegno concettuale all'intero lavoro filosofico (egli riconosce che prassi e progetto sono complementari), ma sembra volergli affidare il primato anzitutto rispetto alla preminenza dell'informazione come "ambiente" della vita contemporanea (quella che egli chiama "infosfera").
Questa tesi offre l'occasione di chiarire il senso del primato che la tradizione ha assegnato alla prassi e che occorre riaffermare, oggi, proprio per orientare le potenzialità della società dell'informazione. Rispetto a tali potenzialità, l'idea del disegno concettuale non offre né i criteri per un giudizio critico (anche in termini sociali) né uno scopo sufficientemente comprensivo per il lavoro filosofico.
La verità ha due lembi: quello epistemico, universale e astratto quanto indubitabile, raccolto intorno alla logica dei concetti *a priori*; e quello pratico, esistenzialmente valido, universale nelle sue massime e individuale nei suoi dettagli perché immerso nell'agire. Fra i due si colloca il progetto: non più pura *theoria* ma non ancora concreta *praxis,* il progetto è essenzialmente *téchne*, attività che conduce alla produzione (*poiesis*) ma non all'azione di cui è intessuto il vivere (*bìos*). Senza il disegno, il vivere non dispiega le sue potenzialità, ma a decidere della sua validità sono, da un lato, le verità teoriche su cui si basa (un progetto basato su teorie false precipita), dall'altro i fini pratici che esso deve servire, il giudizio sui quali spetta all'aristotelica "scienza dei beni umani" – ossia l'etica. Ora, giacché sull'efficacia del disegno è naturalmente regina la tecnica, in base ai criteri dell'efficacia e dell'efficienza, la filosofia si occuperà soprattutto dei due lati estremi, quello della verità degli assunti teorici e quello del bene pratico, per attingere una visione d'insieme in cui anche il disegno prenda un senso.
Circa la tesi per cui si conosce pienamente solo ciò che si fa (*verum factum*), occorre un chiarimento. Il *verum factum* non è tale solo perché è qualcosa di cui siamo stati capaci. Non conosciamo pienamente ciò che facciamo, men che meno ciò che costruiamo. Lo dimostra la storia di molti progetti eclatanti poi svelatisi disastrosi nell'uso: dall'arredamento di design impraticabile agli ascensori dalla pulsantiera cervellotica, dai ponti scivolosi a quelli tragicamente incapaci di resistere all'uso. Al al di là del *fare* c'è un *agire* che vive costantemente le smentite dell'esistenza, che è esposto al vaglio ultimo del reale.
Verum factum – anche nella *Scienza Nuova* di Vico – significa che il vero del *factum* appare nella critica che il fare riceve dall'agire, dalle forze che scatena, dal giudizio della storia, dal

1 Vale la pena di precisare che l'idea di "utente" (del tutto anacronistica se riferita a Platone) indica un utilizzatore cui i costruttori si rivolgono affinché faccia *consumo* dei prodotti, in un'ottica sostanzialmente commerciale. Platone pensa piuttosto a colui che vive il βίος πρακτικός (*bios praktikòs*), la vita attiva, il quale giudica gli artefatti in rapporto ai reali bisogni umani e non si limita a immaginarli. La tesi platonica va letta nel quadro della teoria del mondo come un grande organismo vivente, esibita nel *Timeo* (in particolare, 34 b – 37 d).

fallimento o dal buon esito mostrato dal susseguirsi dei fatti. Fare qualcosa significa sapere come farla, ma non *perché* farla. Un argomento ben congegnato può servire come pretesa giustificazione di un'impresa criminale: il criterio con cui giudichiamo quest'ultima è *al di là* del "disegno" e risiede nei *criteri di valore* di ciò che si fa. Per quanto ardui, sono questi ultimi quelli su cui la disputa filosofica ha veramente ragion d'essere. Non è il primato dell'utile. È il primato del pratico: un giudizio *critico* che riguarda tanto il giusto quanto il bene, senza riduzionismo e senza sottomissione agli obiettivi tecnici, quando non commerciali, cui soggiace il progetto.
Dunque, possiamo fissare un primo punto: la filosofia è pratica del pensiero ed è critica in quanto originariamente inscindibile dall'esistenza, dal *Lebenswelt* e dalla vita stessa. Il suo senso non è disegnare risposte eleganti e facilmente replicabili ai fini della diffusione delle idee; il giudizio cui mira la filosofia è quello di un'*adaequatio* con il reale, ovvero con l'esperienza dell'essere umani. Una risposta basata su un buon disegno non è ancora una buona (una vera) risposta. Nell'incertezza di questo vaglio, perché difficile è la comprensione del reale, un buon disegno è apprezzabile per condurre il pensiero con chiarezza, ma non è tutto. Le risposte fumose, confuse, contorte, idiosincratiche sono certamente da rigettare o riformulare. Ma eleganza e replicabilità *non sono un metodo*. La stessa scienza le apprezza, ma non si basa su di esse. La scienza, come ogni sapere possibile, si fonda sulla non contraddizione e sull'esperienza: queste sono *il metodo*. Il disegno è sostenuto da questi fondamenti ma non ne è la sede né il luogo ultimo. È il luogo in cui tali basi si attivano e si reggono mutuamente, ma solo chi, oltre il disegno, sa leggere sia i suoi presupposti sia i suoi effetti sa ciò che è veramente in gioco.

2. Filosofia del design

Al più, dunque, il "design concettuale" è solo una parte del lavoro filosofico. Ma che cosa si intende, propriamente, per *design*? Floridi sembra usare il termine come una metafora suggestiva che esprime l'idea del "costruzionismo": da un lato, ogni costruzione (argomentativa) risponde a un progetto e – come nell'ingegneria – richiede calcoli (logici) e disegni (tecnici); dall'altro, il design richiama l'idea del pensiero per immagini (le mappe concettuali) e della plastica rappresentazione delle teorie (gli schemi). Di qui la definizione, sopra richiamata, del design concettuale come «arte di [...] disegnare, proporre e valutare risposte convincenti e chiarificatrici» (Floridi 2019, p. 53). Allo stesso modo, Floridi sostiene che «la conoscenza non *descrive* né *prescrive* come il mondo sia, ma l'*iscrive*» (p. 101) e che «il *costruzionismo* ritiene che la conoscenza sia acquisita attraverso la creazione del corretto genere di *artefatti semantici* o, per dirla in altri termini, di modellazione dell'informazione» (p. 110).
Come è noto, il design ha inizio con la Rivoluzione industriale (De Fusco 2009) e su di esso si è esercitata da subito un'ampia riflessione teorica, che Floridi non considera. La filosofia del design offre spunti importanti alla filosofia come design concettuale. Così, per esempio, Vilém Flusser ricorda che «Il design rappresenta il punto in cui convergono grandi idee che, derivando dall'arte, dalla scienza e dall'economia, si sono arricchite e sovrapposte in modo creativo l'una con l'altra» (Flusser 2003, p. 5). La storia del termine, conservata nella polisemia della parola "disegno", indica il riferimento a un'intenzione non sempre commendevole ("un disegno criminale") e anzi è in generale legata al tentativo di tramare un "inganno": «l'intenzione (design) alla base di tutta la cultura è quella di trasformarci subdolamente da semplici mammiferi condizionati dalla natura in artisti liberi» (Flusser 2003, p. 4).
La dimensione estetica e quella economica sono decisive per il design. Come osserva Glenn Parsons, le teorie di Adolf Loos e Walter Gropius offrono idee chiave per la comprensione della razionalità pratica, per definire l'intreccio fra estetica e funzione, per indagare gli ambigui rapporti fra innovazione, aspirazioni artistiche e consumo di massa (Parsons 2015). Nel design

storico, i fattori decisivi sono *quantità, qualità e basso prezzo* (De Fusco 2012, pp. 21-33): la standardizzazione, infatti, consente la creazione di linee di montaggio uniformi; la qualità (o la sua apparenza) si esprime nel rapporto tra forma e funzione – il campo in cui la creatività ha più libero gioco; il basso prezzo è essenziale allo scopo economico, ovvero il largo consumo e la realizzazione di ampi profitti. In quanto evoluzione dell'artigianato nella società industriale avanzata, il design è naturalmente assorbito dagli scopi del mercato, benché mantenga almeno due elementi di resistenza alle logiche meramente commerciali: da un lato, la tendenza a rendere disponibili al più ampio numero di persone strumenti di alleviamento della fatica e di piacevolezza del vivere[2]; dall'altro, la ricerca estetica e il rapporto costante con il gusto dell'epoca, tanto da essere capace, talvolta, di creare le mode anziché inseguirle.
Queste analisi – cui qui purtroppo si può solo accennare – mostrano che il disegno è il luogo della mediazione, o più precisamente della transizione fra teoria e pratica. Tuttavia, proprio per questo esso non può rispondere alle domande fondamentali. Dunque, un progetto o un programma, un software, non è filosofia. Il design concettuale è solo una metafora, che cerca di intercettare l'onda di una moda, quella del *design thinking*, con cui si indica la sintesi di analisi e creatività per lo sviluppo di prodotti innovativi. Non conviene però ai filosofi inseguire le mode del linguaggio, tanto più che, volendo indicare l'esercizio di un pensiero rigoroso e creativo insieme, nella tradizione filosofica i termini certo non mancano, per esempio (fra i molti) l'idea di *Denkungsart* in Kant. Il quale, per altro, ammoniva persino se stesso contro la tentazione di cercare termini alla moda per idee già note: «Escogitare nuove parole là dove la lingua non manca affatto di espressioni atte a indicare dati concetti è uno sforzo puerile di distinguersi dalla folla applicando una nuova pezza su un vestito vecchio, e non in virtù di pensieri nuovi e veri» (Kant 1992, p. 111).
A giudicare il progetto, visto che il suo obiettivo è servire la prassi, è quest'ultima. E questa include i *propri* criteri normativi, che non sono soltanto quelli economici. Tali criteri hanno un rilievo sociale, etico e politico, di cui anche la teoria del management è ormai pienamente consapevole. Il tentativo di ricondurre a unità la nozione di *valore*, per esempio, spetta proprio alla filosofia come capacità di includere il design, fra le altre cose, nella comprensione critica delle pratiche sociali.
L'orizzonte del vero è unitario, ma conoscere e agire sono intrecciati in modo complesso. Il vero speculativo è l'oggetto della componente più elevata delle capacità umane in termini di astrazione, ma il vero pratico, ovvero il vivere bene, è l'oggetto dell'intera vita personale. L'esistenza ha dunque il primato, e precisamente l'esistenza delle persone come individui e come comunità nella storia. Il "disegno" non contiene in sé alcun elemento per giudicare né della verità speculativa né di quella pratica. Dunque, se la filosofia ha da occuparsi dell'esistenza, e non di astrazioni, essa ha da occuparsi degli "usi", ovvero della prassi, non dei progetti. Come appunto sosteneva Platone.

3. Prassi e critica

È per questo che soltanto la prassi è autenticamente critica. Non solo nel senso marxista, che pure è pienamente attuale, ma già nel senso platonico, poi aristotelico e infine kantiano del primato del pratico. Solo ciò che rileva per l'intero della vita delle persone, che si gioca nel ruolo della libertà, è critico.
In altri termini, la critica è possibile solo a partire dal rapporto con l'esistenza. Il progetto struttura i bisogni in funzione di esiti tecnici e commerciali. Lo fa usando il meglio delle conoscenze a disposizione, cioè essenzialmente l'*informazione*, la sua quantità (*big data*), la sua

2 «La forza etica dell'industria moderna è stata un'idea di democrazia dei costumi. [...] È stata un'idea vincente che ha mosso e catalizzato un'intera società». (Manzini 1990, pp. 62-63).

qualità (affidabilità, bellezza), la sua modalità (connessioni necessarie o possibili). Ma una filosofia che si esaurisca in una filosofia dell'informazione è del tutto priva di capacità critica: soggiace all'uso che dell'informazione si fa nella produzione di artefatti. Poiché il design vive di una fondamentale spinta commerciale, il pensiero che esprime sarà più facilmente preda della manipolazione dei bisogni a opera del mercato, piuttosto che essere veicolo di liberazione, come già aveva pienamente intuito Herbert Marcuse (Marcuse 1999). Dunque, nella società dell'informazione (o infosfera), una filosofia come design concettuale sarà di scarso aiuto alla critica dell'ingiustizia e della falsità. Come detto, tuttavia, lo stesso design dispone di elementi di resistenza alla deriva commerciale: gli oggetti di design riusciti si conservano ben oltre il loro impiego. La differenza essenziale fra un'opera d'arte e un oggetto di design è che della prima non è pensabile alcun *utilizzo*; ma, proprio per questo, quando del secondo è decaduto il valore d'uso, ne rimane in evidenza il valore estetico e, in taluni casi, il valore storico come veicolo di emancipazione (si pensi alla parabola delle auto "utilitarie", specie nel nostro Paese).

Il punto cruciale è che il design esprime una capacità critica solo quando una conversione della prassi sociale verso fini inclusivi ed equi guida il processo produttivo, compreso quello della produzione culturale. Tali fini sono la giustizia, la libertà sociale, il rispetto, la solidarietà. Di questi fini il progetto non decide. Anzi, il disegno può essere parte della gabbia d'acciaio e non uscirne mai. Lo dimostra anche la torsione negativa cui è stata sottoposta l'idea di utopia: se essa è solo il disegno di una società perfetta, si converte in distopia e in incubo totalitario. Se si vuole uscire dall'attuale crisi, che è di lungo periodo, il mero disegnare argomenti non basta. È certo fondamentale che si traccino con cura i contorni e le forme di un pensiero non confuso, coerente e consequenziale. Ma a questo mira l'intera tradizione filosofica, ben prima di ogni "design concettuale" e ben dopo che questo sarà dimenticato.

Possiamo dunque concludere: la prassi critica, incluso il pensiero come esercizio della critica è la sostanza e lo scopo della filosofia. Essa può includere il disegno argomentativo, l'informazione, l'ontologia dell'informazione, se si vuole, e persino un'etica dell'informazione. Tuttavia, queste ultime non traggono i propri principi da se stesse, bensì solo dalla *filosofia pratica*, ossia dall'etica, dalla politica e dall'estetica. Persino la filosofia speculativa, per non essere solo il sogno di un visionario, è sottoposta a un fondamentale esame di realtà, all'aderenza alla fenomenologia dell'esistenza umana: se si perde in assiomi logici privi di rapporto con l'esperienza, essa si ferma all'interdetto parmenideo contro il divenire e smarrisce ogni relazione con l'esistenza. Per questo, a Platone può ben perdonarsi il parricidio di Parmenide, perché senza di esso il pensiero si priva di ogni movimento e si svuota di ogni linfa vitale. La trama dell'essere non può essere «secca come un ramo». Solo le parole lo sono. Ma anche queste sanno sempre almeno dire «ciò che *non* siamo, ciò che *non* vogliamo». Il "disegno concettuale" non sa elevare questa protesta. È piuttosto come l'ombra del viandante troppo sicuro di sé, che «la canicola stampa sopra uno scalcinato muro» (Montale 1925).

RIFERIMENTI BIBLIOGRAFICI

Cacciari, M. (2020). *Il lavoro dello spirito.* Milano: Adelphi;
De Fusco, R. (2009). *Storia del design* (26ma ed.). Roma-Bari: Laterza;
De Fusco, R., (2012). *Filosofia del design.* Torino: Einaudi;
Floridi, L. (2019). *Pensare l'infosfera. La filosofia come design concettuale.* Milano: Cortina;
Flusser, V. (2003). *Filosofia del design.* Milano: Bruno Mondadori;
Kant, I. (1992). *Critica della ragion pratica* (1788). a cura di A.M. Marietti, Milano: Rizzoli;
Manzini, E. (1990). *Artefatti. Verso una nuova ecologia dell'ambiente artificiale.* Milano: Domus Academy;
Marcuse, H. (1999). *One-Dimensional Man. Studies in the Ideology of Advanced Society*. Boston:

Beacon Press, 1964; tr. it. di L. Gallino e T. G. Gallino. *L'uomo a una dimensione. L'ideologia della società industriale avanzata*. Torino: Einaudi;
Montale, E. (1925). *Non chiederci la parola.* in Id., *Ossi di seppia.* Torino: Piero Gobetti Editore;
Parsons, G. (2015). *The Philosophy of Design.* New York: Polity Press.

LUIGI VERO TARCA
Università Ca' Foscari di Venezia
tarca@unive.it

INFOSFERA, APERTURA E VERITÀ

abstract

According to the philosophical posture for which every denial is a problem, I have decided, since it lacks here the space to justify such a posture, to propose a dialogue between three fictional characters.
The first part of the dialogue relates to open and closed games. The value of closed answers (§1.1) is supreme (ultimate) because it resides in their undeniability meant as non-negativity. But the undeniability is conditioned, since it is such only insofar as one remains within a certain game. This (§ 1.2) applies both to logical calculation and to empirical evidence, and therefore in general to science. The consequence (§ 1.3) is that philosophical questions, as open, are then unlimitedly *open.*
The second part of the dialogue consists of a philosophical reading of the infosphere. It is possible (§ 2.1) to have an "apocalyptic" reading which, interpreting the infosphere as the final passage of technoscience, interprets it as the space in which even human mind is replaced by an artifice. Therefore (§ 2.2) the problem of "subjectivity" arises: what types of subjects are produced by the technique? And which types of new subjects are producing them?
In conclusion (§ 2.3), the possibility of an alternative reading of the infosphere is announced; it is based on a thought capable of "opening up" even with respect to negation, which up to now seems to be the horizon that thought cannot overstep

keywords

infosphere, technoscience, dialogue, openness, denial.

Phenomenology and Mind, n. 20 - 2021, pp. 184-201
DOI: https://doi.org/10.17454/pam-2016
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Il discorso proposto da Luciano Floridi (2020) è importante per un dibattito filosofico serio e attuale. Trovo indovinate espressioni quali "infosfera" e "design concettuale", come ritengo fondamentale il tema dei "Livelli di Astrazione". Potrei dunque aggiungere molte osservazioni atte a corroborare e arricchire quanto da lui detto, ma credo che in questa occasione sia più opportuno proporre all'Autore alcune questioni di fondo. Sentendo questo, chi legge sarà automaticamente indotto a pensare che io intenda ora manifestare il mio *disaccordo* (Floridi 2020, p. 12) con Floridi muovendo delle osservazioni *critiche* (obiezioni) almeno ad alcuni punti del suo discorso.

Ma il mio *dàimon* filosofico (un lontano parente, almeno a suo dire, del *δαίμων* socratico), avendo un po' l'idea fissa (lui la chiama platonica) del carattere onnipervasivo del bene (il positivo), mi invita ad astenermi, in sede filosofica, dal *negare*, e quindi dal *contrappormi* a qualsiasi posizione. Ciò per via degli effetti nocivi – io li chiamo *necativi* (da *nex*, *necis* = morte violenta, uccisione) – che la negazione comporta (a meno che sia passata attraverso il processo di autopurificazione). Oltre a ciò, ma di conseguenza, devo anche mostrarmi, almeno ai suoi (del *dàimon*) occhi, sostanzialmente distaccato da *ogni* affermazione, dato il rischio che l'affermare qualcosa venga automaticamente inteso come il negare qualcos'altro; rischio che ogni gesto comporta, in particolare poi se è di tipo linguistico e a maggior ragione se è di natura teorica.

D'altro canto, semplicemente *escludere* dalla conversazione filosofica ogni osservazione critica significherebbe sprecare una preziosa occasione di dialogo, tanto più che in questo caso lo stesso Autore invita il lettore a esprimere il proprio disaccordo (Floridi 2020, p. 12), ancorché "ragionevole" (Floridi 2020, p. 51); e giustamente, dal momento che le obiezioni a un discorso finiscono per rafforzarlo, almeno nella misura in cui esso resta confermato persino dal confronto con la sua negazione.

Mancando qui lo spazio per giustificare in maniera adeguata questa mia posizione filosofica in generale,[1] e quindi anche la sua parte riguardante la questione dell'infosfera, ho pensato, per risolvere il problema (come accettare il dialogo pur astenendomi dal negare?), di attribuire a dei personaggi fittizi (A, B e C) una serie di "possibili affermazioni" che quindi, essendo prive di un soggetto reale che ne sia il proponente, costituiscono semplicemente, per parafrasare

1 In relazione alla quale qui mi limito a indicare, a chi fosse interessato, i miei scritti: Tarca 2001; Tarca 2006; Tarca 2016.

liberamente Pirandello, "proposizioni in cerca d'autore". Esse vanno dunque chiaramente distinte dal "mio" discorso filosofico, il quale, almeno in questa sede, consiste semplicemente nella *proposta* di tali discorsi, ovvero nella loro *pro-posizione*. Naturalmente il fatto che io li proponga significa che ritengo giusto prenderli in considerazione, e questo dice pure qualcosa circa quello che io penso; tuttavia bisogna tenere presente che non solo essi costituiscono comunque una rappresentazione *indiretta* del "mio" pensiero, ma soprattutto che spesso le affermazioni sulle quali vale maggiormente la pena di riflettere sono proprio quelle più lontane dal nostro modo di sentire, e a volte addirittura quelle propriamente "eretiche". Insomma, il fatto che io ritenga importante prendere in considerazione determinate affermazioni è ben lungi dall'implicare che io *le faccia mie*, testimonia però la mia attenzione per la posizione filosofica che, avendo cura della verità intesa come interezza, mira a riconoscere il modo in cui questa si rispecchia in *tutte* le proposizioni, sia pure in maniera specifica in ciascuna di esse. Ciò detto, è ovvio, per altro verso, che molte battute presenti nel dialogo riflettono effettivamente, per quanto solo parzialmente, pensieri che sono il frutto di una *mia* elaborazione. Si badi, infine, che anche questo espediente compositivo – all'interno di un contesto antropologico, quale è il nostro, nel quale pensare/parlare equivale a negare, sicché ogni affermazione viene automaticamente vista come implicante, anche solo in forma implicita, una negazione (quanto meno la negazione della propria negazione) – è ben lungi dal risolvere automaticamente la questione filosofica che sto ponendo, tuttavia esso può agevolarne almeno una comprensione iniziale.
Il fine delle pro-posizioni che seguono è dunque quello di favorire una valutazione della proposta filosofica di Floridi, cioè una qualche forma di *giudizio* su di essa. Giudizio che però, intendendo realizzarsi, appunto, secondo *giustizia* (*jus*), dovrebbe far sì che tale *valutazione* si presenti anche come una *valorizzazione*, o una "ri-valutazione" (nella misura in cui fa seguito a osservazioni che potrebbero suonare in parte come critiche); essa dovrebbe infatti costituire una prima chiarificazione delle condizioni alle quali il suo discorso si presenta come dotato del carattere specifico (la verità) che solo può conferire ad esso quel valore e quella considerazione ai quali ogni discorso aspira. Questo almeno nelle mie intenzioni.

1. Giochi chiusi e giochi aperti

1.1. Il valore dei giochi chiusi

DIALOGO
(A) Vorrei proporvi una questione relativa alla distinzione tra domande chiuse e domande aperte.
(B) È un tema fondamentale, ma decisamente complesso.
(A) Per questo preciso subito (in estrema sintesi e banalizzando alquanto) che chiamo qui "chiuse" le domande rispetto alle quale vi è una risposta che non può essere "ragionevolmente" messa in discussione; per esempio: "Se Maria è alta 168 cm. e Paola è alta 171 cm., chi tra le due è più alta?". Chiamo "aperte", invece, le domande alle quali è possibile offrire risposte che, pur essendo contrastanti, sono ugualmente ragionevoli; per esempio: "Chi, tra Maria e Paola, è meglio invitare al cinema?".
(B) Ho capito. Qual è, dunque, la questione che intendevi proporci?
(A) Questa: "Le domande relative alla natura delle domande, e in particolare alla distinzione tra i due tipi di domande (chiuse e aperte), sono chiuse o aperte?".
(B) Io direi che se si assume, come accade anche nel discorso di Floridi, che vi sono domande aperte (Floridi 2020, pp. 23 ss, in particolare p. 25), allora le domande relative alla distinzione tra "chiuso" e "aperto" devono essere almeno in qualche misura aperte. Perché in caso contrario anche lo spazio "aperto" resterebbe in fin dei conti *determinato*, e quindi *definito*, da una logica chiusa, ma allora tutte le domande risulterebbero in qualche senso chiuse sicché domande aperte non potrebbero più essercene.
(C) È un tema davvero molto interessante, che solleva questioni estremamente delicate.

