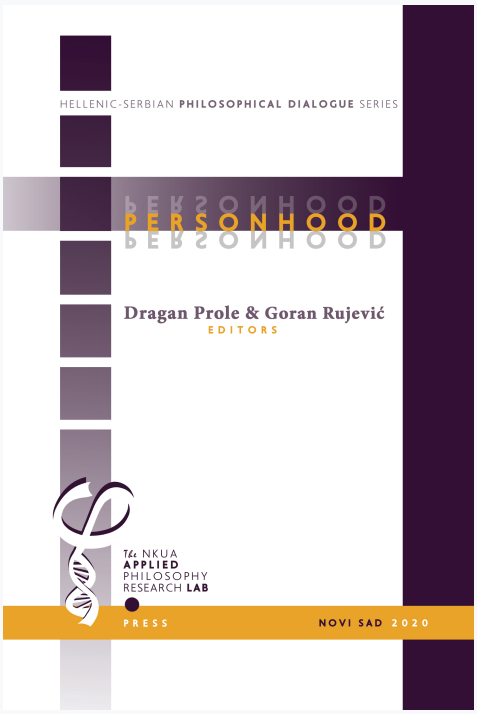
NKUA Applied Philosophy Research Lab Press



Personhood

Dragan Prole, Goran Rujević, George Arabatzis, Una Popović, George Boutlas, Dragan Prole, Panagiotis Kormas, Antonia Moutzouri, Damir Smiljanić, Alkis Gounaris, George Kosteletos, Marica Rajković, Ioannis Ladas

doi: 10.12681/aprlp.49.916

Copyright © 2020, NKUA Applied Philosophy Research Lab Press & Filozofski Fakultet Novi Sad

This work is licensed under Creative Commons Attribution 4.0.

https://epublishing.ekt.gr | **e-Publisher:** EKT | **Downloaded at:** 29/11/2026 21:37:44

HELLENIC-SERBIAN **PHILOSOPHICAL DIALOGUE** SERIES

PERSONHOOD

Dragan Prole & Goran Rujević
EDITORS



The NKUA
APPLIED
PHILOSOPHY
RESEARCH **LAB**

PRESS

NOVI SAD 2020

HELLENIC-SERBIAN PHILOSOPHICAL DIALOGUE SERIES

Personhood

Dragan Prole & Goran Rujević
Editors

The NKUA APPLIED PHILOSOPHY RESEARCH LAB PRESS

Novi Sad 2020

HELLENIC-SERBIAN PHILOSOPHICAL DIALOGUE SERIES

Series Editors
Georgios Arabatzis, Una Popovic, Dragan Prole, Evangelos D. Protopapadakis

VOLUME II
Dragan Prole & Goran Rujević, eds., *Personhood*

Filozofski Fakultet
University of Novi Sad
Dr Zorana Đinđića 1, Novi Sad 21000
Republic of Serbia

&

The NKUA Applied Philosophy Research Lab Press
School of Philosophy, University Campus,
15784 Zografos, Athens, Hellas
E-mail: philosophylab@philosophy.uoa.gr
Web: http://ebooks.epublishing.ekt.gr/index.php/aprl

Editor-in-Chief
Evangelos D. Protopapadakis

Executive Editor
Georgios Iliopoulos

Production Manager & Art Director
Achilleas Kleisouras

© 2020 Filozofski Fakultet, Novi Sad & The NKUA Applied Philosophy Research Lab Press

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of the copyright holders, or as expressly permitted by law, or under terms agreed with the appropriate reprographics rights organizations. Enquiries concerning reproduction outside the scope of the above should be sent to the Faculty of Philosophy, University of Novi Sad, and the NKUA Applied Philosophy Research Laboratory, at the addresses above.
You must not circulate this book in any other binding or cover and you must impose this same condition on any acquirer.

Printed in Novi Sad, Republic of Serbia.
Cover art & layout: Achilleas Kleisouras.

A catalogue record for this book is available from the National Library of Greece.

p-ISSN 2623-3568 • e-ISSN 2623-3576
p-ISBN 978-618-83729-2-4 • e-ISBN 978-618-83729-3-1

doi: https://doi.org/10.12681/aprlp.49.916

HELLENIC-SERBIAN PHILOSOPHICAL DIALOGUE SERIES

Personhood

Table of Contents

Introduction, by Dragan Prole & Goran Rujević 9

Georgios Arabatzis
Byzantine Philosophy, Personhood, and Philosophical Language 13

Una Popović
St. Augustine on Memory and Personhood 29

George Boutlas
The 'Identity Doesn't Matter to Morality' View: Unconditional Third-Person Ascription of Personhood in Kant and Wittgenstein 49

Dragan Prole
The Phenomenological Understanding of the Person: Nietzsche in Husserl's Shadow 85

Panagiotis Kormas & Antonia Moutzouri
Self, Personality and Consciousness: Could Hellenistic Philosophical Approaches Have a Place in Modern Neuroscientific Research? 109

Damir Smiljanić
Personification: A Category Mistake or a Categorial Novum? 121

Alkis Gounaris & George Kosteletos
Licensed to Kill: Autonomous Weapons as Persons and Moral Agents 137

Marica Rajković
Sartre and Personhood 191

Ioannis Ladas
Bioethics and the Person 203

Contributors 217

Introduction

The motives for bringing about the Hellenic-Serbian Dialogue Series originate from the sentiment that mutual relations between Greeks and Sebs far surpass the cultural exchange between Greece and Serbia. Knowing that cooperation does not simply fall into one's lap, but must be initiated by human will and energy, a group of philosophers from Athens and Novi Sad, the proverbial "Athens of Serbia," committed themselves to improving this state of affairs, at least within the confines of their area of expertise, philosophy. None of this would be possible without Evangelos Protopapadakis and his perseverance in wanting to arrange cooperation with colleagues not only from Serbia, but from the entirety of South-eastern Europe. We would also like to take this opportunity to extend our gratitude to the Erasmus Plus Programme, which provided several years of financial support for student and staff exchange between the Department of Philosophy of the University of Novi Sad and the Department of Philosophy of the National and Kapodistrian University of Athens, thereby helping to stabilize this cooperation and turn it into a natural part of our workflow. It is our firm intent to carry out the promise of the title of this series in the following years. Apart from mere cooperation between our two institutions, our aim is also to facilitate an international dialogue that would involve a wide range of thinkers, regardless of their place of employment.

The topic of this second volume is *personhood*, and it was spurred on by the need to investigate the condition of humanity in the twentieth-first century. The volume begins with Georgios Arabatzis' text on the relationship between Byzantine and contemporary philosophy. The leitmotif is the Byzantine notion of *prosopon*, and the contemporary associate is Nietzsche with his idea of "Christian Platonism." Thematically similar

is the paper of Una Popović, which deals with the notion of personhood in the works of St. Augustine as read from the standpoint of young Heidegger, for whom the most interesting task would be to investigate the existential transformation of a pagan life into a Christian one. George Boutlas' article also compares a contemporary and a modern thinker. On the one hand is John Locke's question of relational identity of personality, and on the other Parfit's view that "identity doesn't matter to survival." Dragan Prole's text expounds upon Edmund Husserl's notion of personhood with the help of Nietzsche's notion of "egypticity," which points out the tendency of interpreting temporal phenomena as eternal and immutable. Although the two of them set off from fundamentally different presuppositions, it is soon shown that Nietzsche, the "philosopher of the hammer," and Husserl, the "philosopher of a strict science," produce strikingly similar notions of personhood. Panagiotis Kormas and Antonia Moutzouri investigate the tropes of Hellenistic (i.e. Stoic) understanding of man and contemporary neuroscience. The authors of the text present the lexicon of Hellenistic anthropology in comparison with models that emerged from contemporary studies of the human brain. The article by Damir Smiljanić presents his investigation of the phenomenon of personification, conducted along three main lanes: personification as categorial mistake, as categorial transfer and as hidden vivification. For this purpose, the following philosophical theories proved to be helpful: Ryle's logical analysis, Hartmann's new ontology and Klages' metaphysics of life. Alkis Gounaris and George Kosteletos penned an article motivated by the humanoid Sophia, the first artificially intelligent entity to become a citizen of Saudi Arabia. The perspective used by the authors positions the moral aspects of artificial intelligence front and center and proceeds with the idea that moral capacities presuppose a certain form of personhood. Marica Rajković investigates Jean-Paul Sartre's notion of personhood while pointing out the central role of notions of responsibility and possibility. The author amplifies Sartre's leading issue from the standpoint of the relation

between existence and politics: if man is *zoon politikon*, how is it possible that his fate is not resolved once and for all by gaining political freedom? The final text of the volume is penned by Ioannis Ladas, wherein the notion of personhood is investigated from the standpoint of bioethics. The potential of bioethical understanding of personhood is seen in the complex bioethical approach to life and death decisions: the phenomena of abortion and euthanasia.

With this second volume of the Hellenic-Serbian Philosophical Dialogue Series we have done our best to produce a rich, multi-faceted, broadly scoped, and inspiring book; we wish it becomes for the reader the ideal vehicle for an intellectually stimulating journey.

Dragan Prole & Goran Rujević
November 13, 2020
Novi Sad

Byzantine Philosophy, Personhood, and Philosophical Language

Georgios Arabatzis
National and Kapodistrian University of Athens, Hellenic Open University
E-mail address: garabatz@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0003-4926-9900

Abstract: The article begins with a brief consideration of the preconditions of the modern scientific study of Byzantine philosophy and within it, is given the presentation of some major trends, starting with the examination of Platonic metaphysicism that has been the subject of Nietzsche's strong criticism. Philosophical language, on the limits of this metaphysicism, is an operation of substantiation together with a similar effort for the fragmentation of philosophical disciplines. One is left with the crucial question of the place of realism in the above process, together with the need for an axiological epistemology that in fact develops around the idea of the person, appearing as the constitutive element of a progress toward the light of knowledge.
Keywords: Byzantine philosophy; Platonism; person; language; realism; axiology; epistemology.

I. Position of the problem

Modern research and interest for the study of Byzantine philosophy began only in 1949 (1951) with *La philosophie byzantine* by B. N. Tatakis, published in the series on history of philosophy supervised by the French historian of philosophy Émile Bréhier.[1] Here, one cannot help noticing an issue relevant to modern thought: in short, the modern study of Byzantine philosophy is contemporary to some strikingly modern trends

[1] By P.U.F., Paris.

doi: https://doi.org/10.12681/aprlp.49.919

in philosophy like existentialism or deconstruction. At the same time, Byzantine philosophy in its academic form remained close to the research on sources, *Quellenforschung*, which often is not especially demanding on the philosophical level.

Byzantine philosophy as the culture of Orthodox spirituality and as a reflexive process was strongly related to the movement of existentialism. Its mysticism was thought by some notable Greek thinkers to have much in common with the existentialist precedence of existence over essence.[2] The Orthodox anti-doctrinal mystical spirituality contributed strongly to the establishment of this affiliation. Orthodox personalism that was extended to the content of Byzantine philosophy is strongly related to some forms of existential philosophy. Yet, modern scholarship reconsiders these issues and the existentialism is put under critical test.

II. Platonism

The problem of the philosophical contribution of pseudo-Dionysius Areopagita to the Byzantine thinking arises at this point. The issue became even more crucial after the discovery of the strong ties that connect Dionysius to the Proclusian text. One should ask here: how can a Christian thinker relate to Neoplatonic metaphysics? Dionysius appears distanced from the idea of Christianity as a personal and personalist religion. In his case, one gets the feeling of an "abstract Christianity." Here, henology (the discourse about the One) is an apophatic theology and a negative ontology, a refusal of categorization, a process of purification and, finally, an appeal to God. Is the Good in Dionysius the Christian love or the mystical Eros? Dionysius may have reclaimed Plotinus' school in favor of Christianity. If Dionysius is largely based on Plotinus' school of Neoplatonism, one must de-ontologize Plotinus.

In the above sense, the ontologization refers to cases where the Being is considered a concept covering God and the

[2] See Νίκος Νησιώτης, *Υπαρξισμός και Χριστιανική Πίστη* (Αθήνα: Αρμός, 2019).

beings alike. Above all, God and the beings are founded on a common Axiomatics. To de-ontologize, one must search for the transcendent in the immanent, beyond consciousness, and the relation to the divine must not be limited by the structures of subjectivity. The body must be included in the research and the world must be seen phenomenologically as a donation. Thus one may clarify what being is since immanence as apocalyptic donation seems quite close to the phenomenological reduction. God is at the same time far from representation and the warrant of every representation.[3]

Nietzsche has undertaken the philosophical inversion of Platonism and by that the latter appears not as a disclosure of the Logos but as a historical phenomenon; for Nietzsche, Kant is not in the position to subvert Platonism. The Platonic distinction of the sensible world from Reality is nothing else than the story of a long-running error. Reality is what the philosopher claims to occupy but reality disguised as the world of ideas is, in fact, Plato's private world; Christianity for Nietzsche is this last world in confusion. Kantianism turns the same world into a (moral) imperative and practical spirit. The ideas are thus only suppositions and Axioms and that is what survives from the original Platonic world of ideas. Nietzsche questions the whole axiological process: every new truth is truer than truth in its essence. One should think like an artist and prove the non-truth as truth, the fictional as true. Truth is here only a possible form of truth.[4]

What makes the poverty of axiology is that values are opposed to simple facts, which are thus not values; so facts are non-beings in face of the deontic being of values; yet, the deontic is in fact a will to power and consequently a fiction; this is also relevant for Platonism, which has mythology at the heart of its argument but in a covert manner; here, true science is the Platonist's will to truth, i.e. to power; it is the

[3] The opposite direction may be observed in Lloyd Gerson; see his *Plotinus* (London: Routledge, 1994).

[4] See Sarah Koffman, *Nietzsche and Metaphor*, trans. D. Large (Bloomington, IN: Stanford University Press, 1994).

mysticism and not science that makes Platonism possible; Kantian pure reason is equally a fiction. In view of the above, science is axiological and no thing may be a thing proper; it is not empiricism but the criticism of axiology in comparison to the appearance of things that makes the apprehension of the things possible. Nietzschean criticisms of Kantianism and Platonism converges into the critique of axiology. When ideas turn into values, Platonism falls into crisis, since values are not of the thing but of the judging subject; values are then nothing other than subjective *a prioris*. An idea is, by the same measure, nothing but the condition of the representation of a thing inside General Representationalism. Yet, for Plato, ideas are not only the condition of knowing but also that which permits the things to be what they are; not *a prioris* but real beings. An idea for Plato is not a value but the cause of the things related to it.[5] To be, at the same time, a value and a distinct reality constitutes a mystery, a negative theology. If Plato spoke as a poet then a thing would be a linguistic objectivity, something that allows the thing to become an object. If Plato speaks historically, i.e. on the cultural limits of his time, then he speaks metaphorically and provides his intellectual interpretation as a metaphor.

III. Philosophical language

Philosophy manifests itself in the present tense always as interrogation, project, care, survival or happiness. Academic philosophy promotes a kind of original a-topism through a pretense of ignoring the consequences of thinking beyond pure reflection. No one can entertain the idea of the knowledge of states of things without acknowledging, purely and simply, the reality of fragmented and opposing cultures. This state of fragmentation at its origins cannot be rendered without an expansion of the philosophical language, which is tied to the common speech but desires to extend beyond it, surpassing philosophical academism. The expansive language can thus corroborate a "politics of difference" undertaken under

[5] See *Phaedo* 96a-103a.

various circumstances and, in its richness, would assume the continued communication between fragmented cultures and differentiated philosophical statements.

Henceforth, what is of importance here is the relation between this expansive philosophical language and the question of style. The expansion in itself demands a precise style, renewed for every expansive effort. This is a problem already put forth in Plato's *Ion* and *Phaedrus*, if one tries to understand poetic *mania* as experimentation in style. The problem that appears accordingly is that of aestheticism in philosophy and its subsequent evil, i.e. eclecticism. This would mean a partial abandonment of the Platonic perfectionist model and imply the supplanting of the figure of the philosopher by another intellectual.

Debating philosophical style is thus an equally perfectionist model of a different kind. The renewed style welcomes imitation without compromise and endorses the notion of the plurality of styles, re-inscribing itself in the adventures of *peithô*/persuasion that inaugurated the Sophistic movement and the Aristotelian rhetoric;[6] here, the weak discourse can become strong not by some cunning of reason but as an openness to becoming. The movement is paradoxical, distancing itself from public sensitivity towards a renewed relation to the common through style.

The public today, with its inertia and detachment, excites the philosophical writer not as a demiurge but as a transporter, a *metaphoreas*, hermetical without keyed interpretations, messenger without prophetism. Philosophical language, as the Byzantine philosopher Michael Psellos would have it, conforms to a consequential strategy of the intellectual, to study simultaneously *pantodapôs*, i.e. willing to know in a plurality of modes, and *polymathes*, i.e. willing to learn a plurality of subjects.[7] This is the order of the actuality of thinking in

[6] See Barbara Cassin, *L'effet sophistique* (Paris: Gallimard, 1995).

[7] John Duffy, "Hellenic Philosophy in Byzantium and the Lonely Mission of Michael Psellos," in *Byzantine Philosophy and its Ancient Sources*, ed. K. Ierodiakonou, 139-156 (Oxford: Clarendon Press, 2002, 150).

action. The sociological determinants are not ignored but are listed together with a desire for the new, the *novum*, the avant-garde in the proper sense, i.e. to get one in front of the *faits accomplis*.

The expansive philosophical language is also strongly critical. It is not a mysticism of creative subjectivity, it is not some divine inspiration; it is closely connected and critically related to aesthetics. It is assuming a culture, it attests its various uses, it questions the desire to planify, is actual in pure form without transposing its activity to an objective exteriority.

IV. Philosophical sciences

There is actually a kind of heterotopy in the house of philosophy and in the relations between the different philosophical sciences (i.e. metaphysics, ontology, ethics, aesthetics, political philosophy, logic etc.). The philosophical sciences are facing the deconstruction that is not related to the scientific tradition but to the art and politics of difference. The divisions of philosophical language echo the problem of the violence of categorization and its subsequent confusion. One cannot escape the idea that the division of philosophical sciences stems equally from common sense and speculative mysticism. Each philosophical science is a neutral space between conflicts that is value-neutral and attempts to ignore cultural mediation. Yet, the academic generalizations of the philosophical sciences are also generalizations of philosophical antinomies as Hegel predicted.[8]

---

[8] Hegel writes: "The Antinomies arise, not only in the four objects taken from traditional Cosmology, but in all objects of all genera, in all representations, concepts and ideas. To know this and to seize the objects in the light of this property belongs to the essential part of philosophical study; this property constitutes what will later on be called the dialectical moment of the logical process." Wilhelm Friedrich Hegel, *Encyclopädie der Philosophischen Wissenschaften im Grundrisse*, 3rd Edition (Hamburg: Felix Meiner, 1959), 72, as quoted (and translated) by J. Donaldson, "The Origin of Hegel's Dialectics," *Laval théologique et philosophique* 25, no. 1 (1969):

Philosophical sciences are founded on perception and at the same time distort perception. The actuality of the philosophical discourse is problematized in the paradigm of vision (Platonism) where the present is relativized before the overestimated eternal. The Platonic metaphor of vision equals truth but the value-neutrality of philosophical sciences does not quite match the discerning visibility of things.

According to Richard Rorty, one of the latent goals of higher education after World War II was to repel social sadism.[9] One cannot avoid here the comparison with the "towers of surveillance" put forth by Michel Foucault in his famous *panopticon* example.[10] A critical perspective on philosophical sciences is struggling against two waves of decline in science and in society. This kind of criticism needs something more than common sense and good faith. It requires a critical attention that must surpass the linguistic localizations of common sense.

In any case, the politics of the eternal linear development of philosophical sciences opens a series of questions. One cannot avoid the sentiment that philosophical language is rather formalized. The mapping of the different disciplines obscures its own strategic and tactical orientation. The Heideggerian promotion of the terrestrial element is destined to manifest these territorial confrontations. The quarreling about territory leads to an ever-growing complexity and to rhetoricism. Thus, an effort of deconstruction of the jargon of vision and light has been undertaken. In lieu of the jargon, the reference to the slicing of scientific territories and to some linguistic games such as "orientalism" and "project" become now increasingly crucial.

115-129, 115.

[9] See Richard Rorty, *Achieving Our Country: Leftist Thought in Twentieth-Century America* (Cambridge, MA: Harvard University Press, 1998).

[10] See Michel Foucault, *Discipline and Punish: The Birth of the Prison*, trans. Alan Sheridan (Middlesex: Penguin, 1977), 195-228; pages refer to Chapter 3.3, "Panopticism."

V. Provisional Remark

In this complex setting, the critique of geopolitics of philosophical authorship and the search for an actuality of the philosophical style are at the avant-garde of post-modernity. Para-thetic philosophies, in the Kojèvian sense of anti-rupture philosophies of long running - and one could consider the Byzantine Philosophy to be one[11] – are a clear indicator of this philosophical demand.

One cannot but have the notion of a latent relation between the dialectics of territory and power and the geopolitics of philosophical writing, where the surface of the world matches and/or struggles with the reflexive language beyond the Manichaean splitting between science and ideology. The questioning of the geopolitical rhetoric in the sense given above, through the paradigm of Byzantine philosophy, demonstrates the weak points of the General Philosophical Representationalism initiated by Descartes that Alvin Plantiga thinks is tending toward an epistemic deontology (the search for clear and distinct ideas).[12]

VI. Realism

Until now Byzantine philosophy has been mainly studied in the light of the search for sources (*Quellenforschung*) and in relation to apophatic or mystical theology. One could name both approaches as scientifically esoteric. They are principally interested in conforming to a model of spiritual (in the sense of an interiority in thinking) research. Yet, what would be the chances of reaching towards states of things or "reality" via Byzantine philosophy? In such a search, Byzantine philosophy should relate either to the world of natural kinds and/or to other possible worlds. Yet, the apophaticism, which is the

[11] See G. Arabatzis, "Hegel and Byzantium (With a Notice on Alexandre Kojève and Scepticism)," *Philosophical Inquiry* XXV, nos. 1-2 (2003): 31-39.
[12] See Alvin Plantiga, *Warrant and Proper Function* (New York: Oxford University Press, 1993), vi.

major philosophical theory concerning Byzantine philosophy (in view of the fact that *Quellenforschungen* are in contrast mainly oriented to instrumentality in research), doesn't entail in itself the denunciation of any kind of access to the states of things? Realism in its naïve form means some kind of adequacy between the mind and the things of the world; this cannot be the subject of the history of philosophy alone but implicates the evolution of human knowledge in general. At this point, the relation of metaphysics to realism becomes problematic. As for the Byzantine philosophical tradition, the question is how could one reconcile a warrant epistemology (that for some converges to fideism) with analysis. Here one should look at the relations between epistemology and axiology and the relatively recent rise of the notion of 'true belief' in the analysis of knowledge.[13] Undermining Platonism in favor of knowledge, true belief acquires a position of its own regarding the limits of a pragmatic epistemology, which means that even that which is not valid may still lead to truth. This creates a challenge, for Platonism and beyond, to Cartesian Representationalism and its search for clear and distinct ideas.

Axiology refers to a certain instrumentalism (even in the sense of deontology) that could limit epistemology's ontological aspirations; a belief is by definition of lesser validity claim. In the case of authority, validity and the value of the subject that claims to possess validity are interconnected. Thus subjects may be ivested with authority through their valued claim to validity and by no means thanks to the self-sufficiency of the ideas and notions that they use. All the same, authority is not restrictive of action or, in other terms, is permissive of certain forms of action. After all, a belief may be true and authoritatively valued or true for the sole reason that it is authoritatively valued. Thus, it cannot be untrue for the sole reason that it is authoritatively valued.[14]

---

[13] Edmund L. Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (1963): 121-123.

[14] In Byzantine tradition, the epistemonarches is a title referring to author-

Pragmatic epistemology referring to the idea that a wrong position may still lead to truth is undoubtedly placed on the terrain of social ontology. Social ontology and the affirmation that the substances do exist are co-extensive. The term substance refers to the Greek notion of *hypostasis*. Thus, Plotinus names the three levels of reality (One, Intellect, and Soul) as *hypostaseis*. Among the Greek Fathers of the Church, the term *hypostasis* is used in contrast to the term essence (*ousia*)[15] in order to present the reality of the Holy Trinity - God being an *ousia* and the Father, the Son and the Holy Spirit being the three *hypostaseis*. Among the same writers, the term *hypostasis* and the term "person" (*prosopon*) are interchangeable.[16] In the above sense, any effort to bring the *hypostaseis* closer to the *ousiai* is placed clearly in the general project of philosophical essentialism. It is equally evident that the understanding of *hypostasis* as *prosopon* is rather indicative of a social ontology since the latter term

---

ity knowledge value. The title that Anna attributes to her father the Emperor Alexios is to be distinguished from the traditional term ἀρχιτέκτων which characterizes political mastery. The term belongs to the technical vocabulary of Byzantine monasticism and it means the officer responsible for the discipline. It was used for the first time in a semi-political context a propos the Emperor Manuel I and with regard to his effort to regulate the ecclesiastical affairs; thus, though it was first used in 1157 as an adjective, ten years earlier it referred to Manuel (i.e. about the year 1147). So it is a modern term with regard to the Emperor Alexios. Episteme in its Patristic use goes together with hierarchy. In Michael Italikos, the Art of Royalty, as Episteme, appears as more architectural but in Anna Komnena, the epistemonarches takes the place of the political architect. See Paul Magdalino, "The Pen of the Aunt: Echoes of of the Mid-Twelfth Century in the Alexiad," in *Anna Komnene and Her Times*, ed. Thalia Gouma-Peterson, 15-43 (New York: Garland, 2000), 32.

[15] St. Basil of Caesarea, "Epistula," 236. 6, in Jacques Paul Migne, *Patrologia Graeca* 32, col. 884.

[16] Gregory of Nyssa, "Letter to Peter," in John Behr, *Formation of Christian Theology: The Nicene Faith*, vol. 2 (New York: St Vladimir's Seminary Press, 2004), 419.

in its etymology refers to a face-to-face relation (*prosopon: pros-opsin*). The overlapping or the contradiction of *hypostasis* and *ousia* is responsible, in part, for the quarrel between nominalists and realists during the (mainly Western) Middle Ages. As for Byzantium, Linos Benakis spoke of the conceptual realism of the Byzantines which for him corresponds to that of the Neoplatonists who are the usual source of the former.[17]

In order to be real or effective, the *hypostaseis* must reflect the structure of the world and permit the identification of the individual with the simple; likewise, they ought to be paradigmatized in the worldly things. But, as for the face-to-face relations of the *prosopa*/persons, these must be by definition supra-individualistic. The form of paradigmatization of the *prosopa* as *hypostaseis* may have been the critical question that led to the conception and formation of the "archetypes" in Byzantine iconological thought.

Furthermore, the *prosopon* is manifest as both a relation face-to-face and as a structure of the world. The *prosopon* as pattern may lead one to call the world a "second world" parallel somehow to the sensible world in comparison to the world of ideas of the Platonic project. The notion of the "open structure" is better adapted to the idea of the person in a face-to-face relation; the person may thus be constituent of the real world. On the other hand, the person may be seen through the idea of modern structuralism according to which a structure is basically a difference; the person in this latter case is not a part of the constitutional view of the world but rather inserted in a constructivist philosophy since it would never be equal to itself but always defined by its relation to a limited number of other *prosopa*. Consequently, a person may never be individual or simple as the above prerequisite of the theory of *hypostaseis* posits; but this view of the person is highly counterintuitive at least from the individualistic standpoint.

---

[17] L. Benakis, "The Problem of the Universals and the Conceptual Realism of the Byzantines," in his *Texts and Studies on Byzantine Philosophy*, 107-136 (Athens: Parousia, 2002).

The non-simplicity of the person can be accounted for in a kind of a sociological reminiscence, i.e. the passage from magic to religion. Since the person cannot be totally part of any logical syntax, magic allows for an apprehension of the structural possibilities of it. In magic, a person may be individual but never simple, i.e. it must always be a synthesis individualized and paradigmatized in a topical way, i.e. as a mode of facticity. Its paradigmatic appearance takes place amidst an absolute, supernatural necessity, in the same way that a concrete individual emerges in reality, as such and such and none other. The individual substance appears in a rude, violent way, being what it can be.

In contrast to magic, the religious iconology points to the images of the world. An image is built through the assembly of its parts, which is analogous to the crucial segments of a person/face.[18] The hypostasis as an image clearly facilitates its understanding as individuation; the person thus is not a syntactic unit but an image. What constitutes the logical cohesion of the hypostatic image can be an abstraction. If it is such an abstraction then it could be difficult for one to see how it differs from a rigid structure that would refer to a nominalist state of things. A person can otherwise escape immanence by pointing to an archetype and it can safeguard its validity by insisting on a strict reproduction of the original imprint of an archetype in the world of natural kinds.

The passage from magic to religion, from the supernatural state of affairs to the imagist possibility looks somehow similar to the two ways that Plato proposed in order to explain the causal relation of ideas to things: the method of participation and the method of reflection. This duality cannot become real for Byzantine personhood together with the theory of the separated world of ideas, which are eternal and thus exist prior to the Creator; this cannot be done for obvious reasons. One should speak here of a model-Platonism that makes a cultural structure of long duration

[18] See Plato, *Protagoras* 329 d.

without being absorbed by what Nietzsche calls "Christian Platonism."

If one is to link Byzantine personhood to the apophatic theology, the possibility of an open structure of *hypostaseis* would then be real and effective. In order to do so, one has to escape the Heideggerian onto-theological *Bildungsroman* of the Being. One must resist the Aristotelian legacy of scholasticism and the Neoplatonists' speculative philosophy. Byzantine apophaticism seems to have always rejected the core idea of Western scholasticism as well as the philosophical speculation. There is a Byzantine distrust of dogmatic argument, a discussion that for the Byzantines has been exhausted by the Greek Fathers of the Church, as they genuinely seem to have believed, together with the idea of an unbridgeable difference between contemplation and speculative philosophy. The apophatic theology, nevertheless, is still a theology and not some simple mysticism. The apophatic theology seems to constitute a criticism of the ontological argument for the existence of God and a reference to the Mystery of the Eucharist in the Orthodox liturgy.

The Byzantine personhood would insist that the good and more specifically the person-to-person good is more important than essence; also doing what is good is more important than philosophical foundationalism. Yet, this is not just another forgetting of the Being, as Heidegger would assume, but a means to clarify the relations between beings. Byzantine personhood seems to criticize voluntarism, subjectivism and the vanity of religious sentimentalism. In this sense, the Byzantine personhood finds refuge in the category of the ethico-aesthetic sphere that would inspire the western theologico-philosophical sphere, especially post Kierkegaard.[19]

The ethico-aesthetic sphere seems to contradict the ethico-political approach where the religious element is proto-ethical in the sense that it instructs people the primary moral obligation of obeisance, which is also a *prima facie* political virtue.

[19] Γ. Αραμπατζής, "Το Ηθικο-αισθητικό στη Βυζαντινή Σκέψη και τον Κίρκεγκαρντ," *Εκκλησιαστικός Φάρος* 91 (2010): 231-265.

What is rejected here by the ethico-aesthetic element is the reductionism of the idea that religion is useful for something, i.e. natural theology. The ethico-aesthetic element grants one the possibility to escape the above reductionism by privileging the consideration of the institutional aspect of the question since apophatic theology was and still is part of the institutional discourse. The institutional aspect of the matter also allows one to escape the latent moralism that often lurks in every axiology. The Aristotelian idea that the production of images is important for the cognition-process, stated in *De anima*, III, 3, is a way to relate the classic to modern epistemology. By the same, it facilitates the idea of the *prosopon* as an ontological unit beyond empiricism and divine command theories.

For the purposes stated above, one should consider the apprehension of the world of passion beyond moralism and apart from Aretaic discourse; besides the idea of passion as a moral wrong, one should perceive it also as an intellectual error. The silence produced by any cognitive weakness does not forbid one from the understanding of the emotions as the battlefield of passions. Clearly the personhood cannot be emptied from emotions and the conflicting passions are a way to understand the emotional reality. Also, personhood cannot detach itself from the conscience of a corrupted social world and the harmonious linguistic expression of *Atticism* (as it was practiced in the Byzantine court) may point to sheer and simple hypocrisy. The personhood, in courtly Byzantium, had to express itself as a minor ethics and through an experiential cognition of the passions. This is an ethics of moral nuance and comparison and the basic observation on which it is grounded, seen through self-observation, focuses on the human love for pleasure and the envy; this is an ethics remote from poetical lyricism and based on abstraction and rationality. The moral progress towards the light may have been borrowed from Platonism but is equally a Byzantinist *topos*; what is scrutinized here is the parity of individual conduct. The ethical truth is fundamentally individualistic, moving between two poles, passion and serenity.

The observation of the individual conduct leads to supposition and the elimination process of conjecture after this conjecture guides one to the real causes of the passionate conduct. In the end, the constant observation of behavior in the courtly setting of Constantinople and other major centers of the empire leaves no other actual issue than the retreat from the life of passions, i.e. toward a gentle seclusion into one's self, which does not deny every mundane behavior. The lucidity acquired through the observation of passions does not leave room for any illusion about society: the resistance to passion is the only criterion of any moral stand and it deals mainly with passionate life. Thus, from some universal morality one passes to a strict individual one. The general world of passions that engender imitation leads to the morality of non-imitation directed toward intellectual admiration and philosophical dialogue. This is an additional image of personhood in classical Byzantium following Iconoclasm and the embracing of the axiological/epistemological weight of the images.

References

Arabatzis, G. "Hegel and Byzantium (With a Notice on Alexandre Kojève and Scepticism)." *Philosophical Inquiry* XXV, nos. 1-2 (2003): 31-39.

Behr, John. *Formation of Christian Theology: The Nicene Faith*, vol. 2. New York: St Vladimir's Seminary Press, 2004.

Benakis, L. *Texts and Studies on Byzantine Philosophy*. Athens: Parousia, 2002.

Cassin, Barbara. *L'effet sophistique*. Paris: Gallimard, 1995.

Donaldson, J. "The Origin of Hegel's Dialectics." *Laval théologique et philosophique* 25, no. 1 (1969): 115-129.

Duffy, John. "Hellenic Philosophy in Byzantium and the Lonely Mission of Michael Psellos." In *Byzantine Philosophy and its Ancient Sources*, edited by K. Ierodiakonou, 139-156 Oxford: Clarendon Press, 2002.

Foucault, Michel. *Discipline and Punish: The Birth of the Prison*, translated by Alan Sheridan. Middlesex: Penguin, 1977.

Gerson, Lloyd. *Plotinus*. London: Routledge, 1994.

Gettier, Edmund L. "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (1963): 121-123.

Hegel, Wilhelm Friedrich. *Encyclopädie der Philosophischen Wissenschaften im Grundrisse*, 3rd Edition. Hamburg: Felix Meiner, 1959.

Koffman, Sarah. *Nietzsche and Metaphor*, translated by D. Large. Bloomington, IN: Stanford University Press, 1994.

Magdalino, Paul. "The Pen of the Aunt: Echoes of of the Mid-Twelfth Century in the Alexiad." In *Anna Komnene and Her Times*, edited by Thalia Gouma-Peterson, 15-43. New York: Garland, 2000.

Migne, Jacques Paul. *Patrologia Graeca*.

Plantiga, Alvin. *Warrant and Proper Function*. New York: Oxford University Press, 1993.

Plato. *Phaedo*.

Plato. *Protagoras*.

Rorty, Richard. *Achieving Our Country: Leftist Thought in Twentieth-Century America*. Cambridge, MA: Harvard University Press, 1998.

Tatakis, B. N. *La philosophie byzantine*. Paris: P.U.F., 1951.

Αραμπατζής, Γ. "Το Ηθικο-αισθητικό στη Βυζαντινή Σκέψη και τον Κίρκεγκαρντ." *Εκκλησιαστικός Φάρος* 91 (2010): 231-265.

Νησιώτης, Νίκος. *Υπαρξισμός και Χριστιανική Πίστη*. Αθήνα: Αρμός, 2019.

St. Augustine on Memory and Personhood

Una Popović
University of Novi Sad
E-mail address: unapopovic@ff.uns.ac.rs
ORCID iD: https://orcid.org/0000-0002-3024-1345

Abstract: In his *Confessions*, St. Augustine of Hippo offered his most famous analysis of time. Related to the problem of time is memory, human faculty that allows for our experience of the past and arranges it. Such experience of the past determines the manner in which a human being is aware of objects that are no longer co-present with his own existence, but which are present to his consciousness in a special mode – mediated through his experience of his own existence, as a re-presentation of objects once grasped and known. Memory, therefore, is a concept that refers to the relationship between person and world, and which reveals the world of created beings as a field of relations and interconnections. However, it is also a concept that refers to the person as such, while the experience of the past not only reveals the objects that were once given to the mind and co-present with our own existence, but it also summons our past self and allows for its unification with the present self. Such function of memory is, ultimately, the very condition of personhood, while the person is only possible through unified awareness of particular relations of soul with other beings, self and God. In this essay I will examine St Augustine's understanding of memory and personhood in both of its aspects presented in *Confessions* – the theoretical and the performative one; theoretical aspect being presented with St Augustine's analysis of memory, and the performative one being given with the very composition of *Confessions* as such.
Keywords: St. Augustine; personhood; memory; temporality; *Confessions*; creativity.

The problem of personhood, although philosophically questioned and defined in various different ways, seems to be bound to the heritage of Christian doctrine, inscribed in its first articulations and formulations. One of the most prominent ones is given by St. Augustine of Hippo, author equally devoted to the old realm of

doi: https://doi.org/10.12681/aprlp.49.920

philosophy and to the new horizon of Christianity: philosophy of St. Augustine reflects and transforms ancient ideas towards Christian worldview, which he personally defined in many aspects. Namely, the thought of St. Augustine marked an important change in understanding various philosophical concepts and problems that were already present in ancient philosophy. However, he was also one of the main Christian thinkers to infuse philosophy with another kind of problems and concepts originating from Christian doctrine itself. Moreover, it was the sharp mind and erudition of St. Augustine that allowed for such an infusion to happen on a major scale, since those problems, closely connected to Christian worldview, could not be easily treated and adequately represented in philosophical manner.

It was the personal and private side of Christian doctrine, accentuating the importance of an individual, of personal self in search of personal salvation that was inaugurated as the new main problem of philosophy with St. Augustine. Although St. Augustine did deliver more abstract and more doctrinal works, defending the positions of Christianity against various heretical ideas and philosophical schools, it was this personal side of Christianity that made philosophy of his works and later – philosophy developed under his influence – more Christian in its very essence. The problem of personhood, which is to be addressed in this essay, is one of those problems originating from the fortunate and interesting interconnection between old philosophical and new Christian ways of St. Augustine's thought. Therefore, this problem could be considered as an extraordinary example of the transformation to which philosophy was subjected both in terms of its main issues and its concepts, language and arguments.

The main problem of St. Augustine's thought is – interpreted somewhat from Heideggerian perspective – the problem of the transition from the pagan, non-Christian, to proper Christian life: 'what does it mean to be Christian at all' could be its integral articulation?[1] In St. Augustine's philosophy this

[1] Martin Heidegger, "Augustinus und der Neoplatonismus," in *Phänomenologie des religiösen Lebens*, ed. C. Strube (Frankfurt am Mein:

problem is to be seen in various domains – from the universal perspective, focused in his main work *De Civitate Dei*, to the personal perspective, reflected in various manners and in various works. In this essay I will address one of them, presented in his *Confessions - Quaestio mihi factus sum*.[2]

I. The Problem of Personhood

As one of the problems of St. Augustine's philosophy, the concept of personhood is not to be understood and interpreted from the contemporary perspective of its meaning and function. In my opinion, it should be understood as a consequence of the previously mentioned problem of the transition from the non-Christian to the proper Christian life.[3]

Such life, of course, is not restricted to religious practices – apart from those practices it encompasses the very transition of a human being from a non-Christian to the proper Christian understanding of self and the world.[4] Therefore, the problem of personhood is here to be interpreted as the problem of discovery of the true essence of human being, which is understood as a being constituted by its permanent relationship with God.[5] It could be said that such a relationship between human being and God does not simply reveal the true nature of human essence, but it also reveals the human being as a person – not, for example, as a rational being, political being, being that knows the difference between good and evil, true and false etc.

Understood in such manner, the essence of human being is, according to St Augustine, always present, even in its non-

---

Vittorio Klostermann, 1995), 171-172.

[2] *Confessions* X, 33, 50.

[3] Jeff Nicoll, *Augustine's Problem: Impotence and Grace* (Eugene: Wipf and Stock Publishers, 2016), 142-143.

[4] Martin Heidegger, "Einleitung in die Phänomenologie der Religion," in *Phänomenologie des religiösen Lebens*, ed. M. Jung/T. Regehly (Frankfurt am Mein: Vittorio Klostermann, 1995), 121-122.

[5] Hannah Arendt, *Love and Saint Augustine* (Chicago: University of Chicago Press, 1996), 5.

Christian existence. Therefore, the question here is not whether human existence is constituted through its relationship with God, since Augustine's answer to such question has to be affirmative, for all the dogmatic reasons. The real question here is how are we, human beings, to understand our essence and existence, in order to organize our lives and relationship towards the world according to such understanding? In Foucault's words:

> And we will have moved on to a regime in which the subject's relationship to truth will not be governed simply by the purpose: "how to become a subject of veridiction," but will have become: "how to be able to say the truth about oneself."[6]

The problem of personhood is, therefore, a twofold problem for St. Augustine: a) it is used in terms of defining the essence of the human being as such, and b) in order to be useful for such a definition, the problem of personhood is closely connected to the various possibilities of the experience of self as a human being. Such is the position of the human being for St. Augustine: its essence cannot be defined abstractly because it cannot be lived abstractly – it is lived essence, and therefore it has to be understood and defined from the perspective of corresponding lived experience.

However, such lived experience of human being has always been individual, personal and contingent. Therefore, its interpretation and understanding is challenged, while its medium is conceptual and abstract. In other words, simple concepts and usual philosophical definitions cannot fully grasp lived experience of human being; therefore, if St. Augustine is to give an account of personhood, he has to find new and more flexible ways of its conceptual and verbal articulation. This is what he actually does in the *Confessions*, using this new form of expression to articulate both his own person and the proper way

[6] Michel Foucault, *The Hermeneutics of the Subject. Lectures at the College de France 1981-82* (New York: Pagrave Macmillan, 2005), 362.

of communicating its lived essence with other men, his readers.[7] One can assume that this strategy is a kind of 'performative' account of the human being – his autobiography is not a simple description of his life and experiences, but a kind of retroactive construction thereof. St. Augustine is not describing his life - he is interpreting its meaning.

However, to be able to deliver such an interpretation of his own life, Augustine has to rely on his own, human features and capacities, for it is he – and not something or someone different from him – who is efficiently delivering such an interpretation of (him) self. To be more concrete: he has to employ his reason, for the reason is the most divine aspect of human being, dealing with concepts and conceptual divisions, producing knowledge.[8] Therefore, if there is to be any knowledge of the essence of human being, such knowledge has to be produced by reason and through its activity.

On the other side, reason as such cannot produce knowledge of personhood by itself - it cannot define a priori who Augustine or any other person is. Nor could knowledge of reason as such and its activity be the final step in understanding the essence of human being; Augustine is very clear on this question, for he demands that reason should turn away from its own domain and reflect upon those objects of thought that are clearly above his own capacities – numbers, virtues, beauty, truth as such, and finally God.[9] Those objects are given to our understanding, but have to be differentiated from it, because they reveal the domain of the immutable and the necessary, while our reason and souls are mutable and contingent.

Therefore, since reason alone cannot give us an account of human essence, and because it is itself revealed as contingent

---

[7] Annemaré Kotzé, *Augustine's* Confessions*: Communicative Purpose and Audience* (Leiden: Brill, 2004), 50, 53.

[8] Dewey J. Hoitenga, *Faith and Reason from Plato to Plantinga: An Introduction to Reformed Epistemology* (New York: SUNY, 1991), 98-99.

[9] Augustine, "On the Free Choice of the Will," in *On the Free Choice of the Will, On Grace and Free Choice, and Other Writings*, ed. P. King (Cambridge: Cambridge University Press, 2010), 41-43.

and mutual – *It is clear that each person has his own rational mind*[10] – Augustine has to turn his attention to another human faculty, one that in the fullest sense presents us with our own contingency and mutability, our finitude. Such faculty, of course, is our memory: it presents us with various versions of ourselves, with vivid experience of our mutable nature.[11]

It is, so to say, a fortunate collaboration of reason and memory that allows us to grasp our own essence, our personhood as such.[12] To be true, such collaboration will result in different 'persons' whenever we have different reason and different memory. Therefore, its result - concrete human person (Augustine) or the very concept of personhood abstracted from it - could only be understood as 'formal indications' of human essence, to use Heidegger's term. However, for this essence to be fully grasped, one has to allow for these faculties to step into an interaction.

To rephrase, the self – personhood – can only be revealed if it had previously been lived and understood from such lived experience.[13] The role of reason in such collaboration is to give rise to knowledge – to analyze, to understand, to reveal, to include/exclude and make a coherent unity out of seemingly disparate fragments of memory (while memory is not continuous and freed from disruptions). The role of memory, on the other side, is to offer different materials and fragments, different experiences of self, absolved of any fixed meaning and in need of it: as we have seen, it is not possible for every such fragment of memory to find its place in the final picture, the final story of self.[14] In the words of Hannah Arendt: "The triumph of memory is that in presenting the past and thus depriving it, in a sense, of its bygone quality, memory transforms the past into a future possibility."[15] The final result of the interplay between reason and memory is, in

[10] Augustine, "On the Free Choice of the Will," 43.
[11] Arendt, *Love and Saint Augustine*, 56-57.
[12] Foucault, *The Hermeneutics of the Subject*, 460-461.
[13] Heidegger, "Augustinus und der Neoplatonismus," 195-196.
[14] Arendt, *Love and Saint Augustine*, 46, 68.
[15] Ibid., 48.

Augustine's case, presented with *Confessions* – it is, in a somewhat Wittgensteinian manner, rather *shown* than told.

Nevertheless, such 'performative' manner of constituting and analyzing his person – his own self – is by no means the only manner of understanding and defining personhood which St. Augustine mentions and presents in the *Confessions*. There is at least another one: namely, if the 'performative' manner is to be valid, St. Augustine has to explain why it is so – which features of human being allow for such strategy of re-creation and communication of self and, in consequence, which of them make it legitimate? We have already commented on the question of which human features are understood as the basis for self-understanding; therefore, we are left with the question of the legitimacy of such a project.

The answer to this question is given in the last three books of the *Confessions* – the so called 'philosophical' books, which are written in a more usual philosophical manner and are not presented as a personal confession. As such, they are often considered as a separate inquiry, differentiated from the first part of *Confessions*: in my opinion, this is not so – these books present the very basis on which the 'autobiographical' parts of inquiry are grounded, and therefore all of them have to be considered as parts of the same project.[16]

II. Memory and the Constitution of Personhood

The so called 'philosophical' books of *Confessions* are dedicated to some of the most important questions of Christian doctrine and to their philosophical interpretation. It is commonplace that the problem of time is one of the most prominent Augustine's problems in this context, and that his reflections on time have influenced some of the most important contemporary philosophers, such as Heidegger, for example.

However, if we are to find the problem of personhood in this context, we should read these books not as a consideration

[16] Kotzé, *Augustine's Confessions*, 50; Carl G. Vaught, *Access to God in Augustine's* Confessions: *Books X–XIII* (New York: SUNY, 2005), 104.

of metaphysical structure and origin of the world, but as an inquiry concerning the ontology of human being, questioning the specific manner of human existence.[17] Such existence is to be found in the realm of time – as all created beings, human beings included, exist as temporal beings. The problem of time is, therefore, chosen here as the connection between the usual, metaphysical – and my own, 'personal' and ontological interpretation of Augustine's *Confessions*.

In the context of metaphysics, Augustine sharply differentiates between the being of God and the being of his creatures – God is in eternity, and the creatures are in time.[18] Being in time, they are mutable and contingent – their being is not necessary, they do not have their being on their own grounds, but through the will of God – which is, according to Augustine, also in God, and of his eternal way of being.[19] In such a context, the time is presented as the special way of being, one that is common to all created beings. Time and eternity are, therefore, completely different: eternity is not, according to Augustine, prolonged or still time – they cannot be similar in any way, they are not to be identified or truly compared.[20]

One possible connection between time and eternity is the concept of *the present time* – a moment in the flow of time: in eternity everything is constantly present, whereas in time almost nothing is – in time there is only a passing moment of presence. Time itself is not fulfilled in that moment, but it is extended towards past and future, it has an ecstatic structure; however, in the moment of present time a contingent being is actually present to itself as such and could be aware of its

---

[17] John C. Cavandini, "Time and Ascent in Confessions XI," in *Collectanea Augustiniana: Presbyter Factus Sum*, ed. Joseph T. Lienhard, Earl C. Muller, and Roland J. Teske (New York: Peter Lang, 1993), 177.

[18] James F. Anderson, *St. Augustine and Being: A Metaphysical Essay* (The Hague: Martinus Nijhoff, 1965), 15-16.

[19] Ibid., 17-18.

[20] Ronald J. Teske, *Paradoxes of Time in Saint Augustine* (Milwaukee: Marquette University Press, 1996), 16.

being. Moreover, given that time is of ecstatic structure, such being is present to itself *as contingent*, since it is aware that the moment of present time is passing, that it is not the eternal presence of the divine. Therefore, the real connection between eternity and time here is the concept of presence – not to be understood as dependent on the concept of time, as one of its structural moments. The presence is, consequently, a kind of a signal for the being as such – if there is presence, there is also being. Nevertheless, temporal being is never complete, while it is never present all together in one moment of time.

However, if this is so, how could a human being ever grasp its own personal essence, given that its own being is constantly fleeing from him? If this essence would be traditionally conceived, in terms of eternal and immutable essence of ancient philosophy, in this respect Augustine could rely upon old philosophical strategies and solutions. Nevertheless, Augustine's problem is more complicated – the essence he is searching for is personal essence, personhood we might say. In other words: although this formulation of 'personal essence' seems like an oxymoron, while 'essence' refers to something immutable, and 'person' to something mutable and ever changing, it is exactly such ambiguity and paradoxicality that is here in question and that has to be kept in mind. As we have already seen, Augustine's quest for personal essence is a consequence of Christian doctrine, while only an individual can relate to God, deserve everlasting life, be punished for his or hers sins and so on. Still, Augustine wants to understand – not only to believe, and therefore he has to accept such seemingly nonsensical problems and ideas.