(B) Sì; tra le quali però a me pare fondamentale capire in che cosa consista il pregio peculiare, cioè in sostanza il *valore*, delle domande chiuse.
(A) Le domande chiuse, intese come ho detto, sono definite dal fatto che esse dispongono di una risposta tale che rifiutarla (negarla) è sostanzialmente sintomo di qualcosa che non va, cioè di una sorta di *patologia*: chi rifiuta/nega le *risposte chiuse* (quelle fornite alle domande chiuse) è "irrazionale, confuso o [...] testardo" (Floridi 2020, p. 26). Le domande aperte, viceversa, sono definite dal fatto che le risposte ad esse fornite (cioè le *risposte aperte*) possono essere discusse ("interrogare" (Floridi 2020, p. 27)) e anche rifiutate/negate ("disaccordo" (Floridi 2020, p. 26)) senza che questo costituisca per ciò stesso qualcosa di male (una patologia, un disturbo mentale).
Chiamo quindi "giochi chiusi" quelli passibili di risposte chiuse, e "giochi aperti" quelli caratterizzati da risposte aperte. Si badi che in entrambi i casi le risposte possono essere negate; ma, mentre nei giochi aperti la negazione *fa parte* del gioco (è *interna* ad esso), in quelli chiusi non vi è negazione all'interno del gioco, l'eventuale negazione è *esterna* ad esso.
(B) Possiamo dunque dire che il valore (il carattere *positivo*) delle risposte chiuse consiste nella loro necessità, o, forse meglio, nella loro *innegabilità*?
(A) Sì, possiamo dire così.
(B) Ma in che senso l'innegabilità è un valore?
(A) Nel senso che il "non-negabile", essendo libero dai danni arrecati dalla negazione, è *salvo rispetto al negativo*. E l'esser salvo rispetto al negativo è di per sé un valore (cioè un positivo) sul quale senz'altro sono tutti d'accordo. Questo conferisce *a priori* un valore *universale* all'innegabile, dal momento che, anche senza chiedere a ogni singolo individuo se egli è d'accordo o meno, possiamo far valere per tutti ciò che è innegabile senza temere che esso possa subire alcuna negazione (*contra*-dizione). Proprio questo conferisce all'innegabile un valore fondamentale dal punto di vista etico-antropologico e politico-sociale, appunto perché esso *a priori* può essere considerato *vincolante* per tutti ma in maniera *giusta*, dal momento che farlo valere per ciascuno (*erga omnes*), costituisce un gesto diverso da un atto di prepotenza-violenza, come accade quando quello *erga omnes* si trasforma in un *contra aliquem*, cioè quando ciò che *vale* per tutti in realtà *non ha valore* per qualcuno.
(B) Insomma, quello dell'innegabile è un valore supremo. Perché è un positivo *universale* (riconosciuto a priori da tutti i soggetti) e *perfetto* (salvo rispetto al negativo) in quanto consente al positivo di realizzarsi senza cadere nel negativo.
(A) Sì, però dicevamo che anche delle risposte chiuse si può dare la negazione, solo che questa risulta *esterna* al gioco. Ciò vuol dire che qualcosa si dà come innegabile esclusivamente all'interno di un determinato gioco, e che dunque il suo valore è parziale, *condizionato*: quello che all'interno di un certo gioco si presenta come un valore si può trasformare, all'interno di un altro gioco, in un disvalore. Una mossa vincente nel gioco della dama può essere perdente nella dama "al contrario", dove (visto che nella dama catturare la pedina in presa è obbligatorio) vince il giocatore che costringe l'avversario a catturargli tutti i pezzi. Insomma, le domande chiuse sono sì innegabili, ma solo a determinate *condizioni* (in particolare a condizione che si resti all'interno di un certo gioco).
(B) Nel linguaggio di Floridi si potrebbe forse dire che ogni discorso è condizionato dal Livello di Astrazione (LdA) al quale appartiene. Anche in questo senso il discorso sui LdA si rivela fondamentale.
(A) Sì, ma c'è anche qualcosa di più: *il semplice fatto di fare un certo gioco* può essere vissuto come un disvalore. E addirittura lo stesso fare un gioco *chiuso* potrebbe essere considerato un errore: qualcuno potrebbe farsi dare scacco matto apposta per far capire ai suoi compagni di gioco che egli preferirebbe andare all'aperto a giocare a pallone invece che stare a giocare *al chiuso* (*sic*!).
(B) Potremmo allora dire che ogni risposta chiusa in realtà è solo un momento di una dualità

irriducibile il cui altro momento è costituito dalla dimensione valoriale-soggettiva, quella del *soggetto* le cui esigenze vengono *soddisfatte* dalle risposte chiuse ...
(A) ... che è poi proprio la dimensione che istituisce la chiusura (limitazione/recinzione/isolamento) che rende appunto "chiuso" il gioco.
(B) Certo. Ho chiamato "irriducibile" questa dualità perché ciascuno dei due poli può darsi solo se si dà anche l'altro, così come un figlio può darsi solo se si dà anche un padre, e viceversa. Potremmo insomma dire che si tratta di una a-dualità (nel senso dell'*advaita*).
(A) Sono d'accordo, almeno nel senso che senz'altro la chiusura si dà solo se si dà anche il secondo momento, quello "soggettivo".

1.2. Giochi chiusi e sapere scientifico

(B) Ma non potrebbero esserci giochi capaci di chiudere anche la questione del proprio valore, cioè tali che questo non possa essere contestato? Non penso tanto a dimensioni di tipo religioso, politico o filosofico, quanto – riferendomi di nuovo a Floridi – alla sfera definita dai ragionamenti logico-matematici (LM) e dall'evidenza empirica (qui: EE), cioè in sostanza a quella che nella nostra civiltà è stata chiamata "scienza". Non nel senso che la scienza pretenda di fornire risposte totali e innegabili (infallibili), ché anzi oggi il mondo scientifico è sostanzialmente popperiano-fallibilista, ma nel senso che essa si basa comunque su due pilastri ritenuti incontestabili: la logica (il ragionamento/calcolo) e l'evidenza empirica.
(A) Il problema è che ormai né l'uno né l'altro paiono in grado di fornire alcun sia pur minimo elemento incontestabile, cioè innegabile.
(B) In che senso?

1.2.1 Giochi chiusi e sapere logico-matematico

(A) Si potrebbe dire che dopo Gödel nemmeno di LM si può più dire che esso fornisca risposte chiuse. Perché egli dimostra che LM presenta delle formule *indecidibili*, ma con ciò – e a questo di solito si presta scarsa attenzione – si potrebbe dire che *tutte* le formule, *in un certo senso*, risultano "indimostrabili" e quindi poi, sempre in un senso che andrà precisato, "indecidibili".
(B) Che cosa vuoi dire?
(A) In ogni sistema (*S*) "formale" (capace di esprimere l'aritmetica elementare), corretto (dimostra solo cose vere) e coerente (non contraddittorio), vi è una formula (*g*) che è *indecidibile*, cioè tale che, all'interno di *S*, non è dimostrabile né essa né la sua negazione, perché tali dimostrazioni invaliderebbero il sistema (quella formula, infatti, afferma la propria indimostrabilità). Da questo segue che *S* è *incompleto*, e inoltre (conseguenza fondamentale) che tale incompletezza rende indimostrabile, all'interno di *S*, la sua coerenza (la coerenza del sistema).
Ebbene, il punto sul quale voglio attirare l'attenzione è che allora in *S nessuna* proposizione risulta essere *propriamente* dimostrata, almeno nel senso che la sua eventuale dimostrazione non implica più l'*impossibilità* che si dimostri anche la sua negazione (dato che non si può dimostrare che il sistema è coerente). Per questo, come dicevo, si può affermare che ora *tutte* le sue formule risultano essere *indimostrabili*, almeno in un qualche senso del termine.
(B) A *quale* senso del termine "dimostrazione" ti riferisci?
(A) A quello – che peraltro mi pare costituire l'accezione "normale", e usuale, di tale termine – per il quale se una proposizione (*p*) viene *dimostrata* allora non può essere dimostrata anche la sua negazione (*non-p*).
(B) Capisco: se il tribunale ha *dimostrato* che tu sei colpevole, non può dimostrare anche che sei innocente; e se il calcolo ha dimostrato che "2 + 2 = 4" non può dimostrare pure che "2 + 2 = 5".
(A) Appunto. Nell'accezione usuale, la "dimostrazione" che costituisce il risultato di un "ragionamento" consta di due momenti, che indicherò con α e β. Che una proposizione (*p*) sia *dimostrata* (in *S*1) vuol dire: (α) che *p* è *derivata* applicando le regole di *S*1, e (β) che queste regole non consento di derivare anche la sua negazione (*non-p*). Insomma: α (la *derivazione* di *p* in *S*1)

implica β (l'*innegabilità*, in *S*1, di *p*). Ebbene, dopo Gödel questa *implicazione* viene meno; perché *S*1 non è in grado di dimostrare la propria coerenza.
(B) Ma la coerenza di *S*1 può essere "dimostrata" all'interno un altro sistema (*S*2).
(A) Sì, ma con una "dimostrazione" che, da capo, non può implicare il secondo momento (β2) della dimostrazione, cioè l'innegabilità della dimostrazione (in *S*2) della coerenza di *S*1. Pertanto, il momento β della dimostrazione resta *sempre* indimostrato. Nel senso che resta sempre un βn mancante di dimostrazione. La dimostrazione dell'innegabilità di una proposizione sviluppa così un *regressus in indefinitum*, e non può dunque mai venire chiusa; ma una dimostrazione che non si chiude non ha valore.
(B) Capisco. Possiamo chiamare "pieno" (cioè "completo"), o "proprio" (o anche "forte") questo senso della dimostrazione, cioè quello che si dà quando la *derivazione* di una formula *implica* la sua *innegabilità*. Dopo Gödel una dimostrazione in senso proprio (cioè pieno, forte) non è più possibile in nessun caso, perché nessuna derivazione può più implicare la propria innegabilità, dal momento che anche l'eventuale derivazione di tale innegabilità resta sempre negabile.
(A) D'altro canto questa richiesta di innegabilità non è una pretesa eccessiva, assurda o "metafisica"; direi anzi che è una pretesa "di base" per qualsiasi atteggiamento razionale. L'innegabilità è infatti un aspetto essenziale del *valore* di una proposizione, quindi anche della sua *verità* (sempre attenendoci a un'interpretazione standard, o normale-tradizionale, di questi termini). Che valore potrebbe avere, infatti, una dimostrazione che fosse ridotta alla semplice derivabilità (cioè ad α)? Che verità, insomma, potremmo mai attribuire a una proposizione se non fossimo in grado di escludere che lo stesso valore possa essere attribuito anche alla sua negazione? Del resto, se la non contraddittorietà non costituisse un valore ultimo, perché assumere la coerenza del sistema, e perché tanto impegno nel cercare di garantire la coerenza dei nostri sistemi di conoscenza?
(B) In altri termini, potremmo dire che β esprime il *valore* di ciò (*p*) che in α viene *derivato*, e quindi l'esclusione che si possa dare ciò (la derivazione di *non-p*) che ne comprometterebbe il valore.
(A) In effetti, se una proposizione viene considerata dotata di *valore* per il solo fatto di venire *derivata*, è perché di solito si dà per scontato che, qualora essa venga derivata, *non possa* venire derivata anche la sua negazione. Ma, se questo presupposto (questa assunzione implicita) viene meno, allora di *nessuna* proposizione che sia stata derivata è possibile affermare che essa ha *valore* nello stesso senso in cui ciò poteva essere affermato prima di Gödel.
(B) Sì, nel senso che nessuna proposizione è più dimostrabile *in senso proprio*. Sai che cosa sto pensando, in riferimento anche agli studi del "secondo" Wittgenstein sulla matematica? Si potrebbe dire che dopo Gödel la nozione di "dimostrazione" *cambia significato*. Prima di Gödel si poteva assumere abbastanza tranquillamente che se una formula è derivabile per via logico-matematica allora essa non è negabile (non è derivabile anche la sua negazione); dopo i suoi teoremi, invece, bisogna dire che qualsiasi *p* venga "derivata" non ne sarà mai dimostrata l'innegabilità (naturalmente sempre utilizzando il senso proprio della dimostrazione).
(A) Appunto in questo senso dicevo che dopo Gödel nessuna proposizione risulta davvero dimostrata.
(B) E per quanto riguarda l'indecidibilità? Perché tu dicevi pure che in un certo senso tutte le proposizioni risultano essere indecidibili.
(A) Beh, siccome la "indimostrabilità" (nel senso detto) vale per ogni proposizione, essa vale sia per una qualsiasi *p* che per la sua negazione (*non-p*). Ed è appunto in questo senso che si può dire che nel sistema LM *ogni* proposizione risulta essere indecidibile: non è dimostrabile né essa né la sua negazione (naturalmente sempre nel senso sopra specificato del "dimostrare").
(B) Vale peraltro la pena di osservare che questo è ben lungi dal cancellare la differenza tra le proposizioni che restano *definite* come indecidibili (paradigmaticamente: *g*) e le altre

proposizioni (quelle "normali"). Infatti le prime sono *di per sé* indimostrabili e indecidibili, perché (semplificando al massimo) se fossero dimostrabili il loro stesso esserlo renderebbe contraddittorio il sistema, vanificando con ciò il valore della stessa dimostrazione. Invece la dimostrazione delle altre (quelle "normali") non implica automaticamente una contraddizione; però anch'esse per principio non possono venire né dimostrate né decise, e *in questo senso* tutte possono essere dette *indecidibili* (né dimostrabili né refutabili).
(A) Nota poi che a una conclusione del tutto simile possiamo pervenire anche a proposito della nozione di verità, la quale, dopo Tarski, non può venire espressa formalmente all'interno del sistema (pena il paradosso del Mentitore: *m* = " '*m*' non è vero"). E tutto questo vale naturalmente anche per la macchina di Turing, per ogni algoritmo e per tutto quello che ne segue.
(B) Certo.
(A) Ma la cosa forse più interessante è che allora tale indimostrabilità vale anche per gli stessi teoremi di Gödel, nonché per tutte le conclusioni "filosofiche" che siamo soliti trarre da essi.
(B) Almeno se tutte queste proposizioni sono "dimostrate" tramite *S*, cioè all'interno del sistema LM ...
(A) Sì, ma perché dici questo?
(B) Perché qualcuno potrebbe osservare che tutto quanto è stato detto vale solo per i sistemi "formali", cioè, per dirlo in maniera ultrasemplificata, solo per il ragionamento inteso come un calcolo, un algoritmo; sicché ci si potrebbe chiedere se esso valga *in generale*, cioè per *ogni* sistema "logico" o "logico-matematico". Si potrebbe insomma operare una distinzione tra i sistemi *formali*, quelli per i quali valgono i teoremi di Gödel, e i sistemi *razionali* (così li potremmo chiamare), cioè quelli logici in senso lato che, anche se non sono completamente *calcolabili*, tuttavia sono in grado di esprimere i ragionamenti logici e matematici.
(A) Ah! Capisco. Però anche per i sistemi "razionali" (adotto la tua terminologia) ci si deve chiedere se essi siano in grado di fondare la coerenza (non contraddittorietà) del sistema; cioè se tali sistemi siano in grado di *garantire* che resti esclusa la possibilità che, una volta derivata *p*, sia possibile derivare correttamente anche *non-p*. Perché – da capo per semplificare al massimo questioni assai complesse – si potrebbe dire che prima di Gödel si poteva dare per scontato che un sistema razionale fosse senz'altro coerente/non-contraddittorio, e che, anche se ciò non era ancora stato dimostrato "razionalmente", tuttavia sarebbe bastato un po' di impegno per arrivaci. In un certo senso, prima di Gödel si poteva dare per scontato che il sistema razionale garantisse la propria coerenza, e che quindi tutte le sue conclusioni corrette potessero venire considerate dotate di valore. E per questo sistemi formali di quel tipo potevano essere applicati a *ogni* tipo di questione, potevano cioè essere legittimamente adottati come la base stessa di ogni atteggiamento razionale.
(B) Si poteva insomma dare per scontato che il sistema razionale garantisse l'implicazione tra i momenti che hai chiamato rispettivamente α e β, mentre dopo Gödel questo postulato viene meno. E allora, se non si vuole rinunciare al *valore* della non-contraddittorietà (cioè a β), visto che non è più possibile *nemmeno ipotizzare* che i sistemi formali possano fornire tale garanzia diventa necessario chiedersi *se vi sia* un sistema razionale capace di garantire il proprio valore (coerenza/non-contraddittorietà); e, se sì, quale esso sia.
(A) Una questione spinosissima. Perché – tanto per darne solo una piccola idea – da un lato è del tutto ragionevole assumere che il valore del sistema razionale consista nel suo essere *non contraddittorio*, e che tale carattere sia *innegabile*, perché solo ciò può metterlo in salvo rispetto ai danni di una possibile negazione. Questo è appunto quanto viene garantito dal cosiddetto principio di non contraddizione, il quale può appunto essere inteso come ciò che garantisce che un *risultato* derivato in maniera corretta non possa essere validamente negato.
(B) Sì. In quanto non passibile di negazione, esso è salvo rispetto al rischio di costituire qualcosa di *negativo*. Di qui appunto il *valore* del sistema razionale.

(A) Dall'altro lato, però, la dimostrazione della innegabilità di un *qualsiasi* sistema è probabilmente compito assai più arduo di quanto si possa pensare. A cominciare dal fatto che – sempre per ridurre a una battuta quello che richiederebbe una intera "Enciclopedia filosofica" – si potrebbe dire che la stessa nozione di innegabile è contraddittoria, e che quindi l'affermazione dell'innegabilità di una proposizione è contraddittoria.
(B) Fammi capire.
(A) In estrema sintesi. Se negativo è ciò che viene colpito dalla negazione, e se ciò che nega qualcosa per ciò stesso viene contro-negato da parte di ciò che esso nega, allora tutto ciò che nega qualcosa è anche negato e quindi negativo. Sicché, appunto, pensare a qualcosa di in-negabile (se quello "in" è una negazione di qualcosa) equivale a pensare a qualcosa di negato e quindi di negabile: l'in-negabile è per ciò stesso negabile.
(B) Stiamo entrando in un vortice di pensieri davvero radicale e anche terribile. Perché per esempio qualcuno potrebbe allora concludere che, se assunto come principio innegabile, *il principio di non contraddizione è contraddittorio* (per via di quel "non"). Naturalmente si tratta di interpretare correttamente una proposizione siffatta, tuttavia essa ci aiuta a intravvedere la portata reale delle questioni in gioco.
(A) Il punto è che, fin tanto che non siamo in grado di esibire un sistema razionale capace di dimostrare in maniera conclusiva il proprio *valore* (la propria non contraddittorietà, o innegabilità, o come si dovrà dire), dobbiamo riconoscere che, in mancanza di una siffatta giustificazione, non possiamo più *presupporre* che vi sia da qualche parte un fondamento che garantisca il valore dei nostri "ragionamenti" e delle nostre affermazioni (dei nostri sistemi razionali); e questo vale per *tutti* i nostri ragionamenti e per *tutte* le nostre affermazioni.
(B) Potremmo dunque dire che a questo punto l'attribuzione di valore (nel senso detto: non contraddittorietà, innegabilità, e simili) a un qualsiasi sistema razionale (e a qualsiasi proposizione al suo interno) *presuppone* un criterio di valore *ulteriore* rispetto a quello fornito da qualsiasi sistema razionale. Tale attribuzione *postula* quindi un'assunzione di valore originariamente (definitoriamente) eccedente/trascendente il sistema razionale (compreso naturalmente quello logico-matematico).
(A) Sì, e potremmo allora dire – richiamando i momenti di cui sopra si diceva (α e β) – che a questo punto α è ciò che giustifica l'affermazione di p, mentre β esprime la convinzione, cioè la *fiducia*, che non sia possibile giustificare anche la sua negazione. Ma (almeno fino a quando non si sia in grado di esibire un sistema razionale capace di autogiustificarsi) non si potrà mai ritenere *giustificata* in maniera conclusiva la convinzione che tale *fiducia* sia ben riposta, sicché questa fiducia risulta essere sempre un mero atto di *fede*.
(B) Insomma, ci stai dicendo che ogni dimostrazione contiene sempre un elemento ineliminabile di *fede*. Anche la logica, dunque, è una forma di fede; e anche la razionalità lo è.
(A) Con tutto il rispetto, naturalmente, per chi ha questa fede piuttosto che altre fedi.
Ma il punto è che riconoscere che la scienza è una forma di fede ci costringe a liberarci dall'illusione che essa possegga *un privilegio essenziale* rispetto alle altre forme di fede. Quello che abbiamo visto, infatti, ci libera da un'interpretazione illusoria, ancorché ricorrente, della situazione attuale. Di solito, infatti, si ragiona come se dopo Gödel il sistema razionale avesse sì perso qualcosa (diciamo: la completezza della propria giustificazione), ma questo non compromettesse affatto il suo *valore* (il fatto di poterci affidare tranquillamente ad esso) *almeno per quelle parti* che non risultano colpite da tale perdita. E si tratterebbe poi della maggior parte del sistema; sicché – così si sottintende – il valore del sistema razionale non resta compromesso in maniera sostanziale da quella perdita.
(B) In pratica ragioniamo come se il sistema si dividesse in due parti, una che resta dotata di valore nel senso tradizionale, pieno, e l'altra che invece non gode più di questa proprietà.
(A) Sì, ragioniamo come se il sistema razionale avesse perso il suo valore *completo*, ma non

avesse perso valore *completamente* (cioè in ogni sua parte). Vale la pena di fare qualche paragone, per cogliere meglio il punto. Solitamente si interpreta la situazione come se fossimo in un ristorante dove, essendosi guastata la friggitrice, il menù, mancando delle fritture, non fosse più completo, però ugualmente quello che resta (antipasti, primi, ecc.) fosse più che sufficiente a garantire ai commensali un ottimo pranzo. Oppure è come se disponessimo di una raccolta dei *Canti* del Leopardi non completa perché manca dei Frammenti (incompiuti); oppure ancora come se il treno avesse perso una carrozza durante il viaggio. In tutti i casi di questo genere la "perdita" non impedisce che, almeno in gran parte, venga conseguito in pieno il risultato atteso: rispettivamente un pasto soddisfacente, una buona conoscenza del mondo poetico di Leopardi, il raggiungimento della destinazione da parte dei passeggeri delle restanti carrozze del treno.
(B) Chiaro.
(A) Ma nel nostro caso la situazione è completamente diversa; perché ciò che viene meno è tale che la sua mancanza compromette il valore anche di tutto il resto, ovvero di ciò che resta. Per stare all'esempio del ristorante, è come se ciò che manca fosse la copertura che protegge i cibi da contaminazioni tossiche dovute all'ambiente. Oppure – per fare un paragone che resta nel campo "numerico" – è come se avessimo bisogno di aprire una cassaforte ma la sua serratura funzionasse solo per 9 dei 10 numeri della combinazione: lo sportello della cassaforte resterebbe completamente chiuso. Per metterla un po' sullo scherzoso, è come se ricevessimo delle bellissime lettere d'amore firmate dalla fanciulla di cui siamo innamorati, rispetto alle quali nutrissimo però il dubbio che in realtà esse siano state scritte e inviate non da lei ma da un nostro perfido amico burlone. È chiaro che, per quante lettere successive alla prima ricevessimo, comprese quelle con le quali "la nostra amata" ci rassicura che è davvero lei l'autrice delle lettere precedenti, il nostro dubbio persisterebbe, e anzi la nostra inquietudine crescerebbe di giorno in giorno. In tutti questi casi, insomma, una mancanza, benché quantitativamente minima, compromette *completamente* la possibilità di ottenere un risultato *soddisfacente*, cioè dotato di valore per noi.
(B) Sì, come dice Wittgenstein: se ho dei dubbi circa la veridicità delle notizie riportate da un giornale è inutile che, per verificarla, vada a comprare una seconda copia dello stesso giornale (Floridi 2020, pp. 26-27). Ovvero: se mi trovo in una situazione nella quale ho motivo di dubitare dei certificati che qualcuno mi mostra, è perfettamente inutile che costui mi esibisca un certificato che attesti l'autenticità dei suoi certificati.
(A) Tutto questo ci consente allora di introdurre la distinzione tra incompletezza *quantitativa* e incompletezza *qualitativa*; e per chiarire il senso di tale distinzione proporrei questo paragone conclusivo. L'incompletezza quantitativa è quella che colpisce l'erede il quale, dei 10.000 euro dichiarati nel testamento, in cassaforte ne trova solo 9.000. Peccato! Ma almeno quei 9.000 se li può godere tutti. L'incompletezza qualitativa, invece, è quella che riguarda chi nella cassaforte trova magari tutti i 10.000 euro che gli erano stati promessi dal defunto, ma non trova il testamento. In questa situazione egli non può appropriarsi nemmeno di un centesimo, e con ciò cessa propriamente di essere erede.
(B) Chiarissimo: in tutti questi casi la semplice *incompletezza*, per quanto "minima" dal punto di vista quantitativo, compromette *totalmente* il *valore* di tutto ciò di cui disponiamo (la sua *qualità*, il suo pregio).
(A) Nel nostro caso, essendo l'incompletezza dei sistemi *formali* di tipo qualitativo, accade che anche i sistemi *razionali* perdano *completamente* (quindi a loro volta in maniera qualitativa) il loro valore. Perché viene meno la garanzia che vi sia anche solo una singola proposizione che non possa essere negata, ovvero che non possa rivelarsi essere un *negativo*. Insomma, *tutte* le proposizioni del sistema perdono il loro *valore*, e quindi il sistema perde *completamente* il proprio valore.

(B) Almeno fino a che qualcuno non riesca a proporre un sistema razionale capace di giustificare in maniera conclusiva il proprio valore in modo da salvare almeno qualche proposizione dal rischio della negazione.
(A) Sicché, in conclusione, e tornando al nostro tema, si potrebbe dire che Gödel *apre anche le domande chiuse* del gioco LM – e questo riguarda anche proprio il "concetto matematico" di chiusura (Floridi 2020, p. 32).

1.2.2 Giochi chiusi ed evidenza empirica

(B) Tutto questo, però, non intacca minimamente il valore dell'evidenza empirica (EE), sicché non compromette la possibilità che il discorso venga chiuso per il tramite dell'esperienza.
(A) Di per sé, forse, no; però anche per l'evidenza empirica vale qualcosa del genere. Infatti, alla domanda: "Ci sono *esperienze* che consentono di fornire risposte chiuse?" potremmo rispondere, con Nietzsche: "Non ci sono fatti, ma solo interpretazioni". È vero che questa formula può essere interpretata in maniera da farla risultare immediatamente falsa ...
(B) ... soprattutto per via di quel "non"...
(A) ... sì, cioè come se Nietzsche *negasse* l'esistenza di eventi extralinguistici o extrasoggettivi. Però è possibile anche un'interpretazione diversa, del tipo: "*Chiamiamo* 'fatto' la sintesi di un evento reale e della sua descrizione linguistica"; come quando diciamo: "Chiamiamo 'banconota' la sintesi di un pezzo di carta e del valore che attribuiamo ad esso".
(B) Tradotta "in positivo", insomma, la battuta di Nietzsche, che nella precedente interpretazione suonava "Ogni fatto è *solo* interpretazione (cioè *non è altro che* interpretazione)", verrebbe ora a dire: "Ogni fatto è *anche* un'interpretazione".
(A) Sì, perché quello che essa vuole sottolineare è che ogni descrizione *eccede sempre* la realtà "intangibile-oggettiva" che pure contiene. Ovvero, ogni "verità" presuppone sempre il punto di vista soggettivo (la prospettiva).
(B) Certo, esprimere *lo stesso "fatto"* (per esempio che Giuda bacia Gesù) descrivendolo come "un gesto d'affetto" o invece come un "tradimento" fa molta differenza, perché le due descrizioni *eccedono* ciò che si dà (il bacio); esse dunque in realtà danno luogo a *due differenti "fatti"*.
(A) Sì, certo. Ma anche in casi meno "estremi" di questo la descrizione di un evento presuppone sempre, proprio in quanto fenomeno linguistico, il senso/significato delle parole che vengono usate per descrivere la realtà, e tale dimensione semantica porta comunque con sé le differenti "interpretazioni" implicite nei diversi giochi linguistici e nelle relative forme di vita.
(B) Ti stai evidentemente riferendo al cosiddetto secondo Wittgenstein.
(A) Sì, del resto, se si intende per "domanda" qualcosa di linguistico (e potrebbe forse non essere così?), come si può immaginare che una realtà *extralinguistica* fornisca *di per sé sola* una risposta conclusiva (oggettiva) a una domanda *linguistica*?
(B) In questo senso, dunque, anche le questioni empiriche rimangono costitutivamente aperte.
(A) Come, del resto, dimostrano anche altri aspetti propriamente scientifici del pensiero contemporaneo. Da Einstein, che "relativizza" le kantiane forme a priori della nostra esperienza (spazio e tempo), al principio di indeterminazione di Heisenberg, che mostra l'impossibilità di conoscere i dettagli di un sistema senza perturbarlo, fino alla fisica quantistica, con tutto il grande dibattito che si è aperto in proposito ...
(B) ... dibattito che, "aprendosi", finisce appunto per *aprire* anche le risposte *chiuse* date dall'esperienza, compresa quella scientifica.
(A) Aggiungi infine che nell'epoca contemporanea l'evidenza scientifica è ormai lontana anni luce (qui è proprio il caso di usare questa espressione ...) dall'*indubitabile* esperienza sensibile che accomuna a priori tutti gli umani: i "fatti" che lo scienziato *esperisce* sono ormai quasi solo i risultati forniti da un apparato strumentale estremamente complesso, sicché davvero ogni *fatto* è in realtà il risultato di una costruzione, e cioè di una *interpretazione tecnologica*.

(B) Anche qui, dunque, potremmo dire che, con la tecnica, la nozione stessa di esperienza cambia radicalmente di significato.
(A) E potremmo quindi concludere che in realtà tutte le risposte alle domande scientifiche, sia quelle relative a LM sia quelle relative a EE, sono (in qualche senso: linguistico-interpretativo-antropologico) costitutivamente aperte.
(B) Dovremo allora concludere che *non vi sono* risposte chiuse?
(A) Beh, non ho proprio detto questo, però ...
(B) ... Va bene, questo punto magari lo riprendiamo dopo; perché adesso dobbiamo affrontare un'altra questione, anche tenendo presente che quello di Floridi è ben lungi dall'essere un discorso "scientista".

1.3. Le conseguenze per le domande filosofiche

(A) Certo, ma qual è la questione che volevi affrontare?
(B) Questa: quali conseguenze ha, quello che ci hai appena detto circa il sapere scientifico, per quanto riguarda le domande filosofiche, che sono di per sé aperte (Floridi 2020, p. 25)?
(A) Beh, innanzitutto che ora esse risultano *illimitatamente* aperte. Perché, non essendo più limitate a monte da ragionamenti innegabili o da fatti indiscutibili, l'apertura che le caratterizza investe *ogni loro aspetto*, *in tutti i sensi* e *da tutti i punti di vista*.
Quindi, dato che la filosofia è la dimensione ultima – "*the buck stops here*" (Floridi 2020, p. 9) – e che l'apertura delle risposte filosofiche significa il loro essere esposte al disaccordo, *qualsiasi risposta di qualsiasi tipo* resta sempre esposta alla possibilità di venire negata, cioè *contra*-detta, in ogni suo aspetto e in ogni suo momento. Perché, anche se si parla di "disaccordo informato, razionale e onesto" (Floridi 2020, p. 26, cfr. p. 51 e p. 54), il problema è: ma chi *decide* quando un disaccordo ha queste caratteristiche?
(B) Però così cadiamo in un relativismo o addirittura in un nichilismo totali: *anything goes* (Floridi 2020, p. 38), può valere tutto e il contrario di tutto.
(A) Può darsi ... D'altra parte, come abbiamo visto, a questo punto affidarsi alla scienza come se essa disponesse di un qualche magico punto inattaccabile significherebbe compiere un atto di *fede*, il quale però trasformerebbe la scienza in una sorta di *religione*, la quale poi, essendo *inconsapevole* di ciò e spacciandosi in qualche modo per "verità", verrebbe ad assumere l'aspetto di una *superstizione*.

2. Per una lettura filosofica dell'infosfera

2.1. TecnoScienza e infosfera: uno scenario "apocalittico"?

(A) In effetti, quello che sta accadendo all'umanità può ben essere interpretato in un senso che oggi chiameremmo "catastrofista". E tieni conto che un'interpretazione di questo genere può essere anche il risultato di riflessioni autorevoli, colte e profonde. Pensa per esempio a questa osservazione di Wittgenstein, scritta nel 1947:

> La vera visione apocalittica del mondo è quella secondo cui *le cose non si ripetono*. Non è insensato, ad esempio, credere che l'era scientifica e tecnica sia l'inizio della fine dell'umanità; che l'idea del grande progresso sia un abbaglio, come anche quella che si finisca per giungere alla conoscenza della verità; che la conoscenza scientifica non arrechi nulla di buono o di desiderabile e che l'umanità, mirando ad essa, cada in una trappola. Non è affatto chiaro che non sia così. (Wittgenstein, 1980, p. 106).

(B) Un'interpretazione davvero "apocalittica" ...
(A) Sì, che coinvolge in pieno l'infosfera, la quale potrebbe essere vista come l'ultimo atto della parabola che conduce la TecnoScienza (TS), cioè la tecnica basata sulla scienza, a trasformarsi, in quanto fondata su una fede, nella religione della nostra epoca; e quindi poi in una forma di superstizione, come sopra si accennava.
(B) Perché proprio l'infosfera?

(A) Perché nell'infosfera la *sostituzione* dell'uomo naturale (*Homo sapiens*) con quello artificiale (*Homo technologicus*) diventa *totale*, dal momento che il trasferimento dell'*intera* vita umana (compresa cioè quella "mentale") all'interno di una sfera artificiale *legittima* la costruzione/sostituzione tecnoscientifica persino della mente umana, quindi dell'umano *in toto*.
(B) Che cosa vuoi dire?
(A) La tecnica garantisce la realizzazione degli scopi del soggetto (l'essere umano), e con ciò la soddisfazione di suoi bisogni/desideri. A tal fine essa *sostituisce* le realtà naturali con "artifici" (macchine, procedure ecc.). Il problema, però, è il valore etico-antropologico dell'operare tecnico. La TS, in quanto si fonda sul sapere scientifico, ne condivide il valore. Sicché, fintanto che la scienza viene ritenuta depositaria di un sapere innegabile (chiuso) – certo sempre parziale, quindi provvisorio e rivedibile, ma, almeno in qualche punto, "*unantastbar und definitiv* [incontrovertibile e definitivo]" (per dirla con le parole che Wittgenstein usa nella Prefazione del *Tractatus*) – la TS, che dalla scienza è guidata, eredita quei valori etici/antropologici, caratteristici dell'innegabile, dei quali sopra abbiamo detto. In tal modo essa, potendo considerare pre-giudizialmente risolto qualsiasi problema valoriale (etico-antropologico), è legittimata a procedere implacabilmente alla sostituzione del mondo naturale con quello artificiale. Ebbene, l'infosfera è appunto la fase nella quale la TS, rivolgendosi alla mente umana, si ritiene legittimata a sostituire anche questa con un artificio.
(B) Parlavi di una parabola ...
(A) Sarebbe un discorso lungo ... Diciamo così: nell'antichità, quando la natura umana è considerata qualcosa di immodificabile da parte dell'uomo e quindi di inviolabile, l'innegabilità è garantita dai due pilastri del *lógos* e della sensibilità naturale-universale. Sicché la scienza, nella misura in cui opera esclusivamente sulla base di questa doppia autorità, è sicura di parlare a nome dell'uomo in generale, ovvero dell'umanità, intesa come un insieme omogeneo di individui *sostanzialmente uguali* nei loro tratti essenziali. In questo quadro, l'agire della TS è *definitoriamente* un comportamento *giusto*, perché esso garantisce l'*accordo* dell'uomo con la natura in generale (esperienza) e in particolare con gli altri uomini (*lógos*).
Questa prospettiva, però, pone dei limiti notevoli al sapere scientifico. Come dicevamo, infatti, esso può riguardare soltanto quegli aspetti della realtà che sono *immodificabili* (intangibili) da parte dell'uomo; cioè, emblematicamente, l'"essenza" della natura in generale – lo spazio terrestre (geometria), il movimento degli astri (astronomia), le onde sonore (musica) e simili – e di quella umana in particolare: la ragione (*lógos*) e i sensi (esperienza). Col tempo, l'atteggiamento umano nei confronti della scienza si trasforma radicalmente. Da un lato Francis Bacon enfatizza il carattere operativo e quindi poi tecnico e di potere della scienza, dall'altro lato Galilei rinuncia a conoscere l'essenza ("tentar l'essenza l'ho per impresa [...] vana") e le qualità secondarie, tenendo peraltro fermo l'apparato logico-matematico, quello che consentirà a Newton di sostenere che le sue non sono ipotesi ("*hypotheses non fingo*").
Tutto questo contribuisce a quella "rivoluzione copernicana" (nel senso kantiano) che conduce l'uomo a riconoscere come "reale" solo ciò che corrisponde ai modelli "soggettivi". Cosa che favorisce quel decisivo *rovesciamento* di prospettiva per il quale il problema dell'accordo tra uomo e mondo viene risolto non più conformando i nostri modelli alla natura, bensì sostituendo quest'ultima con un mondo costruito in modo tale da corrispondere a priori ai nostri modelli. In breve, il nuovo atteggiamento (quello della TecnoScienza) ci fa dire: "Se la realtà non corrisponde ai nostri modelli ... tanto peggio per la realtà! La sostituiremo con qualcosa che vi corrisponde". In tal modo ora è la tecnica che assurge a fondamento della scienza, cosicché la TS viene ormai ad essere il luogo della *equivalenza* (logica) delle due. Così essa estende sempre più gli ambiti del proprio intervento, rivolgendosi in maniera sempre più invasiva alla stessa sfera dell'umano.
Il problema diventa esplosivo quando tale intervento giunge fino al punto di toccare/intaccare

gli stessi tratti "essenziali" dell'umano, ma poi addirittura drammatico nella misura in cui viene meno l'innegabilità del sapere scientifico, e, con ciò, anche il suo indiscutibile valore etico. Perché a questo punto cade la legittimazione valoriale dello stesso agire tecnoscientifico. Ci si rende così conto che è un sapere essenzialmente *opinabile* (e quindi arbitrario) quello che sta procedendo a tappe forzate a sostituire lo stesso soggetto umano. Soggetto il quale, peraltro, è l'entità che determina gli scopi/fini la cui realizzazione la tecnica garantisce, tecnica che precisamente da ciò deriva il proprio valore essenziale: la soddisfazione, appunto, dei desideri/bisogni *umani*. Si determina insomma un cortocircuito per il quale la TS, continuando ad attribuire un valore pre-giudiziale (*incondizionato*) al suo agire, si trasforma in un *pregiudizio* a causa del quale essa, *af-fidandosi* totalmente a qualcosa che non dispone di un valore universale, viene a fondarsi su un atto di *fede* e convertirsi così in una religione e poi addirittura in una superstizione, secondo quanto sopra si diceva.
(B) Interessante, ma certo anche decisamente inquietante ...
(A) Sì, perché da questo punto di vista potremmo appunto dire che la scienza della nostra epoca – a differenza di quella antica (*epistḗmē*: Pitagora, Euclide) e anche di quella moderna (*Scientia*: Galilei, Newton) – si sta trasformando nella religione della TecnoScienza, o meglio della *TechnoLogy* (visto che adesso è l'inglese la lingua imperiale), in quanto ritiene legittimo (*giustificato*) imporre a tutti credenze che non sono necessariamente condivise da tutti.
(B) Ma quali sarebbero le credenze della scienza che possono essere ritenute "fedi incrollabili"?
(A) Per esempio il "pregiudizio galileiano" (ma forse già pitagorico) per cui la natura è scritta in caratteri matematici e quindi è *conoscibile* per mezzo di strumenti logico-matematici. Ma anche, più in generale, la fiducia che la realtà sia guidata da "leggi": o di tipo logico (il principio di *non* contraddizione) o di tipo naturale (il principio di causalità, la legge di gravitazione etc.). Ma poi, soprattutto, la convinzione che la conoscenza scientifica costituisca un sapere dotato di un valore (etico) innegabile/assoluto e quindi indiscutibile, dovuto per esempio al fatto di *rispettare* la "natura propria" della natura (*phýsis*).
(B) Ma dove vuoi arrivare? Qui non si salva niente ... Che cosa intendi dire, più precisamente?
(A) Sarebbe un discorso lungo e delicato, come capisci; dovendo limitarci a qualche battuta, direi così. Se su una goccia di olio ne fai cadere un'altra, ottieni sempre una goccia d'olio (sia pure più grande); cioè in natura vale (*anche*) "1+1 = 1". Voglio dire: la matematica, come tutte le scienze formali (logica, informatica etc.) *presuppone* che la realtà sia costituita da oggetti (determinazioni) tali che l'uno *non è* l'altro (due sassolini, due noci, due bit etc.): questi *dati* risultano fissati *pre-giudizialmente*.
Il principio di *non* contraddizione, poi, presuppone che la negazione di una contraddizione sia vera; ma anche questo è discutibile, come mostra esemplarmente il "Mentitore" (Tarski, Priest ecc.). Quanto poi alle leggi di natura, qualcuno (penso per esempio a Raimon Panikkar) potrebbe dire che c'è sì una legge fondamentale della natura, ma essa è ... la libertà.
E si potrebbe aggiungere che "in realtà" ogni gesto (quindi anche ogni azione e ogni processo) modifica il tutto e quindi pure la situazione di partenza. Perché anche ogni "operazione" trasforma inevitabilmente i dati di partenza. *In realtà* accade che la "congiunzione" di 1 femmina con 1 maschio renda gravida la prima, così che in questo caso vale anche 1 + 1 = 3. Appunto perché l'operazione della "congiunzione" trasforma il dato di partenza (1 individuo femmina) in un dato diverso (1 individuo femmina più 1 individuo-feto). Insomma, in generale l'operazione trasforma i dati di partenza, sicché il risultato è *davvero* "vero" solo se include anche gli effetti di tale operazione, i quali però variano a seconda degli oggetti a cui si applicano e alle circostanze in cui ciò accade. Pertanto – concludendo con la questione propriamente etica – anche ogni conoscenza, col suo semplice "rappresentare" la realtà, in verità la trasforma (cioè la rinnova e quindi la "apre"), ed è quindi, di per sé, ben lungi dal rispettarla e quindi dal possedere un valore etico privilegiato.

(B) Sarà, ma ciò non toglie che anche nella realtà naturale ci siano dei "dati", dei "punti fermi".
(A) Certo, ma la realtà naturale è in continua, completa trasformazione, sicché il problema decisivo è proprio la *determinazione* dei "dati", mentre nelle scienze *almeno alcuni* dati sono *presupposti*: essa si atteggia *come se* almeno alcuni "fatti" naturali fossero *indiscutibilmente dati*.
(B) Però certo sarebbe imbarazzante dire che *non ci sono* né dati né risposte chiuse.
(A) Non sto dicendo questo, però certo le scienze esatte possono solo *postulare*, o presupporre, questi "punti fermi".

2.2. L'emergere della dimensione "soggettuale"

(B) Ma allora non potremmo metterla in un modo un po' diverso?
(A) Cioè?
(B) Potremmo forse esprimerci così: gli elementi *chiusi* (dati, risposte ecc.) sono tali nella misura in cui c'è un *soggetto* (in senso lato) che riesce a *fissarli*, e quindi a produrre le *determinazioni* sulle quali poi il pensiero logico-matematico opera.
(A) Sì, si potrebbe anche dire così.
(B) Ma, se è così, qual è allora il criterio in base al quale assumiamo una certa "fissazione" piuttosto che un'altra?
(A) A questo punto direi che è solo la sua efficacia; ma cioè, in fondo, la *potenza* del soggetto che effettua tale "fissazione": le "sostituzioni" che la TecnoScienza opera valgono nella misura in cui *qualcuno* ha il *potere* di farle valere (imporle) universalmente.
(B) Capisco. Però allora si presenta un problema enorme: qual è il motivo che ci fa considerare *legittimo* l'operare della TecnoScienza? Perché, se questa *imposizione* (termine che è poi una possibile traduzione di *epistḗmē*) è fatta *contro* la volontà di qualche soggetto, essa perde il proprio carattere in-negabile, cioè la propria validità pregiudiziale, e con ciò il proprio valore universale, quindi anche etico-antropologico.
(A) Tieni però conto che nell'infosfera la tecnica, che ormai si ritiene autorizzata ad assumere come campo d'azione "*tutto*", costruisce anche i soggetti, sicché in prospettiva essa è in grado di sostituire i soggetti umani "inadeguati" (o "antiquati", per dirla con Günther Anders) con soggetti che corrispondono ai suoi algoritmi.
(B) Il suo obiettivo è dunque quello della *onni-potenza*. E potremmo allora dire a Nietzsche che era ingenuamente ottimista quando parlava di "volontà di potenza", giacché a questo punto parrebbe più corretto parlare di "volontà di onnipotenza".
(A) Si può dire anche così.

* * *

(B) Certo che alla fine ha ragione Floridi a dire che c'è proprio bisogno di una filosofia "aperta" capace di pensare l'infosfera, cioè tutti questi problemi. Come è vero che pare indovinata l'idea della filosofia intesa come design concettuale (Floridi 2020, pp. 97-122).
(A) Sicuramente. In un certo senso è proprio questo il problema. Naturalmente si pone qui la questione di *quale sia* il tipo di pensiero adeguato a capire quello che sta accadendo. Perché, se ci limitiamo al linguaggio dell'infosfera, che è poi quello dell'informatica (Floridi 2020, p. 15; cfr. pp. 129-130: "*inforgs*"), c'è il rischio che il nostro modo di pensare, derivando sostanzialmente da quello scientifico (Floridi 2020, p. 56), ci costringa a restare troppo "chiusi" all'*interno* del mondo tecnoscientifico ("internati" in esso). Anche il discorso di *Pensare l'infosfera* deve prestare attenzione a questo aspetto della faccenda. A parte la centralità di LM ed EE, di cui abbiamo già detto, pensa in particolare all'importanza riconosciuta a Turing, portatore di un "linguaggio universale" (Floridi 2020, p. 22) e considerato addirittura autore di una quarta rivoluzione (Floridi 2020, p. 124, pp. 129-130) ...
(B) ... sicché potremmo, scherzosamente, descrivere l'infosfera come una sorta di "Turing Club" ...
(A) Ah! ah! Pensa poi alla terminologia adottata, fondamentalmente proveniente dall'ambito "tecnoscientifico" (Floridi 2020, p. 56). Mi riferisco sia ai termini peculiari del mondo

informatico – dai LdA (Floridi 2020, pp. 55 ss. e *passim*) alle "variabili tipizzate" (Floridi 2020, p. 58) – sia a quelli derivanti da aspetti del mondo attuale che sono "consustanziali" con l'infosfera: "disponibilità delle risorse" (Floridi 2020, p. 21 e *passim*), "capitale semantico" (Floridi 2020, pp. 134 ss.), "controllo" (Floridi 2020, p. 143 e *passim*), etc. Per tacere del fatto che lo stesso termine che *definisce* tutto il discorso di Floridi (infosfera) deriva totalmente dalla dimensione dell'informazione tecnologica, mostrando così un'appartenenza sostanziale a quella sorta di "metafisica dell'informazione" oggi dominante per la quale la realtà consta di domande-informazioni (Floridi 2020, p. 26), per cui anche la neurobiologia, per esempio, è basata sull'idea che i neuroni si scambino "informazioni" ...
(B) Perché, non è vero?
(A) Sarà anche vero, ma, voglio dire: se hai visto Mario che baciava Paola, sarebbe "veritiero" riferire a Giulia (la moglie di Mario) che ti chiede se hai incontrato Mario e che cosa stesse facendo, dirle che lo hai visto mentre si stava scambiando delle informazioni con Paola? Oppure, sarebbe corretto descrivere il fatto che uno uomo ne ha accoltellato un altro dicendo che i due si sono scambiati delle informazioni? Certo, anche questo sarà vero: anche ricevere un bacio o una pugnalata ti fornisce un'informazione; però, insomma ...
(B) Ah! ah! Questa volta sei tu che fai lo spiritoso. Però direi che quello che osservi non tocca il discorso di Floridi, il quale semplicemente sottintende che questo ormai è il mondo in cui viviamo, sicché solo un linguaggio *adeguato ad esso* è in grado di comprenderlo.
(A) Certo, ma altro è prendere atto del luogo in cui siamo stati trasferiti (internati), altro è *fare automaticamente nostro* il linguaggio e quindi i "valori" di questo mondo ...
(B) ... valori che tu vorresti rifiutare/negare?
(A) No, non ho detto questo. Dico però che, risultando ormai la questione del valore *totalmente* aperta, a questo punto le domande decisive sono quelle che riguardano la dimensione che potrei chiamare *soggettuale*. Uso questo termine per indicare la sfera nella quale, a differenza di quella *soggettiva* nella quale il soggetto è *già dato*, si generano, cioè *si determinano*, le nuove soggettività. E nella quale, quindi, vengono *istituite* ("chiuse") tutte le *determinazioni*; e vengono con ciò fissati gli scopi, i fini, e in generale i valori.