Therefore, if he is to understand his personal essence, Augustine also has to accept that he is never in a position to have a complete identity of self, given that he is a being of time and in time.[21] But now, if this is so, how can we speak of essence here, even in more flexible sense of the term? This could only be possible if we would in fact be in a position to grasp such

[21] *Confessions* 10.8.15.

identity of the self, which would allow us to conceptualize it. However, personal identity cannot be objective one, immutable identity close to eternal essences of the traditional philosophy; it has to be 'subjective' identity, open to personal change and ever fleeing from one's grasp. Of course, human essence can be abstractly defined in terms of human being as an image of God, but such definition is not enough for Augustine – it presents us only with a sign, a 'formal indication' of real and factical personal life.

Personal essence and subjective identity, then, should be explained in terms of temporality of human being as such. Therefore, giving the account of the problem of time, St. Augustine also comments on memory – this very special human feature that allows us to experience our own changing nature, our own temporality. Connection between time and memory is very interesting: namely, memory is the mode of presence of past beings and events for us – the mode of our own consciousness that allows for the concept of past time – or of past being.[22] In an influential and important passage from the *Confessions* Augustine says that recollection of past beings does not put forward the past things themselves, but the images of those things which were immersed into soul – and the words which were formed out of those images.[23] Immediately after, Augustine connects this account on past things with his recollection on himself: his own childhood, he says, does not exist in the past, but in the present moment of time - through his memory it is currently in front of his mind.

Therefore, our past – versions of ourselves lived in past – are not objective features of ourselves, existing independently from our self-consciousness. Rather, they are real only if they are present to our current consciousness: although they are 'things of past,' they can only be real as something existing in present time. Thus, being that they are given via memory, it should be concluded that memory is a special feature of ours

[22] *Confessions* 10.17.26; Vaught, 48-49.

[23] *Confessions* 10.8.15; 10.15.23.

that allows us to be aware of our present and our past selves at the same moment.[24] That is, to have experience of our present and past selves opened at the same instance of time: such opening should further be interpreted as a point of rupture in the idea that human being could be understood as a being of finished and completed essence.

In other words, such opening represents not only the opening of various experiences of self, but also – and moreover – the opening of the possibility of self as such. Personal self, personhood, is never to be understood as a simple identity, but always as a relationship between at least two versions of self – the present one and the past one.[25] Although we are present to ourselves only in the moment of present time – be it even in the mode of our past selves – nevertheless such presence is never to be given in a form of pure and full presence of one person – actual personality; if it would be so, than we would be like God. On the contrary, such presence is fundamentally constituted by the gap between the conscious self and the past self, which represents its memory. This gap is a mark of our temporality and non-identity, our modificability – but it is also a realm in which such modificability is to be actualized.

The fact that our present and past selves are opened for us at the same moment of time reveals that our self is a matter of choice – a matter of free will. There is no predefined identity of anyone's person; it is something to be lived and actualized, something to be chosen over and over again. There are various possibilities of choice here, various selves that one can choose to become; of course, according to St. Augustine, there is only one true choice, the one that would guide our choosing with regard to our fundamental relationship with God. However, here we do not want to stress the problem of criterion of said choice, but the conditions of its possibility; as we have seen, memory is the most important one.

---

[24] *Confessions* 11.11.13.

[25] Kenneth B. Steinhauser, "The Literary Unity of the *Confessions*," in *Augustine. From Rhetor to Teologian*, ed. J. McWilliam, T. Barnes, M. Fahey, and P. Slater (Waterloo: Wilfrid Laurier University Press, 1992), 24.

If we accept our previous findings, we can now conclude that memory is not only a passive container of material to be recollected and gathered by the active and actual, present self. On the contrary, it is in itself active – creative and formative, influencing the final result of choice of the personal essence. Memory is in itself selective, while we do not remember every single event we've lived through and which we've experienced. Therefore, even if we take memory to be a passive container of materials to be formed and organized by reason, the most active and divine-like feature of human being, we would be compelled to treat it as failing to fulfill its purpose, because it could never deliver all of our past experiences for reason to choose freely among them. If there is choice to be made, then it is made on unstable grounds – and it is made by the joint interactive relationship of memory and reason.

Further, the fact that memory cannot deliver fullness of past experiences leaves us with the above-mentioned problem: resulting personal identity is always contingent and therefore can be put in question and once again reconstructed at any given moment. However, this should not be understood as a negative feature of human being, but as a constitutive feature – human being is a being of time.[26] Its constitutive and reflexive relationship with itself, opened by the presence of the actual and the past self, reflects the fundamental relationship ontologically inscribed in human being, namely relationship between us and God. Therefore, its identity is in a mode of becoming – not in the mode of full metaphysical or the empty formal identity.

The role of memory in such context is, as we have seen, undeniable – it is in fact delivering both content (material and the form) choice of the self. If it weren't so, then one could have an idea of his own personality only as a philosopher, rationally approaching to self, but not in ordinary mundane life. Of course, conscious use of reason in the process of self-becoming is an ideal case, presented by Augustine, while reason is the feature that can reveal this process in its – and our – essence. However,

[26] Teske, 38.

the true realm of freedom and choice is offered by the very temporal mode of our being, brought about through memory, through (our) consciousness that we were someone else and that we could once again become someone else – that we can choose who we are.[27]

Left to its own operations, memory has no principle of choosing – only the choice as such – and it can deliver our identity as a result of any of various possible interconnections of our past experiences. That is why the reason is important, while only reason can offer the true understanding of the abstract essence of human being, and then apply it to concrete circumstances.[28] Only reason can show that we are images of God, and that our choice of self should be governed by this fundamental relationship; but for it to be actualized, it has to be chosen over and over again. In Hannah Arendt's words:

> The very fact that man has not made himself but was created implies that the meaningfulness of human existence both lies outside itself and antedates it. [...] Hence, to "return to God" is actually the only way in which a created thing can "return to itself."[29]

III. Concluding remarks

St. Augustine's concept of personhood, as we have previously seen, lays in the very heart of his philosophy. The concept of personhood, namely, represents the very connection between the new Christian problems and ideas of St. Augustine on one side, and the old philosophical conceptual and argumentative framework of his thought on the other side. This problem is closely connected to the understanding of human soul in new

---

[27] Memory, therefore, transcends itself, and because of that it points us beyond our own being. See Vaught, 63.

[28] Paige E. Hochschild, *Memory in Augustine's Theological Anthropology* (Oxford: Oxford University Press, 2012), 151-152.

[29] Arendt, 50-51.

Christian manner, since Christian soul cannot be pure and abstract form of human being, as later quarrels concerning Averroism clearly show. Therefore, such soul has to be personal one, has to retain all the individual moments and specific feature of a person living its life, in order to be considered responsible for its choices, good and bad deeds.

Now, if such understanding of human soul is to be philosophically analyzed, if it is even possible to accept it as a philosophical problem, then philosophy used for such analysis has to be subjected to a specific transformation. The question is more than important, given that the salvation of personal soul is in the very center of Christian doctrine; therefore, positive answer to such question would also define the possibility of rewarding collaboration between philosophy and Christianity in general. Nevertheless, the very transformation of philosophy that is here required should also be legitimized and proven as philosophical in its core, since mere proclamation of the possibility for Christianity and philosophy to be connected is not valid enough to persuade any rational mind.

Dealing with this problem, St. Augustine opted for a twofold strategy. Firstly, to be able to analyze his own soul, his own person in terms of philosophy, but yet honoring its Christian meaning and function, he had to reach for a rather unusual mode of philosophical writing. Namely, he presented us with his autobiography, with his own life told and explained in a mode of unique story of personal growth and development, diverging in various ways, but essentially leading towards baptism and transition from non-Christian to Christian self. In other words, St. Augustine presented us with a personal story organized in such a manner to clearly show all important aspects of both human life and retroactive knowledge of its fundamental meaning and truth. By extracting those essential aspects of human existence out of background horizon of individual life, Augustine aims for the impossible: to sketch, with a single stroke, both universal structure of human being, which applies to any individual human being, and the fact that

there are only individual human beings, singular persons with their personal and unique lives. Such strategy, finally, allows him to present his *Confessions* not only as a personal confession, but also as a philosophical treatise dedicated to the inquiry of human nature as such.[30]

However, Augustine's efforts in this respect are merely presenting the results of already acquired knowledge of self. No matter how natural and simple those parts of *Confessions* seem to be to their readers, they are not some plain personal confession, put to paper as a consequence of a uninterrupted mind flow. They are much more than that: this is carefully organized text, with philosophical point apart from doctrinal one. Most of all, they are clearly written as a reconstruction of personal life which is to give it its proper meaning: interpretative line of Augustine's life is to be seen both in the final point of the autobiographical books – namely, that human person is not to be realized in full without actual and engaged relationship with God,[31] and in the manner in which text was organized, since every single detail in it has its meaning and purpose, defined by the very same final point of Augustine we've mentioned.

If they would be considered only as personal confession, *Confessions* would need no additional elements, and St. Augustine could have stopped his writing at any given point. However, they are not only personal confession, but the presentation of knowledge of 'personal essence,' that is of knowledge of lived human essence, gained through analysis of personal self-awareness. Now, given that it is exactly knowledge – and not just some random explication of self – that is here in question, Augustine has to deliver some arguments in its favor; he has to show that his strategy is legitimate. Nevertheless, he cannot do that in the usual philosophical manner, using already known philosophical means, since the knowledge he gained is too interconnected with rather contingent and concrete life events.[32]

---

[30] Hochschild, 142.

[31] *Confessions* 10.17.26.

[32] Steinhauser, 25.

Therefore, he reaches for another possibility: he legitimizes his project by explaining its origins, its conditions and, consequently, the very fact that it could not be legitimized in any other way.

In my opinion, non-autobiographical books of the *Confessions* are meant to do just that: to offer legitimatization of autobiographical books and the project St. Augustine realized with them. Namely, his presentation of the creation and of the difference between God and his creatures in terms of eternity and temporality ought to impose clear restrictions on possible line of conclusion – so the mind would accept the fundamental temporal constitution of its own character. If our mind is also temporal in its essence, although it can contemplate on eternal and immutable things, it has to consider that the knowledge it can deliver is also temporal and mutable – especially if it is knowledge of human (and its own) essence, which is to be understood as temporal. Therefore, if such knowledge is to be delivered not only in terms of stating that human essence is mutable, but in full expression of the content of that statement, manner in which it is presented also has to indicate the fundamental temporality of its subject. Autobiographical mode of writing is, thus, proved as a proper manner of verbal expression of human essence, since it is always personal and individual human soul.

However, Augustine was not satisfied with just that – he also presented the concept of memory and its analysis, as the constitutive feature of human being allowing for both our awareness of ourselves in our temporal and mutable nature and our ability to use that awareness to gain proper philosophical knowledge of our nature and its character.[33] Memory is, therefore, *conditio sine qua non* of any self-understanding, be it ordinary non-theoretical semi-unconscious understanding of ourselves as having some kind of past, present and future,[34] or be it highly theoretical and philosophical knowledge of temporality as the fundamental mode of human being. To

[33] Hochschild, 144.

[34] Teske, 46.

give an account of memory, then, is to give a legitimatization of the whole project of *Confessions*, in both their Christian and philosophical character.[35] The fact that such account was itself incorporated into a project to which it was supposed to give legitimatization should, finally, suggest that St. Augustine considered it crucial for his main goal – to show that Christianity and philosophy are closely interconnected.[36]

The account of memory St. Augustine presents us with is, however, rather unusual. As we have already seen, memory is to be understood as an active faculty, not as a storage container for past experiences. Memory is not only active, but also creative faculty of human being, for it organizes, creates and recreates various experiences into a more or less coherent web we consider to be our person. Such a web, of course, could be reconstructed at any given moment merely by enlarging it with some new fragments of memory, or by diminishing it through discarding some fragment. Such possibilities are endless, and they are always presented in memory and through memory.[37] However, freedom of memory to make any of such connections, although it presents the very foundation of personhood according to St. Augustine, is not able to lead us to proper knowledge of our personal essence; only reason can do that. Memory, on the contrary, can deliver anything but the knowledge – it is a mode of self-experience, temporally orientated towards past.[38]

Nevertheless, its creative capacities can be used for a more dignified purpose: for achieving life which is led according to the truth of oneself as a human being. Namely, once reason delivers proper knowledge of our human essence and its mode of being, such knowledge can retroactively be applied to organize and define our awareness of ourselves; it can become a principle regulating workings of memory, as presented with

[35] Hochschild, 1.
[36] Vaught, 56-57.
[37] Vaught, 58.
[38] Cavandini, 179.

Confessions. The problem of personhood in philosophy of St. Augustine is, therefore, always to be considered in both of its aspects – the theoretical and contemplative one, as well as the practical, performative and lived one.

References

Anderson, James F. *St. Augustine and Being: A Metaphysical Essay*. The Hague: Martinus Nijhoff, 1965.

Arendt, Hannah. *Love and Saint Augustine*. Chicago: University of Chicago Press, 1996.

Augustine. "On the Free Choice of the Will." In *On the Free Choice of the Will, On Grace and Free Choice, and Other Writings*, edited by P. King. Cambridge: Cambridge University Press, 2010.

Cavandini, John C. "Time and Ascent in Confessions XI." In *Collectanea Augustiniana: Presbyter Factus Sum*, edited by Joseph T. Lienhard, Earl C. Muller, and Roland J. Teske, 171-185. New York: Peter Lang, 1993.

Heidegger, Martin. "Augustinus und der Neoplatonismus." In *Phänomenologie des religiösen Lebens*, Gesamtausgabe Abt. 2: Vorlesungen, Bd. 60, edited by C. Strube, 160-302. Frankfurt am Mein: Vittorio Klostermann, 1995.

Heidegger, Martin. "Einleitung in die Phänomenologie der Religion." In *Phänomenologie des religiösen Lebens*, Gesamtausgabe Abt. 2: Vorlesungen, Bd. 60, edited by M. Jung, and T. Regehly, 3-159. Frankfurt am Mein: Vittorio Klostermann, 1995.

Hochschild, Paige E. *Memory in Augustine's Theological Anthropology*. Oxford: Oxford University Press, 2012.

Hoitenga, Dewey J. *Faith and Reason from Plato to Plantinga: An Introduction to Reformed Epistemology*. New York: SUNY, 1991.

Kotzé, Annemaré. *Augustine's* Confessions*: Communicative Purpose and Audience*. Leiden: Brill, 2004.

Nicoll, Jeff. *Augustine's Problem: Impotence and Grace*. Eugene: Wipf and Stock Publishers, 2016.

Steinhauser, Kenneth B. "The Literary Unity of the *Confessions*." In *Augustine. From Rhetor to Teologian*, edited by J. McWilliam, T. Barnes, M. Fahey, and P. Slater, 15-30. Waterloo: Wilfrid Laurier University Press, 1992.

Teske, Ronald J. *Paradoxes of Time in Saint Augustine*. Milwaukee: Marquette University Press, 1996.

Foucault, Michel. *The Hermeneutics of the Subject. Lectures at the College de France 1981-82*. New York: Palgrave Macmillan, 2005.

Vaught, Carl G. *Access to God in Augustine's* Confessions: *Books X–XIII*. New York: SUNY, 2005.

The 'Identity Doesn't Matter to Morality' View: Unconditional Third-Person Ascription of Personhood in Kant and Wittgenstein*

George Boutlas
National and Kapodistrian University of Athens
E-mail address: gboutlas@phs.uoa.gr
ORCID iD: https://orcid.org/0000-0002-1898-2845

Abstract: In this paper we are going first to expose certain theories of conditional recognition of personal identity, from Locke's relational account of identity to Parfit's identity-doesn't-matter-to-survival view. In the second part we will focus on philosophical conceptions of unconditional ascription of personhood to others, examining especially Kant and Wittgenstein's references on that matter which both seem to face skepticism of other persons as redundant. That kind of skepticism is a greater scandal for philosophy than objects' skepticism as Stanley Cavell puts it. We will conclude with Kant and Wittgenstein's respective views based on their scarce references, labelled together as identity-doesn't-matter-to-morality view. This unconditional acceptance of the third-person ascription of personhood to others as intentional objects of our moral concerns, will satisfy our common beliefs and practices, while recognizing others' personhood as a brute fact.
Keywords: other minds; personal identity; identity doesn't matter view; brute fact; Kant; Wittgenstein.

* An account of unconditional third-person recognition of personhood in Kant, was initially presented in Georgios Boutlas, and Stelios Virvidakis, "Advance Directives and Personal Identity," *Bioethica* 4, no. 2 (2018): 17-32, as also in our common presentation with Stelios Virvidakis, "The Incompetent as Person. Tracing the Personal Identity of Other in Kant and Levinas," in the 13th World Conference on Bioethics Medical Ethics and Health Law, Jerusalem, Israel, November 27-29, 2018. I am grateful to Stelios Virvidakis for our discussions on that matter.

doi: https://doi.org/10.12681/aprlp.49.921

> Thus the persistence of the soul, merely as an object of inner sense, remains unproved and even unprovable, although its persistence in life where the thinking being (as a human being) is at the same time an object of outer sense, is clear of itself.
> Immanuel Kant, *Critique of Pure Reason*

> My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul. The human body is the best picture of the human soul.
> Ludwig Wittgenstein, *Philosophical Investigations*

The philosophical discourse on personhood pertaining mainly to the first-person recognition of personal identity attempts to determine the conditions under which someone recognizes himself as identical through time and is thus morally and legally responsible for his own acts as a person. The third-person recognition of the other person in the modern context of other minds problem, is not encountered as such by ancient Greek philosophy. That problem appears in the context of the orthodox empiricist conception of personal identity and renders the gap between first and third-person ascriptions of mental states unbridgeable and the consequent difficulty to ascribe moral duties between the self and the others (who can be automata or Martians or dummies) insuperable. Patricia Kitcher says that in the case of other minds, we must make a dubious inference to assert than anyone else even has a mind, whereas, in our own case, the ascription of mental states is "immune to error," or at least "immune to error through misidentification."[1] There are several bioethical implications of the third-person recognition of other persons problem. How can we ascribe personhood and moral status to others when they may be comatose, demented, unable to communicate any information or terminally ill patients represented by a proxy or by advance directives, etc.? In the first part of this paper, we are going to expound certain theories of *conditional recognition of personal identity*, either by first-person's standpoint (Locke's *relational* view and late Parfit's

[1] Patricia Kitcher, "On Interpreting Kant's Thinker as Wittgenstein's 'I,'" *Philosophy and Phenomenological Research* LXI, no. 1 (2000): 48-49.

identity doesn't matter view) or by self-*narrative criteria of personal identity* (Frankfurt, Dworkin, and MacIntyre's views) as also by third-person's standpoint community orientated criteria (Schechtman's *anthropological* view and Rorty's *relativistic* view). In the second part, we will go on investigating philosophical conceptions of *the unconditional ascription of personhood to others*, examining some scarce references in Kant and Wittgenstein.[2] In concluding, we will ascribe an alleged similar view to Kant and Wittgenstein, as *identity-doesn't-matter-to-morality* view. This view, facing recognition of the other human as a brute fact[3] is tuned with our common beliefs and practices and bestows personhood to others without empirical conditions.

I. Conditional recognition of personal identity. From Locke's *relational account of identity* to Parfit's *identity-doesn't-matter-to-survival* view

There are several theories of conditional recognition of personal identity. We will place them here in three groups: the *relational account of identity* and the *self-narrative or self-constructed criterion of personal identity* which are both first-person views on identity, and finally third-person views which demand certain

---

[2] Kant and Wittgenstein made claims about the "unknowability" of cognitive subjects and there is a vast bibliography connecting their claims in the context of the 'unknowability of the subject of thought.' Henry Allison, John McDowell, Gareth Evans, Jonathan Hacker, John McDowell, Quassim Cassam, Ralph C. S. Walker, T. E. Wilkerson are included by Patricia Kitcher in this influential tradition assimilating Kant's position on the thinking subject to Wittgenstein's 'I' (Ibid., 33-35). Here we are going to investigate the narrower space of their views on third-person recognition of other persons or souls.

[3] In contemporary philosophy, a brute fact is a fact that has no explanation; see G. E. M. Anscombe, "On Brute Facts," *Analysis* 18, no. 3 (1958): 69-72. Also Barry Smith, and John Searle, "The Construction of Social Reality: An Exchange," *The American Journal of Economics and Sociology* 62, no. 1 (2003): 285-309, where Searle developed Anscombe's concept of brute facts distinguishing between physical facts and social or institutional facts.

conditions for the identification of personhood (the *relativistic* and the *anthropological view*).

Locke's criterion of the *relational account of identity* is the first attempt to render identity a *forensic term* with a broader moral echo with normative implications. According to Locke, person as a forensic term is "appropriating actions and their merit; and so belongs only to intelligent agents capable of a law, and happiness, and misery."[4] His conception of personal identity is exclusively recognized by one's self-reflection, if her consciousness "can be extended backwards"[5] by remembering her thoughts and experiences, being dependent on certain relations through time, which is the reason it is called *relational account of identity*. That conception triggers by its structure quasi-science fiction mental experiments, like the resurrection hypothesis by Locke himself[6] or the *fission problem* structured by Derek Parfit in contemporary philosophy.[7] We call the *relational account of identity "*the orthodox approach."

Parfit, in the context of *Psychological View of personal identity* which comes straight down from Locke's relational account,

---

[4] John Locke, "Of Identity and Diversity," in *Personal Identity*, ed. John Perry (Berkeley: University of California Press, 1975), 50-51.

[5] Ibid., 39.

[6] In case of resurrection, I will be someone to whom my present consciousness extends so this resurrected person will be me even though he might have a different body than I have now (Ibid., 44) This conception is dualistic, rendering the human body unimportant for self-identity. Locke says that if I cut my little finger and my consciousness adhered to it, "that would be the same self which was concerned for the whole body yesterday, as making part of itself, whose actions then it cannot but admit as its own now" (Ibid., 46).

[7] Parfit uses a thought-experiment echoing Locke's severed finger, called *fission*. According to this, I donate each one of my identical triplet brothers one of my functional duplicates brain hemispheres because their brains have been irreversibly damaged. What has happened to me? Have I survived? Are my brothers both me? Can they be identical persons? etc.; see Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), 254-255.

introduces q memories that do not presuppose identity, as a solution to Joseph Butler's objection to Locke.[8] By q memories,[9] Parfit wants to establish a strong psychological connectedness, consisting of overlapping chains of significant numbers of direct psychological connections like memories, intentions, beliefs/goals/desires.[10] Parfit is a reductionist considering that the facts about persons correspond to physical facts about the body the brains and mental events.[11] We will come again to late Parfit's view in this paper, as it evolved in *the identity- doesn't -matter- view*.

The second influential group of theories is grounded on the *self-narrative criterion of personal identity*. We could put together here Harry Frankfurt, Ronald Dworkin, Alasdair Mac Intyre, all adopting some way the equation of personal identity with self-story telling. The moral agent creates or possesses an inner *Bildungsroman* unifying temporally, morally, and legally her moral life with coherence and intelligibility that it could not have otherwise.

In "Freedom of the Will and the Concept of a Person" Harry Frankfurt declares that "one essential difference between persons and other creatures is to be found in the structure of a person's will" but humans are not alone in making choices and other species "even appear to engage in deliberation and to make decisions based upon prior thought." What is different between them is that "humans are able to form what I shall call 'second-order desires' or 'desires of the second-order.'"[12] This Frankfurt's early view has evolved in a later "essential character

---

[8] Butler claimed that memory presupposes identity so memory just reveals to me my identity relation to some past experience and cannot constitute that relation. In David Shoemaker, "Personal Identity and Ethics," in *The Stanford Encyclopedia of Philosophy,* ed. Edward N. Zalta (Summer 2019 Edition), https://plato.stanford.edu/entries/identity-ethics/.

[9] Parfit, *Reasons*, 207.

[10] Shoemaker, "Personal Identity," 5.

[11] Parfit, *Reasons*, 210-211.

[12] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68 (1971): 6.

view" according to which "motives are internal to the self when they are essential to the subject's volitional nature" and "a person acts autonomously only when his volitions derive from his essential character."[13] David Velleman, criticizing Frankfurt, says that his conception of self is appealing because of his *idealization* of the way we are that makes it like the way we wish we could be. The motives moving that self are irresistible, and "are in concert rather than in conflict" so the self "will not be divided against itself". Frankfurt believes "that the well-constituted self is wholehearted rather than ambivalent."[14]

Ronald Dworkin's 'integrity view' echoes Harry Frankfurt's authenticity view with a strong element of self-narrative too. According to Dworkin, integrity view "recognizes that people often make choices that reflect weakness, indecision, caprice, or plain irrationality" and that any plausible integrity-based theory of autonomy must recognize its consequences for a particular person on a particular occasion.

> Autonomy encourages and protects people's general capacity to lead their lives out of a distinctive sense of their own character, a sense of what is important to and for them. Perhaps one principal value of that capacity is realized only when a life does in fact display a general, overall integrity and authenticity.[15]

---

[13] Harry Frankfurt, "Autonomy, Necessity, and Love," in *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1998), 132. Frankfurt though differentiates his conception of autonomy by Kantian autonomy which he thinks is impersonal as an expression of 'pure will.' Its commands "are issued by no one in particular" (ibid.). The volitions he renders important for self-identity or "authenticity" are not tied to a law-like universal moral law, they just need to have a perfectly idiosyncratic character. "Even though a person's interests are contingent they can belong to the essential nature of his will" (ibid.,135).

[14] David Velleman, "Identification and Identity," In his *Self to Self Selected Essays* (New York: Cambridge University Press, 2007), 341.