* * *

(B) Questione che però non mi pare affatto esclusa nella prospettiva di Floridi .... Anzi, a me pare che la questione del design concettuale dica qualcosa di significativo proprio in questa direzione. Comunque capisco la tua esigenza; e allora dimmi: quale tipo di *Homo*, secondo te, viene messo in produzione all'interno dell'infosfera? Quali tipi di soggetto vengono prodotti? E, soprattutto, qual è la natura dei soggetti che operano la sostituzione dell'*Homo naturalis* con questi nuovi soggetti?
(A) Eh ... ma sono problemi molto più radicali di quelli che solitamente ci poniamo. Anche perché, volendo metterla in maniera scherzosa, e prendendo lo spunto da una famosa battuta di Mark Twain («Se votare servisse a qualcosa non ce lo lascerebbero fare»), potrei dire: "Se pensare l'infosfera servisse a qualcosa (per esempio a capire davvero quello che sta accadendo) non ce lo lascerebbero fare".
(B) Ah! ah! Buona la battuta, anche se un po' provocatoria ... Ad ogni modo capisco il problema: la filosofia adeguata a pensare il tempo presente è quella che *determina* (conosce) la/e soggettualità operante/i, cioè le diverse forme di vita che emergono nell'epoca dell'infosfera, e in particolare quelle dominanti.
(A) Sì, e una distinzione decisiva è quella tra la "soggettualità" che *opera* la tecnosostituzione umana e i soggetti umani che invece la subiscono. I soggetti che appartengono alla prima categoria potremmo chiamarli "sovraumani" (o "dèi"), perché essi dispongono della meta-tecnica, che da un lato progetta e gestisce i "soggetti umani" e dall'altro lato genera e costituisce dei soggetti *nuovi*: i meta-soggetti.

(B) Quindi non ha poi torto Floridi quando, parlando di Platone, valorizza il ruolo del costruttore rispetto a quello dell'utente (Floridi 2020, pp. 102 ss.). E, a ben pensarci, anche la sua trattazione del tema della telepresenza (Floridi 2020, pp. 64 ss.) ha a molto che fare con tale questione.
(A) Credo di sì, ma vorrei sottolineare che, a questo proposito, il problema centrale è quello del *potere*, che ormai è il potere *sulla vita*. Già vent'anni capitava di pensare, ragionando con alcuni amici: "Internet è la fine della libertà di parola per gli umani". A prima vista pare un'affermazione assurda, perché sembra vero tutto il contrario: Internet ha incrementato al massimo le possibilità di comunicare; però sta di fatto che il trasferimento dell'intera comunicazione umana, sia personale che pubblica, all'interno di uno spazio universale e *privato* consegna nelle mani dei pochi padroni di tale spazio un diritto sostanziale su tutto quello che accade al suo interno, quindi anche sulla mente degli umani. Ciò consente a costoro di procedere alla *costruzione* della "mente" (conoscenze, emozioni, sentimenti, relazioni ecc.) degli uomini "naturali", i quali, in quanto le loro menti sono in potere altrui, possono ben essere chiamati i "mente-catti" (o "de-menti").
(B) Insomma: "sovraumani" contro "mente-catti": altro che infosfera, forse dovremmo parlare di "infernosfera" ....
(A) Ah! ah! Tieni però conto che quello che ho tracciato è un quadro molto schematico ...
(B) ... e soprattutto questo discorso non tiene conto della democrazia, la quale pone dei limiti al potere ...
(A) ... a meno che, al contrario, lo assolutizzi in maniera totale e irreversibile. Giacché in democrazia chi gestisce il potere è legittimato ad agire *in nome del popolo*, cioè di tutti, anche quando opera contro la volontà di qualcuno. Del resto, anche la democrazia, alla fine, *presuppone*, come *dati*, i soggetti umani. Per ridurre di nuovo il tutto a una battuta, potrei dire: nell'infosfera i "sovraumani democratici", dopo aver costruito (e quindi programmato) le menti dei soggetti umani, chiedono (tramite le elezioni) a questi loro "prodotti" che cosa desiderano e da chi vogliono essere rappresenti, dopo di che impongono ai politici così eletti di eseguire le indicazioni "liberamente" espresse dal popolo ...
(B) Quindi un pensiero adeguato a pensare l'infosfera deve porsi anche il problema della democrazia. Ma, tornando al problema della soggettualità, quali scenari tu intravedi?
(A) Mah! Al momento quelli visibili a occhio nudo paiono essere tre. 1) La *neospeciazione*: l'*Homo technologicus* si separa dagli uomini naturali come a suo tempo l'*Homo sapiens* si è distaccato dagli altri ominidi. Un gruppo di uomini "eletti" – o, forse meglio, "*seletti*", visto che sono il risultato di una selezione naturale – prende il sopravvento su tutti gli altri animali (compresi gli "uomini naturali"). 2) I *tecnoanimali* (grandi animali postantropici): nascono nuovi organismi "animati" (per esempio i "tecnosauri", ma non solo), certamente derivanti dalla vicenda umana e quindi dalla stirpe dell'uomo, ma ormai distanti, anche dal punto di vista morfologico, strutturale e in generale fisico, dall'animale umano. 3) Le *nuove forme di vita* (*vita technologica*): fanno la loro comparsa forme di vita talmente nuove da non replicare più, nemmeno in grande, la figura dell'animale.
Queste tre linee evolutive, pur distinte tra di loro, sono verosimilmente destinate a mischiarsi; cosa che contribuisce a rendere ancora più complesso il discorso filosofico, il quale incomincia proprio qui, dove si pongono questi problemi ...
(B) ... sui quali, però, Mark Twain direbbe che non ci sarà consentito di prendere la parola ...

2.3. La possibilità di un'altra lettura dell'infosfera

(C) ... allora, fintanto che è ancora possibile parlare, ne approfitto per ...
(A) Ah! Bentornato nel dialogo! Che cosa vuoi dirci?
(C) L'ultima ipotesi (la 3) dischiude addirittura la possibilità che emergano forme di vita *nuove* rispetto alla vicenda animale e anche umana. "*Quanto* nuove?", vorrei chiedere. Insomma: è possibile anche una lettura dell'infosfera diversa da quella "apocalittica" che hai presentato?

(A) Su quali basi?
(C) Stavo ancora pensando alla questione iniziale, quella relativa alle domande aperte; e mi verrebbe da chiedere: ma *quanto* possono essere aperte le domande filosofiche? Perché, se "aperto" vuol dire esposto al "disaccordo" (Floridi 2020, p. 51), il quale implica sempre una qualche forma di negazione, allora in realtà ogni domanda resta *definita* dalla negazione, quindi in fondo *chiusa* rispetto ad essa: il gioco della negazione finisce per essere inoltrepassabile ...
(B) ... e a Nietzsche potremmo allora dire che c'è ancora un "ultimo dio" rimasto in vita: la negazione.
(C) Ah! ah! Buona! ... Ma, se la negazione è inoltrepassabile, allora anche il negativo (ciò che è determinato dalla negazione) lo è, e quindi anche il *necativo* (l'insieme dei danni arrecati dal carattere nocivo della negazione) risulta insuperabile.
(B) D'altro canto, come si può anche solo immaginare di oltrepassare la negazione e il negativo, visto che negare la negazione vuol dire confermarla? Come si può pensare di uccidere il Dio-negazione se uccidere è *necare* e la necazione è in qualche modo il nocciolo della negazione?
(C) Giusto! Bravo; questo, però, risulta inevitabile perché tu presupponi che la *differenza* rispetto al negativo costituisca sempre e comunque una *negazione* (del negativo o comunque di qualcosa a questo connesso). Prova invece a distinguere la differenza dalla negazione, cioè a pensare la differenza "pura". A partire da questo pensiero potremmo forse allora avanzare una lettura del tempo presente la quale *vede* – invece della "rottamazione" del mortale, cioè della negazione del *necato/necatore* (l'animale mortale e mortifero) – lo schiudersi di una situazione del tutto diversa. Per esempio la conclusione della storia della *necazione*, cioè del mortale, e l'apertura quindi di una dimensione nella quale da criterio di valore universale (ciò che vale per tutti) funge, invece dell'onnipotenza (onni-pre-potenza: *polemos*), il *fatto* che tutti i soggetti sono *effettivamente* (*di fatto*) d'accordo. Un criterio cioè per il quale il gioco è considerato *chiuso* solo quando *tutti* i soggetti sono *soddisfatti* dalle risposte ...
(A) Calma, calma ... Con tutti questi filosofemi ...
(B) ... ci stai trascinando nella "filosofemisfera" ...
(A) Ah! ah! No, comunque davvero mi pare che questi tuoi discorsi, oltre tutto palesemente "utopistici", ci stiano davvero portando fuori strada.
(C) Può essere, ma che cosa, *ormai*, ti autorizza a *chiudere* la conversazione che si sta aprendo?
(A) Beh, anche solo il fatto che il tempo a nostra disposizione è finito.
(C) Ah! ... Beh, sì, certo, mio caro Eutifrone ... Adesso "hai fretta, e devi andare via", e quindi dobbiamo interrompere il dialogo ...
(A) In ogni caso, anche se tu riuscissi a presentare una lettura diversa, diciamo "positiva", dell'infosfera, comunque anch'essa resterebbe, *ormai*, semplicemente una tua *opinione*. Per quale motivo allora dovremmo preferirla a quella "apocalittica"?
(C) "Una mia opinione ...", dici. E se, nel mondo delle domande *totalmente* aperte, *vero* fosse proprio il discorso *migliore* tra tutti quelli possibili?
(B) Cioè "*id quo* melius *cogitari nequit*", per dirla con la formula anselmiana ritoccata però mediante una evidente "licenza filosofica".
(C) Sì, potremmo anche dire così. Perché se "vero" è ciò che è confermato persino dalla propria negazione ...
(A) ... eh no! Non ricominciamo, il tempo è proprio finito ... E poi tieni conto che questa tua prospettiva sarebbe comunque lontanissima da quella del libro di cui stavamo parlando.
(C) Ne sei sicuro? In realtà a me essa parrebbe non solo del tutto coerente/compatibile con il discorso di Floridi, ma per molti versi addirittura ad esso molto vicina. Pensa anche solo al carattere costitutivamente aperto delle questioni filosofiche; ma poi anche al fatto che il discorso sui Livelli di Astrazione e sulla "variabile tipizzata" valorizza l'originaria *duplicità* di

ogni discorso, cioè in sostanza la costitutiva eccedenza-trascendenza del discorso rispetto alle determinazioni date.
(B) Però nel tuo discorso tali indicazioni vengono radicalizzate anche in senso etico-valoriale, in particolare ponendo il problema dell'apertura persino rispetto alla negazione e quindi anche rispetto a una lettura semplicemente matematico-quantitativa della realtà.
(C) Sì, ma tieni conto che tale "svolta qualitativa" mi pare comunque molto vicina all'interpretazione della filosofia come design concettuale.
(B) In effetti, in entrambi i casi filosofo è colui che *dà forma* (conferisce verità = valore universale) alle esperienze/pensieri di tutti, piuttosto che colui che *subisce passivamente* dei *dati* che sarebbero appunto disponibili prima della *costruzione* di una concettualità adeguata. Penso per esempio ai temi del "costruzionismo" (Floridi 2020, pp. 118 ss.), dello stesso "capitale semantico" (Floridi 2020, pp. 134 ss.), soprattutto se pensato in relazione alla "risorsa intellettuale" (Floridi 2020, p. 25), e a quello delle "risorse noetiche" (Floridi 2020, p. 43). Ma potrei richiamare anche molti altri punti. Solo per fare qualche esempio (e citando in maniera disordinata): il "minimalismo" (Floridi 2020, p. 114) e anche il "liminalismo" (Floridi 2020, p. 116); il superamento delle ("dicotomiche") soluzioni booleane (Floridi 2020, p. 116), nonché la denuncia di un certo "vandalismo" humeano (Floridi 2020, p. 28, poi pp. 29 *passim* e p. 51), probabilmente connessa alla presenza di "tracce della logica hegeliana" (Floridi 2020, p. 116).
(C) E tutto questo mi pare in grande sintonia con un'altra fondamentale esigenza che percorre tutto il libro di Floridi, cioè quella che il discorso filosofico, pur mancando di una *chiusura*, sia qualcosa di diverso da una posizione relativistica (per esempio p. 116; cfr. p. 59, p. 94) o addirittura irrazionalistica (Floridi 2020, p. 143; cfr. p. 19).
(A) Insomma, adesso ci vieni a dire che condividi tutto quello che c'è scritto nel libro di Floridi ...
(C) In un certo senso sì, naturalmente. Però parlerei piuttosto di *con/divisione*; intendendo evidenziare, con questo espediente grafico, la circostanza che la vera condivisione lascia la libertà (a ciascuno dei soggetti dialoganti) di restare fedeli alle proprie specifiche proposizioni, solo invitandoli tutti a procedere in maniera *condivisa* anche laddove le differenze di prospettiva li conducono a percorrere strade diverse, cioè a procedere in modo da *dividersi* ma in maniera *con*-sensuale ...
(B) ... va bene, direi che con questa proposta di ... come l'hai chiamata? ... Ah! Sì, di *con/divisione* possiamo proprio concludere ...
(C) ... dove naturalmente il "con-cludere", dato che *in verità* il "chiuso" va distinto dal "*non* aperto", significa *aprire* uno spazio filosofico *nuovo* ...

[A questo punto le parole vengono coperte da alcune, sia pur discrete, risate, da frasi praticamente indecifrabili e da vari rumori di fondo (sedie spostate e altro). Subito dopo la registrazione del dialogo si interrompe].

REFERENCES

Floridi L. (2020). *Pensare l'infosfera. La filosofia come design concettuale*, Milano: Raffaello Cortina;
Tarca L.V. (2001). *Differenza e negazione. Per una filosofia positiva*. Napoli: La Città del Sole;
Tarca L.V. (2006). *Quattro variazioni sul tema negativo/positivo. Saggio di composizione filosofica*. Treviso: Ensemble '900;
Tarca L.V. (2016). *Verità e negazione. Variazioni di pensiero* (a cura di Th. Masini). Venezia: Cafoscarina Editrice;
Wittgenstein L. (1980). *Vermischte Bemerkungen* (1977), tr. it. a cura di M. Ranchetti, con il titolo *Pensieri diversi*, Milano: Adelphi.

FREE CONTRIBUTIONS

FREE CONTRIBUTIONS

Shiva Rahman
On the Existential Significance of ‘Readiness Potentials’

Marco Viola
Three Varieties of Affective Artifacts: Feeling, Evaluative and Motivational Artifacts

SHIVA RAHMAN
Independent Researcher
shiva9995115836@gmail.com

ON THE EXISTENTIAL SIGNIFICANCE OF 'READINESS POTENTIALS'

abstract

Could there be a balanced philosophical stance capable of accommodating the scientific facts pertaining to free will without compromising the ideal of human freedom and autonomy? A stance that can render intelligible the inferences emerging from the factual analysis of free will in terms of the phenomenon called 'Readiness Potential'(RP), at the same time, existentially upholding the ideal of freedom? In the present paper, an attempt will be made to bring to light such an existential phenomenological perspective implicit in the philosophy of the French phenomenologist Maurice Merleau-Ponty. In the analysis, some of the most relevant scientific facts pertaining to RPs and the corresponding scientific inferences as to the conception of 'free will' will be taken into account with a view to see how in this phenomenological scheme they could all be intelligibly accommodated. Once this is achieved, a unique version of compatibilism inherent in Merleau-Ponty's thought will also be traced

keywords

free will, readiness potential, lived body, consciousness, temporality

Phenomenology and Mind, n. 20 - 2021, pp. 204-227
DOI: https://doi.org/10.17454/pam-2017
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

I. Introduction

Philosophically, the concept of 'free will' derives its significance from the more comprehensive theme of human freedom and autonomy. Therefore, any scientific discovery relevant to the concept may have its due influence on the corresponding philosophical debate, but without any overshadowing effect on its origins. However, the general philosophical tone of the debate on freedom during the last few decades has been excessively influenced by the unprecedented progress of neuro-scientific investigations into free will, registered over the same period. This becomes evident in the role that the neuro-phenomenon of 'Readiness Potentials' (RP) has assumed lately in the philosophical discussions of human freedom and autonomy.

This trend, characteristic of 'scientism' of our age, has had such an impact on the philosophical enquiry into the ideal of freedom, that the latter has often been virtually reduced to a factual analysis of 'free will' in terms of RPs. Thus, in the fact versus value analysis of the notion of freedom, the factual aspect seems to have been dominating for a while now. Consequently, paradigm-neutrality that is supposed to be the hallmark of philosophical approach seems to be lacking here.

In fact, this is inevitable a situation in view of the transformation that our social life and intellectual outlook have undergone in response to the scientific and technological advancements of the last few decades. Also, it has to be admitted that unless and until scrutinized by science, even so foundational a value such as freedom or autonomy maybe considered untenable in the times that we live in. That being so, we cannot detach totally the neuro-scientific facts pertaining to 'free will' from an analysis of the theme of freedom and autonomy anymore. However, on this line of thought, if we make room for the scientific findings in the analysis and synthesize them with the theme of human freedom, a libertarian may still argue that it is possible for human agents to act freely independent of any necessitating causes.

It is in this context that an enquiry into a balanced philosophical stance, which accommodates the scientific facts (irrespective of their implications) without compromising the ideal of human freedom and autonomy, becomes pertinent. A stance that can intelligibly accommodate the inferences emerging from the factual analysis of 'free will' in terms of RPs, at the same time, existentially upholding the ideal of freedom. In the present paper, we will bring to light such an existential phenomenological perspective implicit in the philosophy of the French phenomenologist Maurice Merleau-Ponty, wherein facts and values are harmonized existentially.

In our analysis, we will take into account, on a representative basis, some of the most relevant

scientific facts pertaining to RPs and the corresponding scientific inferences and see how in the said phenomenological scheme they could all be harmonised with the ideal of human freedom. The criterion of relevancy in our choice of the scientific analyses is the stress on the nexus between consciousness and temporality- which, as we will see shortly, is crucial in the philosophical stance at hand.

2. Scope of the Concept of Free Will

What is meant by 'free will' for our purposes is the first thing to be clarified.
As we have already indicated, we will be treating it as a derivative of the more comprehensive theme of human freedom and autonomy. And when it comes to freedom and autonomy, we in fact have a whole conceptual spectrum available as to what they mean. At one extreme of this spectrum, we have the broadest folk-psychological ideas of human capacities such as that allowing one to control his/her action and behaviour, change one's habits and routine, overcome addictions and predilections etc. and the graver course of deciding or changing the way of one's life as a whole by consciously considering and exercising one's willpower in doing so. On the other end of the spectrum, we have the narrowest version of expressing freedom and autonomy, which consists in responding to and consciously carrying out specific experimental instructions in the capacity of subjects volunteering in a neuro-scientific experiment[1]. In between these two extremities[2] are placed the other views of religious, socio-legal[3] and psychological origin.
A practical definition of free will which would resonate with the whole spectrum mentioned, without any bias, may be that:

> free will is a feature of cognitive/motivational systems with sufficient complexity to compute the likely effects of alternative (sequences of) hypothetical actions, and to compare them with respect to moral and personal preferences of many kinds before deciding what to do (Boden, 2011, p. 163).

But the interesting fact is that when it comes to conceptualising the idea, such an unbiased approach is often not materialized. The findings made in such experimental settings as that forming the latter extreme of the spectrum mentioned usually have a sweeping impact on the whole spectrum, irrespective of the conceptual limits and practical limitations of such findings. And consequently, we often have a narrowed-down conception of free will based on and catering to such findings alone, without any regard to the existential situation of mankind reflected in the remaining stretch of the spectrum.
The above mentioned fact has been substantiated in the works of a number of researchers. Alfred Mele, in particular, has had a long pursuit in bringing this fact to light. In a recent article (Brass *et al.*, 2019) he has consolidated his arguments for his stance, in continuation of what he has argued in his *Effective Intentions*. Agreeing with Mele on his conception of 'overt actions' as actions 'that essentially involve peripheral bodily motion', our approach in the present paper as well, will resonate his point as to the insufficiency of scientific experiments

1 These come under what Alfred Mele calls 'overt actions'. This notion will be examined shortly.
2 For an interesting account on how these extremities may actually be forming part of a spectrum, see (Smith, 2011).
3 It is well acknowledged that there is an ongoing debate about how, and to what extent, empirical findings pertaining to free will may affect the long established judicial procedure and our long standing outlook on culpability. See for example: (Hodgson, 2009; Morse, 2003; Nestor, 2019).
We do not intend to enter that debate here. However as will be seen towards the conclusion, we will be supporting in a unique way the camp upholding free will.

in demonstrating such actions as lacking free will. While his arguments are formulated around the idea of the difference between 'proximal' and 'distal' intentions, in our endeavour, we will be developing our own arguments for a similar view deriving them rather from the phenomenological scheme that we have adopted.[4]

Another point that Mele makes in the article in defence of his position is the role that moral and practical involvement of the agent may be playing in rendering an overt act free. This, in turn, depends on the view one takes of the idea of free will. Adhering to the definition of free will that we have adopted above, we too will be accounting for such role. However, our treatment of subject won't be merely as a 'cognitive agent' but rather as an embodied agent.[5]

Apart from the mentioned overshadowing effect of scientific findings, an important factor contributing to the prevalence of the aforesaid scenario is the view upheld by the majority of scientific community that free will is not compatible with determinism. In the philosophical literature pertaining to compatibilism, however, there are formidable arguments against this view[6], formulated taking into account the other aspects in the spectrum mentioned that reflect the existential plight of mankind.

It has to be noted here that the scientific findings pertaining to RPs so far do not prove determinism. Therefore, a strict compartmentalisation in terms of compatibilism and non-compatibilism is not yet theoretically possible. So when we use the term 'compatibilism' in the present context, what we mean is a tentative position which upholds free will in tune with the latter set of arguments mentioned.

It is in these circumstances that it becomes utmost important to set an appropriate conceptual framework within which the idea of 'free will' can be invoked and debated meaningfully. This has to be done reflecting on the existential situation of man, as well as taking into account the said delimiting role of scientific discoveries, thereby reckoning the implications of both compatibilism (in the above sense) and determinism.

A certain conceptual framework proposed by Eddy Nahmias seems to meet this requirement and therefore could serve as an ideal starting hypothesis for our purposes. He proposes the following framework as a sensible one within which recent evidence from psychology and neuroscience may challenge free will avoiding subscription to extreme views:

> (R1) Free will requires that one's actions properly derive from decisions or intentions that one has at some point consciously considered, or at least that one would accept, as one's reasons for acting.
> (R2) Science is showing that our actions do not properly derive from decisions or intentions that we have consciously considered or would accept as our reasons for acting. Rather, our actions are produced by other factors, and we rationalise them after the fact.
> (R3) So science is showing that we do not have free will (Nahmias, 2010, p. 353).

4 This will be done based on the distinction that is drawn between 'perceptual' and 'intellectual' consciousness in that scheme, and the interpretation of RPs in such a scenario. Therein, overt actions will become part of the perceptual and motor capacities of the 'lived body'- wherein no conscious preference or action control is involved. Whereas controlling behaviour that is the hallmark of free will and subjectivity will be founded in intellectual consciousness - which in the scheme coincides with temporality. Drawing on some of the most relevant empirical findings as to the deep seated connection between the neuro-phenomenology of RPs and the invocation of time consciousness, we will substantiate our phenomenological proposition, 'time is subject'.

5 In a unique Merleau-Pontyan version of compatibilism that we will be developing towards the end of the paper, we will have our defence of this stance in terms of the phenomenological notion of embodied intersubjectivity.

6 See, M. McKenna, Compatibilism In E. Zalta (ed.), The Stanford Encyclopedia of Philosophy URL: http://plato.stanford.edu/entries/compatibilism.

Adopting this conceptual framework, let us analyze the first premise forming it and see the implications of the same. Obviously, it connects the phenomenal aspect of free will i.e. action, with the non-phenomenal aspects - decisions and intentions, in a derivative mode, as either:

i) the consequence of the subject's conscious consideration, or (at the least)
ii) the subject's reasons for acting.

We can immediately sense that the former alternative above has set the narrower paradigm for scientific enquiry into free will and has consequently been motivating rigorous empirical research. Whereas the latter has set the broader paradigm for the enquiries into human freedom and autonomy and has consequently been inspiring rather a sociological as well as psychoanalytical approach, alongside the empirical one.[7] Hence it seems proper to deal with these two scenarios separately.
In what follows, focusing first on the former alternative, we will briefly survey the empirical endeavors chosen based on the criterion of relevancy set at the end of the last section. After analyzing their inferences, we will compile the tenable implications thereof and bring to light the relevance of the same to Merleau-Ponty's phenomenological scheme of human embodiment and perception. We will then elaborate this scheme and show how the same can accommodate intelligibly the empirical findings analyzed, without resorting to any deterministic stance. Once this is achieved, we will take up the latter alternative mentioned (i.e. subject's reasons for acting) and show how the same could be made an intelligible phenomenon in Merleau-Ponty's philosophical approach towards human freedom and autonomy viewed as a unique version of compatibilism.

3. A Survey of Relevant Neuro-Scientific Enquiries

In this task, we have as a starting point in the vast literature of relevant empirical research, path breaking neuro-scientific discoveries to be found in the work of British neuroscientist Benjamin Libet. Basically, Libet's work enquires into the relation between timing of intention and brain activity. Libet had as his source of inspiration, the discovery of what was called 'Bereitschaft-potential' or 'readiness potential' (RP), by Kornhuber and Deecke (1965). They had found that, performance of 'self-paced' voluntary acts was preceded by a slow electrical change recordable on the scalp at the vertex, and the onset of this electrical indication of certain brain activities preceded the actual movement by up to one second or more.
Through a self-designed ingenious experiment, Libet went on to analyze the implications of the said finding on human free will. In the experiment, the subjects were to report a 'clock-time' (*Figure 1*) at which he/she was *first aware* of the wish or urge to act, where the act consisted in a simple flick or flexion of the wrist at any time they felt the urge or wish to do so. These voluntary acts were to be performed capriciously, free of any external limitations or restrictions (Libet *et al.*,1982).

7 This has been inevitable in the post-Freudian and post-existentialist intellectual world. After Freudian interventions, it has become virtually impossible to attribute full fledged autonomy to 'persons' or to take their 'reasons' for voluntary actions at face value without invoking a psychoanalytical approach as well alongside the main course of enquiry- whether scientific or otherwise. Similarly in the post-existentialist world, autonomy of subject, whether factual or axiological has been re-cast from the perspective of the subject's embeddedness in her socio-cultural as well as ecological environment. However, it shouldn't be thought that we have two water-tight compartments available here for exclusive treatment. As we have hinted, it's rather just a matter of priority. As we will shortly see, even in the empirical approach the psychoanalytical norms and insights may play a crucial role. (See below the discussion of the work of Schlegel *et al.*).

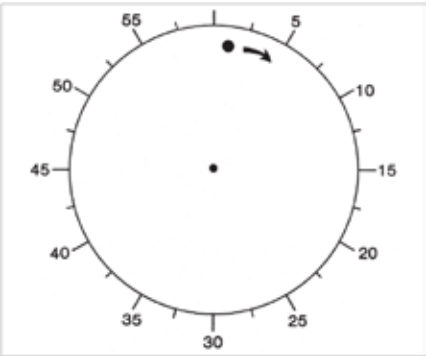


Figure 1
Oscilloscope 'clock'. Spot of light revolves around periphery of screen, once in 2.56 sec. (instead of60 sec. for a sweep-second hand of a regular clock). Each marked off 'second' (in the total of 60 markings) represents 43 msec. of actual time here. The subject holds his gaze to the centre of the screen. For each performed quick flexion of the wrist, at any freely chosen time, the subject was asked to note the position of the clock spot when he/she first became aware of the wish or intention to act. This associated clock time is reported by the subject later, after the trial is completed. (From Libet *et al.*,1983).

The clock set for the experiment was much faster than the usual clock, in order to accommodate time differences in the hundreds of msec (fig.2). Each RP was obtained from an averaged electrical recording in 40 trials. In each of these trials the subject performed the sudden flick of the wrist whenever he/she freely wanted to do so. After each of these trials, the subject reported W, the clock-time associated with the first awareness of the wish to move (Libet *et al*., 1983). The results of many such groups of trials are diagrammed in fig. 2.

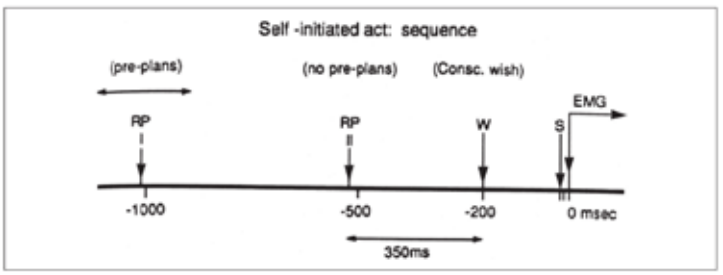


Figure2.
Diagram of sequence of events, cerebral and subjective, that precede a fully self-initiated voluntary act. Relative to 0 time, detected in the electromyogram (EMG) of the suddenly activated muscle, the readiness potential (RP)(an indicator of related cerebral neuronal activities) begins first, at about –1050 ms. when some pre-planning is reported (RP I) or about –550 ms. with spontaneous acts lacking immediate pre planning (RP II). Subjective awareness of the wish to move (W) appears at about –200 ms., some 350 ms. after onset even of RP II; however, W does appear well before the act (EMG). Subjective timings reported for awareness of the randomly delivered S (skin) stimulus average about –50 ms. relative to actual delivery time. (From Libet, 1989.)

In Libet's own words, the outcomes of the experiment summarized thus:

> For groups in which all the voluntary acts were freely spontaneous, with no reports of rough preplanning of when to act, the onset of RP averaged –550 msec. (before the muscle was activated). The W times for first awareness of wish to act averaged about 200 msec.- for all groups. This value was the same even when subjects reported having preplanned roughly when to act! If we correct W for the –50 msec. error in the subjects' reports of timings of the skin stimuli, we have an average corrected W of about –150 msec. Clearly, the brain process (RP) to prepare for this voluntary act began about 400 msec. before the appearance of the conscious will to act (W). This relationship was true for every group of 40 trials and in every one of the nine subjects studied. It should also be noted that the actual difference in times is probably greater than the 400 msec; the actual initiating process in the brain probably starts before our recorded RP, in an unknown area that then activates the supplementary motor area in the cerebral cortex. The supplementary motor area is located in the midline near the vertex and is thought to be the source of our recorded RP (1999).

The overall implication of the outcomes being that the volitional process is initiated unconsciously.
However, in sequels to the original work, Libet has continued his enquiry, coming out with significant rejoinders to the initial findings. In the latest rejoinder (Libet, 1999) that emerged out of a similar set of experiments specifically aimed at studying the vetoing power of humans as to the conscious urge or wish for action, he has observed that the existence of a veto possibility is not in doubt. In this experiment, there were no recorded RPs with a vetoed intention to act. However, while vetoing an act at a pre-arranged time, a large RP has been observed to precede the veto. This, according to Libet, signifies that the subject had indeed been preparing to act, even though the action was aborted by the subject. Therefore he comes to the conclusion:

> The awareness of the decision to veto could be thought to require preceding unconscious processes, but the content of that awareness (the actual decision to veto) is a separate feature that need not have the same requirement (Libet, 1999).

And limiting his earlier inferences, infers now that the conscious function could still control the outcome and therefore freewill is not excluded. However, he observes:

> These findings put constraints on views of how free will may operate; it would not initiate a voluntary act but it could control performance of the act. The findings also affect views of guilt and responsibility (ibid.).

Taking the said 'awareness' to be the hallmark of human agency, we shall see below in the discussion of our phenomenological scheme, the existential significance of the functional difference between 'the awareness of the decision' and the 'content' of such awareness that Libet draws in the above excerpt. Even before that, however, we have to see the further development and refinements of these proposals of Libet's at the hands of some other neuroscientists so that we can arrive at a reasonable set of inferences pertaining to the neuro-phenomenological significance of RPs.
Though these enquiries implicitly presume identity of brain states with consciousness[8] – a

8 This observation is not specifically derived from Libet, but is of generic import. It could be argued that Libet did not find any neural correlate of 'conscious will' but merely observed that specific brain states precede the time of

hallmark of almost every neuro-scientific enquiry – they are of particular interest to us because of the direct and specific nexus that they try to establish between subjective consciousness and temporality with reference to the phenomenology of RPs. As hinted above, in our phenomenological scheme, time is that phenomenon which is crucially connected with, or rather identical with subjective consciousness. Therefore the experiments are chosen keeping the said nexus in mind.[9]

We will start with one of the most recent works on the significance of RPs carried out by Schlegel *et al.*, (2015) for this experiment occasions a re-visit of the conceptual framework of free will that we have opted. In their discussion of the empirical results, we can find a more concrete formulation of the two alternatives to the idea of free will that we have derived from that framework. Also, their observation clearly sets the paradigm of the whole genre of empirical research associated with Libet's work. It runs thus:

> In discussions of will it is useful to distinguish distal acts of willing (e.g. willing to take part in an experiment and what it entails) from proximal acts of willing (e.g. willing to move one's finger during a particular trial of that experiment). Experiments in the tradition of Libet, including our own described here, test whether acts of proximal conscious will play a causal role on each trial where a movement is made. What we and Libet and his followers have studied is whether the proximal will to make a movement at a particular time - what Mele (1992) calls a "proximal intention" to move - plays a causal role in the sequence of events that include the RP and the subsequent motor act (Schlegel *et al.*, 2015, p. 201).

Precisely, inquiry into this 'proximal intention' and its causal role referred to have characterized the experiments in the tradition of Libet. And in the above excerpt, the experimenters have confessed to this fact. Hence our categorization of such a course of enquiry in our analysis under the head, 'free will in action as the consequence of the subject's conscious consideration' clearly stands vindicated.[10]

Just as we have observed, our latter category, i.e. 'free will as the subject's reasons for acting'[11], having been not tested so far on similar lines, has rather attracted a sociological or psychoanalytical approach. This fact too has been highlighted by the same researchers in their concluding remarks:

> Given our own and Libet's data, it is reasonable to conclude that proximal conscious willing is not a necessary cause of action in the kinds of cases our experiment tested. But it is important to acknowledge that there are other scenarios where conscious willing might cause movement, as described above. The findings here are restricted to a very special class of actions, so they should not be generalized into broad conclusions

conscious awareness. But it has to be noted that for some phenomenon 'to precede' some other, there should already be a scaffold of temporality existing- whether conscious or not. And without implicitly presuming some kind of identity and continuity of the 'willing subject' with such a temporal scaffold, Libet couldn't have proposed the said 'precedence'. In this sense, the observation applies to Libet as well.

9 It may be noted that there are many other empirical endeavors following the same line. See for example, the crucial work carried out by Schurger et al. available at: https://www.pnas.org/content/109/42/e2904
We consider our choice of experiments to be thematically representing all such efforts.

10 It may be argued that both proximal and distal intentions can be guided by 'considerations'. But the question here is whether such 'consideration' is proximally conscious or not.

11 It may also be argued that the notion of distal intention does not necessarily translate into the notion of reason for acting. However it does translate into reason for 'not acting now'- which too is a manifestation of free will.

> about all actions or about free will in general. Future experiments should move beyond the Libet tradition to test whether distal conscious intentions and willing play a causal role in subsequent actions (2015, p. 202).

With that observation, let us analyze what made them reasonably conclude that 'proximal conscious willing is not a necessary cause of action'.
Briefly, the experiment consisted in making the subjects watch specifically edited and sequenced video clips over a predetermined time interval, during which they would be subjected to hypnotic induction. The hypnotic induction included a post-hypnotic suggestion instructing subjects to squeeze a stress ball in their left or right hand, according to the red arrow on the screen, once during each video clip at a time of their choosing. This way both RP and lateralized RP (LRP) were accounted for.[12]
The readings from the experiment showed that both an RP and LRP occurred even when subjects performed a motor act without being conscious of having commanded it, for due to the post-hypnotic suggestion, they perceived that the ball squeezes happened due to external forces, thus indicating that the RP and LRP may be unrelated to the subjective experience of intentional movement. This is how they concluded: "While the exact nature of the neural processes reflected by the RP remains unclear, we can conclude that those processes are not specific to conscious willing" (Schlegel *et al.*, 2015, p. 201).
As we noted earlier, the work of Schlegel et al. typifies a neuro-scientific experiment wherein psychoanalytical methods and principles are at work. We have already commented on the inevitability of adopting such an approach. However, doubts may be cast as to whether the fact that subjects did not report that the action was voluntarily executed (but was rather imposed by an external force) is really enough to infer that these were not voluntary actions? The response to this will be that, when it comes to deciding 'voluntary nature' of an action from a phenomenological point of view, the necessary condition is the subject's conscious experience of such action, which is lacking here. It could be demonstrated that such conscious experience may be flawed, thereby proving that it is not a sufficient condition.[13] However, since our approach is set to be phenomenological, we may stick to the former view and reasonably infer from the conclusion of Schlegel et al. the following hypothesis:

> H1: The experiments in the tradition of Libet may not be sufficient to establish the nexus of the neuro-phenomenology of RPs with the conscious phenomenology of free will.

It may be recalled here that in his original experiment, Libet was not trying to explain the phenomenology of free will, but rather to establish a possible temporal connection between such experience and the onset of the neuro-phenomenon RP. It is exactly the same aspect that we are problematising here, and not the 'why' question as to such experience.
Now we shall take up for discussion two more empirical results shedding light on the same stance. Unlike the experiment just discussed, they do not involve any psychological technique like hypnotic induction in drawing the conclusions, but remain strictly analytical in their approach.
In the first experiment, Libet's experimental setup was altered in such a way that the subjects

12 While the RP is symmetric and represents the early part of the motor activity preceding movements, the LRP represents a later phase and is lateralised to the hemisphere contra-lateral to the side of movement.
13 See for example, (Wegner & Thalia, 1999).

had to press a button with one of their hands, the choice of which depended on a visual cue (C.S. Herrmann et al., 2008). They found that the RP set in before the stimulus was presented-thereby indicating that the RP was compatible with different outcomes. Also the participants reacted appropriately to the cue.

Hence their conclusion is that, RP cannot be decisive as to which of the two alternative courses of action available (right-hand vs. left-hand movement in this case) are executed in any such experimental setup. Rather, it seems to reflect a general expectation or an unspecific motor preparation for action (of both hands in this case).Technically, this result provides evidence against what they call the stronger hypothesis of Libet, viz. "the brain decides to initiate or, at least, to prepare to initiate the act before there is any reportable subjective awareness that such a decision has taken place" (Libet, 1985).

However, as we can readily see, it does not rebut what is called the weaker hypothesis viz. 'the initiation of the free voluntary act appears to begin in the brain unconsciously, well before the person consciously knows he wants to act!' (Libet, 1999). It may be noted that in order to defend our hypothesis H1, evidence against the stronger as well as the weaker hypothesis of Libet is needed. Therefore, let us discuss now, another recent empirical result which seems to provide such evidence. And in fact this result is more significant to the paradigm of free will that we have set so far.

The results of experiments conducted by Susan Pockett and Suzanne Purdy (2010) of University of Auckland, confirm the stance that we have just seen above as to the status of Type II RP (i.e. LRP) with regard to free will argument. They too rebut the 'strong hypothesis' of Libet by way of a similar experimental strategy. But their analysis of the Type I RP has been ingenious and just as we have observed, that has put the legitimacy of even the 'weak hypothesis' at stake.

The difference in their approach was both conceptual and psychological. They asked the question whether an urge is different from a decision, and based on that point, set their experimental procedure. Libet's experiment was repeated but the experimenters compared the subjective time reports elicited by Libet's original instructions which emphasized spontaneity, with those elicited by a new set of instructions. The new set of instructions was designed to eliminate spontaneity and focus all of the subjects' attention in the pre-movement period on a definite decision about which of two fingers to move. This required the subject to add two numbers, a different pair for each trial, which appeared in the centre of the Libet clock. "If the sum was odd they were to press one key. If the sum was even they were to press an adjacent key. After each trial they were asked to report the instant of their decision about which key to press" (Pockett & Purdy, 2010, p. 39).

Our observation above that their approach had been conceptually and psychologically innovative becomes evident from the following excerpt:

> In these experiments subjects were not given a choice of whether to report "urges", "wantings" or "decisions". In the trials emphasizing spontaneity, only the word "urge" was used - the words "wanting" or "decision" were not mentioned. In the decision trials the words "urge" and "wanting" were not mentioned: the subject was asked only to report the instant at which they decided which key to press. To eliminate any subconscious bias either on the part of the subject or on the part of the experimenter, only completely naïve subjects who had never even heard of Libet's experiments were studied, and no training sessions (where the experimenter might unconsciously have reinforced a desired result) were given. As a further attempt at achieving unbiased accuracy we also inserted an accept/reject step, so that immediately after each trial the subject had the opportunity to reject that trial if they felt they had lost concentration

> momentarily and had to guess their reported time (Pockett & Purdy, 2010 pp. 39-40).

Given this scenario, they tested the hypothesis:

i. the experiments on spontaneous urges would replicate Libet's result, whereas,
ii. in the experiments on definite decisions the reported instant of decision would be shifted back in time to the start of the RP.

Just as anticipated, both Libet's original finding and the first part of the hypothesis were confirmed. However, the second part of the hypothesis,- i.e. for decision trials the reported instant of conscious decision would be shifted back in time to the start of the RP- was not only not entirely fulfilled, it rather gave way to an overall revision of the role of RP in decision trials.
The decision trials produced either no RPs at all, or RPs that started at the same time as or after the reported decision time. Further, these latter readiness potentials were both smaller and radically shorter than the "normal" RPs recorded during spontaneous movements. The most secure conclusion from these experiments in the experimenters' own words: "...is that the ERPs (event related potentials) associated with decision-related movements are different from the ERPs associated with urge-related movements" (Pockett & Purdy, 2010 p. 40).
In the decision trials, according to them, the subject's attention in the time period immediately before the movement had been completely taken up by performing the necessary calculations. Therefore the subject had no spare capacity to spend on anticipating the arrival of a "spontaneous" urge. Given this scenario, "there were no early RP components - and often no RPs at all" (ibid.).
Accordingly, they inferred that the early part of a standard RP may be more related to expectation or readiness than to specific preparation for movement. More importantly, according to them, a second implication of the results was that, even if one disagrees with the conclusion that RPs are associated with general readiness rather than movement *per se*,

> ...it may not be particularly valid to base any conclusions about the conscious or unconscious nature of *decisions*, as opposed to spontaneous urges, on Libet's experimental data. Decisions are different from urges (Pockett & Purdy, 2010 p. 44).

Thus they have proposed their results to be restricting the scope of Libet's findings to spontaneous urges, and this is done on the assumption that decisions are different from urges. Gathering from the experiments discussed so far, the inferences that have been supportive of our hypothesis H1, we have the following set of points:

1. Voluntary action can be based either on 'decisions' or on 'spontaneous urges', where 'decisions' are different from 'spontaneous urges'.
2. In the case of conscious decisions, there are no early RP components activated and often no RPs at all.
3. It is in the case of 'spontaneous urges' that Libet's postulations based on RP have any legitimate applicability, if at all.
4. Even in that case, neuroscientists are not yet certain about what all brain areas are active and what all specific processes are involved in the preparatory period prior to a spontaneous voluntary movement.
5. Notwithstanding this, there has been consistent empirical evidence that those processes need not be specific to conscious willing and action, so that RPs are likely to

> be neither necessary nor sufficient for voluntary action.

The first three points above are a direct consequence of the last experiment analyzed, wherein the conceptual difference drawn between 'decisions' and 'spontaneous urges' has been demonstrated empirically. And we have seen on a representative scale, certain experiments in Libet's tradition that have paved way to the said conceptual innovation. Also, it was from them that we formulated our hypothesis H1. Now, based on our existential experience as practical human agents if we subscribe to the conceptual difference mentioned, we can readily see the first three points to be substantiating our hypothesis H1. In addition to this, a perusal, from a practical point of view, of the various experiments in Libet's tradition- not only the ones we have taken up for discussion, but others as well- will convince us as to the validity of the points 4 and 5 above.

As such, we can comprehend from the above compilation of points that empirical neuroscience has not yet been able to explain the phenomenon called RP, though the efforts at de-linking the phenomenology of it from the question of free will have succeeded to a certain extent.

All the same, the fact that RPs do occur conjoined with human action has not been rebutted yet. Then the questions remain: What are RPs ? What do they stand for? And what existential significance do they assume in the psycho-physical being of humans? These are the questions that we will address from an existential-phenomenological perspective towards the end of the paper. Even before that, we have to analyse a few more empirical studies that will orient us towards that end.

4. A Recent Stance

As indicated above, efforts are still on towards having a more satisfactory account of the neural processes reflected by the RPs and their implications on human action and life at large. And now we are bound to discuss such a proposition reached by K.S. Baker *et al.* (2011) as a result of a set of neuroscientific experiments rigorously planned in the light of the facts noted above, and towards answering exactly the same question posed above -a fact clearly expressed in the following excerpt from the abstract of their work published:

> The initiation of voluntary action is preceded by up to 2 s of preparatory neural activity, originating in premotor and supplementary motor regions of the brain. The function of this extended period of pre-movement activity is unclear. Although recent studies have suggested that pre-movement activity is influenced by attention to action, little is understood about the specific processes that are involved in this preparatory period prior to voluntary action. (p. 3303)

Just like other experiments in Libet's tradition, in this one too, RPs were recorded averaged from EEG activity as participants made voluntary self-paced finger movements. However, the experimenters manipulated the processing resources available for action preparation using a certain strategy. The strategy adopted was that of combining a voluntary movement task with a secondary task in which the relative difficulty or "load" was varied parametrically, thereby limiting the degree to which resources could be allocated or engaged in the preparation for voluntary action. The two "load" tasks that were involved were namely perceptual and cognitive load tasks.

The perceptual load task involved the detection of pre-specified target letters amongst a rapid sequential stream of distractor letters, with load manipulated by increasing the visual similarity of target and distractor letters, thereby engaging limited-capacity selection processes.

The cognitive load task was a version of the n-back task designed by Owen *et al.* (2005) in which

participants were presented with a stream of letters and were required to match the current stimulus with the one presented up to two items ago, thereby placing substantial demands on working memory.
Participants executed sequences of simple, self-paced finger movements once every 4-5 s while their EEG was recorded to measure the neural activity associated with planning for voluntary action. They were instructed explicitly not to count or otherwise try to judge the time elapsed between each movement. While performing these movements, participants also completed either the perceptual load task or the cognitive load task.
Overall, they found that:

> pre-movement neural activity was not influenced by the availability of selective attentional resources in the perceptual load task, whereas it was significantly reduced when cognitive resources were limited under conditions of high versus low working-memory load (Baker *et al*., 2011, p. 3312).

And have come to the conclusion that their findings suggest that the earliest components of planning for voluntary action, and corresponding pre-movement neural activity, are strongly influenced by cognitive control processes that share resources with working memory.
Based on previous neuro imaging studies, and their own recent work, they have proposed that it is particularly movement timing processes - such as deciding when to move, or cognitive control involved in endogenously orienting attention in time ready for action - that modulate the early neural activity prior to voluntary action. They, therefore, suggest that, when performing voluntary self-paced movements, participants focus on deciding the right moment to initiate their movements, relying on cognitive control mechanisms for selection and endogenously orienting attention in time (ibid.). And finally, we have their most important suggestion:

> We suggest that the earliest component of pre-movement neural activity, evident in the readiness potential, represents cognitive rather than obligatory motor planning processes, perhaps representing the cognitive control processes needed for focusing or orienting attention in time to that crucial point of movement execution, readying the system for action (ibid.).

Why this suggestion is important to us, we will see in the next section. Before that, corroborating these findings and suggestions, we have the results from another experiment (what may be called a sequel to the present one) conducted towards the end of the same year by Baker *et al.* (2012). In the experiment EEG was recorded throughout a time reproduction task in which participants replicated the interval between two tones with two button-press actions. The first action, i.e. the beginning of the reproduced interval, was somewhat incidental to the task of time reproduction and required minimal attention to the time of initiation, while the second action required explicit attention to the time of initiation.
Here is the abstract of their findings:

> Pre-movement neural activity preceding the first, relatively unattended movement was greatly reduced in amplitude and almost absent in the early stage, in contrast with readiness potentials typically seen prior to voluntary movement. Neural activity preceding explicitly timed movements was significantly larger, with effects emerging in the early component of pre-movement activity over frontal and right frontal scalp regions (Baker *et. al*., 2012 p. 715).

Based on these findings, the experimenters have proposed that attention to movement timing, i.e. the process of orienting attention in time towards the moment of movement initiation, is a key component of voluntary action preparation that is reflected in the early-stage neural activity typically seen prior to voluntary movement.

5. Relevance of the Findings to the Phenomenological Scheme

The overall empirical method involved in the first of the last mentioned two experiments has been that of specifically examining the influence of perceptual and cognitive resources on readiness potential amplitudes i.e. the preparatory neural activity, originating in pre-motor and supplementary motor regions of the brain. This method was opted by the experimenters, keeping in mind the fact that the function of such RPs remained unclear and hence, with the object of shedding some light on the same. And they have come out with the suggestion that 'the earliest component of pre-movement neural activity, evident in the RP, represents cognitive rather than obligatory motor planning processes, perhaps representing the cognitive control processes needed for focusing or orienting attention in time to that crucial point of movement execution, readying the system for action'.