[15] Ronald Dworkin, *Life's Dominion: An Argument about Abortion, Euthanasia, and Individual Freedom* (New York: Alfred A. Knopf, 1993), 224.

Contingent values without any demand for universal or even communal acceptance are established as the material of a self-constructed personal identity deserving respect of her autonomy, while a severely incapacitated person has "presumably lost the capacity that it is the point of autonomy to protect" so "recognizing a continuing right to autonomy for [them] would be pointless."[16] This approach to personhood becomes somehow relational too because it yields integrity (as a presupposition of autonomy) under certain relations between persons and certain capacities they have.

Midway between the views already mentioned and the unconditional third-person recognition of other persons in Kant and Wittgenstein that we are going to investigate in the second part, we meet the relativistic (the question of what it is to be a human being doesn't matter) or the anthropological view (relation between our practical concerns and personal identity). They share a third-person view of the person, but they put it under the scrutiny of empirical conditions that regulate interpersonal relations according to each view.

According to Richard Rorty, we should be better off if we ceased even to ask the philosophical question of what it is to be a human being, echoing the "identity doesn't matter' thesis of late Parfit. Our concern with the needs and the fate of others rests on the imaginative capacity for identification with them which is dependent on facts historically developed differently in different communities. It is not the Humean qualities of human-animal implanted in us by nature, but a moral concern as members of communities within which our way of communicating and linguistic use of 'we' and 'us' create this special community-dependent concern for others.[17] He recognizes the tension between poetry and philosophy as "a tension between an effort to achieve self-creation by the recognition of contingency and an effort to achieve universality

[16] Ibid., 225.

[17] Cora Diamond, "The Importance of Being Human," *Royal Institute of Philosophy Supplement* 29 (1991): 39.

by the transcendence of contingency."[18] Rorty sides with the poets' side towards a permanent indeterminacy of the question about what it is to be a human being. The priests, the philosophers, the empirical scientists, tried to answer that question making the same claim. "They were going to inform us what we really are, what we are compelled to be by powers, not ourselves. They would exhibit the stamp which had been pressed on *all* of us."[19] These are conceivings we don't need. Facing a person of another community far different than ours or somebody linguistically unable to communicate with 'us' like an incapacitated person, we cannot say she is human. Rorty accepts a third-person recognition of other human beings in the context of the same linguistic game, leaving all other possibilities of human interaction in the dark. Cora Diamond says that

> despite the differences between Rorty and the Orthodox [meaning empiricists like Peter Singer who devalue mentally retarded], he is in an interesting way *with* them; both he and they fail to give an adequate account of possibilities of moral responsiveness to the retarded because both he and they, though for different reasons, will not attach to being human *the significance that it has in much moral thought* [our emphasis here].[20]

The so-called *anthropological* view introduced by Marya Schechtman recognizes capacities like Lockean ones together with the capacities we acquire as members of communities and families, during a process that keeps going through our lives, participating so in the unity of the persons we are. Schechtman's view is as community-dependent as Rorty's but she will defend

---

[18] Richard Rorty, *Contingency, Irony, and Solidarity* (Cambridge, New York: Cambridge University Press, 1989), 25.

[19] Ibid., 26.

[20] Diamond, "The Importance," 53.

incapacitated persons by saying "in real-world the vegetative individual in the hospital bed is Father, who worked hard his whole life and traveled the world before he took ill." This is a third-person life-narrative, which, like all other narratives, presupposes a Lockean person whom the storyteller-society can watch acting throughout life, even if now that she is disabled her story continues as a sequel of her Lockean life. In attempting to justify our common beliefs in a philosophical context, this view also fails to offer a robust normative ground for the concept of person.

We will, at last, examine late Parfit's views that will drive us to the next section. Parfit remains a reductionist in *Reasons and Persons* (1984), but in the "The Unimportance of Identity" (1995) he claims that except being a reductionist he also is a "realist about importance" and from that he concludes that "personal identity is not what matters."[21] Late Parfit claims that "most of us believe that we should care about our future because it will be our future. I believe that what matters is not identity but certain other relations."[22] In this context, he distinguishes between the "Argument from below" expressed as: "personal identity cannot be rationally or morally important. What matters can only be one or more of the other facts in which personal identity consists"[23] and the "Argument from above" according to which: "even if the lower-level facts do not themselves matter, the higher-level fact may matter… the lower level facts have a derived significance."[24] This distinction has strong moral implications. According to late Parfit, probably I won't survive fission, but it is as if I will have survived. Although Parfit refers to the rational and moral unimportance of personal identity he focuses especially on survival. We can call this the *identity-doesn't-matter-to-survival*-view.[25]

---

[21] Derek Parfit, "The Unimportance of Identity," in *Identity*, ed. Henry Harris (Oxford: Oxford University Press, 1995), 33.

[22] Ibid.

[23] Ibid., 29.

[24] Ibid.

[25] Shoemaker, "Personal Identity," 9.

All of the mentioned above approaches of personal identity, whether relational or self-narrative or relativistic or even anthropological, are reductive. They reduce the strongly normative concept of the human person to empirical facts like mental capacities, desires, memories, traits, ways of life, choices, linguistic games, interpersonal relations, etc. leading to empirical contingency, while the quest for strong normativity needs unconditional universality. Being reductive on qualified conditions they face the accusation of *idealization* as Velleman puts it.[26] Each one of them requires special conditions that cannot be met by everyone when 'met by everyone' is a central moral quest for the normative role of the concept of person, taking into account that the respect of the dignity of persons is the cornerstone of most democratic states' constitutions. Onora O'Neill, talking about idealization in moral philosophy, believes that certain contemporary moral theories like 'abstract liberalism' (whether 'deontological' or utilitarian), handle certain issues badly not because of abstraction but because they almost always idealize specific conceptions of the human agent that are admired and feasible in certain privileged circumstances. "Genuine abstraction, without idealization, is, however, the route rather than the obstacle to broad scope."[27] She concludes that "idealization masquerading as abstraction produces theories that appear to apply widely, but which covertly exclude those who do not match a certain ideal or match it less well than others. Those who are excluded are then seen as defective or inadequate."[28]

Another accusation that all these views face, is that of *skepticism*. Kant who was fighting skepticism of the outer world, which in his opinion "remains a scandal of philosophy and universal human reason" (*CPR* XXXIX),[29] didn't face the

---

[26] Velleman, "Identification," 341.

[27] Onora O'Neill, "Justice, Gender, and International Boundaries," In *Quality of Life*, eds. Martha Nussbaum, and Amartya Sen (Oxford: Clarendon Press, 1993), 304.

[28] Ibid., 309.

[29] Immanuel Kant, *Critique of Pure Reason,* ed. and trans. Paul Guyer, and

skepticism of other minds (a philosophical discourse which didn't exist yet) which he would probably also condemn as a scandal. Ancient skeptics did not extend their *investigation* (*σκέψις*) in issues like other persons' mind. Their investigation does not meet modern skepticism on that matter. They discuss nothing for this problem while they make the distinction between affections of the mind and affections of the world.[30] Ancient skeptics should probably have discarded empiricist quantifications of conditions demanded by the concept of the person as *dogmatism* and would put them under the test of the *sorites paradox*.[31]

Parfit, after adopting his *identity doesn't matter to survival* thesis, argues for a diachronic personal identity as a special case of psychologically based survival, using sorites-like arguments.[32] What he attempts is to shake the trust in personal identity's determination by criteria of physical continuity or psychological continuity, or a combination of the two (i.e. his early thesis).[33]

Allen Wood (Cambridge, New York: Cambridge University Press, 1998), 121.

[30] J. Warren, "Precursors of Pyrrhonism: Diog. Laert. 9.67–73," in *Pyrrhonian Skepticism in Diogenes Laertius,* ed. K. Vogt (Tübingen: Mohr Siebeck GmbH and KG, 2015), 105-121.

[31] The sorites [heap] paradox formulated broadly in the following way: It seems that no single grain of wheat can make the difference between a number of grains that does, and a number that does not, make a heap. Therefore, since one grain of wheat does not make a heap, it follows that two grains do not; and if two do not, then three do not and so on. This reasoning leads to the absurd conclusion that no number of grains of wheat make a heap. The puzzle undermines the certainty on vague terms which can be assessed by quantitative characteristics. See Dominic Hyde, and Diana Raffman, "Sorites Paradox," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2018 Edition), https://plato.stanford.edu/archives/sum2018/entries/sorites-paradox/.

[32] J. M. Goodenough, "Parfit and the Sorites Paradox," *Philosophical Studies* 83 (1996): 113.

[33] Parfit makes a thought experiment: Scientists replaced 1% of the cells of my brain and body. I should still continue to exist, even if slightly handicapped in some way. He next proposes a spectrum of operations

Parfit's late view, nevertheless, remains reductionist, dualist, and focusing on the first-person view of self. We are going now to investigate the *identity doesn't matter to morality view* as we labelled the unconditional third-person ascription of personhood to the other, as a unified body-and-mind person, with strong moral implications.

II. The unconditional ascription of personhood to others. Third-person recognition of personhood in Kant and Wittgenstein

After having examined in short certain personal identity theories from Locke on, we will proceed to investigate the views of the two mentioned above philosophers who seem to declare that we don't need empirical criteria or conditions, to ascribe personhood to other humans. They both seem to consider personhood as a brute fact (we have an indisputable third-person view of the other) in non-empirical practical terms.

a. Kant's thinker

There are very few short references in Kant's work that are explicitly or implicitly referring to the third-person recognition of a person.

There is a hint in the 'Paralogisms of Pure Reason' (in the 'Refutation of Mendelsson's proof of the persistence of

---

in which the scientists' activities become more and more extensive; at the far end, they undertake an operation that replaces 99% of the cells of my brain and body leaving only 1% of the original cells in place. In the very last operation, even this 1% is replaced. There is now none of the original physical matter left. If we suppose, as seems reasonable, that my identity is sustained through the replacement of 1% of my cells, and 2% of my cells, and so forth, it seems plausible to believe that my identity continues to be sustained through the replacement of 98%, and then 99%, and finally 100% of my cells. The conclusion should be unacceptable to a believer in the physical criterion of personal identity as it asserts the diachronic continuation of personal identity in a situation with no physical continuity whatsoever (Ibid., 114).

the soul') of the third-person recognition just in the physical appearance of the other:

> Thus the persistence of the soul, merely as an object of inner sense, remains unproved and even unprovable, although its persistence in life, where the thinking being (as a human being) is at the same time an object of outer sense, is clear of itself. (*CPR,* B415)

In (A359-60), Kant also notes that

> The expression that only souls think would be dropped; and instead, it would be said, as usual, that human beings think, that the same being that as outer appearance is extended is inwardly (in itself) a subject, which is not composite, but is simple and thinks. (*CPR,* A359-60)

In the third Paralogism of personality, Kant illustrates the contrast between first- and third-person views with a thought experiment:

> The identity of person is therefore inevitably to be encountered in my own consciousness. But if I consider myself from the standpoint of another (as an object of his outer intuition), then it is this external observer who originally considers *me* as in *time*; for in apperception time is properly represented only *in me*. Thus from the I that accompanies –and indeed with complete identity- all representations at every time in my consciousness, *although he admits this I* [our emphasis], he will still not infer the objective persistence of my Self… so the identity that is necessarily combined with my consciousness is not therefore combined with his consciousness, i.e. with the outer intuition of my subject. (*CPR*, A362-363)

Kant does not explain how the third-person viewer admits this 'I', as if it needs no explanation, as if it was a brute fact. In the second Paralogism, Kant admits:

> This is obvious: if someone wants to represent a thinking being, then he must substitute his own subject for the being he wants to consider (which is not obvious in any other species of investigation). (*CRP,* A353-354)

In all these quotations there seems to be a solid belief in human beings as composite (outer and inwardly) simple things that think, whose conscience cannot be intuited, but are certainly considered thinkers by a common way of understanding on the part of their observers. So far in the first Critique, Kant declares that in thinking, subjects do not intuit a self (*CPR,* A 107). "Consciousness itself is not a representation, differentiating an object."[34] All we know are the 'formal conditions' for any representations which are not sufficient to make inferences about the self's constitution (*CPR,* A 398). This is the outcome of the Paralogisms, where Kant does not deny the possibility of raising the question of identity with regard to the self. What he denies is raising it exclusively from the first-person perspective of the rational psychologist. Kant locates rational psychologist's mistake exactly in the contrast between first- and third-person views which is illustrated above in A362-363. From a third-person perspective evaluating my personhood, there can be no outer intuition of my Self persisting in time, but the outer observer *admits this I* of the other (me). From the first-person, the question cannot be asked because 'I' is presupposed as a formal condition of thinking, while by the third-person it can be asked as related to an object of outer sense, but the 'I think' is no longer there. [35]

[34] Kitcher, "On Interpreting," 45.

[35] Henry E. Allison, *Kant's Transcendental Idealism. An Interpretation and Defense*

Kitcher believes that "Kant saw no philosophically interesting asymmetry between first and third-person ascriptions of mental states"[36] and consequently "he was oblivious to the kinds of worries that generate philosophical anxiety about other minds."[37] By substituting his own subject for the being he wants to consider, he didn't mean something like the argument from analogy that Mill used on the same subject,[38] but just a common way of understanding. To "substitute his own subject for the being he wants to consider" in A353-354 is something so simple "as we might understand an unobserved linden tree through those we have observed."[39] Other humans are persons by the common way we recognize them and there can be no serious philosophical anxiety about them being robots or aliens etc.

All these short references in the third-person recognition of thinking beings in *CPR* seem to prepare the treatment of persons as the unquestionable intentional objects of morality in practical reason. We can make the remark here that Kant uses the same bridging practice between the speculative and the practical Reason in the case of another central Critical idea namely the transcendental *freedom* which in *CPR* was taken

> in that absolute sense in which speculative reason needed it, in its use of the concept of causality, in order to rescue itself from the antinomy into

(New Haven and London: Yale University Press, 2004), 344-345.

[36] Kitcher, "On Interpreting," 49.

[37] Ibid.

[38] J. S. Mill uses the argument from analogy writing: "Other humans have feelings like me, because they have bodies like me, which I know (in my own case) to be the antecedent condition of feelings; and secondly, they exhibit the acts and other outward signs, which in my own case I know by experience to be caused by feelings." J. S. Mill, *An Examination of Sir William Hamilton's Philosophy* (London, UK: Longmans, Green and Co., 1865), 208-209.

[39] Kitcher, "On Interpreting," 49.

> which it unavoidably falls when it wants to think the *unconditioned* in the series of causal connection… only problematically, as not impossible to think, without assuring it objective reality… and plunge it into an abyss of skepticism. (*CPrR*, 5:3)

But in the *CPrR* where

> its reality is proved by an a apodictic law of practical reason, constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason; and other concepts (those of God and immortality), which as mere ideas remain without support in the latter, now attach themselves to this concept and with it and by means of it get stability and objective reality, that is, their *possibility* is *proved* by this: that freedom is real, for this idea reveals itself through the moral law. (*CPrR*, 5:3-4)

Kant reserved the same treatment to the Self, [40] from its initial acceptance in *CRP* as a phenomenon to its grounding use in *CPrR* as a noumenon where it will be a thing in itself, following the general affirmation of the objective reality of the categories applied to noumena which was denied in the first *Critique* (*CPrR*, 5:6). In the *Critique of Practical Reason* (*CPrR*) Kant pinpoints

> the paradoxical requirement to make oneself as subject of freedom a noumenon but at the same, with regard to nature, a phenomenon in one's own empirical consciousness. (*CPrR*, 5:6)

[40] The Self as the 'I' or the 'I think' have a very extensive treatment in the first Critique where the 'I' is the protagonist. What has a 'short treatment' in the first Critique, is the third-person recognition of the other human who will take the leading role in the practical Reason, as the other person becomes the intentional object of the moral law, knowable as a thing in itself.

This happens through

> the union of causality as freedom with causality as natural mechanism, the first of which is established by the moral law, the second by the law of nature, and indeed in *one and the same subject* [my emphasis], the human being, is impossible without representing him with regard to the first as a being in itself but with regard to the second as an appearance, the former in pure, the latter in empirical consciousness. Otherwise the contradiction of reason with itself is unavoidable. (*CPrR*, 5:6)

So, in the practical sense, it is mandatory to face the self as a thing in itself while in the *CPR* it was only known as a phenomenon. In other words, the two of them can coincide in *one and the same subject* under the moral demands of practical reason, a statement that was originally made at the Paralogisms, where

> the concept of personality, just like the concepts of substance and of the simple, can remain (insofar as it is merely transcendental, i.e. a unity of the subject which is otherwise unknown to us, but in whose determinations there is a thoroughgoing connection of apperception) and to this extent this concept is also necessary and sufficient for practical use. (*CPR*, 365-366)

The self has a long and extended treatment throughout all the first Critique. Patricia Kitcher claims in "Kant's Paralogisms" that "the discussions of the Paralogisms chapter depend on and complement the account of the self defended in the Transcendental Deduction."[41] Kant seems to identify the 'I' of the

[41] Patricia Kitcher, "Kant's Paralogisms," *The Philosophical Review* 91, no. 4 (1982): 515.

paralogisms with the 'I' of apperception so the Transcendental Deduction where the unity of apperception is introduced is the place where the Kantian materials of construction of the thinker are found.[42] In the Deduction, he is defending the self from Hume's attack who denies any relation of existential dependence among mental states, while he is in complete agreement with him about the failure of introspection to divulge a continuing self.[43] In the first Paralogism "like Descartes, Kant believes that any mental state, a fortiori any judgment, must be attributed to a self, which we can call the subject of the judgment, but unlike Descartes, Kant conceives of this self not as a simple substance."[44] So far my Self is unknown as an object of inner intuition but my mental states can be attributed to my Self which identifies with the I of the unity of apperception. But "this I, he or it (the thing) that thinks" (*CPR,* A346/B404) according to Patricia Kitcher is not empty but consists of "faculties of sensibility, understanding, a productive imagination, and reason that operate in various ways in the combination of representations."[45] This reach-in faculties "I" which I cannot recognize as a simple substance, I can nevertheless attribute to the other without any criteria, just by admitting this "I" to him as to myself (*CPR*, 362) representing him as a thinking being by substituting my own subject for the being I want to consider (A 353-354). No "other minds" problem seems to emerge in the Kantian context and recognizing other humans as persons seems to be a brute fact, a common human practice that cannot be an object of empirical or philosophical research.

The hint of this can be found in the 'Paralogisms of Pure Reason,' in the claim "since the thinking being (as a human being) is at the same time an object of outer sense" (*CPR,*) and the claim that "the expression that only souls think would be dropped; and instead, it would be said, as usual, that human beings think"

[42] Ibid., 523.

[43] Ibid., 524.

[44] Ibid., 526.

[45] Kitcher, "On Interpreting," 48.

(*CPR,* A359-60), which both have an existential essence. Allison says that Kant has already included an existential dimension in his account of apperception early, in B-Deduction, where he remarks in B157 "I am conscious of myself not *as* I appear to myself, nor as I am in myself, but only *that* I am."[46]

Peter Strawson argues that one of the weaknesses of Kant's exposition is "that he barely alludes to the fact that our ordinary concept of personal identity does carry with it empirically applicable criteria for the numerical identity through time of a subject of experiences (a man or human being) and that these criteria, though not the same as those for bodily identity involve an essential reference to human body," but Kant does not ignore it as it is evident in *CPR,* B415[47] (interpreting Kant as saying in this quotation that we need physical criteria for reidentifying persons). Kitcher believes that Strawson's interpretation of this section of the *CPR* is problematic because there is no mention of physical criteria for reidentification in these passages, and "he [Strawson] regards this as the overall message of the chapter"[48] while she takes this Kantian reference "to be simply that during life bodily continuity is the usual way to determine continuity of the self, not that bodily continuity is 'criterial evidence' for self-identity."[49] There is ground though against her interpretation, as bodily criteria can be the only evidence, becoming critical for the persons that exist at the borderline of personhood, (demented, without consciousness, etc.) and it is difficult to say that Kant would exclude them as persons from the moral territory. Kant's conception of human nature emerges from mingled, different criteria, rational, empirical, anthropological, representing his equal interest in all these items expressed in his writings, a fact that

---

[46] Allison, *Kant's Transcendental Idealism,* 351.

[47] Peter Strawson, *The Bounds of Sense. An Essay on Kant's Critique of Pure Reason* (London and New York: Routledge 1990), 164.

[48] Patricia Kitcher, *Kant's Transcendental Psychology* (New York, Oxford: Oxford University Press, 1990), 266-267.

[49] Allen Wood, "Kant on Duties Regarding Nonrational Nature," *Aristotelian Society Supplementary Volume* 72 (1998): 189.

the rationalistic interpretations of Kant have difficulty to accept. Allen Wood in "Kant on Duties Regarding Nonrational Nature" says that Kant recognizes three original predispositions of our nature. 'Humanity' is the capacity to set ends according to reason. 'Animality', which includes our instinctual desires promoting our survival, reproduction and sociability, 'personality' which is our rational capacity to give moral laws and obey them (*APPW* 7:321-324)."[50] Wood, although arguing pro logocentric ethics, which grounds all duties on the value of humanity or rational nature, accepts that "of course we should respect rational nature *in* persons, and this means respecting the persons themselves. But… we should *also* respect rational nature *in the abstract*, which entails respecting fragments of it or necessary conditions of it, even where these are not found in fully rational beings or persons."[51]

Taking into account that Kant rewrote the "Paralogisms of Pure Reason" for the second edition of the Critique (the only chapter of the Dialectic which he rewrote), and he spent four Parologisms on the self or the soul alone, and one antinomy only for each Idea of the World, Freedom, God, we must take it for granted that the thinker is the real protagonist of the first critical enterprise. And if someone wants to identify another human as a thinking being "he must substitute his own subject for the being he wants to consider" (*CRP,* A353-354) recognizing him without criteria of personhood as the intentional object of moral law.

b. Wittgenstein's "eine Einstellung zur Seele"

In Part II, Section iv of *Philosophical Investigations*, Wittgenstein writes:

> "I believe that he is suffering." – Do I also *believe* that he isn't an automaton?
> It would go against the grain to use the word in both connexions. (Or is it like this: I believe that he is

[50] Ibid., 202.
[51] Ibid.

> suffering, but am certain that he is not an automaton? Nonsense!)
> Suppose I say of a friend: "He isn't an automaton." – What information is conveyed by this, and to whom would it be information? To a *human being* who meets him in ordinary circumstances? What information *could* it give him? (At the very most that this man always behaves like a human being, and not occasionally like a machine.)
> "I believe that he is not an automaton," just like that, so far makes no sense.
> My attitude towards him is an attitude towards a soul [eine Einstellung zur Seele]. I am not of the *opinion* that he has a soul."
> "The human body is the best picture of the human soul." (*PI*, Part II, iv, p.178)

This section that has attracted philosophical attention as illuminating Wittgenstein's attempt to investigate the grammar of the soul, deals with pain expression, souls, human beings, automata. Wittgenstein places the discussion of the soul in the context of the metaphysician's conception of the essence of language as a private inner activity of a disembodied subject. The world 'automaton' has a long presence in metaphysics of the soul.[52] Peter Winch believes that Wittgenstein,

---

[52] René Descartes considers animals are mere automata. In Part Five of the *Discourse on Method*, Descartes writes: "I made special efforts to show that if any [automatons] had the organs and outward shape of a monkey or some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals" (AT VI, 56; CSM I, 139); see René Descartes, *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny (Cambridge: Cambridge University Press, 1985-1991). J. S. Mill writes also "experience obliges me to conclude that there must be an intermediate link, which must either be the same in others as in myself or a different one. Thus, they are either *alive* or *automatons*;" see J. S. Mill, *An Examination of Sir William Hamilton's Philosophy* (London: Longmans,

against empiricist behaviorists' and dualists' conceptions, who considered the other man as a conscious being only under certain beliefs about him, seems to say here that this is not a matter of holding a belief or opinion but a matter of having a certain "attitude" towards him.[53] The two "connexions" here are 1) I believe that he is suffering, 2) I also *believe* that he isn't an automaton. Why believing that he is not an automaton makes no sense? It seems that what Wittgenstein claims is that we deal with two different things here. Beliefs or opinions about others being in pain and attitudes towards them as human persons, as souls. And the nonsense is to believe that those two different terms are interdependent in a logical way, although they have a totally different position in the space where they are placed. What kind of statement is to say that someone is not an automaton, what is the practical outcome of this opinion except in the special cases where someone could face an automaton?[54] In our common everyday practice, the other is a human, a soul, and this is not an opinion that we shape out of an assessment of her external behavior that resembles ours, using either the argument from analogy or the one from inference.[55] Her being a soul just appears as a result of our attitude towards a soul.

Winch compares the use of the word "Einstellung" in Section iv with the one in *PI* Part I, §310 :

Green and Co., 1865), 208-209.

[53] Peter Winch, "Eine Einstellung zur Seele," *Proceedings of the Aristotelian Society, New Series* 81 (1980-1981): 2.

[54] Under certain circumstances, there still remains the *possibility* of seeing others as automata and this has mainly to do with the connections that draws the person who is looking. "[…] the skeptic about other minds presents her problem as one of *knowledge* – as if what we needed was more *evidence* of some kind, something that (*per impossibile*) would allow us to go *beyond* the other's (mere) body, or maybe *through* it, thus reaching a 'naked soul.'" See Jonadas Techio, "Seeing Souls: Wittgenstein and Cavell on Other Minds," *Conversations: The Journal of Cavellian Studies* (2013): 78.

[55] For Mill's *argument from analogy* see ref. 38. The *argument from inference* claims that the best hypothesis we can make is that other people have minds and they are not machines.

> I tell someone I am in pain. His attitude to me will then be that of belief; disbelief; suspicion; and so on. Let us assume he says: "It's not so bad." – Doesn't that prove that he believes in something behind the outward expression of pain?
> His attitude is a proof of his attitude. Imagine not merely the words "I am in pain" but also the answer "It's not so bad" replaced by instinctive noises and gestures. (*PI,* Part I, §310)

Winch says that "to be clear what a belief (e.g.) that someone is in pain comes to...[we] should look at the whole range of behavior, demeanor, facial expression, etc. in which such verbal expressions are embedded."[56] By "His attitude is a proof of his attitude" Wittgenstein is not rejecting his belief in something, but his belief "in something behind the outward expression of pain," something like a Cartesian self. By this rejection, Wittgenstein does not declare that all he believes is that the other person is behaving in a certain way. "His belief concerns someone to whom he has "eine Einstellung zur Seele" and this helps to make his belief what it is"[57] a human being for whom "[t]he body is the best picture of the human soul" (*PI*, Part II, iv, p.178).[58] The expression of another's suffering and my belief that he suffers, is confined to a particular occasion, and the generalization of particular occasions (beliefs about his mental states in different times) cannot lead to the belief that he is not an automaton which is "a view of the kind of being he in general is."[59] Him being the kind of being that he is, is the condition of him having mental states. We take it for

---

[56] Winch, "Eine Einstellung," 3.

[57] Ibid.

[58] This reference has an impressive resemblance with the Kantian "since the thinking being (as a human being) is at the same time an object of outer sense." (*CPR,*)

[59] Winch, "Eine Einstellung," 6.

granted that he is having several states of consciousness from our attitude towards him as an attitude towards a soul. There cannot exist a radical separation of soul and external expression of mental states, that would presuppose a private language. Two references in "Einstellung zur Seele" make Winch sound as if he is considering an attitude as a brute fact. One is: "There is no question here of an attitude which I can adopt or abandon at will… it is a condition I am in vis-a-vis other human beings *without choosing* [my emphasis] to be so."[60] The other one is when he interprets Simon Weil's phrase in *The Iliad, Poem of Might* "The human beings around us exert just by their presence a power" as meaning that

> our characteristic reactions towards other people are not based on any theory we have about them, whether it is a theory about their states of consciousness, their likely future behavior, or their inner constitution.[61]

Both these claims refer to Wittgenstein's usually applied phrase "part of the natural history of mankind." Wittgenstein, according to Winch, rejects the empiricist list of states of consciousness needed to ascribe personhood to others by using the attitude towards a soul which seems for him to be 'the way we do it' as part of the natural history of mankind.

Cavell makes a distinction parallel to belief-attitude in "Knowing and Acknowledging." Facing the problem of other minds, he claims that 'the problem' is not a matter of *knowledge,* but rather of *acknowledgment*.

> Your suffering makes a *claim* upon me. It is not enough that I *know* (am certain) that you suffer – I must do or reveal something (whatever can be done). In a word, I must *acknowledge* it, otherwise I do not know what '(your or his) being in pain' means.[62]

[60] Ibid., 11.

[61] Ibid., 8.

[62] Stanley Cavell, *Must We Mean What We Say?* (Cambridge: Cambridge

My failure to acknowledge your pain consists of "soul-blindness."[63] The outcome of not acknowledging the other as a soul is "my avoidance of him, call it my denial of him."[64]

It seems that Wittgenstein's attitude does not rest in beliefs about the other, that she has pain etc. but the other way round. Our attitude toward her as having a soul creates the grammatical space for beliefs about her mental states. Edmund Dain concludes in the same line of thought that "human being, a soul, just is on this account what stands at the center of these forms of talk, just is the kind of thing that provides footholds for these concepts, and the kinds of behavior they make intelligible."[65] He compares section iv attitude with section 284 of the *Philosophical Investigations* pertaining to the difference between attitudes towards what is alive and what is dead.

> Look at a stone and imagine it having sensations. – One says to oneself: How could one so much as get the idea of ascribing a *sensation* to a *thing*? One might as well ascribe it to a number! – And now look at a wriggling fly, and at once these difficulties vanish,

University Press, 1976), 263.

[63] Stanley Cavell, *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy* (Oxford: Oxford University Press, 1979), 378. Wittgenstein's notions of aspect and meaning-blindness are probably the sources of Cavell's notion of soul-blindness. The former being the failure to see something *as something* (*PI, Part II*, xi §257) the latter the failure to distinguish between different meanings of a word (*PI, Part II*, xi §262-3). In a way soul-blindness contains both aspect and meaning-blindness. Wittgenstein compares soul-blindness with aspect-blindness in *PI*, *Part I,* §420: "Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another; the cross-pieces of a window as a swastika, for example." Techio, "Seeing Souls," 72.

[64] Cavell, *The Claim*, 389.

[65] Edmund Dain, "Wittgenstein on Belief in Other Minds," assessed November 9, 2019, https://www.academia.edu/22598240/Wittgenstein_on_Belief_in_Other_Minds.

> and pain seems able to get *a foothold* here, where before everything was, so to speak, too *smooth* for it. (*PI, Part I,* §284)

Is it our beliefs or scientific data ("ascribe it to a number") that which makes us believe that a stone is not suffering but wriggling fly is, or our attitudes towards them? Dain concludes that the difference between attitudes towards what is alive and what is dead "is not a matter of what is true or false in a narrow sense, so much as it is a matter of what it makes sense to say, of the concepts that find a grip, a purchase, and the forms of behavior that are available to us as a result."[66] In the paper "Do We Believe in Other Minds?" on the same subject, Dain defends the opinion that we don't really believe in other minds, but our attitude towards a soul means that "our understanding of others as having minds lies in our basic modes of behavior in relation to other human beings, and the kinds of things that we can say about them."[67] This interpretation seems to be resting on behavior or having to do with our social practices, while Wittgenstein's 'attitude' seems to be something more basic, something outside, or at the borderline of the space of rules and practices. The attitude seems to be more primitive than practices in language games. In *Remarks on the philosophy of psychology* this is stated more explicitly:

> Here it is a help to remember that it is a primitive reaction to take care of, to treat, the place that hurts when someone else is in pain, and not merely when one is so oneself – hence it is a primitive reaction to attend to the pain-behaviour of another, as, also, *not* to attend to one's own pain-behaviour. (*RPP*, §915)

---

[66] Ibid.

[67] Edmund Dain, 'Do We Believe in Other Minds?" *Austrian Ludwig Wittgenstein Society* 36 (2016): 45-47.

> What, however, is the word "primitive" meant to say here? Presumably, that the mode of behaviour is *pre-linguistic*: that a language-game is based *on it*: that it is the prototype of a mode of thought and not the result of thought. (*RPP*, §916)

Maybe attitudes are based on this kind of certainty we acquire when we reach bedrock where our justifications are exhausted:

> If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: "This is simply what I do." (*PI, Part I,* §217)

But how does Wittgenstein's attitude towards other souls lead us to the moral space, the way the Kantian embodied thinker leads us to the categorical imperative?[68] In the context of pain-feeling of an embodied subject he remarks:

> […] only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (*PI, Part I,* §281)

From these, possible feelings come out about who "sees; is blind; hears; is deaf; is conscious or unconscious."

> How am I filled with pity *for this man?* How does it come out what the object of my pity is? (Pity, one may say, is a form of conviction that someone else is in pain.) (*PI, Part I,* §287)

Pity leads us to the grammar of moral space where human beings recognize each other not *because* of the pain expressions

[68] Especially the second formulation, the so-called *formula of humanity*: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means." (*GMM*, 4:429)

alone but because their attitude towards other souls makes them recognize their pain and have spontaneous and primitive reactions of sympathy for them. If rocks or speaking automata had pain, even expressing it, could we feel pity for them? (Maybe yes but it would be out of moral space). We could not have towards them the attitude that we have towards human beings.

Attitude towards a soul makes Wittgenstein as well as Kant radical opponents of dualistic conceptions of human nature, mainly of utilitarian origin, that thrive in the contemporary bioethical field.[69] Wittgenstein, much like Kant, does not discriminate or distinguish among persons, he counts as living human beings the unconscious, the blind, the deaf. It's not the contingent expression or ability to express mental states that has a moral impact on us but our permanent not empirically formed attitude towards souls that make us spell our moral vocabulary no matter how disabled the others can be, whether physically or mentally.

There is a vast bibliography connecting Kant and Wittgenstein through their claims about the "unknowability" of cognitive subjects and "unknowability of the subject of

[69] Peter Singer's 'opinion' that "the fact that a being is a human being, in the sense of a member of the species Homo sapiens, is not relevant to the wrongness of killing it; it is, rather, characteristics like rationality, autonomy, and self-consciousness that make a difference. Infants lack these characteristics. Killing them, therefore, cannot be equated with killing normal human beings, or any other self-conscious beings" faces human nature under the dualistic evaluation of mental capacities as important and the bodily remnants of a human being without those capacities as redundant. His 'opinion' conflicts with common beliefs on the value of human persons as both body and mind that is expressed in Wittgenstein's 'attitude towards a soul' when we face other humans; see Peter Singer, *Practical Ethics* (Cambridge: Cambridge University Press, 1993), 182. For a detailed discussion on the ethics of infanticide from antiquity till present time see Evangelos D. Protopapadakis, *From Dawn till Dusk: Bioethical Insights into the Beginning and the End of Life* (Berlin: Logos Verlag, 2019), 21-48.

thought."[70] Kitcher believes that Kant's epistemic analyses that lead to a model of the cognitive self have nothing to do with Wittgenstein's remarks:

> There is no such thing as the subject that thinks or entertains ideas.... (*TLP*, 5.631)
>
> The subject does not belong to the world: rather it is a limit of the world (*TLP*, 5.632)

According to Kitcher

> Kant's claims about the special status of the 'I think' by appealing to the early Wittgenstein makes sense only if there are some real affinities between the former's ways of thinking about the synthetic a priori and the latter's attitude towards the inexpressible.[71]

But Kant posited his synthetic a priori claims in transcendental deduction, while Wittgenstein wanted non-logical claims only to be "shown." Comparing Wittgenstein's 'I' with Kant thinkers is a discussion that we can't entangle with here, a discussion much more extensive than our original intention to compare their views on the third-person ascription of personhood to others, where our focus is limited. In this much more limited space, both conceptually and textually, because of the very few references of the two philosophers on that matter, we discover a convergence in their opinions. Their references have in common an anti-skeptical, anti-dualist position, rejecting the possibility of facing other humans as automata, or non-souls, or aliens, etc. in everyday practice.

Cavell in "What is the Scandal of Skepticism" claims that objects' skepticism is acceptable. "I [can] object to your claim to know by saying for example, that 'you don't see the back half of

---

[70] Kitcher, "On Interpreting," 33-35. See n. 2

[71] Ibid., 41.

the object'… the case of the material object is argumentative."[72] But the skepticism of the other person is "too trivial almost to mention" because everybody knows others because of their behavior, or their contact "or the subtler movements of the body, especially the face."[73] Cavell seems to suggest that we know that the other is not transparent, but we always recognize her in the way we do it in our form of life. The unconditional recognition of the other person for Cavell bears a similarity to Wittgenstein's vision of the metaphysical ungroundedness of language, of our ability to go on without concepts.[74]

III. Conclusion

Our initial question was about the possibility of ascribing personhood and moral status to non-ideal agents. Comatose, demented, unable to communicate any information or terminally ill patients represented by a proxy or by advance directives, are residents of the 'personal identity's twilight' zone. How can personhood be preserved at the dawn of personal identity? The orthodox empiricist dualist conceptions of personal identity as the *relational* account, together with the *narrative*, the *anthropological*, and the *relativist* view cannot perform that task as they are grounded on the contingency of different scientific beliefs, different societies, different narratives, and customs. Late Parfit's *identity-doesn't-matter-to-survival*-view, although initially offering an unconditional view of personhood surviving several bodily changes, is essentially dualistic and has many empiricist assumptions (i.e. equating persons to their brains) failing to answer our question. Kant and Wittgenstein's views on an unconditional third-person acknowledgment of others' personhood that we investigated, as the *identity-*

[72] Stanley Cavell, *Philosophy the Day after Tomorrow* (Cambridge, MA, and London, UK: The Belknap Press of Harvard University Press, 2006), 149.
[73] Ibid., 149-150.
[74] Ibid., 135.

doesn't-matter-to-morality- view, seem to satisfy our common beliefs and practices leaving space for other minds problem only in special circumstances where it can be raised (i.e. the society in 'The Matrix' movie).

Skepticism of other minds is not only a scandal as Cavell puts it, but is also a philosophical tendency in bioethics, where the utilitarian rationale ascribes degrees of personal identity to humans estimating accordingly their lives as worth living or not, as it is in the case and euthanasia or infanticide. Kant and Wittgenstein seem to take the third-person unconditional ascription of personhood to others as a brute fact, leaving no space for discriminating among kinds, or classes of persons. Cora Diamond in "The Importance of Being Human" suggests that we could put imagination in action to engage our moral concerns with disabled persons by trying to imagine the kind of lives they live. We should engage with an "imaginative elaboration of what it is to have a human life" as a "response to *having a human life to lead* – to what we find strange or dark or marvelous in it – may be seen as present in actions, thoughts, talk, feelings, customs."[75] She proposes it as an answer to the empiricist evaluations or classifications of human beings which "deny the existence of imaginative shaping of meaning, and [...] treat thought about morality as capable of going on without loss in a context emptied of all intimacy with such imaginative *shapings*."[76]

The persistence of the soul in life where the thinking being (as a human being) is at the same time an object of outer sense, is clear of itself (*CPR* B 415) as the human body is the best picture of the human soul (*PI, Part II, iv, p.178*).

[75] Diamond, "The Importance," 48.

[76] Ibid.

Abbreviations

Kant's writings are cited using the following abbreviations:

CPR Critique of Pure Reason
CPrR Critique of Practical Reason
APPW Anthropology from a pragmatic point of view

Wittgenstein's writings are cited using the following abbreviations:

RPP Remarks on the Philosophy of Psychology
PI Philosophical Investigations
TLP Tractatus Logico-Philosophicus

References

Allison, Henry E. *Kant's Transcendental Idealism. An Interpretation and Defense*. New Haven and London: Yale University Press, 2004.

Anscombe, G. E. M. "On Brute Facts." *Analysis* 18, no. 3 (1958): 69-72.

Boutlas, Georgios, and Stelios Virvidakis. "Advance Directives and Personal Identity." *Bioethica* 4, no. 2 (2018): 17-32.

Cavell, Stanley. *Must We Mean What We Say?* Cambridge: Cambridge University Press, 1976.

Cavell, Stanley. *Philosophy the Day after Tomorrow*. Cambridge, MA, and London, UK: The Belknap Press of Harvard University Press, 2006.

Cavell, Stanley. *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy*. Oxford: Oxford University Press, 1979.

Dain, Edmund. "Do We Believe in Other Minds?" *Austrian Ludwig Wittgenstein Society* 36 (2016): 45-47.

Dain, Edmund. "Wittgenstein on Belief in Other Minds." https://www.academia.edu/22598240/Wittgenstein_on_Belief_in_Other_Minds.

Descartes, René. *The Philosophical Writings of Descartes*. Translated by J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny. Cambridge: Cambridge University Press, 1985-1991.

Diamond, Cora. "The Importance of Being Human." *Royal Institute of Philosophy Supplement* 29 (1991): 35-62.

Frankfurt, Harry. "Autonomy, Necessity, and Love." In *Necessity, Volition, and Love,* 129-141. Cambridge: Cambridge University Press, 1998.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1971): 5-20.

Goodenough, J. M. "Parfit and the Sorites Paradox." *Philosophical Studies* 83 (1996):113-120.

Hyde, Dominic, and Diana Raffman. "Sorites Paradox." *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta (Summer 2018 Edition). https://plato.stanford.edu/archives/sum2018/entries/sorites-paradox/.

Kant, Immanuel. *Anthropology from a Pragmatic Point of View*. In *Anthropology, History, and Education*, translated by Robert B. Louden. Cambridge, New York: Cambridge University Press, 2007.

Kant, Immanuel. *Critique of Practical Reason.* In *Practical Philosophy*, edited and translated by Mary G. Gregor. Cambridge, New York: Cambridge University Press, 1999.

Kant, Immanuel. *Critique of Pure Reason.* Edited and translated by Paul Guyer, and Allen Wood. Cambridge, New York: Cambridge University Press, 1998.

Kitcher, Patricia. "Kant's Paralogisms." *The Philosophical Review* 91, no. 4 (1982): 515-547.

Kitcher, Patricia. "On Interpreting Kant's Thinker as Wittgenstein's 'I.'" *Philosophy and Phenomenological Research* LXI, no. 1 (2000): 33-63.

Kitcher, Patricia. *Kant's Transcendental Psychology*. New York, Oxford: Oxford University Press, 1990.

Locke, John. "Of Identity and Diversity." Reprinted in *Personal Identity,* edited by John Perry, 33-52. Berkeley: University of California Press, 1975.

Mill, J. S. *An Examination of Sir William Hamilton's Philosophy.* London, UK: Longmans, Green and Co., 1865.

Parfit, Derek. "The Unimportance of Identity." In *Identity*, edited by Henry Harris, 13-46. Oxford: Oxford University Press, 1995.

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press,1984.

Protopapadakis, Evangelos. *From Dawn till Dusk: Bioethical Insights into the Beginning and the End of Life*. Berlin: Berlin Logos Verlag, 2019.

Rorty, Richard. *Contingency, Irony, and Solidarity*. Cambridge, New York: Cambridge University Press, 1989.

Schechtman, Marya. *Staying Alive, Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: Oxford University Press, 2014.

Shoemaker, David. "Personal Identity and Ethics." In *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2019/entries/identity-ethics/.

Singer, Peter. *Practical Ethics*. Cambridge: Cambridge University Press, 1993.

Smith, Barry, and John Searle. "The Construction of Social Reality: An Exchange." *The American Journal of Economics and Sociology* 62, no. 1 (2003): 285-309.

Strawson, Peter. *The Bounds of Sense. An Essay on Kant's Critique of Pure Reason*. London, New York: Routledge, 1990.

Techio, Jonadas. "Seeing Souls: Wittgenstein and Cavell on Other Minds." *Conversations: The Journal of Cavellian Studies* 1 (2013): 63-84.

Velleman, David. *Self to Self Selected Essays*. New York: Cambridge University Press, 2007.

Warren, J. "Precursors of Pyrrhonism: Diog. Laert. 9.67-73." In *Pyrrhonian Skepticism in Diogenes Laertius,* edited by K. Vogt, 105-121. Tübingen: Mohr Siebeck GmbH and KG, 2015.

Winch, Peter. "Eine Einstellung zur Seele." *Proceedings of the Aristotelian Society, New Series* 81 (1980-1981): 1-15.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Basil Blackwell, 1986.

Wittgenstein, Ludwig. *Remarks on the Philosophy of Psychology*, vol. I. Edited by G. E. M. Anscombe, and G. H. von Wright, translated by G. E. M. Anscombe. Oxford: Basil Blackwell, 1980.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Translated by C. K. Ogden, with an introduction by Bernard Russel. Mineola, NY: Dover Publications, 1999.

Wood, Allen. "Kant on Duties Regarding Nonrational Nature." *Aristotelian Society* Supplementary Volume 72 (1998): 189-210.

The Phenomenological Understanding of the Person: Nietzsche in Husserl's Shadow

Dragan Prole
University of Novi Sad
E-mail address: proledragan@ff.uns.ac.rs
ORCID iD: https://orcid.org/0000-0001-7352-4583

Abstract: The article deals with the phenomenological notion of person. The basic hermeneutical intention is to show the similarity of Nietzsche's and Husserl's "anthropology," that is, of what characterizes man. The first part of the paper presents Nietzsche's critique of "Egypticism," that is, the thesis of eternal human nature, what is unchangeably, essentially human. Consciousness is determined by historicity, time, space, its moving nature, and something like that cannot have a permanent, unchanging constant. In the second part of the paper, the author shows why Husserl's phenomenology represents the most productive legacy of Nietzsche's imperative "To the sea, philosophers!" The third part brings a metaphorical variation on the theme of Ecce Homo "Why is phenomenology destiny?" Modern teaching about personality insists on radical change, in which the phenomenological science of the necessity of abandoning the natural attitude can be a valuable guide.
Keywords: Nietzsche; Husserl; person; egypticity; natural attitude.

The term personhood is certainly not one of the more recognizable phenomenological concepts. Today, the methodological concepts such as epoche, reduction and variation, and *termini technici*, for example eidos, intentionality and constitution are much more in focus than personhood. The unfortunately named, and only occasionally mentioned term of *Wesenschau* was studied far more than the idea of personhood. An irrelevant and insignificant term drew more of readers' attention and incited more interpretative effort than the core, but not as conspicuous, quaint idea of personhood. This information might seem unusual when we realize that

doi: https://doi.org/10.12681/aprlp.49.922

the phenomenological understanding of personhood owes its existence to the extremely fierce discourse. Very few terms provide an insight into the construction of a specific phenomenological break with tradition. In it, it is possible to see the signs of a modified understanding of man, where the roots of those changes still reach deep into tradition. Phenomenological person will be neither substance, nor a purely natural being, nor a psychological reflex of the current. In a word, Husserl found the notion of "human essence," "human nature" and similar platitudes, ever-present in the pseudo-scientific argumentation, as deeply foreign. From the phenomenological perspective, this natural-scientific tendency to suppress and erase man's historicity seems quite unusual, as it assumes that the timeliness of consciousness is reduced to the naturality of a plant or a rock. It is as if the development of natural sciences emboldened, in the long term, the attempts to finally "scan" and solve the riddle of humanity. Anthropologism, psychologism, scientism, naturalisms, philosophy of the mind, are just different names of the same modern strategy. Each of them gives a small contribution to the "petrification" of man, that is, the convincing simulation of inalterability and absolute stability where constant flow and movement are exclusively present. Nietzsche calls such a strategy "egypticity," calling out the philosophers for the tendency to de-historize the object of their examination, making mummies out of their concepts. The philosopher is another name for the talent to transform a living phenomenon into a dead thing:

> You want to know what the philosophers' idiosyncrasies are? [...] Their lack of historical sense for one thing, their hatred of the very idea of becoming, their Egypticity. They think that they are showing respect for something when they de-historicize it, sub specie aeterni, II - when they turn it into a mummy. For thousands of years, philosophers have been using only mummified concepts; nothing real makes it through their hands alive.[1]

[1] Friedrich Nietzsche, *The Anti-Christ, Ecce Homo, Twilight of the Idols, and*

Husserl adds that egypticity cannot be reduced to dehistorization, as it creates an entire attitude, ironically called "the natural attitude." Even though nominally natural, that attitude is emphatically anti-life. Furthermore, the dominance of the natural-scientific world is not possible without the suppression of the living world. Due to that, the reign of the natural attitude is at the same time a document of self-oblivion of the personal worldly life.[2]

Phenomenology of the person is not a creation of a specific time and space. There's no eternal, permanent or essential personality, but only an attempt to "tear" such a personality from self-understanding, the everyday milieu in which it is built with the logic of the natural attitude. Aware of the actual spiritual ambience in which the "scientific" modernity understands personality, phenomenology is initially forced to "isolate," that is, to suspend all publicly endorsed convictions. Therefore, the phenomenological idea of personhood does not start with assertion, but with a negation. The man is not a substance, there are no stipulations that would repeat themselves as an unavoidable constant. Ontologically, *the man is not, but is only becoming* – "it is not understandable why the world is natural, but also a personal factum of specified development. Incomprehensible: the specified order of feelings, personal distinctions."[3]

I. Can there be truth as consent without permanent human nature?

The relationship between the understanding of the man and understanding of the truth is very interesting. Where the conviction that man's essence is permanent and unalterable

Other Writings, trans. J. Norman (New York: Cambridge University Press, 2005), 166-167.

[2] Lothar Eley, *Die Krise des Apriori in der transzendentalen Phänomenologie Edmund Husserls* (Den Haag: M. Nijhoff, 1962), 105.

[3] Edmund Husserl, *Grenzprobleme der Phänomenologie. Analysen des Unbewusstseins und der Instinkte, Metaphysik, Späte Ethik*, Husserliana Band XLII, ed. R. Sowa, and T. Vongehr (Dodrecht: Springer, 2013), 18.

is prevalent, and his "nature" tacit and defined, is the place where the truth is perceived as the consent of the substantial subject and the substance, which presents the object of its understanding. As is the case with personhood, the truth's comfort zone is in dissent. rather than in consent. If we assume that the former term of truth as *adequatio* necessarily came from the understanding of human nature as permanent, constant, completely understandable and determinable, the question imposed is if the *adequatio* can ever be a reliable criteria in situations where the subjectivity is determined as an *ultimate existence*, and with that constantly variable?

Husserl's phenomenology announces the primary course of philosophy of the 20th century. Most of its insights have had their foundation in the idea of *Mehrmeinung*, which tells us that the object is always something more than what we have thought or known. No matter how much we think, no matter how good we get to know it, the object is always something that is more than what was thought of it. The same logic works in reverse on the phenomenological concept of personhood. This means that personhood is always something more that the knowledge of it. Learning about *Mehrmeinung* is valid even when it is about our own personhood. Due to that, the idea of self-knowledge is unavoidably illusionary and impossible.