In the second experiment, the specific contribution of cognitive processes, proposed to be occurring during the earliest stage of planning voluntary actions has been investigated more thoroughly. This was done keeping in mind the suggestion from the first set of findings, i.e. attention to the timing of movement is a key voluntary process contributing to early-stage cortical activity. And the experimenters have come out with the proposition that 'attention to movement timing, i.e. the process of orienting attention in time towards the moment of movement initiation, is a key component of voluntary action preparation that is reflected in the early-stage neural activity we typically see prior to voluntary movement'-which clearly is confirmative of their earlier suggestion.

Now, this is only a proposition. Also, the demarcation of cognitive faculties from perceptual faculties has always been a norm of the neurosciences in terms of empirical as well as analytical convenience. However, in the conceptual schemes prevalent in the Philosophy of Mind, the said demarcation has not been taken so seriously except in the Merleau-Pontyan phenomenological alternative, which we have set as our standard. In this phenomenological scheme, ontological primacy is attributed to perceptual consciousness, whereas intellectual consciousness is conceived to be generative of temporality and thereby subjectivity in a derivative manner. Viewed from this perspective, the above findings and suggestions could be seen to be assuming immense existential significance. Under the Merleau-Pontyan scheme of human embodiment and perception we will be able to existentially harmonize these 'facts' pertaining to free will with that essential 'value' idealized in the conception of human freedom. Towards this end, we will first try to have a comprehensive grasp of this phenomenological scheme, and in that light, take up the findings again.

6. Merleau-Ponty's Scheme of Intellectual- and Perceptual-Consciousness.

In his phenomenological account of human embodiment and perception, Merleau- Ponty draws a crucial distinction between perceptual and intellectual consciousness. Such distinction is drawn based on the peculiar phenomenology of our perceptual experience. From the phenomenology of sense perception, according to him, it is evident that it takes place purely in the spatial dimension of one's being where the subject of perception remains anonymous and opaque. Whereas subjectivity at the level of perception is attributed retrospectively, thereby invoking the temporal dimension and historicity of one's being. Thus, in that account, while perceptual consciousness is conceived to be identical with the spatial dimension of existence, intellectual consciousness is identified with the temporal dimension.

This scheme of perception is founded upon the idea of human subjectivity as a necessarily embodied point of view, wherein the body is itself already the concrete agent of all the

perceptual acts. Merleau-Ponty terms this conceptualization of embodiment, 'lived body'[14]. According to him, the lived body is the synthesizing point of perceptual events. And this synthesis happens always in the present. And in this process, the lived body projects around the present, 'a double horizon of past and future'- thereby giving the perceptual events a historical orientation.

While explicating the phenomenology of sense experience in his magnum opus, *Phenomenology of Perception*[15], Merleau-Ponty puts forth the idea thus:

> Perception is always in the mode of the impersonal 'One'. It is not a personal act enabling me to give a fresh significance to my life. The person who, in sensory exploration, gives a past to the present and directs it towards a future, is not myself as an autonomous subject, but myself in so far as I have a body and am able to 'look' (*PP*, 279).

In other words, the body 'takes possession of time; it brings into existence a past and a future for a present' (ibid.), and thereby creates time instead of submitting to it. And all this happens as a correlate of its eternal perceptual being which, as we have stated already, is impersonal and anonymous.

Now, the 'history' so created is not a genuine history in the temporal sense. "Rather than being a genuine history, perception ratifies and renews in us a 'prehistory' " (ibid.) That is, though the perceptual synthesis marks the present, the realization of such synthesis doesnot take place in the present. In other words, perception fails to realize the synthesis of its object simultaneously. And in this fact is contained the genesis of time.

> It fails at this moment to realize the synthesis of its object, not because it is the passive recipient of it, as empiricists would have it, but because the unity of the object makes its appearance through the medium of time, and because time slips away as fast as it catches up with itself (ibid.).

Thus unlike the empiricist view point of human embodiment as a passive recipient of sensory inputs, in the Merleau-Pontyan scheme, body takes up an active role in perception, but fails to simultaneously realize this synthesis because this synthesis itself is generative of time, or simultaneity for that matter, the medium through which the object has to make its appearance. Merleau-Ponty admits it to be true that one finds, through time, later perceptual experiences interlocking with earlier ones and carrying them further. But he reminds us that nowhere in such experience does one enjoy absolute possession of oneself by oneself, since "the hollow void of the future is for ever being refilled with a fresh present" (ibid.).

Thus subjectivity as a temporal synthesis of the perceptual experience of the object appears always on the horizon of such experience. Now, such formation of subjectivity and the synthesis of the object is not any culmination but an ongoing and everlasting dialectical process, for every synthesis is both exploded and rebuilt in (or by the generation of) time.

> There is no related object without relation and without subject, no unity without unification, but every synthesis is both exploded and rebuilt by time which, with

14 This concept forms the foundation of all his philosophy. Since an exhaustive account of the same is impossible within the limits of this paper, we touch only on the necessary and relevant aspects for our purposes. For a quick grasp, see the relevant sections of: (Baldwin, 2003).

15 Hereafter referred to as *PP*. We confine the scope of our discussion to the ideas of the thinker contained in this seminal work.

> one and the same process, calls it into question and confirms it because it produces a new present which retains the past. The duality of *naturata* and *naturans* is therefore converted into a dialectic of constituted and constituting time (*PP*, 279-280).

Viewed from the subjectivity point of view, this dialectic amounts to the ratification and renewal of a prehistory in terms of perception, and viewed from the temporality point of view, the hollow void of the future is refilled with a fresh present. Later on, while dealing exclusively with temporality, Merleau-Ponty further clarifies this inextricable nature of time and subjectivity thus: "To analyse time is not to follow out the consequences of a pre-established conception of subjectivity, it is to gain access, through time, to its concrete structure" (*PP*, 477).
And this exactly is what he does there. He analyzes time and space and lays bare the 'inception of subjectivity' in one's being- purged of any prejudiced conception of the same. He asserts that in the present and in perception, one's being and consciousness are at one. However, this is so, not because one's being is reducible to the knowledge one has of it, or it is clearly set out before her, but because perception is opaque.

> ... for it brings into play, beneath what I know, my sensory fields which are my primitive alliance with the world.- but because 'to be conscious' is here nothing but 'to-be-at . . .' ('être à . . .'), and because my consciousness of existing merges into the actual gesture of 'ex-sistence' (*PP*, 493).

He reminds us the fact that "it is by communicating with the world that we communicate beyond all doubt with ourselves"(ibid.). According to him, "we hold time in its entirety, and we are present to ourselves because we are present to the world" (ibid.). And hence for him, time itself is subjectivity.

> Time is 'the affecting of self by self '; what exerts the effect is time as a thrust and a passing towards a future: what is affected is time as an unfolded series of presents: the affecting agent and affected recipient are one, because the thrust of time is nothing but the transition from one present to another. This ek-stase, this projection of an indivisible power into an outcome which is already present to it, is subjectivity (*PP*, 494-5).

Thus, according to Merleau-Ponty, perception is not a cognitive activity, but the realization of a generic and primitive 'self-world relation' characteristic of human existence, in terms and by means of the 'lived body' and instances of objective reality. And this takes place not in the full light of the day. Rather, what the sentient subject aims at in perception is recognized only blindly, in virtue of the lived body's familiarity with it and the subtle mechanics of the body schema.[16]
Also, in this scheme, 'existence' or 'subjective consciousness' which is sine qua non for cognitive faculties is equivalent to temporal consciousness. And this is accomplished only by being in the world strictly embodied and engaged in action, voluntary or otherwise. Now the very phenomenology of such 'being in the world' objectively spoken is the 'happenings' in the world

16 An elaboration of this concept is beyond the scope of this paper. Briefly, body as a 'system of present positions, as well as one open to an infinite number of equivalent positions directed to other ends' (*PP*,163) - poised and ready to anticipate and incorporate a world prior to the application of concepts and the formation of thoughts and judgments is what he calls 'body schema'.

that such embodiment is subject to, whether pertaining to one's own body or otherwise. But such happenings can only be there for such an 'existence'. The moment such an 'existence' grasps and seizes of such a 'happening', marks the upsurge of time, and simultaneously the upsurge of subjectivity. Rather it is only as time that such upsurge of subjectivity can happen, for the world is already there with the embodiment, as a pre-reflective communication, called 'perception'. Thus in the Merleau-Pontyan scheme, human action in general, doesn't necessarily presume subjective consciousness, or 'free will' in the technical sense for that matter. Only when the action is voluntary, again in the technical sense so far used, subjective consciousness is posited in the mode of creating temporality. And for that very reason, its phenomenology will strictly be retrospective for the 'subject' so constituted.

7. Existential-Phenomenological Interpretation of the Experimental Results

We now attempt an interpretation of the experimental results discussed so far, with reference to the phenomenological scheme explicated above. For this, it would suffice to focus on the inferences from the last two experiments that we discussed, as these experiments were presented as the culmination of our selective survey of scientific literature revolving around the phenomenon of RP and its bearing on the free will debate. Once again we remind ourselves that the general criterion for the selection of empirical data has been their stress on the nexus between consciousness and temporality. However in the last two experiments, we have seen such nexus being established based on a demarcation of cognitive faculties from perceptual faculties- which scheme conceptually resonates with the Merleau-Pontyan scheme of embodiment and perception that we have elaborated. Based on this resonance, let us try to understand the existential significance of the findings.

The inference from the first experiment suggests that the earliest component of pre-movement neural activity, evident in the RP, represents cognitive rather than obligatory motor planning processes. And the inference from the second experiment confirms this suggestion with more specific details. There, attention to movement timing, i.e. the process of orienting attention in time towards the moment of movement initiation, is proposed to be the key component of voluntary action preparation reflected in the early-stage neural activity that is typically seen prior to voluntary movement.

The first inference above could be seen to be making clear sense in the light of Merleau-Pontyan conception of human embodiment as the 'lived body' and its existential phenomenology. As we have seen already, perceptual and motor capacities of the lived body are accomplished anonymously without any subjective intervention. So RPs could consistently be conceived as having no role in these processes. Now this stance can accommodate the whole of neuro-scientific findings that we surveyed so far. However through them RPs have been confirmed to be real phenomena as well. Hence it remains to be explained what their real significance might be. We turn to this task now.

We have seen the suggestion from the first experiment, that the earliest component of pre-movement neural activity perhaps be representing 'the cognitive control processes needed for focusing or orienting attention in time'. And in the second experiment we have seen a probable confirmation of the same. Also we are convinced by now that there is ample neuro-scientific evidence for the scientists to put forth and defend such a proposition. However, there seems to be a conceptual redundancy in their propositions, inasmuch as they propose 'the cognitive control processes' to be needed for 'focusing or orienting attention in time'. This redundancy can be exposed simply by asking the question: 'orienting' whose attention? As a matter of fact, such redundancy is typical of the entire empirical tradition of neuroscience inasmuch as any attempt at answering this question immediately invokes what is notoriously known as the 'homunculus riddle'. So, within the present conceptual scheme unconsciously followed by neuroscientists and other stakeholders in the area of brain and mind research,

such propositions seem to end up in some kind of an infinite regress. And this happens even when the experimenters have impressed us by making their findings empirically rigorous and logically inevitable, as has been the present case.

So what is wrong in the conceptual scenario? Wherefrom does such redundancy creep in? Just as we have hinted above, it is from nowhere other than the traditional conception of human embodiment and the resultant wrong idea of the 'mental'. The 'mental' when identified with the 'conscious', and thereby 'selfhood' and 'freedom' are attributed to it, such regression of 'homunculi' is bound to emerge. However, when the 'mental' is acknowledged, as we have explicated above, to have two aspects to it- namely, perceptual and intellectual, such discrepancy can be checked.

We have seen that, in the Merleau-Pontyan scheme, through such an approach the problem is sensibly addressed and resolved, so that there does not arise any such redundancy at all. In that scheme, the perceptual aspect which is essentially spatial in nature is conceived to be that passive and unconscious 'natural self' upon which we may construct any idea of 'selfhood' and 'freedom', only retrospectively through the creation of time. And such sense of temporality is what subjective consciousness is. Therefore, the question, 'orienting whose attention in time?' doesnot arise there at all, because creation of the temporal dimension itself marks attention or the creation of subjective consciousness. Rather, it is only as time that the upsurge of subjectivity can happen.

Thus, it is the temporal aspect of human existence that is crucial in invoking subjectivity and retaining it, and the whole idea of free will is founded on subjectivity so invoked. And given this scenario, we can readily make sense of the RPs and thereby the whole sequence of neuro-scientific experiments that we have related so far. Seen in the light of this existential scheme, RPs could safely be conceived as neuro-phenomenological markings of the invocation of temporality or subjectivity. And interestingly, the same fact has been specifically inferred in the second experiment above.

8. Reasons and Persons in a Merleau-Pontyan Version of Compatibilism

We had started our discussion by observing the derivative connection between the phenomenal aspect of free will i.e. action, and the non-phenomenal aspects - decisions and intentions as either:

i. the consequence of the subject's conscious consideration, or (at the least)
ii. the subject's reasons for acting.

By now, our explication of the theme of free will seems to be mature enough for a revisit of this bifurcation. As we had promised, the first part of it has been addressed from an empirical point of view, briefly surveying the legacy of related empirical research as well as discussing some of the most relevant and contemporary research findings. We culminated the discussion by seeing for ourselves how accommodative of those findings has the Merleau-Pontyan idea of human subjectivity been. While proposing the said bifurcation, we had observed that the latter alternative had been inspiring rather a sociological and psychoanalytical approach, in contrast to the dominant empirical approach of the neurosciences as to the former. Let us now focus on this latter aspect with a view to bring to light a unique version of compatibilism that is inherent in the Merleau-Pontyan thought.

As we have seen, in the Merleau-Pontyan scheme, we can conceive any idea of 'selfhood' and 'freedom' only on the basis of that 'natural self' which is formed by the embeddedness of the lived body in the world. Now, the so formed autonomy, in an analogous way is embedded in the life-world of inter-subjectivity. Hence we will first try to have a better grasp of the notion of subjectivity in the context of the life-world and inter-subjectivity.

We have seen how the phenomenology of temporality invokes subjective consciousness. Once invoked, interestingly, it has the potential of assuming the whole of one's being. We have Merleau-Ponty's account of the situation and its bearing on the idea of free will as follows:

> Consciousness holds itself responsible for everything, and takes everything upon itself, but it has nothing of its own and makes its life in the world. We are led to conceive freedom as a choice continually remade as long as we do not bring in the notion of a generalized or natural time (*PP*, 526).

But in the Merleau-Pontyan phenomenological scheme, we have seen that there is no natural time, if we understand thereby a time of things without subjectivity. And this is the natural consequence of the conception of temporality and subjectivity in that scheme. Therefore the situation mentioned above is unavoidable. There is, however, at least a generalized time- he acknowledges. And this is what the common notion of time envisages- "the perpetual reiteration of the sequence of past, present and future"(ibid.). However as we have seen in Sec.6 above, the phenomenology of such temporality is peculiar:

> It is, as it were, a constant disappointment and failure. This is what is expressed by saying that it is continuous: the present which it brings to us is never a present for good, since it is already over when it appears, and the future has, in it, only the appearance of a goal towards which we make our way, since it quickly comes into the present, whereupon we turn towards a fresh future (ibid.).

Thus, this time is not the objective time of a 'cogito'. Rather it is

> ...the time of our bodily functions, which like it, are cyclic, and it is also that of nature with which we co-exist. It offers us only the adumbration and the abstract form of a commitment, since it continually erodes itself and undoes that which it has just done (ibid.).

This form of temporality is what we have analysed so far with the help of neuroscientific data. We have postulated RPs to be marking the neuro-phenomenology of the same. In the Merleau-Pontyan scheme, this is the foundational and generalized time which facilitates the inception and support of subjectivity known in the philosophical circles as the 'For itself'. Thus, as we have already seen, this subjectivity is foreshadowed by the 'natural self' (or the so called 'In itself' in the philosophical circles) formed of the lived body existentially embedded in the world. And according to Merleau-Ponty,

> As long as we place in opposition, with no mediator, the For Itself and the In Itself, and fail to perceive, between ourselves and the world, this natural foreshadowing of a subjectivity, this prepersonal time which rests upon itself, acts are needed to sustain the upsurge of time, and everything becomes equally a matter of choice, the respiratory reflex no less than the moral decision, conservation no less than creation (ibid.).

Now, this exactly is what has happened in the philosophical circles. They have totally ignored the said natural foreshadowing of subjectivity and taken for granted what they call 'free will'. And then, everything is seen as a matter of choice. But as far as the scheme that we have adopted, we must remind ourselves with Merleau-Ponty that: "...consciousness attributes

this power of universal constitution to itself only if it ignores the event which provides its infrastructure and which is its birth" (ibid.).
We have seen this 'event' to be 'perception' as it is felt as the embeddedness of the lived body in the world. And according to Merleau-Ponty, "A consciousness for which the world 'can be taken for granted', which finds it 'already constituted' and present even in consciousness itself, does not *absolutely* choose either its being or its manner of being" (*PP,* 526-27).
What then is freedom?- he immediately raises this question and responds. And in his response we have an abstract statement of the version of compatibilism that we are trying to trace. It runs thus:

> To be born is both to be born of the world and to be born into the world. The world is already constituted, but also never completely constituted; in the first case we are acted upon, in the second we are open to an infinite number of possibilities. But this analysis is still abstract, for we exist in both ways *at once*. There is, therefore, never determinism and never absolute choice, I am never a thing and never bare consciousness (*PP,* 527).

In our exposition of the Merleau-Pontyan scheme, we have seen the details of the event of the birth of consciousness referred to. And we already know that for a subjectivity so formed, the world cannot be taken for granted, for it is always in and of the world, that too as an intersubjectivity. Here is Merleau-Ponty's account of the cohesion of an intersubjective life and a world in terms of the phenomenology of the 'generalized time' that he has defended:

> If the subject made a constant and at all times peculiar choice of himself, one might wonder why his experience always ties up with itself and presents him with objects and definite historical phases, why we have a general notion of time valid through all times, and why finally the experience of each one of us links up with that of others. But it is the question itself which must be questioned: for what is given, is not one fragment of time followed by another, one individual flux, then another; it is the taking up of each subjectivity by itself, and of subjectivities by each other in the generality of a single nature, the cohesion of an intersubjective life and a world (*PP*, 525).

We have already seen how time 'is the taking up of each subjectivity by itself', and how, that way, it resolves human action to be 'the consequence of the subject's conscious consideration'. Now we will see in some detail, how time is the taking up of subjectivities by each other in the generality of a single nature- 'the cohesion of an intersubjective life and a world'. And therein we will locate what we have termed: 'subject's reasons for acting' and thereby unveil a unique version of compatibilism contained in the Merleau-Pontyan phenomenological scheme.
We have him continuing:

> The present mediates between the For Oneself and the For Others, between individuality and generality. True reflection presents me to myself not as idle and inaccessible subjectivity, but as identical with my presence in the world and to others, as I am now realizing it: I am all that I see, I am an intersubjective field, not despite my body and historical situation, but, on the contrary, by being this body and this situation, and through them, all the rest (ibid.).

Now, that marks the whole difference of the scheme. Human subjectivity for Merleau-Ponty is an intersubjective field, not despite its embodiment and historical situation, but, on the contrary, by verily being such embodiment and its situation. Now, how is this synthesis

carried out? He answers that it is by taking up the present. By 'taking up a present', one draws together and transforms one's past- altering its significance, freeing and detaching oneself from it. But one does so only by committing oneself 'somewhere else' (*PP*, 528).
This 'somewhere else' is facilitated in the normal case by the life-world in general. And subject's reasons for actions are to be sought in that inter-subjective field. In the peculiar context of psychoanalytic treatment, the inter-subjective situation is more specific and consequently the retrospective reasoning is more obvious. Just like how he brings out the phenomenology of perception through analyzing the so called 'pathologies'[17], Merleau-Ponty brings out the significance of inter-subjectivity in the realization of freedom by analyzing this peculiar context.
According to him, psychoanalytical treatment does not bring about its cure by producing direct awareness of the past. Rather this is accomplished primarily by binding the subject to his doctor through new existential relationships. On the part of the patient, it is not a matter of giving scientific assent to the psychoanalytical interpretation, and discovering a notional significance for some situation in the past. Rather, it is a matter of reliving that situation as significant. And this the patient succeeds in doing "only by seeing his past in the perspective of his co-existence with the doctor. The complex is not dissolved by a non-instrumental freedom, but rather displaced by a new pulsation of time with its own supports and motives" (*PP*, 529).
Thus the general notion of 'intersubjective field' is instantiated in the context of psycho-analytic treatment as well. Here the field is formed as a binding of the subject to his doctor through new existential relationships. The field so formed facilitates for the patient, a re-living of the past situation by way of seeing it in the perspective of his co-existence with the doctor. And in that process, 'the complex is not dissolved by a non-instrumental freedom, but rather displaced by a new pulsation of time with its own supports and motives'.
We have just seen that this 'new pulsation of time' is what is creative of the consciousness of free will (in the form of subjectivity) in tune with the first aspect of our conceptual bifurcation. Hence, the above observation applies not only in the context of psychoanalysis, but in general to all cases of 'coming to awareness'. "...they are real only if they are sustained by a new commitment. Now this commitment too is entered into in the sphere of the implicit, and is therefore valid only for a certain temporal cycle" (ibid.).
Is it leading to determinism and there is no freedom possible? Or, don't we have any choice? 'Of course yes' - Merleau-Ponty would answer. But the choice which we make of our life is always based on a certain givenness- he would emphasize. "My freedom can draw life away from its spontaneous course, but only by a series of unobtrusive deflections which necessitate first of all following its course - not by any absolute creation" (ibid.).
If that be so, going back to the second aspect of the bifurcation, what is the significance of the 'subject's reasons for acting' in the Merleau-Pontyan scheme? And what is the bearing of such reasons on freedom?
This indeed is what is hinted at by referring above to the 'supports and motives' of the pulsation of time. What do they stand for? Can we identify them with the speculative ideas of this or that sociological or psychoanalytical theory? Or can they be merely the phenomenology of adaptive mechanisms of natural evolution experienced subjectively? Or, as has been the fashion in the traditional philosophical and theological circles, can they be the 'passions of the soul' or 'the reasons of a transcendental subjectivity'?
In all these alternatives, except the biological one, one fact is conveniently ignored. It is

17 See for example Schneider's case discussed in (*PP*, 118-228)

the fact that, man as we analyze him, is essentially an intersubjective being, and thus his psychological and historical existence are inextricably intertwined. Once we try to tear asunder the latter of these aspects from the former, we will be compelled to resort to those alternatives for an explanation of the former. And only in that context do such theories have any relevance at all. But what is the lived reality?
Merleau-Ponty continues:

> All explanations of my conduct in terms of my past, my temperament and my environment are therefore true, provided that they be regarded not as separable contributions, but as moments of my total being, the significance of which I am entitled to make explicit in various ways, without its ever being possible to say whether I confer their meaning upon them or receive it from them (*PP*, 529).

The integral nature of embodied existence and its oneness with the world is again stressed as the existential foundation of personhood. The meaning of actions, in terms of historicity and temperament, then assumes an ambiguous status. But this quality of existence is something constructive rather than self defeating. It gives the 'person', what Merleau-Ponty calls a certain 'style'.

> I am a psychological and historical structure, and have received, with existence, a manner of existing, a style. All my actions and thoughts stand in a relationship to this structure, and even a philosopher's thought is merely a way of making explicit his hold on the world, and what he is (ibid.).

Let us have a re-look at what we termed 'subject's reasons for acting'. Just as we have noted already, any attempt at explicating this idea is bound to attribute an absolute innateness to the 'subject' and hence to the 'reasons'. However when we analyze it in the light of the observations quoted above, we can readily see that such innateness is not so innate indeed. Rather, it is defined and sustained by the transcendent motivations of the life-world. And then,

> The fact remains that I am free, not in spite of, or on the hither side of, these motivations, but by means of them. For this significant life, this certain significance of nature and history which I am, does not limit my access to the world, but on the contrary is my means of entering into communication with it (ibid.).