There is no complete "knowing thyself," as the entirety of life is, at best, getting to know thyself: "I can be more and different than the I as an apercetive unity [...] Nobody 'knows' themselves, no one 'knows' what they are, they get to *know themselves*."[4]

There is no overlap of words and things, which means reality can never truly be translated into a concept. Modernity seriously considers the contingent of the order of things. Indeterminacy is its constitutive principle, which has consequences on the recognition of the unknowability of human personhood, that

[4] Edmund Husserl, *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Phänomenologische Untersuchungen zur Konstitution*, Husserliana Band IV, ed. M. Biemel (Den Haag: M. Nijhoff, 1952), 252.

is, on the "unpredictability of human affairs."[5] If the modern era had a doorway, on them we could find the first part of Ludwig Wittgenstein's *Tractatus*, according to which the world is everything that is accidental (*Die Welt ist alles, was der Fall ist*).[6] The idea which belongs more to poetry than to the philosophical way of thinking, from a rationalistic perspective, fits perfectly into the phenomenological horizon. Subjectivity in the outer relations seems predictable, reliable and calculable, but is in itself the very opposite: unpredictable, unreliable, incalculable. The Uncertainty principle is well emphasized by Husserl's idea of a principal and inevitable contingency of world facts: "every factum, including the factums of the world are, as factums, contingent."[7]

Speculative thinking assumes mindfulness, and not the contingent of the thought. If a factum of the world is not mindful, but contingent, then it is necessary to find a different thought approach. The "guy upstairs" can hardly be of assistance. Instead of speculative understanding, Husserl chooses description. Unlike the euphoric, already pre-defined criticism of phenomenology *à la Adorno*, the point of description is not in the conservation of things. There's nothing easier than to banalize description – it shows everything as it is, and it does that because it is an expression of the hidden attempt to preserve everything as it is, and not allow for any changes to be made. Description? But a new name for an old bourgeois style of philosophizing, like phenomenological certainty: "Exuberance toward raw factness does not prevent us from accepting the world of things from being accepted 'as is.'"[8]

---

[5] Gerhard Gamm, *Der unbestimmte Mensch. Zur medialen Konstruktion der Subjektivität* (Berlin, Wien: Philo, 2004), 11.

[6] Ludwig Wittgenstein, Tractatus logico-philosophicus (Sarajevo: V. Masleša, 1987), 26.

[7] Edmund Husserl, *Erste Philosophie II. Theorie der Phänomenologischen Reduktion*, Husserliana Band VIII, ed. R. Boehm (Den Haag: M. Nijhoff, 1959), 50.

[8] Theodor W. Adorno, *Metakritik der Erkenntnistheorie. Studien über Husserl und die phänomenologische Antinomien* (Stuttgart: W. Kohlhammer, 1956), 141.

II. The observer and the point of observation are variables

Far from representing a pure apparition, description in phenomenology takes care of the transcendent character of subjectivity. Unlike the idea, which was self-implied for centuries, phenomenology final breaks with the idea of a stable, founded point of philosophical observation. If there ever was a thinker who literary applied Nietzsche's anti-Cartesian imperative. "To the sea, you philosophers!"[9] that would be Edmund Husserl. Even though phenomenological terminology is vibrant with Descartes' favorite concept – foundation (Begründung), that concept is with Husserl above all reminiscent of Plato's philosophical argumentations. Instead of searching for solid, secure foundations on which to build up philosophical knowledge, for Husserl, *Begründung* means *logon didonai*, the revelation, philosophical "settling of the scores."

Like the mobile camera of the Russian Dziga Vertov, Husserl's consciousness is mobile, instead of Archimedes' stable point, it is guided by kinesthesia, the sense of movement. Despite being corporeal and movable, Husserl's consciousness is, unlike the camera, expressly variable. In its flow we see the imprint of transcendence, the transformation of I, which happens in accordance with the essential lawfulness – "otherness of the final subjectivity which is inherit to it with its inner necessity."[10] Description is then not the display of something changeable from the perspective of the immovable, nor is the display of the changeable with the help of the mechanism that is fixed from inside but can change the point of view. Transcendence, like the otherness of subjectivity, becomes the immanent part of the description within the confines of egology, that is, the phenomenological explication of ego through ego. With Husserl,

---

[9] Friedrich Nietzsche, *Fröhliche Wissenschaft*, § 289, Kritische Studienausgabe, Band 3, ed. G. Colli, and M. Montinari (München, Berlin: DTV/De Gruyter, 1980), 530.

[10] Walter Schulz, *Der Gott der neuzeitlichen Metaphysik* (Pfullingen: G. Neske, 1957), 29-30.

the power of transcendence is explored in the first person, personhood is perceived in a constant change, *the change of the living world inevitably forces the change of personality*: "In that constant shift of man's living world, it is obvious that people as personalities change too, insofar as they correlatively have to obtain new characteristics."[11]

It is probably not by chance that the method of description is for both the Munich and Göttingen circles of phenomenologist the fundament of phenomenology. Being that description contains the inevitability of personal convictions, that is, the personal experience of the truth, it acted as a practical cornerstone for the idea that understood "the truth constituted on the source givens as the authentic truth."[12] Personal convictions and source givens with Husserl are placed in the function of a struggle against psychologisms and naturalism. Unlike its contemporaries, James, Russell, Mach, phenomenology fiercely opposed every attempt of the naturalization of consciousness, being that it also has naturalization of ideas as a consequence, or the naturalization of ideals and norms. Contrary to the fanatical obsession with psychology which Husserl called out throughout his career, phenomenology is conceived in the accidental, uses description, insists on the personal convictions, and does all that for the establishment of essential absolutes and apodictic laws. A similar constellation can be found when discussing the relationship between the individual I and the general sense of personhood:

> Although the personal I is individual, we can determine according to universal eidetic laws what this I is, an I which can be comprehended only through living familiarization with an actual cogito.[13]

---

[11] Edmund Husserl, *Cartesianische Meditationen und Pariser Vorträge*, Husserliana Band I, ed. S. Strasser (Den Haag: M. Nijhoff, 1973), 162.

[12] Ernst Tugendhat, *Der Wahrheitsbegriff bei Husserl und Heidegger*, 2nd Edition (Berlin: W. de Gruyter, 1970), 229.

[13] Rudolf Bernet, Iso Kern, and Eduard Marbach, "The 'I' and the person," in *Edmund Husserl. Critical Assessments of Leading Philosophers*, Volume IV, ed. R. Bernet, D. Welton, and G. Zavota (London, New York: Rout-

In short, phenomenology probably owes its success to the strategy of using the means of positivism and empiricism only to the extent necessary to leave those positions. In its initial phase, phenomenology only seemingly shares radical empirical starting points, while is as far from them as possible in the later stages. This is not about Husserl initially accepting the position of empiricism to later abandon it. The idea of transcendent retroactive determines what we perceive as empirical, until the inverse is valid, that the transcendent in all forms comes from the empirical. Accordingly, Husserl's transcendental motive always determines in advance what we would like to denote by the notion of natural attitude.[14]

III. Phenomenological confrontation with the accidental

In order to be able to say anything about the truth in "maritime conditions," devoid of foundation and support, we must first face the coincidence of each particular experience. Husserl's phenomenology does this by pausing, restraining, "bracketing" the validity of the immediately experienced. In Nietzschean spirit, phenomenology starts with the idea that there are more idols in the world than there is reality. Because of this, it insists on a certain immunity to the immediate reality, convinced that it is only at a distance from it that there is room for thought. Being that transcendental subjectivity alone is not related to the immediately existing, phenomenology is necessarily a transcendental philosophy. Philosophy that calls for unconditionality necessarily assumes immunity to immediacy. Apart from the confrontation with the conditionality, *phenomenology demasks the false presentation, that is, teaches that the experience contents can have different meaning from the one normally attributed to them*. In short, after the denunciation of

ledge, 2005), 311.

[14] Jean Grégori, "L'attitude personnaliste entre naturalité et transcendentalité – le problème du 'quotidien' dans Ideen II d'E. Husserl," *Arhe* II, no. 4 (2005): 24.

the apparent absolute of the given the objectively experienced, comes the pointing out of everything that was not visible or accessible before the denunciation. Husserl's terms of disclosure (Enthüllung) and Heidegger's idea of truth (Unverborgenheit), indicate a plain that was initially hidden, and which only becomes accessible through the demasking of everything it hides. Through that, what was invisible and unknown becomes visible and known. Yet, phenomenological description has nothing to do with going behind the obscure, but recognizable veil. Unlike the naïve search for the projected depths, description did not want to explore what hides beneath the fallible and unreliable surface. Simply put, the phenomenological method can be perceived as a thought construction which is preceded by deconstruction. Due to oversaturation with historical sense, Husserl tacitly shared Nietzsche's belief that a powerful dismantling is required wherever you wish to set up something.

In contrast to his time, Husserl did not believe that philosophical rationality is possible if one insists on the fascination with the "facts" or "empirical data." His unwritten *Untimely Meditations* could have been ironically titled "About the harms of positivism to life" or "Immediacy as a tutor." If the fundamental impulse of modernity in phenomenology is alive and vital anywhere, then it is in the view that the truth of an object of thought lies not in approaching, but in moving away from its immediate givenness. Distancing oneself from the immediacy still does not imply nearing the dept and the beyond in the traditional sense. Contemporary thinkers no longer believe in a world of essences which is somewhere "beyond" the visible, tangible, experienced. The contemporariness is principally against the traditional dichotomies of the deep/surface, truth/lie, essence/simulacrum, reality/phenomenon. Instead of these, "vertical" binary oppositions in which one side is implicitly positive and desirable, and the other implicitly negative and undesirable, the spirit of modern philosophy cares far more about the "horizontal" treatment which does not work with the higher and lower, but with the different, mutually irreducible and incomparable

methods. In Husserl's terms, with the different "attitudes," that is, suprapersonal configurations of the relationship between consciousness and the world. Still, even in phenomenology, the unbridgeable difference, built on the axis of the philosophical-unphilosophical, transcendental-mundane, is still very much alive. Not all ideas are equally important, and it would not be wise to claim them equal. Some allow for a meaningful and responsible life, while others promise only stagnation in immaturity and naïve dependency. Likewise, dogmatic, self-comprehensible exaltation of the transcendental in relation to the mundane is not instructive, since the transcendental attitude, if the individual in it relies solely on himself, may well end in alienated pride:

> Closing oneself within the confines of the natural attitude (naturalization) is alienation, but alienation is also the retreating of the subject into itself, the tearing under a subjectivist (Cartesian) form.[15]

Only when it realizes what it owes to other forms, can the phenomenological idea of personhood expect to escape the traps of alienation and vanity.

IV. Phenomenology is the restoration of personhood

Starting with the *Idea I*, Husserl's transcendental phenomenology was understood as a step outside an idea, for the sake of assimilation and appropriation of another. Becoming a philosopher inevitably means changing oneself, becoming a different person. Husserl tied his philosophical program explicitly to the idea of *renewal*. Renewal is mostly tied to the legacy of the philosophical institution. In the spirit of his avant-garde contemporaries, Husserl's phenomenology rejects tradition, but only through accepting the task of revealing and accomplishing its basic motives and tendencies. Unlike the avant-gardists, who

[15] Rober Legro, *Ideja humanosti, Izdavačka knjižarnica Zorana Stojanovića* (Sremski Karlovci, Novi Sad: J. Popov 1993), 225.

paradoxically attempted to create their own tradition through the various break gestures, Husserl's paradox is in the idea of abandoning tradition through the realization of its original motives and goals. Because of that, it is recommended to carefully listen to Husserl's suggestion with which he started his lecture at Sorbonne. At first glance, it is typically German and protestant. Very much in the spirit of Fichte's idealism, according to which the "Philosophy wisdom (sagesse) is the philosophizer's quite personal affair."[16] However, when the famous Jena professor claimed that philosophy is not a piece of furniture, that its choices do not depend on the aesthetical, but character criteria, he above all, said that we cannot expect an unfree person to choose idealism as his modus of thinking: "A person indolent by nature or dulled and distorted by mental servitude, learned luxury, and vanity will never raise himself to the level of idealism."[17] In short, to think freely, I must be free. Philosophy will not set me free, as I will always choose an unfree philosophy as an unfree man.

If we could sum up Fichte's morals into a single idea: my choice of philosophy is dictated by the type of man I am - then the noted affinity in the case of phenomenology related mostly to the changes in one's own personality. Husserl attempts to disrupt Fichte's rigid and irreconcilable differences between the free and the unfree. Instead of it being predetermined, it would be far better for the line separating them to be dynamic and changeable. Thought that, the change with Husserl does not imply the advance of something unknow or completely new.

Change gives name to renewal in the sense of living in accordance with the ideal of true, true humanity. This idea of "true humanity" is not a determined substance, but a continuous effort of change and self-abandonment for the sake of conquest of something new: "At every moment here, humanity has been overcome, the idea

[16] Edmund Husserl, *Cartesian Meditations. An Introduction to Phenomenology*, trans. D. Cairns (Den Haag: M. Nijhoff, 1960), 2.

[17] Johann Gottlieb Fichte, "Erste Einleitung in die Wissenschaftslehre," in *Johann Gottlieb Fichtes Sämmtliche Werke* Band I, ed. I. H. Fichte (Berlin, 1845/46), 434.

of 'overman' has become the highest reality."[18] Paradoxically, change thus brings together both future hopes and aspirations of the past. Renewal, according to Husserl, encompasses both the individual and the human community. As such, the renewal is the highest theme of ethics, the science of the essence of possible ways of life in an a priori generality.[19] As an idea, renewal offers a common link between the specific and the general, between the individual and the world community. Within that horizon we catch the glimpse of the person of the higher degree. Husserl makes a strict distinction between the solipsistic-individual purposes and achievements on one side, and the common purposes and achievements on the other. In doing so, common purpose has a completely different spiritual meaning from that which can be obtained by the action of an individual subject as part of a community. Along those lines, the personalities of the higher order become "officials,"[20] in a sense that the community is represented as a single individual.

The state is for Husserl the will of the whole in which the singular individual becomes the "person of a higher order." In short, compared to the collective personhood, the personhood of the individual as an empirical subject is irrelevant for Husserl. Instead of dealing with the individual, renewal of philosophy for Husserl is possible only if he manages to pull off the actual objectivization of the universal human sense, and not the description of an arbitrary, accidentally picked empirical type.

Every individual awareness is interpreted by genetical phenomenology as a living history of its own making. Every

---

[18] Friedrich Nietzsche, *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, trans. J. Norman (New York: Cambridge University Press, 2005), 130.

[19] Edmund Husserl, "Fünf Aufsätze über Erneuerung," in *Aufsätze und Vorträge (1922-1937)*, Husserliana Band XXVII, ed. T. Nenon, and H.-R. Sepp (Dodrecht, Boston, London: Springer, 1989), 20.

[20] Edmund Husserl, *Einleitung in die Ethik. Vorlesungen Sommersemester 1920/1924*, Husserliana Band XXXVII, ed. H. Peucker (Dodrecht: Springer, 2004), 359.

personality has its own "positive potentiality,"[21] its entire previous existence is summed up in a moment with its willingness to go straight to the point, to act in a very specific manner. If the factum of the world is accidental, then the willing actions certainly are not. If we recall that Husserl's contemporary Rilke called for reform movements by asking his readers to alter their way of life, phenomenology could be read as a certain guidebook for the transformation of life from an unphilosophical one to a philosophical. Such reading would become even more interesting if we were to subject phenomenology to the strategy of denunciation. Its essential question must be: can the founder of phenomenology be also accused of false pretenses? What is the concept of personhood that the phenomenological transformation into a philosopher offers us?

V. Why is phenomenology destiny?

Careful readers did not miss the almost existential sound of Husserl's sentences, in which the possibility of philosophical explanations and settling of the scores is presented as the only available option without which it is not possible, or conceivable to live: "Many statements, which consider 'life threatening' and one 'can't keep on living,' point in that direction."[22] The encounter with Husserl's texts will testify to the fact that in life it is possible to have different choices, but in time, it becomes clear that any possibility of choice for a phenomenologist is but imaginary. Being that there is no other methodology thanks to which it is possible to obtain similar results, the orientation is reduced to only one – phenomenological route. If the phenomenological route is the necessary, or even the only possible route, it is by no means natural. Husserl did not find anything more senseless

[21] Edmund Husserl, *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Phänomenologische Untersuchungen zur Konstitution*, Husserliana Band XXXVII, ed. H. Peucker (Dodrecht: Springer, 2004), 255.

[22] Ferdinand Fellmann, "Lebenswelt und Lebenserfahrung," *Archiv für Geschichte der Philosophie* (Berlin: Walter de Gruyter, 1987), 90-91.

that the spontaneous shift from the natural to the philosophical attitude. Therefore, he could not even imagine an option in which someone is simply "born" to philosophy: "No one can be born a philosopher for substantive reasons. Anyone can start only as a natural, non-philosophical man."[23] If everyone starts as a non-philosopher, how is it that only a minority decides to pursue philosophy while a vast majority remains in the natural default? Husserl's methodology is marked by certain educational lessons. The explicit mention of the "educational possibilities" of the phenomenological reduction relates mostly to the possibility of accepting the change in the attitude. More precisely, what it actually means to accept an attitude, to pursue it, abandon it, and to adopt another attitude:

> […] the natural attitude is not the only possible attitude [...] the *educational part of the phenomenological reduction* is that it first and foremost makes us more receptive for the understanding changes in the attitude.[24]

Plato knew that nothing could be known in a cave about what a cave was, and that it was therefore necessary to get out of it. Understanding the change of the attitude is necessarily tied to the capacity of the abandonment of the usual human point of view. Still, it seems that the necessity of leaving the world to which we are tied with our senses was not something the ancient philosopher was able to explain. The regime of images was for him the natural starting point but the myth of the cave does not hold any suggestion which could help us understand why the noted individual decided to leave the cave, what is the thing that separated him from all the other tenants of the cave. Ontologically and epistemologically, the image in Plato's dialogues is that of

---

[23] Edmund Husserl, *Phenomenological Psychology. Lectures. Summer Semester 1925*, trans. J. Scanlon (The Hague: M. Nijhoff, 1977), 34.

[24] Edmund Husserl, *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Phänomenologische Untersuchungen zur Konstitution*, Husserliana Band IV, ed. M. Biemel (Den Haag: Springer, 1952), 179.

unity of the being and the non-being, the illusionary products, but his dialectic did not offer the necessity of the exit, nor did it show why it is necessary to reach a certain saturation with the world of images: "Nothing is harder than making the offer of freedom more appealing."[25]

Unlike Plato, whose fugitive from the world of images cannot carry out his escape on his own, but needs others to "forcefully pull him out into the sunlight,"[26] Husserl in the phenomenological reduction recognizes the individual path into freedom, that is, in the transcendental subjectivity, as only it points to the absolute, to the source of the being not tied to the existing. Although at the end of his *Cartesian Meditations* he cites St. Augustine's thesis that truth resides within man, Husserl was by no means concerned with offering one of the many variations on Judeo-Christian anthropology. The foul and corrupt corporeality is with him simply invalid as the position of the impeccable purity of the immortal soul, eventually saved by grace of God. It is without doubt that the phenomenological subject does not care about being made of this world. The adjective *mundane* in Husserl's register contains the pejorative connotation still nurtured in the English language tying the phrase to the banal, prophane, crude and earthly. Mundane interests are synonymous for a life dictated by trivialities, whose time goes by in dealing with the incidental and irrelevant, while at the same time neglecting the crucial and significant.

Mundane affinities necessarily force a man into becoming an amateur, layman and commoner. The connection between the earthly goods and values for a Christian signify a life of sin, but for Husserl they point to a *naïve life*. Both are essentially alienated, the first one from God, and the second from true humanity. The lack of the commoners is in the missed opportunities for repentance and salvation, and lack of the naïve subject is in the

---

[25] Hans Blumenberg, *Höhlenausgänge* (Frankfurt am Main: Suhrkamp, 1996), 87.

[26] Platon, *Država* 516a, trans. A. Vilhar, and B. Pavlović (Beograd: BIGZ, 1993).

fixation with the ordinary, the natural attitude. For a Christian and phenomenologist alike, the appreciation of the common suffers from the lack of change. Despite the possibility to become different, the commoner chooses to remain the same. Still change which the Christian expect is significantly different from the one requested by the phenomenologist. The antonym of *mundanity* for Husserl is no longer *holiness*, but the *phenomenological attitude*. The successful change of the attitude in phenomenology does not lead to salvation or eternal life, but to the only possible rational and responsible living. There's no more talk of salvation, only philosophizing. The crucial thing is that the choice of phenomenology is not a matter of theoretical orientation, but relates to the existential, life question. Does this imply that the phenomenologist took the place in which the saint once stood? Is it not then, from the Christian perspective, the phenomenological subject the typical representative of pride?

VI. Is egology necessarily the surrender to pride?

The Christian sensibility would certainly never approve of the philosophy which declares itself as egology. The phenomenologist would have been the ideal typical representative of pride, as he puts his subjectivity above everything, even above God. On the other hand, phenomenology does not recognize any intentional object other than the concrete modes of givens or fantasy produced variations of those givens. To such an attitude, every religion must seem naïve, as it is founded on the unprovable and unverifiable hypothesis:

> Within our actual experience we do not encounter divinity anywhere, and so exclude the questions of God as a transcendence of a different type than the transcendence of the objects of empirical sciences.[27]

[27] Zagorka Mićić, *Fenomenologija Edmunda Huserla. Studija iz savremene filozofije* (Beograd: F. Pelikan, 1937), 74.

The questions whether God exists or not, what our relationship with him would be, are necessarily excluded for methodological reasons. The primacy of the ego is, for phenomenology, indisputable, as it precedes all other eidetic necessities. With that, the ego is not just a contingent, individual being, created by chance and in the unpredictable social, economic and historical conditions. Every ego is the unique and inimitable, but is, despite that, marked by the essential necessities, which characterize the essence *qua* essence.

The existence of ego is absolute, and even if all the world givens were to vanish, that would not mark the end of ego. In short, ego can be without the world, while the world cannot be without ego. Ontologically, ego is the personhood in the sense of the unity of many, being a person for Husserl means being aware of one's center, the power of unity. Still, self-awareness alone is insufficient for personhood. Empirical self-awareness is, in particular, insufficient, as for reaching the person from ego, you need something else. Aside from the awareness of self, of one's convictions, desires, *social acts* like enjoyment, analogue apresentation and others are necessary.[28] Even though the discourse of egology can falsely point us in that direction, personhood cannot be conceived as a transcendental Robinson Crusoe, as it simply cannot function without other ego:

> It is only with empathy and the constant orientation of empirical reflection onto the psychic life which is apresented along with the other's Body and which is continually taken Objectively, together with the Body, that the closed unity, man, is constituted, and I transfer this unity subsequently to myself.[29]

---

[28] René Toulemont, *L'essence de la sociéte selon Husserl* (Paris: PUF, 1962), 230-231.

[29] Edmund Husserl, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy II. Studies in the Phenomenology of Constitution*, trans. R. Rojcewicz, and A. Schuwer (Dodrecht, Boston, London: Kluwer, 1989), 175.

Therefore, one cannot think of the transcendental ego as a superior, vain unity beyond the multitude of other subjects. Though it may seem confusing at first, the conclusion, according to which "the transcendental and intersubjective with Husserl do not go hand in hand,"[30] implies the superiority of the transcendental (intersubjective). We over the transcendental I. Contrary to the logic of common sense or the natural attitude, according to which the awareness of self prepares the way for the awareness of others, one of the crucial insights of Husserl's phenomenology is that "the experience of the world is not a private experience but the experience of the community."[31] However, this experience is, like any other, constituted as self-experience, in where lies the entire paradox of the phenomenological position. *Others are before me, but they cannot exist independent of me, they are initially there as givens in me*.

Experience of other is the necessary self-experience, it becomes understandable thanks to the phenomenological *inspectio sui*. The experience of apresentation, acting with-present, with Husserl is usually reduced to the awareness that the presence of others is analogous to my own, and that, based on the insight into others, I myself exist. Everything relies on the variations of selfness, the experience of the foreign presents the modification of me. Opposed to that, pride is based on the construction of the self as the superior unity, which sees as a disturbing factor every possible instance of the excellence of the other.

> A prideful person lives inside himself with a light source, even if that light is sometimes blindingly reflected by external objects. Those objects occlude the physical and spiritual qualities of that person […] all that, for pride, is out of the question. A prideful

[30] Walter Brüning, "Der Ansatz der Transzendentalphilosophie in Husserls Cartesianischen Meditationen," *Zeitschrift für philosophische Forschung* XX (1966): 195-196.

[31] Edmund Husserl, *Formale und transzendentale Logik. Versuch einer Kritik der logischen Vernunft* (Halle: M. Niemeyer, 1929), 209.

> person does not say: 'I am the one representing so and so, and who did this and this', on the contrary, he says just 'I am Me.'[32]

Fichte's notion of subject turns out to be the ideal typical representative of pride, but a revolutionary, rebellious version which no longer accedes to opportunism.[33]

Unlike a prideful person, the phenomenological notion of personhood sees that without others, there is no intimate personal world. What the V *Cartesian Meditation* rejects as principally impossible, in the attitude of pride becomes a regular ordeal. Instead of finding the testimony of self in the intersubjective mutuality, the prideful subjectivity creates a self-relationship through excluding the possibility of being molded in exchange with others. Through the illusion of the inner primordial world which is supposed to be the "source," meaning that it genetically precedes every objectivity, pride functions by closing into itself. However, unlike shame, in which the individual retreats into itself as a way of defending from the binding generalities, norms and ideals, pride functions by "leaving itself" in the attempt to break and deny the validity of any generality, norm or ideal.