Once this realization happens in one's being, then it is by being unrestrictedly and unreservedly what one is at present that one has a chance of moving forward. Or put in temporal terms, it is by living one's time that one is able to understand other times.
Thus, as Merleau-Ponty insists it: "...by plunging into the present and the world, by taking on deliberately what I am fortuitously, by willing what I will and doing what I do, that I can go further" (ibid.).
According to him, one can miss being free only if one tries to bypass one's natural and social situation by refusing to take it up instead of assuming it in order to join up with the natural and human world (ibid.). And this is so not because of any sociological necessity either. Rather, this is owing to the peculiar existential makeup of mankind, that we have elaborated so far. "Nothing determines me from outside, not because nothing acts upon me, but, on the contrary, because I am from the start outside myself and open to the world" (*PP*, 530).
Thus, existentially we need reasons and justifications because right from the beginning we are outside ourselves and open to the world. Verily as Merleau-Ponty puts it, all explanations of

one's conduct in terms of one's past, one's temperament and one's environment are therefore true, provided that they be regarded not as separable contributions, but as moments of one's total being consisting in a certain psychological and historical structure. And all one's actions and thoughts stand in a relationship to this structure but in an ambiguous manner. One cannot say for sure, whether he/she confers the meaning upon the structure or receives from it. Given this scenario, what do 'reasons' in the technical sense stand for? They are no more the free choices and plans emanating from an innate private 'self' or subjectivity for sure. Rather they are a kind of human responsiveness to the existential situation in which he/she finds him/herself embedded right from the beginning. And exactly this existential situation is what is creating any possibility of freedom at all, provided that he/she remains responsive, rather than refusing to take up such natural and social situation. This is why Merleau-Ponty says that by taking on deliberately what one is fortuitously that one can be free. It is a dialectical engagement rather than a positing act.

9. Conclusion

Thus in a unique version of compatibilism traceable in the thought of Merleau-Ponty, the source of reasons for human action is no more an enigma to be searched somewhere in the complex neuronal network or explained away by some ingenuous sociological or psychoanalytical theory. Nor is it situated in some innate private realm of the mental dimension of humans. It is to be sought in the very existence of each individual in the world as a thrownness into it and embeddedness in it which consists of psychological as well as historical aspects, natural as well as social aspects. As we have seen, in the Merleau-Pontyan scheme of embodiment and perception, such thrownness and embeddedness are materialized in the unique spatio-temporal phenomenology of individuals as 'lived bodies'. And in that phenomenology, the neuro-phenomenon of RPs could be intelligibly accommodated.

REFERENCES

Baker, K. S., Mattingley, J.B., Chambers, C.D., Cunnington, R. (2011). Attention and the readiness for action. *Neuropsychologia*, 49, 3303–3313;

Baker K. S., Piriyapunyaporn, T., Cunnington, R. (2012). Neural activity in readiness for incidental and explicitly timed actions. *Neuropsychologia*, 50, 715-722;

Baldwin, T. (Ed.) (2003). *Maurice Merleau-Ponty : Basic writings*. London, Routledge;

Boden, M. A. (2011). The Philosophies of Cognitive Sciences In J. Garvey (Ed.) *The Continuum Companion to Philosophy of Mind* (pp. 151-170). London, Continuum International Publishing Group;

Brass, M., Furstenberg, A., Mele, A. (2019). Why neuroscience does not disprove free will. *Neuroscience & Biobehavioral Reviews,* 102, 251-263;

Herrmann, C. S., Pauen, M., Byoung-Kyong Min., Busch, N.A., Rieger, J.W. (2008). Analysis of a choice-reaction task yields a new interpretation of Libet's experiments. *International Journal of Psychophysiology,* 67, 151-157;

Hodgson D. (2009). Criminal Responsibility, Free Will, and Neuroscience. In Murphy N., Ellis G.F.R., O'Connor T. (Eds.) *Downward Causation and the Neurobiology of Free Will.* Understanding Complex Systems Series. Berlin, Springer. doi.org/10.1007/978-3-642-03205-9_14;

Kornhuber, H. and Deecke L. (1965). Hirnpotentialanderungen bei Willkurbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pfluegers Arch Gesamte Physiol Menschen Tiere,* 284, 1-17;

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly,* 64, 354-361;

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioural and Brain Sciences,* 8, 529-566;

Libet, B. (1989). Conscious subjective experience vs. unconscious mental functions: A theory of the cerebral processes involved. In R.M.J. Cotterill (Ed.) *Models of Brain Function*. New York, Cambridge University Press;
Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies,* 6 (8-9), 47-57;
Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain,* 106, 623-42;
Libet, B., Wright, E.W., Gleason C.A. (1982). Readiness potentials preceding unrestricted spontaneous pre-planned voluntary acts. *Electroenceph.&Clin. Neurophysiology,* 54, 322-5;
Libet, B., Wright, E.W., Gleason C.A. (1983). Preparation-or intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroenceph.&Clin. Neurophysiology,* 56, 367-72;
McKenna M. (2020). "Compatibilism" In E. Zalta (Ed.), The Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu/entries/compatibilism;
Mele, A. R. (1992). *Springs of action: Understanding Intentional Behavior*. New York, Oxford University Press;
Mele, A. R. (2009). *Effective Intentions.* New York, Oxford University Press;
Merleau-Ponty, M. (1962). *Phenomenology of Perception.* Colin Smith (Tr.), London, Routledge & Degan Paul;
Moore J, G. (2016). Criminal Responsibility and Causal Determinism. *Washington University Jurisprudence Review,* 9(1). Retrieved from http://openscholarship.wustl.edu/law_jurisprudence/vol9/iss1/6;
Morse, S. J. (2003). Inevitable mens rea. *Harvard Journal of Law and Public Policy,* 27, 51-64;
Nahmias, E. (2010). Scientific Challenges to Free Will. In C. Sandis & T. O'Connor (Eds.), *A Companion to the Philosophy of Action*. (pp. 345-356) Wiley-Blackwell;
Nestor P. G. (2019). In defense of free will: Neuroscience and criminal responsibility. *International journal of law and psychiatry*, *65*, 101344. https://doi.org/10.1016/j.ijlp. 2018.04.004;
Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping,* 25(1), 46-59;
Pockett, S., Purdy, S. (2010). Are voluntary movements initiated preconsciously? -The relationships between readiness potentials, urges and decisions. In Walter Sinnott-Armstrong & Lynn Nadel (Eds.), *Conscious Will and Responsibility: A Tribute to Benjamin Libet.* New York, OUP. doi:10.1093/acprof:oso/9780195381641.003.0005;
Schlegel, A., Prescott, A., Sinnott-Armstrong, W., Roskies, A., Tse, P. U., Thalia Wheatley, T. (2015). Hypnotizing Libet: Readiness potentials with non-conscious volition. *Consciousness and Cognition* 33,196-203;
Smith, B. C. (2011). Folk Psychology and Scientific Psychology In J. Garvey (Ed.) *The Continuum Companion to Philosophy of Mind* (pp. 102-032). London, Continuum International Publishing Group;
Wegner, D. M., Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will: What cognitive mechanism makes us feel as if we are acting consciously and willfully? *American Psychologist* 54(7), 480-492.

MARCO VIOLA
Università degli Studi di Torino
marco.viola@unito.it.

THREE VARIETIES OF AFFECTIVE ARTIFACTS: FEELING, EVALUATIVE AND MOTIVATIONAL ARTIFACTS[1]

abstract

Inspired by the literature on extended/scaffolded mind, a debate concerning the contribution of extra-bodily resources to our (extended) emotions is recently gaining traction. Within this debate, inspired by the literature on cognitive artifacts introduces the notion of "affective artifacts", indicating those objects that exert persistent effects on our feelings, possibly altering our self. However, by focusing on feelings, this notion neglects other facets of emotional episodes. Following Scarnatino's tripartition between feeling, appraisal, and motivational theories of emotion, I present three varieties of affective artifacts: Feeling, Evaluative and Motivational Artifacts.

keywords

extended mind, emotions, feelings, cognitive artifacts, extended emotions, extended affectivity.

1 I am grateful to Francesco Bianchini, Giovanna Colombetti, Marco Fasoli, Richard Heersmink, Juan Loaiza, Valentina Petrolini, and Giulia Piredda for their useful comments.

Phenomenology and Mind, n. 20 - 2021, pp. 228-242
DOI: https://doi.org/10.17454/pam-2000
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

© The Author(s) 2021
CC BY 4.0 Rosenberg & Sellier
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

THREE VARIETIES OF AFFECTIVE ARTIFACTS:

1. Introduction

Over the last twenty years, several philosophers of mind have entertained the possibility that extra-cranial and extra-bodily resources contribute to mental states. According to the Extended Mind Hypothesis based on Parity Principle (Clark and Chalmers, 1998), extra-bodily objects (e.g. a notebook) can play a constitutive role in implementing mental states (e.g. beliefs), provided that certain conditions are met (e.g. that the notebook is a source always available to, and trusted by, the user). Other flavors of the Extended Mind emphasizing the Complementarity Principle (Menary, 2006), or similar theses such as the Scaffolded Mind hypothesis (Sterelny, 2010) also recognize the contribution of extra-bodily resources to mental processing, although they tend to assign them a supporting role rather than a constitutive one. Regardless of the role that external objects are playing (be it constitutive or merely supporting), the fact that they contribute to cognitive tasks qualifies them as cognitive artifacts (Heersmink, 2013; Norman, 1993).

Recently, philosophers began to wonder whether not only 'epistemic' mental states, such as beliefs, but also affective phenomena, such as *emotional episodes,* could be extended to external resources (Carter *et al.*, 2016; Colombetti, 2017; Colombetti and Roberts, 2015; Slaby, 2014; Piredda, 2019), or at least scaffolded by them (Colombetti and Krueger, 2015; Griffiths and Scarantino 2009; Krueger 2016b; Stephan *et al.*, 2014. For some discussion on the differences between the two approaches see Krueger and Szanto, 2016; Maiese, 2019). The list of external resources evoked within this debate includes other human agents (Leon *et al.*, 2017), primary intersubjectivity (Candiotto, 2016), affective affordances (Caravà and Scorolli, 2020), affective atmospheres (Slaby, 2016), affective arrangements (Slaby *et al.*, 2019), and social norms and institutions (Gallagher, 2003; Griffiths and Scarantino, 2009; see also Ekman, 1972 on display rules).

However, in this paper I am only concerned with a specific kind of external resources: namely, *artifacts* involved in emotional episodes (roughly corresponding to what Krueger [2018] calls "material scaffolding"). The literature abounds with examples of such objects, including music players (Colombetti and Krueger, 2015), autobiographical diaries (Colombetti and Roberts, 2015), and Catholic confessionals (Griffiths and Scarantino, 2009). In a recent paper, Piredda (2019) offers a philosophical analysis of such objects, which she dubs 'affective artifacts'. Inspired by the literature on cognitive artifacts (e.g. Heersmink, 2013; Fasoli, 2018), she construes affective artifacts as objects that have "the capacity to alter the affective condition of an agent, thus contributing to her affective life" (2019, p. 7); she adds that they are "objects with which the agent entertains a constant and persistent affective relationship"

(p. 8), thus contributing to the construction of one's extended self (see also Candiotto and Piredda, 2019); and she specifies that "should we lose such a personal affective artifact, this would alter our affective condition" (Piredda, 2019, p. 8).
This notion is a welcome addition to the literature, as it deals with some central aspects of our affective lives, namely our feelings. But does it account for every possible facet of emotion – or, more modestly, of occurrent emotional episodes? In this paper, I aim at broadening the scope of the notion of affective artifacts, showing that they can meaningfully interact with other facets of emotional episodes.
To do so, I will adopt a rather simple notion of artifact, bracketing many of the controversies that surround the metaphysics of artifacts (see for instance Franssen *et al.*, 2014; Preston, 2018). The Stanford Encyclopedia of Philosophy offers the following ecumenical definition of artifacts: "tools that have been designed for, and are commonly employed for, some specific function" (Preston, 2018). While the relation between proper function (the function tools have been designed for) and system function (the function tools are actually employed for) is a central knot in many discussions in the philosophy of artifacts, for the sake of simplicity in this paper I will readapt the definition above smoothing the conjunction into an inclusive disjunction:

> Artifacts $=_{def}$ tools that have been designed for, *or* are commonly employed for, some specific function.

As I will discuss, the kind of function they play will then offer a criterion according to which they can be sorted out.
My discussion will proceed as follows. First, in section §2 I present Scarantino's distinction between three clusters of emotion theories, which put the emphasis on different features of emotional episodes: the appraisals, the physiological and phenomenal feelings, the motivation push toward certain behavioral outcome. Following his tripartition, I sort out three varieties of artifacts, depending on which of the abovementioned features they primarily affect (§§3—5). Finally, I sum up and make some general considerations (§6).

2. Dissecting emotions to sort out affective artifacts

Paraphrasing Augustine's famous quote about time, Fehr and Russell once remarked that: "Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows" (Fehr and Russell, 1984, p. 464). When back in 1981 Kleinginna and Kleinginna surveyed the literature looking for a unified definition of emotions, they found 92 distinct definitions (along with 9 skeptical statements that argued against their definability), to which they added their own. Neither a consensus has emerged over the following decades. Indeed, while many theories agree on the properties that emotions in general *tend to* have, as well as on many properties that specific emotions most likely instantiate, they often disagree on which of these properties are *essential* and which are *contingent*.
To put some order into the debate, Scarantino (2016; 2018) has taken into account several prominent theories of emotion across the centuries (both scientific and philosophical), and sorted them into three traditions – perhaps at the price of some simplifications. Each tradition privileges some criteria for sorting emotional episodes into specific emotion kinds (e.g. for distinguishing episodes of disgust and fear); and for establishing what is an emotion and what is not. They are the following:

a. *The evaluative tradition.* Theories belonging to this tradition regard some kind of evaluation about salient situations or stimuli as a central feature of emotion. Importantly, this tradition branches into two distinct approaches – i.e. the constitutive and the causal.

According to the former, these core evaluations are themselves emotions (e.g. Nussbaum's neostoicism [Nussbaum, 2001]), whereas for the latter evaluation is but a necessary condition among others[1]. While these theories have the merit of nicely accommodating the normative dimension of emotions, a common objection raised against them is that they sound overly intellectualistic (but see Colombetti, 2007). Yet, they have trouble accounting for recalcitrant emotions: an example would be those situations in which a stimulus that we would not consciously consider dangerous still evokes reactions typical of fear (notably, in phobias).

b. *The feeling tradition*. According to feeling theories, an emotion is essentially a feeling. This set obviously encompasses Jamesian and neo-Jamesian theories that vindicate a pivotal role for the autonomic system's activities, as well as older and newer theories that rather emphasize phenomenal feelings (e.g., Descartes and Hume; Kriegel, 2014; LeDoux and Brown, 2017). Both a merit and a flaw can be ascribed to theories of this tradition, particularly with regard to their phenomenal variants. On the one hand, assigning a prominent role to feelings possibly makes these theories closer to the everyday notion of emotion than the alternatives (e.g. Panksepp, 2000; but see Scarantino, 2016, fn. 3). On the other hand, the well-known problems of introspection (Schwitzgebel, 2019) make it hard to operationalize them.

c. *The motivational tradition.* Theories within the motivational tradition construe emotions as specific patterns of behavior, or at least as dispositions to enact behavioral patterns. This set includes a broad range of accounts, including behaviorism (here Scarantino enlists Watson and Skinner), evolutionary psychology (e.g. the affect program tradition), but also social constructionist account (according to which emotions are culturally-shaped constructs that serve some social purpose). This tradition is rather popular among scientists because behavior is easier to measure than phenomenal states, as well as allowing for cross-species generalizations (see Adolphs, 2017). One problem with these approaches (at least from the philosophical standpoint) is that typically they fall short of accounting for the intentionality of emotions. We are afraid *of something*, we feel guilty *about something*, and so on; but patterns of behavior (or dispositions to enact them) arguably lack this aboutness.

To put it roughly, theories from the evaluative tradition classify emotions on the basis of their input or triggering conditions; those from the feeling tradition on the basis of the inner mental and/or bodily processing; whereas motivational ones on the basis of the behavioral output or function of an emotional episode.

While harsh ontological disputes are being currently fought over which features shall count as the essential properties for emotions, some scholars (notably, Scherer, 2005) ecumenically acknowledge that emotions are complex phenomena with multiple components, neither of which more essential than others. Within the debate on extended emotions, Stephan *et al.* (2014), as well as Krueger and Szanto (2016) seem to endorse this ecumenical approach and invite to focus on the extension of specific components of emotions.

Moreover, within the embodied and enactive tradition, other scholars argue that, while conceptually separable, these components are but different flips of a same coin: for instance, Colombetti (2007) proposes to construe appraisals in terms of motor planning (rather than in the disembodied and intellectualistic ways in which they are typically construed). Caruana

1 Unsurprisingly, Scarantino mentions Scherer (e.g. 2005), who champions a multi-componential account of emotions, a representative herald of this approach; see the main text below.

(2019) leverages on neuroscientific evidence of a same neural basis for experience and expression to reject an ontological seizure between the two.
In this paper I shall not commit to any theory of what features – if any – are essential to emotions. Nor I shall commit to the claim that these features are ontologically separable. Rather, my aim is to discuss how each of these features – the feelings, the appraisals, the behavioral outcome – gets influenced, and sometimes perhaps even replaced, by some artifacts. Based on which of these three features is influenced (or mimicked), I propose to speak of *evaluative, feeling* or *motivational artifacts*, respectively.
As a matter of fact, the various features of emotions are highly intertwined. Hence, an artifact's influence over one facet of emotion is likely to reverberate onto the others. This may make the ascription of the artifacts to one class troublesome. Imagine the following scenario: a Geiger counter informs me about the presence of radiation in some environment (appraisal). This would probably make me sweat (bodily feeling) and willing to leave the area as quick as possible (behavioral disposition). *Prima facie*, we may think that the Geiger counter falls within all three classes of affective artifacts.
A similar situation may likely generalize to virtually every artifact. But then, if every artifact may belong to all three classes, is my three-fold classification pointless? Not really. Indeed, just like I do not commit to any deep ontological separation between the three sets of features of emotion, neither I do commit to any strong *ontological divide* between the three classes of affective artifacts, as if they were mutually exclusive. If someone is willing to argue that the Geiger counter eventually belongs to all three classes, I have no *apriori* reasons to deny this. More modestly, by describing this tripartition I propose a conceptual tool, which can be employed to shed light on some hitherto underappreciated aspects of affective artifacts.
Nevertheless, affective artifacts can be sorted out in the three classes on a weaker, merely epistemic reading. This can be done by establishing which features are primarily or more directly affected by the artifact. In the abovementioned scenario, the effects of the Geiger counter upon our feelings and motivations are only indirect, as they are mediated by its capacity to alter our appraisal of the situation, enabling us to make an otherwise unavailable evaluation. Thus, it shall be enlisted among the evaluative artifacts. Similarly, even though some evidence suggests that one's feeling may impact the appraisals of new stimuli (Schechter and Singer, 1962; see also Petrolini and Viola, 2020), an artifact that alters appraisals by targeting our feelings shall be enlisted among feeling artifacts.
Finally, notice that the class an affective artifact (primarily) belongs to is agent-relative. A motivational artifact that plays a communicative role for agent A will likely count as evaluative artifact for one or more agents B (and C, etc.). I shall come back to this issue in §6.

3. Feeling Artifacts

To feeling theorists, the essence of emotion is some bodily or phenomenal feeling. As these theories are likely the closest to commonsense views of emotions, most scholars in the debate on extended emotions arguably assume some form of the feeling theory. In order to stress the environmental resources that play a role in emotion processing, Slaby explicitly speaks of "'tools *for feeling'* as analogous to [extended mind]'s 'tools for thinking'" (2014, p. 44. Italics mine). Less starkly, Colombetti and Krueger talk about the affective-ladenness of handbags for some (usually Western) women:

> [a] handbag—including its contents—functions as a highly portable, self-styled collection of technologies specifically chosen *for regulating affect*: charms and tokens for good luck and peace of mind, *which influence* one's appraisal of, and ability to cope with, specific situations (2015, p. 7. Italics mine).

The italicized words are meant to stress that, while the authors surely acknowledge that handbags can exert some influence over its wearer's evaluations and motivations, these effects are presented as byproducts ('which influence') of its main and direct effect ('for regulating affect').

The idea of artifacts interacting with feelings is nicely captured by the characterization of affective artifacts recently offered by Piredda (2019). To restate, Piredda defines affective artifacts as objects that have "the capacity to alter the affective condition of an agent, thus contributing to her affective life" (2019, p. 7), and "with which the agent entertains a constant and persistent affective relationship" (p. 8), and whose loss "would alter our affective condition" (*ibid.*).

Just like the authors above, Piredda correctly acknowledges that the artifacts' impact on feeling reverberates into actions and reactions (Candiotto and Piredda, 2019; Piredda, 2019). However, in her discussion, feelings (rather than other aspects of emotions) are primarily at stake, as shown for instance in the following passages:

> Photographs are an example of affective artifacts, as they elicit memories of past events and, at the same time, move something within us: they make us *feel something* (e.g. happiness, sorrow, sadness) (Piredda, 2019, p. 3. Italics mine).
> a doll, a stuffed animal or any other toy can be considered an affective artifact for a child that always wants to keep this object with her in order *to feel good*, or to remain calm (*ibid.* Italics mine).

Piredda's constual of affective artifact may then be taken to constitute an insightful characterization of the class of feeling artifacts. In more general terms, and following the definition of artifacts provided in §1, I propose the following definition:

> Feeling Artifacts $=_{def}$ tools that have been designed for, or are commonly employed for altering (i.e. regulating, inhibiting, enhancing) some bodily and/or phenomenal state.

Classic examples include wedding photo books (which arguably evoke pleasant feelings), gothic cathedrals (meant to instill awe and admiration), movies and other objects of art, or even roller coasters (prompting some mix of fear and amusement). However, some artifacts originally conceived and built for a different purpose may end up becoming a feeling artifacts. For instance, "someone could invest affective value in a post-it that was accidentally in her pocket when she received a life-changing phone call" (Piredda, 2019, p. 11). But redeployments need not be idiosyncratic: cases of social redeployment also exist. For instance, think about vintage cars: their original function of serving as vehicles is overshadowed by their affective (and aesthetic) value. The social nature of this fact is witnessed by the existence of a market and of institutionalized meetings for vintage cars owner[2].

Moreover, the literature on cognitive artifacts usually disregards objects lacking representational properties (but see Heersmink, 2013) and tends to prefer enduring objects over consumables. Yet, as long as one's theory recognizes the centrality of arousal in emotion (as most do: see Feldman Barrett and Russell, 1999), consumable substances such as coffee

2 However, as redeployment of some artifact becomes widespread and institutionalized, its novel purpose may eventually subside the original one as "primary function". Similar challenges are the bread and butter of the philosophy of artifacts (see Preston, 2018).

and marijuana will fall within the definition, as they are thought to increase (Smith, 2002) and decrease (Russell and Mehrabian, 1977) arousal.
It is worth stressing that, since bodily and phenomenal feelings do require a body and a mind, they can be *scaffolded*, but never totally *offloaded*, to artifacts (cf. Stephan *et al.*, 2014, p. 74). This requirement might be loosened for evaluative artifacts, and even more so for motivational artifacts, as we will see below.

4. Evaluative Artifacts

Theories within the evaluative tradition put judgments of the form "X is dangerous" or "I deem my behavior Y blameworthy" at the core of their account. Dating back to stoicism, evaluative theories used to be in vogue among (analytic) philosophers, since they allow to treat emotions as one more kind of propositional attitude, along with beliefs and desires (see Scarantino, 2016, fn. 3).
According to evaluative theories, emotions *are* (*or* at least *involve*) judgments about something, depending on whether they assign a constitutive or causal role to such judgments. Importantly, only constitutive approaches take claims about judgments to be necessary and sufficient conditions for an emotion, thus representing an alternative to theories belonging to different traditions. On the other hand, as Scarantino puts it: "[t]he causal approach [...] is at least in principle compatible with the idea that emotions are essentially feelings or motivations, because it only claims that emotions are caused by evaluations while (at least potentially) remaining agnostic on what emotions are" (2016, p. 15). However, Scarantino also describes some theories that fall somewhere in the middle between the two approaches. These theories regard judgments (or 'appraisals') as being both a cause and a proper part of a multicomponent emotion (here again the reference is to Scherer [2005]). Therefore:

> Evaluative Artifacts $=_{def}$ tools that have been designed for, or are commonly employed for supporting or replacing a human agent in assessing some salient situation or stimuli.