Radically observed, pride is the existence without the existing, and therefore, the attitude of pride can bless the destruction of all things, even if they are symbolical or imaginary. In this, pride is not just different, but contrary positioned to the phenomenological attitude, and the idea of phenomenological personhood. The attitude of pride needs no one, it does not recognize any potentiality of interaction, and therefore excludes the constitutive horizon of the things between us, *entre nous*. Pride is marked by "broken intentionality," in it the noesis projects itself sky high, without caring to establish real contact

---

[32] Aurel Kolnai, *Ekel, Hochmut, Haß. Zur Phänomenologie feindlicher Gefühle* (Frankfurt am Main: Suhrkamp, 2007), 68.

[33] Dragan Prole, "Metamorfoze gordosti. Od Aristotelovog samoodnosa do Kafkinog stida," *Gordost*, Adresa (Novi Sad, 2014), 16.

with the experience givens. Pride does not reflect the things that constitute and neglects the noema. Due to that, the others and the world are for a prideful person only what he wants to see in them, while the reflection of the given, reductions and variations are simply irrelevant. Phenomenological attitude represents the methodological path which starts with the testimony of the other to build a "world of personalities, their achievements, the kingdom of freedom,"[34] while pride remains a synonym for the tendency of the arbitrary I to, beyond all other I's, present itself as the absolute and indisputable. Even though the ideas of the *Übermensch*, the will to power and the "experience of seven loneliness"' point to the prideful nature of the "free spirit," the subjectivity it represents is anything but absolute. The Protean character of human existence, unlike the prideful immunity towards the existing, insists on participation with it. To live means to go towards things and phenomena, the Dionysian man: "He enters into any skin, into any affect: he constantly transforms himself."[35] The modern idea of personhood brings the radicalized teaching of change. Twists and turns are no longer interpreted as unique, imitable events in life. On the contrary, they are human life.

References

Adorno, Theodor W. *Metakritik der Erkenntnistheorie. Studien über Husserl und die phänomenologische Antinomien*. Stuttgart: W. Kohlhammer, 1956.

Bernet, Rudolf, Iso Kern, and Eduard Marbach. "The 'I' and the person." In *Edmund Husserl. Critical Assessments of Leading Philosophers*, Volume IV, edited by R. Bernet, D. Welton, and G. Zavota. London, New York: Routledge, 2005.

Blumenberg, Hans. *Höhlenausgänge*. Frankfurt am Main: Suhrkamp, 1996.

[34] Edmund Husserl, *Einleitung in die Ethik. Vorlesungen Sommersemester 1920/1924*, 313.

[35] Nietzsche, *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, 197.

Brüning, Walter. "Der Ansatz der Transzendentalphilosophie in Husserls Cartesianischen Meditationen." *Zeitschrift für philosophische Forschung* XX (1966): 195-196.

Eley, Lothar. *Die Krise des Apriori in der transzendentalen Phänomenologie Edmund Husserls*. Den Haag: M. Nijhoff, 1962.

Fellmann, Ferdinand. "Lebenswelt und Lebenserfahrung." *Archiv für Geschichte der Philosophie*. Berlin: Walter de Gruyter, 1987.

Fichte, Johann Gottlieb. "Erste Einleitung in die Wissenschaftslehre." In *Johann Gottlieb Fichtes Sämmtliche Werke* Band I, edited by I. H. Fichte. Berlin, 1845/1846.

Gamm, Gerhard. *Der unbestimmte Mensch. Zur medialen Konstruktion der Subjektivität*. Berlin, Wien: Philo, 2004.

Grégori, Jean. "L'attitude personnaliste entre naturalité et transcendentalité – le problème du 'quotidien' dans Ideen II d'E. Husserl." *Arhe* II, no. 4 (2005): 7-24.

Husserl, Edmund. "Fünf Aufsätze über Erneuerung." In *Aufsätze und Vorträge (1922-1937)*, Husserliana Band XXVII, edited by T. Nenon, and H.-R. Sepp. Dodrecht, Boston, London: Springer, 1989.

Husserl, Edmund. *Cartesian Meditations. An Introduction to Phenomenology*, trans. D. Cairns. Den Haag: M. Nijhoff, 1960.

Husserl, Edmund. *Cartesianische Meditationen und Pariser Vorträge*, Husserliana Band I, edited by S. Strasser. Den Haag: M. Nijhoff, 1973.

Husserl, Edmund. *Einleitung in die Ethik. Vorlesungen Sommersemester 1920/1924*, Husserliana Band XXXVII, edited by H. Peucker. Dodrecht: Springer, 2004.

Husserl, Edmund. *Erste Philosophie II. Theorie der Phänomenologischen Reduktion*, Husserliana Band VIII, edited by R. Boehm. Den Haag: M. Nijhoff, 1959.

Husserl, Edmund. *Grenzprobleme der Phänomenologie. Analysen des Unbewusstseins und der Instinkte, Metaphysik, Späte Ethik*, Husserliana Band XLII, edited by R. Sowa, and T. Vongehr. Dodrecht: Springer, 2013.

Husserl, Edmund. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Phänomenologische Untersuchungen zur Konstitution*, Husserliana Band IV, edited by M. Biemel. Den Haag: M. Nijhoff, 1952.

Husserl, Edmund. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Phänomenologische Untersuchungen zur Konstitution*, Husserliana Band XXXVII, edited by H. Peucker. Dodrecht: Springer, 2004.

Husserl, Edmund. *Phenomenological Psychology. Lectures. Summer Semester 1925*, translated by J. Scanlon. The Hague: M. Nijhoff, 1977.

Husserl, Edmund. *Formale und transzendentale Logik. Versuch einer Kritik der logischen Vernunft*. Halle: M. Niemeyer, 1929.

Husserl, Edmund. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy II. Studies in the Phenomenology of Constitution*, translated by R. Rojcewicz, and A. Schuwer. Dodrecht, Boston, London: Kluwer, 1989.

Kolnai, Aurel. *Ekel, Hochmut, Haß. Zur Phänomenologie feindlicher Gefühle*. Frankfurt am Main: Suhrkamp, 2007.

Legro, Rober. *Ideja humanosti, Izdavačka knjižarnica Zorana Stojanovića*. Sremski Karlovci, Novi Sad: J. Popov 1993.

Nietzsche, Friedrich. *Fröhliche Wissenschaft*, Kritische Studienausgabe, Band 3, edited by G. Colli, and M. Montinari. München, Berlin: DTV/De Gruyter, 1980.

Nietzsche, Friedrich. *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, translated by J. Norman. New York: Cambridge University Press, 2005.

Platon. *Država*, translated by A. Vilhar, and B. Pavlović. Beograd: BIGZ, 1993.

Prole, Dragan. "Metamorfoze gordosti. Od Aristotelovog samoodnosa do Kafkinog stida." *Gordost*, Adresa. Novi Sad, 2014.

Schulz, Walter. *Der Gott der neuzeitlichen Metaphysik*. Pfullingen: G. Neske, 1957.

Toulemont, René. *L'essence de la sociéte selon Husserl*. Paris: PUF, 1962.

Tugendhat, Ernst. *Der Wahrheitsbegriff bei Husserl und Heidegger*, 2nd Edition. Berlin: W. de Gruyter, 1970.

Wittgenstein, Ludwig. *Tractatus logico-philosophicus*. Sarajevo: V. Masleša, 1987.

Zagorka, Mićić. *Fenomenologija Edmunda Huserla. Studija iz savremene filozofije*. Beograd: F. Pelikan, 1937.

Self, Personality and Consciousness: Could Hellenistic Philosophical Approaches Have a Place in Modern Neuroscientific Research?

Panagiotis Kormas
National and Kapodistrian University of Athens
E-mail address: pkormas@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0001-7767-567X

Antonia Moutzouri
National and Kapodistrian University of Athens
E-mail address: amoutzouri@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0001-7078-3037

Abstract: The philosophers of the Hellenistic period devoted more attention to the foundations of knowledge than to knowledge itself, shifting the philosophical concern to the study of the individual and bringing the notion of the 'self' and the interrelated concepts of personality and consciousness under the spotlight. Both Epicurean and Stoic philosophers established a materially oriented system; animus and hegemonikon, correspondingly representing the seats of consciousness for the two schools, are the faculties that govern the human actions by ascribing personalised mental representation to the stimuli received by the sense organs. Today, neurosciences investigate the material substrate of consciousness and study its manifestations in the brain activities. Human consciousness, as approached by the Hellenistic philosophical theories seems relevant within the modern inquiries of what is the 'self,' of how the individual becomes aware of his sensory world or of the relation between emotion and intellect.
Keywords: Stoics; Epicureans; psyche; consciousness; self; mind; neural; brain.

Consciousness has been at the centre of interest throughout the evolution of philosophy, epitomized today in what is widely known as "the mystery of the brain." Consciousness is a complex function that allows a person to mentally be aware of the situation in which

doi: https://doi.org/10.12681/aprlp.49.923

he is at a certain point of time with the help of knowledge of his personality as well as to take hold of an idea of his future status.

Etymologically and conceptually there is a connection between the notions of consciousness and conscience. The Latin origin of the terms is 'conscientia,' whose etymology has been complex but it seems to be a translation of the Greek term 'syneidesis' which means the knowledge that one shares with oneself.[1] Today, the two terms are used in a different meaning; conscience "denotes the activity or the ever-vigilant readiness of a faculty of internal moral feeling or judgment," while consciousness "denotes the content as well as the activity of an ongoing and at bottom involuntary psychological reflection encompassing all of our actual experience."[2]

Being at the interplay between mental content and physical substrates, consciousness is at the core of what has been known as the Mind-Body problem. It is philosophically approached from one of the two standpoints; on the one hand, mind can be considered a bodily construct, conceived in terms of physiology and, on the other hand, mind is a subjective, "introspectable" element.

The distinctive importance of the Hellenistic philosophies in the frame of the present-day interdisciplinary study of the mind and its processes lies in the fact that, historically, it is the period when the study of consciousness and mindfulness, of introspection, of awareness of the functions of the self, actually commenced. This shift in the concern of philosophers from knowledge itself to the possibility of knowledge and the means by which truth may be discovered, has been characterized as an epistemological turn.[3]

---

[1] Udo Thiel, *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume* (New York: Oxford University Press, 2011), 8.

[2] Michel Weber, and Anderson Weekes, *Process Approaches to Consciousness in Psychology, Neuroscience, and Philosophy of Mind* (Albany, NY: State University of New York Press, 2009), 76.

[3] Jacques Brunschwig, "Introduction: The Beginnings of Hellenistic Epis-

The Stoic's view on the transcendence of cosmos within a human is profound. While, at first, they had adopted the Cynics' indifference against the goods of the external world, in fact the Sage's self-sufficiency (*αυτάρκεια*) becoming an indelible feature of their ethical doctrine, they quickly managed to moderate the Cynics' radical naturalism and to underscore the unity and self-efficiency (*αυτοτέλεια*) of the individual soul. Personality becomes a deterministic principle.

As Brad Inwood has written, "there was never a monolithic ideal of life for the Stoics."[4] This can also be understood by the set of moral rules, the so-called 'kathekonta,' which serve as a guideline – for those who are not yet wise – to behave appropriately and righteously under some typical circumstances. The caveat, however, is that these rules cannot be uniformly applied to any occasion. The difference of the ideal Sage is that the virtuous actions are not the result of following inflexible rules but are optimised by the particularities of the unique personality of the agent in action.

Marcus Aurelius' *Meditations* are written in a dialogic discourse, for private use as deliberations of a man with himself. The individual is called to reflect by himself in order to be free from perturbations. Similarly, Lucretius employs meditative elements in his work *De rerum natura*. The meditative exercises should result in "an attitude of mind."[5]

The virtuous character acquires a protagonistic role in moral decision-making processes, reflecting the increased interest in the individual. Richard Sorabji discusses the role of identity or

---

temology," in *The Cambridge History of Hellenistic Philosophy*, eds. Keimpe Algra, Jonathan Barnes, Jaap Mansfeld, and Malcolm Schofield (Cambridge: Cambridge University Press, 1999), 229.

[4] Brad Inwood, "Rules and Reasoning in Stoic Ethics," in *Topics in Stoic Philosophy*, ed. Katerina Ierodiakonou, 95-127 (New York: Oxford University Press, 1999), 126.

[5] Michael Erler, "Physics and Therapy. Meditative elements in Lucretius' *De rerum natura*." https://www.dwc.knaw.nl/DL/publications/PU00010720.pdf.

persona in stoic ethics.[6] Moral decisions should be based not only on – common to us all – human rationality, but also on the unique persona of the individual who is to make the decision. Each and every individual has formed a persona on the basis of the different nature compared to others. This difference is that which creates variability in moral decisions, explaining why a decision might be right for one but not for others, even under the same circumstances. Apart from the example of Cato's suicide, which, according to Cicero, is a rightful decision only in the case of Cato's persona, Sorabji makes reference to Epictetus, who also acknowledges the need of conforming decisions to the nature of the particular agents who carry them.

The Stoics sought after a physical explanation of the processes of sensing information, of transmitting it to a central ruling faculty and of the subsequent processing and, thus, experiencing. 'Hegemonikon' is the single entity that explains the governing of all physical faculties. The hegemonikon is the seat of reason, logos, i.e. God, the universe's controlling 'pneuma.' With the hegemonikon, the stoics create a model of consciousness in which a person can deliberate with his inner being.[7] For Epicureans, the seat of consciousness is the animus. Both hegemonikon and animus capture the notion of a monistic self that actively engages as a whole with all living experience and ascribes personalized mental representation to each experienced object giving the sense of privateness and introspection.

Hegemonikon is not just the faculty that transforms the stimuli of the different organs into sense-perceptions, but it also transforms emotions into volitionary manifestations. Hence, the true essence of this 'syn-eidenai' – whose work is to uniformly perceive and configure - is, per se, intelligence (nous).

The stoic Sage, even if not managing to avoid the emotional irritations originating from the outside world, will resist by

[6] Richard Sorabji, *Self. Ancient and Modern Insights about Individuality, Life and Death* (Chicago: The University of Chicago Press, 2006), 157-162.

[7] Vivienne Brown, "The Dialogic Experience of Conscience: Adam Smith and the Voices of Stoicism," *Eighteenth-Century Studies* 26, no. 2 (1992): 238.

means of reason and will not allow emotions to be converted into passions. The transcendence of the world is actually the transcendence of one's own impulses. Passions are false judgments and not a consequence of irrational origin. Virtue is accomplished by the development of personal reason and not by some kind of mastering of non-rational parts of the soul.

Both Epicurean and Stoic philosophers argued in favour of certain criteria of truth against which all opinions should be examined. For the Epicureans, the criteria of truth are sense-perceptions, passions (pathe), preconceptions (prolepseis) and the 'presentational (phantastikai) applications of the mind;'[8] for the Stoics, the criterion is the 'cognitive impression' (kataleptiki phantasia). But it is sensations (aistheseis) from which the path of information from the outside world originates. The stoic 'phantasia,' operating as a representational function, is the initiation mechanism of the cognitive functions.

The stoic phantasia is a state of consciousness, contrary to the raw data coming from the senses before reaching the hegemonikon. The transferring of raw data from the sense-organs to the ruling faculty is not something for which a person is aware of; however, the outcome of the interaction between the psyche and the physical objects, which according to the stoic physics, leaves a "stamping" or creates an alteration to the soul, is a state of consciousness. Again, for Epicureans, the phantasia (representation or impression) is any event of the senses which is inscribed in consciousness. It is created upon the reception of a sensory stimulus. Phantasia establishes a direct connection between the sensory organs and the objects of reality and is the mechanism by which perception and cognition are explained. However, phantasia cannot be thought of independently of the intellect (the reason), as it is inactive and not capable of constructing concepts. Logic intervenes, through the thin and kinetic atoms and aligns the irrational phantasia with the stored experience of memory, giving it meaning.

Underlying the Stoics' theory of passions or emotions is the conscious control that is involved in the process of evaluating

[8] Diogenes Laertius, *Vitae philosophorum*, X 31

and assessing by the hegemonikon. The soul's essence is manifested through the rational competence/capability to deny assent to impulses and this is an idiosyncratic rivalry admitted by the Stoics to exist in the psychic life. The impulses of the senses are excluded by the soul which is part of the cosmic Logos.

Keeping things at a distance and using our assent, we ensure a robust independence of our personality; this doesn't mean that a person can avoid the joy or pain that destiny reserves for him, but it means that he becomes independent of the trajectory of things keeping his self-sufficiency proudly intact.

Rationality, of course, plays the principal role as it makes sure that the individual character will practice virtue. Under this perspective, the one and only good is virtue and the one and only bad is the dominance of passions over rationality. All other things are, from an ethical point of view, indifferents. At this point, however, Stoics introduce a further evaluation of goods, albeit subordinate to the goodness of virtue. This secondary distinction is between preferred and dispreferred indifferents.

The mechanism of shaping the individual course of reason and consciousness by each human being is described by the theory of *oikeiosis*. Oikeiosis is a predominant concept in the Stoic philosophy, indispensable and unifying element of their moral psychology and ethics. It can be seen as an evolution of the Classical period's philosophical injunction "gnothi seauton" ("know thyself"), a practice calling for self-consciousness. For the Stoics, the concept of self-consciousness goes beyond the knowing of oneself to becoming aware of and relating to one's environment.[9]

It is a term enclosing multiple meanings making it rather difficult to translate. Its etymological root is the work *oikos*, meaning "house," including the persons who belong to the household; its cognate adjective is *oikeios*, referring to those who are members of one's household or to the objects that

[9] Ali Kashani, *Radical Generosity. Resisting Xenophobia, Considering Cosmopolitanism* (London: The Rowman & Littlefield Publishing Group, 2019), 13.

one possesses. Stoics often used the middle verb *oikeiousthai* to suggest making something familiar and making something one's own, with emphasis on the claim rather than the possession itself. Blundell has described oikeiosis as "the process by which we recognize our natural affinity first to ourselves and subsequently to various features of our environment, which we pursue as being oikeios or 'belonging to us."[10] A list of English translations include 'appropriation,' 'familiarization,' 'affinity,' 'well-disposedness towards,' 'attachment,' 'propensity.' Appropriation has been the one mostly used, although it does not capture the concept of personal affinity.[11]

Oikeiosis is not a static psychological state but entails a natural aspect of human evolution, an unending process of change. The starting point is the innate impulse of all animals, including humans, for self-preservation; the relationship here is the one between an animal and itself, specifically its constitution (systasis).[12] Gradually, through human maturation, we progress to a 'rational mode of existence,' which means that we move from valuing natural advantages to valuing reason in its own right and, thus, acting according to it. Sociability develops from self-affiliation, with one becoming aware that all human beings are members of the same human community. The personal and social aspects of oikeiosis go hand in hand, promoting the idea of a world citizen (cosmou politis).[13]

Hierocles visualizes the various relationships with self and others as a series of concentric circles around one's own intelligence, *dianoia*. Body and the associated material advantages form the innermost circle, practically in contact with the centre.

---

[10] Mary Whitlock Blundell, "Parental nature and Stoic *Οἰκείωσις*," *Ancient Philosophy* 10, no. 2 (1990): 221-242.

[11] Jacob Klein, "The Stoic Argument from Oikeiosis," *Oxford Studies in Ancient Philosophy* 50 (2016): 143-200.

[12] Wayne M. Martin, "Stoic Self-Consciousness. Self-Comprehension and Orientation in the Stoic Theory of Oikeiosis," September 28, 2006, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.405.2438&rep=rep1&type=pdf.

[13] Kashani, *Radical Generosity*, 13.

As extending outwards, each circle represents a declined level of relationship affinity, beginning with family and friends and reaching the whole mankind.[14] Hierocles actually visualized the phenomenon now known as proprioception. Human beings and animals as well, need a kind of familiarity with their perceptual organs for their effective usage.

The norms that outline the conditions enabling a person to work towards moral development are, essentially, the medical norms that define the psychological health in human beings. *Oikeiosis* is therefore the psychological component in the principle governing the "journey" of a person leading, ultimately, to virtue.[15] This perspective of *oikeiosis* becomes even more evident when acknowledging it as a form of self-consciousness; it provides an organism with the normative orientation in the environment.[16]

The doctrine of oikeiosis is closely related to the hegemonikon, in that it provides a description/characterisation of the psychic faculty that governs human action. In children (pre-rational humans), as in non-rational animals, oikeiosis drives self-preservation through self-perception. When the child reaches the age of 7 or 14, it is in the position to become capable to regulate impulses and guide actions by reason. In this case, self-perception is transformed into a sophisticated set of conceptions. This possession and employment of concepts, which is the distinction from the animals, is due to the governing faculty, the hegemonikon.

The sensations allow us to access knowledge that concerns specific observable facts, but to gain knowledge

---

[14] Brad Inwood, "Hierocles: Theory and Argument in the Second Century AD," *The Society for Ancient Greek Philosophy Newsletter* 115, no. 1 (1983): 115-136.

[15] Lawrence Becker, "Human Health and Stoic Moral Norms," *The Journal of Medicine and Philosophy* 28 (2003): 221-238.

[16] Wayne M. Martin, "Stoic Self-Consciousness. Self-Comprehension and Orientation in the Stoic Theory of *Oikeiosis*," September 28, 2006, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.405.2438&rep=rep1&type=pdf.

that leads to wisdom requires more and, at this point, both philosophical schools point out preconceptions (prolepseis). The development of reason is made possible because of the existence of preconceptions, which may be innate, i.e. residing in the human cognitive architecture, but they arise naturally and are not active before the impression of sensations takes place. Preconceptions are constituents of reason whose function is to interpret what the senses perceive.

The theory of cognitive development that takes place in phases, depending on the human's age, is nuanced as in the sophisticated work of Jean Piaget, who established Developmental Psychology. In this theory, preconceptions are rudimentary forms of understanding reality which are replaced by increasingly reasoned and structured concepts.

So, the teleological success or failure depends entirely on the integrity and sophistication of the hegemonikon, which, in turn, through the mechanisms of assent, governs human actions. A failure to conform to nature stems from a defect in the coordinator faculty.[17] However, hegemonikon is not metaphysically or causally independent of the whole body.

Lombardini distinguished between objective and subjective oikeiosis to discriminate variations in what can be considered valuable for a human individual; although virtue is the objective terminal of human moral development and, thus, it could be postulated – in concept – that each and every individual would count as valuable that which accords with nature, in reality, the human beings consider valuable what they perceive as appropriate and advantageous for themselves. The multitude of different subjective standpoints of valuing what is appropriate explains the spectrum of "points of view." There exist as many sets of beliefs as human beings, and what distinguish them are the individual criteria of value.[18]

---

[17] Jakob Klein, "Nature and Reason in Stoic Ethics" (PhD Diss., Cornell University, 2010), https://ecommons.cornell.edu/bitstream/handle/1813/17104/Klein,%20Jacob.pdf;sequence=1.

[18] John Lombardini, "Stoicism and the Virtue of Toleration," *History of*

Today, consciousness is acknowledged as retaining a personal character throughout the lifespan of an individual and is an essential background for the advanced intellect. The functions of the mind, as well as the perpetual development are based on constant conscious self-evidence. The brain creates an internal representation of reality which is constantly faced with stimuli from the external reality, available through the senses. The input is intertwined with the current brain activity and complex spatial and temporal re-arrangements take place. Hence, a dynamic system is in place for which the relationship between function and structure is by itself inadequate to understand the brain and its evolvement through life.[19]

Consciousness is also closely associated with neuroplasticity, interconnected in a 2-way relationship, promoting the ever-learning mechanism of the brain. The complex networks in the human brain are subjected to continuous re-organisation following changes in consciousness, which are caused by either input from the senses or by changes in the internal states of mind.[20]

The dominant feature of the stoic and epicurean theories about the ethical consciousness is that they entail an internal realization of the person himself and not a kind of arrangement with the external environment. The success in avoiding ethical deviations and distinguish between good and evil is not a result of metaphysical processes but is a mental work of the intellect. In this frame, human consciousness is of significant value.

Political Thought 36, no. 4 (2015): 664-665.

[19] Egidio D' Angelo, and Claudia Gandini Wheeler-Kingshott, "Modelling the Brain: Elementary Components to Explain Ensemble Functions," *Rivista del Nuovo Cimento* 40, no. 7 (2017): 304.

[20] Jean Askenasy, and Joseph Lehmann. "Consciousness, Brain, Neuroplasticity." *Frontiers in Psychology* vol. 4, no. 412 (2013): 1-10.

References

D' Angelo, Egidio, and Claudia Gandini Wheeler-Kingshott. "Modelling the Brain: Elementary Components to Explain Ensemble Functions." *Rivista del Nuovo Cimento* 40, no. 7 (2017): 297-332.

Askenasy, Jean, and Joseph Lehmann. "Consciousness, Brain, Neuroplasticity." *Frontiers in Psychology* 4, no. 412 (2013): 1-10.

Becker, Lawrence. "Human Health and Stoic Moral Norms." *The Journal of Medicine and Philosophy* 28, no. 2 (2003): 221-238.

Blundell, Mary Whitlock. "Parental Nature and Stoic O□ κείωσις." *Ancient Philosophy* 10, no. 2 (1990): 221-242.