To my knowledge, within the debate on extended emotions only Carter *et al.* (2016) explicitly frame their discussion on the background of the evaluative theory. More specifically, they target theories according to which judgments are (at least) part of the emotion. They argue that, if some cognitive judgment is a necessary ingredient for emotion, and if our cognitive processes are scaffolded by external resources, it follows that some emotions can be partially scaffolded by external resources.
Interestingly, rather than evoking novel types of artifacts, Carter and co-workers exploit Clark and Chalmers' (1998) classic example, a notebook. However, they elaborate on the scenario so that the information retrieved from the notebook triggers some emotional judgment. In one of their examples, Lauren, a vocalist, overhears a conversation about the bad performance that a singer gave at a given café last Saturday. Since her biological memory informs her that *she* was the one playing at that café last Saturday, she gets angry. In a similar situation, Lauren*, Lauren's alter-ego with impaired memory due to her Alzheimer disease, would rely on her trustworthy notebook to gather the same information. Thus, for Lauren* it is the notebook – rather than her memory – that prompts an emotional reaction. Noticeably, here the notebook is first and foremost a cognitive artifact employed to scaffold or subside memory. Then, *a fortiori*, it might also count as an evaluative artifact *qua* cognitive artifact.
An example along the same lines is evoked by Stephan and colleagues (2014, p. 74):

> consider Otto's cousin Arnold, an autistic person incapable of directly perceiving and recognizing the emotional states of others in social interactions. If Arnold is equipped with a headset camera connected to a computer running a program for decoding human emotional states, his appraisal system is supplied with online information in real time about the emotional states of his interaction partners.

Further examples may be found. For instance, xenophobic parties' propaganda poster depicting members of some ethnic group as dangerous may scaffold appraisals of fear in some observers. A handbook on mushrooms might count as an evaluative artifact for disgust, as it scaffolds judgments about which foods should not to ingested. On the other hand, a road sign may work as an evaluative artifact for fear, as it informs about some specific danger along the road. Similarly, metal detectors in airports and other public places scaffold the guards in evaluating the dangerousness of passengers; good and bad marks in exams might be seen as artifacts suggesting that a certain performance is worthy of pride and shame, or even happiness and sadness.

Just like cognitive artifacts may be further classified based on which kind of contribution they provide to cognitive tasks (Fasoli, 2018), evaluative artifacts may be distinguished based on how they contribute to our appraisals. Some, like road signs, merely scaffold our judgments by emphasizing some information that we may fail to notice without them. Others, like the Geiger counter mentioned at the end of §2, or Lauren*'s notebook discussed above, provide information that would be otherwise unavailable, thus playing a constitutive role. Still others, like the mushroom handbook, might even substitute the role of our emotional judgment altogether. A scholar committed to the constitutive variant of the evaluative theory may want to embrace the counterintuitive thesis that the judgments contained in the handbook are themselves emotions; *whereas* the majority of scholars will rather say that the non-emotional judgments in the book prevent the triggering of disgust, and *vicariate* its role by eliciting the corresponding behavior (i.e. avoid ingestion) via a non-emotional route[3].

The distinction between synchronic and diachronic artifacts introduced by Griffiths and Scarantino (2009) seems particularly relevant for evaluative artifacts. Usually, feeling and motivational artifacts can influence an emoter's behavior only when they are in a synchronic coupling with her. Instead, evaluative artifacts can either work synchronically, by informing or substituting a specific appraisal, or diachronically, by shaping her future appraisal via some learning mechanism (being it explicit or not). The notebook employed by Lauren* or the metal detectors are examples of evaluative artifacts that work synchronically, whereas the xenophobic party propaganda posters exert their effect diachronically. Still other artifacts, such as the mushroom book, can work either way: we can study it and keep it in our library when we look for mushrooms, thus exploiting its appraisal-shaping potential diachronically; or we can bring it with us and check it anytime we find a mushroom, to check its edibility synchronically – although apps for smartphones are under development to provide quicker mushrooms-judgments.

The diachronic effect of external resource on appraisal is best seen if we move from artifacts to something more abstract, such as cultural norms. While in this paper I mostly

3 Accepting that the very statements printed in the mushroom handbook *constitute* genuine instances of disgust may sound counterintuitive. However, this conclusion seems to logically follow from the adoption of *both* the constitutive version of evaluative theories of emotion (according to which disgust boils down to judging something as noxious) *and* the thesis of extended emotion (according to which emotions need not be constrained within an agent's body). Hence, such conclusions may be employed as intuition pumps against the conjunctions of these two theses.

focus on physical objects, socio-cultural norms are probably the most interesting kind of external resource to take into account when examining affective artifact from an evaluative perspective (Griffiths and Scarantino, 2009; Slaby *et al.*, 2019). An interesting example here may be represented by Ekman's (1972) Neurocultural Theory. Within this theory, most discussions focus on the controversial *display rules*, that is, a set of social norms that prescribe when it is appropriate to show the facial expression of some emotion and when it is best to inhibit it. Yet, Ekman also predicted that cultural norms influence the appraisal of the same elicitors in different circumstances. For instance, confronted with the news that their son died on the battlefield, a modern mother and one from ancient Sparta might react in different ways. Arguably, the former would cry and the latter would smile proudly, because their cultural norms lead them to appraise a similar situation in terms of different emotions, such as despair and pride. That said, most scholars would be recalcitrant to treat something as abstract as social norms as an artifact. Therefore, even if the investigation of how social norms scaffold emotional judgment undoubtedly represents an interesting endeavor, this is a task for another day.

5. Motivational Artifacts

Roughly speaking, motivational theories see emotions as means to do something. They might identify emotions with certain behaviors (as behaviorists do), or more prudently with an increased disposition to enact them (notably, this is the case of Frijda [1986]). Many of these theories are inspired by evolutionary considerations, and maintain that "each emotion evolved to deal with a particular, evolutionarily recurrent situation type" (Tooby and Cosmides, 2008, p. 117), such as foraging and mating. In other words, emotions play some teleological functions which are pivotal for the survival of the individual and/or the species. The paradigm case of emotions on this view are the so-called basic emotions, whose working recalls that of Fodorian modules (Griffiths, 1990), since they (can) bypass consciousness, arguably by exploiting some subcortical route to achieve fast-and-frugal reactions (Tamietto and DeGelder, 2010)[4]. However, the motivational tradition also includes theories that essentially see emotions as tools playing some social roles. For instance, the facial expressions that Ekman-friendly theorists viewed as vestigial byproducts of some evolutionary behavior may be reinterpreted as tools for social communications (Fridlund, 1994. For a discussion see Griffiths and Scarantino [2009]. For an evolutionary view that also embeds considerations about social roles, see Shariff and Tracy [2011]). Accordingly:

> Motivational Artifacts $=_{def}$ tools that have been designed for, or are commonly employed for supporting or replacing a human agent in some (evolutionary or socially) relevant life task[5].

4 It might be worth noting that such kinds of phenomena are the most problematic ones for evaluative theories: indeed, scholars from these traditions sometimes argue that evaluations need no conscious cognitive processing and establish by *fiat* that every behavior is mediated by cognitive processes. Yet, in doing so they are likely to overstretch *ad hoc* the notion of judgments (see the instructive exchange between Lazarus [1982] and Zajonc [1984]. For a general discussion, see Scarantino [2010]).

5 Some have questioned whether such putative cases of motivational artifacts actually have to do with emotions. Indeed, the notion of "(evolutionary or socially) relevant life task" has vague and porous boundaries. Hence, the class of motivational artifacts risks to be over-inflated. I have a double rejoinder to this worry. On the one hand, recall that my classification has no ambition to capture some deep ontological truth about affective artifacts. More modestly, it is a conceptual toolbox aimed at prompting reflections on how several features of our affective lives can be scaffolded/offloaded onto artifacts. As such, the broad scope of this class of affective artifact may represent a virtue rather than a flaw. On the other hand, parallel to the case of evaluative artifacts (see fn. 3), such examples represent several cases of *reductio* of the conjunction of the extended emotion thesis *plus* the motivational theory of emotion.

To the best of my knowledge, thus far the extension of the motivational properties of emotions via some artifact has only been discussed by Stephan and colleagues (2014; but see Viola, 2020a). Building on the aforementioned case of Arnold, the autistic person with a camera that decodes the emotions from faces of the people he is speaking with, they go on imagining that his futuristic device may also elicit some action tendencies, e.g. suggesting when is time to interrupt a conversation.

But the boundaries of the class of functional artifacts can be further extended by considering not only what happens *upstream* but also *downstream* of a given emotional behavior, i.e. if we take into account those objects that interact with the possible behavioral outcomes of emotions, or that vicariate them. As soon as we consider such objects, we no longer need to imagine science fiction scenarios in order to meet motivational artifacts. In fact, they are already pervasive in our societies.

A most obvious case is disgust. To some psychologists, it is a part of the so-called behavioral immune system (Schaller and Park, 2011), whose scope is to prevent infections by avoiding possible sources of contamination. Indeed, during the Covid-19 pandemic, many countries made compulsory the use of facemasks. As I argues elsewhere (Viola, 2020b), facemasks may be interpreted as a part of technological immune system, enhancing the behavioral immune system represented by disgust.

Moving to less intuitive cases, take anger. Anger is often construed as a tool shaped by evolution to overcome obstacles and consists in preparing enraged agents to undertake aggressive behaviors, e.g. by mobilizing energy for boosting physical prowess. If so, then melee weapons such as clubs or axes are paradigmatic anger-enhancing artifacts, as they are built precisely to enhance the effectiveness of anger outbursts. However, ranged weapons such as bows or rifles recruit other mechanism to achieve the same objectives. In a way, they render the fight-enhancing function of "anger" obsolete, evolutionary speaking, and perhaps even detrimental (an angry sniper is probably less effective than a cold-blooded one).

Alternatively, consider the social function of anger, whose expression is sometimes taken as a means to intimidate enemies, so as to win a confrontation without undertaking actual risks. Again, historical warfare provides good examples of artifacts that mimic anger expression, from war paint to scary ornaments on armors or carriages. More recently, a (in)famous example of redeployed anger-mimicking artifact is provided by the atomic bomb, whose original purpose as weapon faded in favor of that of deterrence since the Cold War. Other examples of artifacts that mimic the intimidating function of anger might include the gallows in medieval towns, the Black Spot in the novel *Treasure Island*, or (moving from classical cases to creative redeployments) the horse head that some member of the Corleone family delivers to Frank Woltz to intimidate him in *The Godfather*. Notice that successful communication involves (at least) two agents, i.e. a sender a receiver. Whenever a motivational artifact fulfils a social role for the sender, it counts as an evaluative artifact for the receiver(s) – recall that the ascription of some artifact to a class is user-dependent.

Other nice examples concern fear. Traditionally, fear is conceived as a mechanism whose main function is to avoid threats (Adolphs, 2013). Throughout history, we invented plenty of tools to help us avoid specific threats. Some of them even to avoid threats in our place – to the point that they may quench our fear, as if we offload the need to feel afraid onto them. In ancient times, towns were surrounded by walls in order to prevent human or animal attacks. However, as technology develops, new threats emerge. Keeping that in mind, the plastic covers of electric wires also work as a straightforward case of artifacts onto which we offload the task of avoiding direct contact with electric wires (also because, evolutionary speaking, we are poorly equipped for appraising wires as a threat). Smoke detectors

with automatic water sprinklers rather represent a paradigmatic example of integrated evaluative-motivational artifact; and so do cars with automatic braking systems. Importantly, some artifacts *inhibit* or *prevent*, rather than enhance or vicariate, the biological or social tasks usually carried out by emotions. Policemen capturing some alleged dangerous criminal during a riot usually restrain him with handcuffs to inhibit both fearful and raging behaviors (e.g. fleeing and fighting back). You may also redeploy sunglasses as tools that prevent people from knowing that you are crying out of sadness. Or they can be used to prevent poker players to notice whether you like your hand (just beware of reflecting lenses!). In a sense, when thus used, sunglasses become a sort of artifactual analogue of Ekman's display rules. Again, if the evolutionary purpose of lust is to drive organisms to mate and breed, then a chastity belt may work as a paradigmatic artifact for contrasting the drives brought about by it (whereas if one settles for preventing breeding, condoms and other contraceptives might also do). Finally, when Ulysses' crew tied him up to the mast of the ship in order to prevent him from reaching the sirens, they redeployed the ropes (originally meant for governing the ship) as a motivational artifact for containing his lust. While undoubtedly many objects qualifying as motivational artifacts have had their share of attention by disciplines studying material culture, their connection with emotions is still underexplored. If we conceptualize human emotions as tools forged by evolution or culture to undertake some task, we can appreciate the relevance of motivational artifacts by trying to address the following question: what would happen to humankind if we only had to rely on these tools, i.e. emotions, barring the possibility to complement them with artifacts?

6. Conclusive remarks

In this paper, I aimed at broadening the notion of affective artifact, which primarily accounts for tools that alter *feelings,* so as to include artifacts that alter other facets of emotion: namely, *evaluation* and *motivation*. To do so, I have started from a deliberately noncommittal definition of artifacts as "tools that have been designed for, or are commonly employed for, some specific function". Then, I have classified affective artifacts according to which functions they play with regard of our affective lives, i.e. depending on whether they primarily alter/vicariate our *feelings*, our *evaluations*, or our *motivations* (see tab. 1 for a summary).

It is worth recalling that this categorization of affective artifacts is agent-relative, so that a same affective artifact may belong to different classes depending on which agents are considered. Notably, the medieval gallows mentioned in §5 count as a motivational artifact for a lord, as they allow him to display his rage *in absentia*, whereas they may count as an evaluative artifact for his servants. Noticing this paves the way for analyzing the phenomenon that Slaby (2016) dubbed *mind invasion*. In his view, a majority of the debate on extended cognition hinges on what he calls the *user/resource model*, i.e. the assumption that extension are due to "a fully conscious individual cognizer ("user") who sets about pursuing a well-defined task through intentional employment of a piece of equipment or exploitation of an environmental structure ("resource")" (Slaby, 2016, p. 5). Such optimistic expectation has also dominated the debate on extended affectivity thus far. However, he warns us that sometimes the persons who deployed the external resources might differ from the cognizer/emoter on which these exert their effects. In this case, rather than being extended, the cognizer/emoter's minds are *invaded*. Among the other things, thinking in terms of affective artifacts, and adopting an agent-relative classification, may offer a useful heuristic tool to those interested in exploring the political import of mind invasions.

According to...	**Feeling theories** *(e.g. Hume; James, Kriegel, 2014; LeDoux and Brown, 2017)*	**Evaluative theories** *(e.g. Nussbaum, 2001; Scherer, 2005)*	**Motivational theories** *(e.g. behaviorism; Fridlund, 1994; Frijda, 1986)*
...emotions are	Bodily/phenomenal states	Evaluations about some salient stimulus	(Dispositions to enact) patterns of behaviors to address evolutionary or social life tasks
A corresponding artifact is a ...	**Feeling Artifact**	**Evaluative Artifact**	**Motivational Artifact**
	General working definition of an artifact tools that have been designed for, or are commonly employed for [some specific function]		
	...altering (i.e. regulating, inhibiting, enhancing) some bodily and/or phenomenal state	...supporting or replacing a human agent in assessing some salient situation or stimuli	...supporting or subsiding a human agent in some (evolutionary or socially) relevant life task
Examples	Handbags (Colombetti and Krueger, 2015), wedding rings (Piredda, 2019), coffee, marijuana	Notebook (Carter *et al.*, 2016), xenophobic propaganda poster, mushroom handbook, metal or smoke detectors	Sunglasses, weapons, war paint, self-braking cars, plastic covers for electric wires, Ulysses' ropes

Table 1. A summary of the three varieties of affective artifacts, based upon the tripartite classification of emotion theories by Scarantino (2016; 2018).

REFERENCES

Adolphs, R. (2013). The biology of fear. *Current biology, 23*(2), R79-R93;
Adolphs, R. (2017). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social cognitive and affective neuroscience*, *12*(1), 24-31;
Candiotto, L. (2016). Extended affectivity as the cognition of primary intersubjectivity. *Phenomenology and Mind*, (11), 232-241;
Candiotto, L., & Piredda, G. (2019). The Affectively Extended Self: A Pragmatist Approach. *HUMANA. MENTE Journal of Philosophical Studies*, *12*(36), 121-145;
Caravà, M., & Scorolli, C. (2020). When Affective Relation Weighs More Than the Mug Handle: Investigating Affective Affordances. *Frontiers in Psychology*, *11*, 1928;
Carter, J. A., Gordon, E. C., & Palermos, S. O. (2016). Extended emotion. *Philosophical Psychology, 29*(2), 198-217;
Clark, A., Chalmers, D. (1998). *The extended mind. analysis, 58*(1), 7-19;
Colombetti, G. (2017). Enactive affectivity, extended. *Topoi, 36*(3), 445-455;
Colombetti, G., Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology, 28*(8), 1157-1176;
Colombetti, G., Roberts, T. (2015). Extending the extended mind: the case for extended affectivity. *Philosophical Studies, 172*(5), 1243-1263;
Ekman, P. (1972). Universal and Cultural Differences in Facial Expressions of Emotion. In J. Cole

(Ed.), *Nebraska Symposium on Motivation*, 1971, vol. 19. Lincoln: University of Nebraska Press;
Fasoli, M. (2018). Substitutive, complementary and constitutive cognitive artifacts: Developing an interaction-centered approach. *Review of Philosophy and Psychology, 9*(3), 671-687;
Fehr, B., Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of experimental psychology: General, 113*(3), 464-486;
Barrett, L. F., Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current directions in psychological science, 8*(1), 10-14;
Franssen, M., Kroes, P., Reydon, T. A., Vermaas, P. E. (2014). *Artifact kinds*. Dordrecht: Springer;
Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press;
Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press;
Gallagher, S. (2013). The socially extended mind. *Cognitive systems research*, *25*, 4-12;
Griffiths, P. E. (1990). Modularity, and the psychoevolutionary theory of emotion. *Biology and Philosophy, 5*(2), 175-196;
Griffiths, P. E., Scarantino, A. (2009). Emotions in the wild: The situated perspective on emotion. In P. Robbins, M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 437–453). Cambridge: Cambridge University Press;
Heersmink, R. (2013). A taxonomy of cognitive artifacts: Function, information, and classes. *Review of philosophy and psychology, 4*(3), 465-481;
Kleinginna, P. R., Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion, 5*(4), 345-379;
Kriegel, U. (2014). Towards a new feeling theory of emotion. *European Journal of Philosophy, 22*(3), 420–442;
Krueger, J. (2016a). The Extended Mind and Religious Cognition. In N. K. Clements (ed.), *Religion: Mental Religion. Part of the Macmillan Interdisciplinary Handbooks: Religion series,* (pp. 237–254). London: MacMillan;
Krueger, J. (2016b). Musical worlds and the extended mind. In *Proceedings of A Body of Knowledge - Embodied Cognition and the Arts conference CTSA UCI*. Available online at: https://ore.exeter.ac.uk/repository/handle/10871/28061;
Krueger, J. (2018). Schizophrenia and the scaffolded self. *Topoi*, 1-13, 10.1007/s11245-018-9547-3Krueger, J., Szanto, T. (2016). Extended emotions. *Philosophy Compass, 11*(12), 863–878;
Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American psychologist, 37*(9), 1019.1024;
LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences, 114*(10), E2016-E2025;
León, F., Szanto, T., Zahavi, D. (2017). Emotional sharing and the extended mind. *Synthese*, doi: 10.1007/s11229-017-1351-x;
Maiese, M. (2019). Extended Affectivity, Reconsidered. *Andy Clark and His Critics*. In: Colombo, M., Irvine, E., & Stapleton, M. (Eds.), *Andy Clark and His Critics*. Oxford: Oxford University Press, 56-68;
Menary, R. (2006). Attacking the bounds of cognition. *Philosophical Psychology, 19*, 329–344;
Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge: Cambridge University Press;
Panksepp, J. (2000). Emotions as natural kinds within the mammalian brain. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions, 2nd ed*, New York: Guildford Press, 137–156;
Petrolini, V., & Viola, M. (2020). Core Affect Dynamics: Arousal as a Modulator of Valence. *Review of Philosophy and Psychology*, 11(4), 783-801;
Piredda, G. (2019). What is an affective artifact? A further development in situated affectivity. *Phenomenology and the Cognitive Sciences*, doi: 10.1007/s11097-019-09628-3;

Preston, B. (2018). Artifact. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = https://plato.stanford.edu/entries/artifact/;
Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology*, *22*(1), 1-19;
Russell, J. A., & Mehrabian, A. (1977). Environmental effects on drug use. *Environmental psychology and nonverbal behavior, 2*(2), 109-123;
Scarantino, A. (2010). Insights and blindspots of the cognitivist theory of emotions. *The British Journal for the Philosophy of Science, 61*(4), 729-768;
Scarantino, A. (2016). The philosophy of emotions and its impact on affective science. In Barrett, L. F., Lewis, M., and Haviland-Jones, J. M. (Eds.), *Handbook of emotions, 4th edition* (pp. 3–65). New York: Guilford Publications;
Scarantino, A. (2018). Emotion. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = https://plato.stanford.edu/entries/emotion/;
Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, *69*(5), 379-399;
Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current directions in psychological science*, *20*(2), 99-103;
Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social science information, 44*(4), 695-729;
Schwitzgebel, E. (2019). *Introspection*. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = https://plato.stanford.edu/entries/introspection/;
Shariff, A. F., Tracy, J. L. (2011). What are emotion expressions for?. *Current Directions in Psychological Science, 20*(6), 395-399;
Slaby, J. (2014). Emotions and the extended mind. In C. von Scheve, and M. Salmela (Eds.), *Collective emotions. Perspectives from psychology, and sociology* (pp. 32–46). Oxford: Oxford University Press;
Slaby, J. (2016). Mind invasion: Situated affectivity and the corporate life hack. *Frontiers in Psychology*, 7, 266. doi: 10.3389/fpsyg.2016.00266;
Slaby, J., Mühlhoff, R., & Wüschner, P. (2019). Affective arrangements. *Emotion Review*, *11*(1), 3-12.Smith, A. (2002). Effects of caffeine on human behavior. *Food and chemical toxicology, 40*(9), 1243-1255;
Sparrow, B., Liu, J., Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776-778;
Stephan, A., Walter, S., Wilutzky, W. (2014). Emotions beyond brain and body. *Philosophical Psychology, 27*(1), 65-81;
Sterelny, K. (2010). Minds: extended or scaffolded?. *Phenomenology and the Cognitive Sciences, 9*(4), 465-481;
Tamietto, M., De Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience, 11*(10), 697-709;
Tooby, J., Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In Lewis, M., Haviland-Jones, J. M., Feldman-Barrett, L. (Eds.), *Handbook of Emotions, 3rd edition* (pp. 114-137). New York: The Guilford Press;
Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, 39, 117–123;
Viola, M. (2020a). From Emotions to Artifacts: Four Modes of Fulfilling Life-Relevant Tasks. In Bertolotti, T.W. (Ed.), *Cognition in 3E: Emergent, Embodied, Extended* (pp. 99-112). Cham: Springer;
Viola, M. (2020b). L'altra faccia de/con la mascherina sanitaria. Leone, M. (Ed.), *Volti virali* (pp. 207-236), Turin: FACETS Digital Press.

Editorial Team
Raffaele Ariano, Università Vita-Salute San Raffaele
Bianca Bellini, Università Vita-Salute San Raffaele
Francesca Boccuni, Università Vita-Salute San Raffaele
Emanuele Caminada, KU Leuven
Laura Caponetto, Università Vita-Salute San Raffaele
Bianca Cepollaro, Università Vita-Salute San Raffaele
Marco Di Feo, Università Vita-Salute San Raffaele
Greta Favara, Università Vita-Salute San Raffaele
Francesca Forlè, Università Vita-Salute San Raffaele
Alfredo Gatto, Università Vita-Salute San Raffaele
Giuseppe Girgenti, Università Vita-Salute San Raffaele
Barbara Malvestiti, Università di Bergamo
Carlo Martini, Università Vita-Salute San Raffaele
Nicole Miglio, Università Vita-Salute San Raffaele
Giacomo Petrarca, Università Vita-Salute San Raffaele
Francesca Pongiglione, Università Vita-Salute San Raffaele
Maria Russo, Università Vita-Salute San Raffaele
Elisabetta Sacchi, Università Vita-Salute San Raffaele
Virginia Sanchini, Università di Milano and KU Leuven
Sarah Songhorian, Università Vita-Salute San Raffaele
Silvia Tossut, Università Vita-Salute San Raffaele

International Advisory Board
Lynne Baker, University of Massachusetts, Amherst, USA († 2017)
Jocelyn Benoist, Université de Paris 1- Sorbonne, FRANCE
Shaun Gallagher, University of Memphis, USA
Vittorio Gallese, Università degli studi di Parma, ITALY
Margaret Gilbert, University of California, Irvine, USA
Sara Heinämaa, University of Jyväskylä, FINLAND
Barbara Herman, University of California, USA
John Horton, Keele University, UK
Neil Levy, University of Melbourne, AUSTRALIA
Dieter Lohmar, Universität zu Köln, GERMANY
Susan Mendus, University of York, UK
Glyn Morgan, Syracuse University in New York, USA
Jean-Luc Petit, Université de Strasbourg, FRANCE
Sonja Rinofner Kreidl, Graz Universität, AUSTRIA
Dan Zahavi, Københavns Universitet, DENMARK
Wojciech Żełaniec, Uniwersytet Gdański, POLAND