Brown, Vivienne. "The Dialogic Experience of Conscience: Adam Smith and the Voices of Stoicism." *Eighteenth-Century Studies* 26, no. 2 (1992): 233-260.

Brunschwig, Jacques. "Introduction: The Beginnings of Hellenistc Epistemology." In *The Cambridge History of Hellenistic Philosophy*, edited by Keimpe Algra, Jonathan Barnes, Jaap Mansfeld, and Malcolm Schofield. Cambridge: Cambridge University Press, 1999.

Inwood, Brad. "Hierocles: Teory and Argument in the Second Century AD." *The Society for Ancient Greek Philosophy Newsletter* 115, no. 1 (1983): 115-136.

Inwood, Brad. "Rules and Reasoning in Stoic Ethics." In *Topics in Stoic Philosophy*, edited by Katerina Ierodiakonou, 95-127. New York: Oxford University Press, 1999.

Kashani, Ali. *Radical Generosity. Resisting Xenophobia, Considering Cosmopolitanism*. London: Rowman & Littlefield, 2019.

Klein, Jakob. "Nature and Reason in Stoic ethics." PhD Diss., Cornell University, 2010. https://ecommons.cornell.edu/bitstream/handle/1813/17104/Klein,%20Jacob.pdf;sequence=1.

Klein, Jacob. "The Stoic Argument from Oikeioōsis." In *Oxford Studies in Ancient Philosophy* 50, edited by Victor Caston, 143-200. Oxford: Oxford University Press, 2016.

Lombardini, John. "Stoicism and the Virtue of Toleration." *History of Political Thought* 36, no. 4 (2015): 643-669.

Martin, M. Wayne. "Stoic Self-Consciousness. Self-Comprehension and Orientation in the Stoic Theory of *Oikeiosis*." http://citeseerx.ist.psu.edu/viewdoc/download?-doi=10.1.1.405.2438&rep=rep1&type=pdf.

Sorabji, Richard. *Self. Ancient and Modern Insights about Individuality, Life and Death*. Chicago: The University of Chicago Press, 2006.

Thiel, Udo. *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. New York: Oxford University Press, 2011.

Weber, Michel, and Anderson Weekes. *Process Approaches to Consciousness in Psychology, Neuroscience, and Philosophy of Mind*. Albany, NY: State University of New York Press, 2009.

Personification: A Category Mistake or a Categorial Novum?

Damir Smiljanić
University of Novi Sad
E-mail address: dr_smiljanic@ff.uns.ac.rs
ORCID iD: https://orcid.org/0000-0002-5791-620X

Abstract: In this paper the author considers the problem of whether personification, the process of treating something without personal characteristics as if it were a human person, is based on the misuse of the category of personhood. A very serious historical form of this kind of categorial gerrymandering is so-called anthropomorphism. In order to shed light on the mentioned problem the author thematizes personification in three ways: personification as a categorial mistake, as a categorial transfer and as a hidden vivification. For this purpose, the following philosophical theories will be helpful: Ryle's logical analysis, Hartmann's new ontology and Klages' metaphysics of life. At the end of the paper the author pleads for an integrative approach in the philosophical theory of personhood.
Keywords: personification; anthropomorphism; categorial mistake; categorial transfer; vivification; Gilbert Ryle; Nicolai Hartmann; Ludwig Klages.

I. Introduction

Both within and without the philosophical context, the word 'person' is surrounded by a nimbus of sanctity which was a guarantee for its frequent usage in metaphysical and ethical theories (especially in deontological ones). Furthermore, it can be designated as a guiding concept of philosophical anthropology. This is nothing extraordinary because the human being defines themself[1] as a being which possesses a Self, which means that it is able to

[1] In order to avoid the difference between male and female individuals I will use the pronoun 'themself' when it is necessary to refer to self-reflective activities of a human being.

doi: https://doi.org/10.12681/aprlp.49.924

attribute the capability of self-consciousness to themself. The human being (the person) knows that they are the subject of their thoughts, wishes and feelings but also the subject of their actions and the initiator of interactions with other human beings, i. e. persons. This state of affairs has a theoretical and a practical consequence. Both are relevant for a philosophical investigation of the phenomenon of *personhood*. But there is also an aspect of this phenomenon that could be of systematic importance for the philosophical investigation – the so-called *personification*, the semantic (maybe also pragmatic) process of treating something that does not possess a Self as personal or quasi-personal, literally *making it personal*.

II. On anthropomorphism

From a historical point of view, the procedure of personification can be seen in a similar kind of projection of human characteristics onto something non-human or super-human – the so-called *anthropomorphism*. In ancient myths and legends gods are presented in a humanlike form: they look like humans, have similar thoughts, they intend and realize their actions in the same way as humans do, show feelings and passions, enjoy pleasures, deal with disappointments and express their emotions in different types of situations. This problematic attribution of human characteristics to divine beings induced some thinkers to scathing criticism. Xenophanes noted that if horses and lions had hands they also would create pictures and statues of gods in the likeness of the form they themselves possess. This critique of anthropomorphism anticipates the very ideas of the later formulated criticism of religion represented in the works of such thinkers as Ludwig Feuerbach and Karl Marx.

In the Christian tradition the ancient and pagan polytheism is refuted primarily because of the anthropomorphic personification of divine beings. Although the Christian believer tries to keep their faith aloof from anthropomorphism, they take the risk of regression to this alienated form of consciousness by using nouns and verbs that describe human

mental states and ways of planning and realizing actions when they want to render God's ideas and intentions. In the age of *Enlightenment* the situation has been turned upside down – God is now disempowered, and (enlightened) man is deified. This is a different way of understanding "the dialectics of Enlightenment."

But this is not the end of the story – by which I mean the history of anthropomorphism and its critical evaluation. These days, anthropomorphism is unmasked as anthropocentrism, a sort of *speciesism*. In present bioethical debates the application of human standards of value for the purpose of specifying the moral status of non-human beings is criticized as tendentious and injudicious so that it must be refused or corrected. It seems that man no longer wants to show themself godlike and – instead of this – declares solidarity with animals by making them humanlike. Not only (the imagined) gods have personal traits – now the animals also have the right to be treated as persons.

III. Personification as a categorial problem

After these preliminary considerations of anthropomorphism we must immerse ourselves in the problematic matter. For this purpose, we can choose the standpoint of *categorial analysis* conceived as a method of classifying phenomena under concepts not only in ontological theories but also in theories which find their point of departure in the linguistic turn. First, we must emphasize that it does not make sense to describe the person themself as a category. It is more correct to designate *personhood* as a category. A man or woman can be subordinated under the category of personhood, and then we state that he or she is a person.

The main characteristic of a person is their *self-consciousness*. In a discipline such as bioethics this way of determining personhood is refuted by the argument that there are some humans like babies, people who suffer from Alzheimer's disease or vegetate in a state of coma – people who are still to be treated

as persons not only because of the social environment in which they are embedded but also due to other characteristics which are further intrinsic components of personhood (attributes like sensibility to pain or responsiveness to stimuli from their environment). Even if the characterization of a person as a self-conscious subject may be grasped as one-sided, it has the advantage of pointing out a seeming difference between humans and other beings. (We now leave aside the crucial problem whether it is an essential account of distinguishing beings or just a speciesistically motivated projection of arbitrary traits of human beings.)

Returning to the problem of anthropomorphism, i. e. the way non-human beings are described by means of human characteristics, we can notice that these beings are outfitted with some abilities which imply self-consciousness (for example: thinking, planning and realizing actions, communication with other beings etc.). When Zeus, the mightiest god in ancient Greece, is enraged by sacrilegious or reckless practices of the mortals, he *considers* which kind of punishment could be necessary to rebuke the contumacious race; finally, he *chooses* the adequate punishing procedure and *instructs* other gods or human executioners to put his will into practice. The mentioned (mental) activities (consideration, deliberation, issuing commands and so on) are undertaken on the fundament of self-consciousness of a personal or, at least, person-like being. It is the privilege of persons to be aware of their thoughts, intentions, wishes and projects of action. If the ancient immortals are persons too then they possess the same capabilities – and, vice versa, the possession of these abilities qualifies them for being deemed as persons.

Self-consciousness is a categorial moment within the human condition – one of its most important. It is a constitutive personality trait. The traditional philosophy has stylized it to the *conditio sine qua non* of human personhood: from the ancient definition of man as *animal rationale* and the Christian doctrines of man as the image of God to the theories of action and communication in contemporary thought, philosophers hang

on to the figure of person as the owner of self-consciousness. It is the guarantee for attributing dignity to human persons. So far, so good! But what happens in the case where these characteristics of human persons are transferred to other beings? In other words, what are the theoretical implications of the categorial transfer of personhood outside the sphere of humanity? Is personification a correct way of categorial transfer?

There are at least two possibilities to answer this question. One answer is skeptical, the other rather affirmative. One can come to these answers from different points of view: either from a logical perspective or from an ontological one. The first answer is based on the philosophical critique of language, the second results from a realistic approach in ontology based on the idea of levels of reality. It is time to raise the question of whether personification represents rather a *categorial novum* than a case of *category mistake*. This is the crucial problem I want to discuss in my paper.

a. Personification as a categorial mistake (Ryle)

In the tradition of nominalism we can find breadcrumbs of language criticism regarding the usage of general concepts. This trend is continued with John Locke's description of the abuse of words and later with the neopositivistic refutation of metaphysical language. Finally, this kind of philosophizing has found its adequate manifestation in critical analysis of language represented in *analytic philosophy* which dominates today, not only in the Anglo-Saxon context. In his work *The Concept of Mind* (1949), one of the most renowned proponents of analytic philosophy, Gilbert Ryle, introduced *category mistake* as a specific topic. This kind of fundamental mistake "represents the facts of mental life as if they belonged to one logical type or category (or range of types or categories), when they actually belong to another."[2] Having in mind the procedure of Cartesian

[2] Gilbert Ryle, *The Concept of Mind* (London: Hutchinson & Co., 1966), 16. Cf. also Gilbert Ryle, "Categories," in Gilbert Ryle, *Collected Papers*.

metaphysics to hypostatize two different sorts of substance – *res extensa* and *res cogitans* – Ryle wants to show that this dualism is based on a category mistake by treating mind as an autonomous substance instead of reconstructing it just as a collective name for a set of activities which can be registered by the help of our senses and which, considered in aggregate, constitute a behavior of a certain person (including the dispositions which cannot be perceived by our senses). The so-called "Dogma of the Ghost in the Machine" is one of the most prominent examples in philosophy for making a category mistake. According to the "grey eminence" of philosophy in Oxford the key for clarifying this conceptual confusion can be found in the means of logical analysis of language: "The exhibition of these absurdities will have the constructive effect of bringing out part of the correct logic of mental-conduct concepts."[3]

What does it mean when someone is making a category mistake when talking about persons? Obviously, we can recognize this case when someone uses concepts from one categorial sphere to describe beings, events and processes which belong to a different categorial sphere. For example, if one says that their dog *decided* to take one path in a situation where many paths are passable, they actually misinterpret the behavior of this animal by using categories which are applicable only to human behavior, which means that they lose sight of the fact that the dog's behavior is in particular directed by its olfactory sense and presumably not by even a rudimental rational way of decision-making. The problems compound when we try to use categories of human behavior for the purpose of describing mental states and activities of superhuman beings. It could also be identified as a case of category mistake when it is said that the ancient god Zeus, irritated by the impudence of Prometheus who stole the fire and gave it to the humans, *made the decision* to punish

Volume 2. Collected Essays 1929-1968 (London, New York: Routledge, 2009), 178–193 (reprinted from *Proceedings of the Aristotelian Society*, vol. XXXVIII, 1938).

[3] Ibid., p. 23.

the enchained titan by sending an eagle to feed on his liver. Why would this description of state of affairs be interpreted as a category mistake? Because it is disputable whether terms which describe human characteristics could be used without a hitch to reconstruct the behavior of superhuman beings like gods, demons or fairies, i. e. creatures about whose nature and potential sphere of action we do not have any empirically verifiable knowledge. Things get even more complicated when we give consideration to the problematic ontological status of these beings – after all, we do not know if they even exist. In other words, it is the typical case of a so-called "*fallacy of misplaced concreteness*"[4] when we deal with fictive subjects or objects as if they were real and accessible to our experience, in other words, when something extremely abstract is taken into consideration as something very concrete. From the logical point of view, it is very arguable to treat nonhumans and superhumans like autonomous persons. This kind of personification inevitably must fail.

b. Personification as a categorial transfer (Hartmann)

How could personification be scrutinized from an ontological standpoint? In this context I want to refer to Nicolai Hartmann's project of "New Ontology" primarily because I think this theory could be useful when the problem of personification is considered in a categorial manner. Hartmann took the view of *critical realism* and presented a *pluralistic* approach in ontology. The idea of being is not unitary here – it is composed of several *layers* or *strata*: the inorganic (inanimate), the organic (biological), the psychical and the spiritual stratum.[5] The fundamental strata

[4] This is a term introduced in the philosophical discussion by Whitehead. Cf. Alfred North Whitehead, *Science and the Modern World* (Cambridge: Cambridge University Press, 1929), 49-70 (Chapter III: The Century of Genius). Cf. also Alfred North Whitehead, *Process and Reality. An Essay in Cosmology* (New York: The Free Press, 1978), 7-8.

[5] Cf. Nicolai Hartmann, *New Ways of Ontology*, translated by Reinhard C. Kuhn (Westport, Connecticut: Greenwood Press, 1975), 43-53 (Chapter V: *The Stratified Structure of the World*).

of the world are determined by specific laws and regularities.[6] One of them is the law of *categorial novelty* which postulates that there is a specific novelty in the higher stratum which cannot be reconstructed simply as a sum of the categorial moments which belong to the lower ones. It is something new and not reducible to already given components; because of this it is called a 'categorial novum."[7] For example, the self-organized metabolism is the novum which appears in the organic stratum and cannot be found in the inorganic one. The crucial question is whether this ontological *ansatz* could be fruitful for the categorial analysis of the phenomenon of personification.

What kind of criteria must be regarded if a being should be denoted as a person? According to Hartmann's theory of strata (*Schichtenlehre*) it must be enfolded by categories of the highest stratum – the *spiritual* stratum. It is the region of three forms of spirit: the *personal*, the *objective* and the *objectivated* spirit. The most important categorial moments of personal spirit are consciousness, will, foresight, teleological activity and liberty. A person is aware of their thoughts, wishes and feelings, they can plan their activities and realize intended purposes, finally – what is especially important from an ethical viewpoint – they are disposed to the idea of liberty. The human being[8] is participating in the spiritual stratum: as an individual they are implied as having a personal spirit, as a member of a community they are among other individuals in the medium of objective

[6] Cf. op. cit., 73-83 (Chapter VIII: *The Strata Laws of the Real World*).

[7] "The recurrence of lower categories never determines the character of the higher stratum. This character always rests on the emergence of a categorial novelty which is independent of the recurrent categories and consists in the appearance of new categories. The modification of the recurring elements is contingent upon the emergence of novelty." Ibid., 76.

[8] It must be mentioned that Hartmann advocates an *integrative* view of the human nature: "The nature of man can be adequately understood only as the integrated whole of combining strata and, furthermore, as placed within the totality of the same order of strata which, outside of man, determines the structure of the real world." Ibid., 121-122.

spirit and they produce artefacts such as books, paintings, sculptures etc. which are manifestations of the objectivated spirit (mind). When we want to describe a human person, we can use extensively most of these categories without making a categorial mistake. Problems begin to arise if we make a step underneath (or even above) the spiritual stratum and use the categories of personal spirit to grasp quasi-personal abilities and activities of non-human beings.

How could personification be reconstructed from the perspective of Hartmann's new ontology? I think that this could be done in two different ways. The first way is similar to Ryle's linguistic criticism – personification can be refused as a form of illegitimate category transfer. For example, if one says that their dog decided to choose one path and not the other, they use the category of decision-making which is an activity of human will and thereby transfer a category from the spiritual layer to the lower one where psychic phenomena are considered. Even more problematic is the transfer of a category from a lower to a higher stratum although Hartmann assumes different possibilities of transformation of "lower" categories at a higher level (for this purpose he is distinguishing between *superinformation* [*Überformung*] and *superimposition* [*Überbauung*][9]). In a fictional (mythological) context the proposition 'Zeus wants to punish Prometheus' can be understood without problems. But does this proposition have a sense in our real world? Does a godlike Zeus really exist? The spiritual stratum is the highest one in Hartmann's stratified view of the real world. There is no layer above the spiritual one. Even though ancient and other divine beings are part of the so-called objectivated spirit because of their appearance as characters in myths, from Hartmann's neo-ontological point of view nothing can be said about the ontological status of these beings. A major problem considered within this philosophical approach is the difficulty of representing ontic forms without grounding them in the lower layers of reality. In other words, Hartmann

[9] Cf. Nicolai Hartmann, *New Ways of Ontology*, 78-79.

excludes the possibility of 'levitated' layers which would be lacking any categorial contact to other levels. It is notable that Nicolai Hartmann has never written a book about philosophy of religion.

It seems that personification must be unmasked as a form of category mistake also from the perspective of Hartmann's critical realism.[10] But there is a second possibility to understand this phenomenon without refusing it as senseless. This interpretation is dealing with the very concept of person. Namely, if we understand this concept as a stratified structure, we can then recognize some of its categorial moments in the lower layers of reality as well. Maybe it is problematic to ascribe personal characteristics inherent to humans also to animals, but if we recognize some abilities like sensibility or pain susceptibility as belonging to the scope of personhood then we can treat all beings which, for example, feel pain as persons. Pain sensitivity is something that humans share with animals due to the fact that it is a category which connects the organic and the psychical stratum. It is no wonder that this insight is often used in contemporary bioethical debates to argue the case for *animal rights*. The main argument is that animals should be treated like persons because they can feel pain and someone who tries to harm or even to kill them transgresses their right to be treated with respect. Anyway, for someone who is defending the concept of animal rights on the basis of *pathocentrism*, it could be possible to take advantage of some of Hartmann's ontological ideas in order to strengthen their position.

---

[10] Hartmann listed many categorial errors in his article "How Is Critical Ontology Possible? Toward the Foundation of the General Theory of the Categories, Part One," translated by Keith R. Peterson, *Axiomathes* 22 (2012): 315-354. An extended version of his critique of ontological "prejudices" can be found in Nicolai Hartmann, *Der Aufbau der realen Welt*. Grundriß der allgemeinen Kategorienlehre (Ontologie Band 3), 3. Auflage (Berlin: De Gruyter, 1964), 61-156.

c. Personification and vivification (Klages)

Finally, I want to sketch a further sense of personification that might provoke a degree of odium among certain people who are subscribing to a strong realistic view of the world. This could be understood as a wider sense of the term 'personification.' In a stricter sense, personification refers to all kinds of living beings, no matter whether they are mortal (humans or animals) or immortal (gods, angels, demons and the like). But having in mind our emotional responsiveness triggered by external impact we tend to phrase that sunshine *pleases* us, rain is *boring* us by its monotonous sound, boredom is *killing* us, some painted figures seem like they're amicably *smiling* at us etc. It seems that in this case some personal qualities, i. e. qualities like pleasantness, boringness, deadliness, happiness etc. have been ascribed to non-living entities such as weather conditions, emotional states or drawings. If the first case of personifying fictive beings like demons or fairies is already confusing, what should be said about the mentioned case? Hence the confusion reaches its climax.

Personification of non-living beings or events by ascribing humanlike qualities to them can rightly be refused from a logical point of view and considered as a notable category mistake. But there is a way to recognize one positive effect of such personification. It can namely be treated as a vehicle for renewing our linguistic means in order to describe the impressions that similar beings or events induce in minds of humans. This is not simply the case of the so-called "transfer of meaning." Maybe it is rather the opposite case that this kind of personifying things and processes makes the forming of metaphors or other "transferred" terms possible at all.

An impassionate proponent of this interpretation of personification was the German philosopher Ludwig Klages, best known for the radical criticism of occidental rationalism presented in his opus magnum *The Spirit as Adversary of the Soul* (1929□ 1932). As the title of his work suggests, he transvalued the role of soul in contrast to the overvaluation of the spirit.

The reality of the world is given in images[11] and because of the ability of humans to perceive very lively impressions on account of the influence of these images on their soul, the world view is permeated with interpretable expressions. Therefore, every movement or change within perceived things, events or situations has its individual signature which triggers a specific reaction in the observer who is interpreting it as a quasi-personal trait. For example, if one says: "The wind gently tousled Mary's hair," then gentleness (tenderness) is a quality which is being ascribed to human persons but it seems that the wind (a natural process) also can generate similar effects as a hand movement. But in our everyday opinion, characteristics of human activities (i. e. activities of human persons) can also be paraphrased as symptoms of non-human agents. Taking into consideration that there is a subthreshold nexus between occidental rationalism and personification, Klages prefers the retreat – in terms of Hartmann's ontology: he tries to step back from the higher stratum to the lower one.[12] These two tendencies – personification and vivification[13] – both could be dismissed as a sort of metaphysics or mysticism.[14] But they stand for the potential of our language to "depict" reality as

[11] According to Klages the reality is *ipso facto* reality of images. Cf. Ludwig Klages, *Der Geist als Widersacher der Seele*, 6., ungekürzte Auflage (Bonn: Bouvier Verlag Herbert Grundmann, 1981), 801-1248 [*Die Wirklichkeit der Bilder*].

[12] Cf. the comparison with *botanic* symbols and metaphors in Ludwig Klages*, Der Geist als Widersacher der Seele*, 1308. Klages emphasizes the reverse perspective, namely "that here rather the human person is taken back into the more elemental sphere of the vegetative than the tree [this is his foil for comparison, D. S.] personified" (cf. op. cit., 1310 [translated by the author, D. S.]).

[13] Klages describes the essence of personhood as *dichotomous*: the person hovers between the poles of *spirit* and *life*, the two conflicting forces in his metaphysical system. Cf. op. cit., 61-76 ("Die Zwiespältigkeit der Person").

[14] They also could be related to the magical worldview which dominates in elemental communities (this relation is the research object of the social and cultural anthropology).

something that is affecting us and not as something separated from our personal experience. The world is touching us – and we respond to its effects by reflecting this in our allusive language. Personification and vivification have more in common than it might seem at first sight.

IV. Conclusion

In the end, I want to summarize the results of my consideration of personification in a few statements. (1) Personification is based on a category mistake if some higher-order categories are transferred to a lower layer of reality and then used to characterize abilities and activities of non-human beings. (2) Personification is also a sort of category mistake if it consists in the use of categories whose purpose is to specify the behavior of entities the being of which is transcending the well-known layers of reality (at least the four above-mentioned strata presented in Hartmann's ontology). This kind of personification is more a projection of human attributes onto something unreal than a description of entities in the real world. (3) Personification can furthermore be understood as a manifestation of categorial novum if the categories, which serve to picture the character and behavior of human persons, can be explained as a kind of "superformation" – or at least "superposition" – of categories already existing in the lower strata. It also needs to be pointed out that categories of the lower stratum *per se* can be sufficient for qualifying a non-human being as a person. (4) It is also possible to personify entities and processes in the non-living (physical) world if it is shown that the quality of impressions elicited in other subjects by these entities is nearly the same as the quality of impressions which other (human) persons evoke in ourselves. In like manner there is the possibility to recognize qualities of "lower" forms of life in human characteristics, actions and behavior patterns on condition that the impressions of former and latter mentioned processes resp. beings are look-alikes. The category descent or ascent could be justified

by means of the theory of *modifying predicates*.[15] (5) Finally, the essence of personhood should not be explicated by means of just one distinction, no matter how important it is (for instance self-consciousness). The very state of personhood is *stratified* so that an additional effort is needed to describe it adequately. Like in other cases, an *integrative* approach is more fruitful than a narrow-minded reductionism.

References

Hartmann, Nicolai. *Der Aufbau der realen Welt*. Grundriß der allgemeinen Kategorienlehre (Ontologie Band 3). 3. Auflage. Berlin: De Gruyter, 1964.

Hartmann, Nicolai. *New Ways of Ontology*. Translated by Reinhard C. Kuhn. Westport, Connecticut: Greenwood Press, 1975.

Hartmann, Nicolai. "How Is Critical Ontology Possible? Toward the Foundation of the General Theory of the Categories, Part One." Translated by Keith R. Peterson. *Axiomathes* 22 (2012): 315-354.

Klages, Ludwig. *Der Geist als Widersacher der Seele*. 6., ungekürzte Auflage. Bonn: Bouvier Verlag Herbert Grundmann, 1981.

König, Josef. *Sein und Denken*. Studien im Grenzgebiet von Logik, Ontologie und Sprachphilosophie 2., unveränderte Auflage. Tübingen: Max Niemeyer Verlag, 1969.

Ryle, Gilbert. *The Concept of Mind*. London: Hutchinson & Co., 1966.

Ryle, Gilbert. *Collected Papers*. Volume 2. Collected Essays 1929-1968. London, New York: Routledge, 2009.

[15] This kind of logical theory is presented in Josef König's work *Being and Thinking* (1937). Modifying predicates articulate the so-called *certain impressions* which are equivalent to *impressions of something certain*. Cf. Josef König, *Sein und Denken*. Studien im Grenzgebiet von Logik, Ontologie und Sprachphilosophie, zweite, unveränderte Auflage (Tübingen: Max Niemeyer Verlag, 1969) (1st Edition: 1937), especially 1-80.

Whitehead, Alfred North. *Science and the Modern World*. Cambridge: Cambridge University Press, 1929.

Whitehead, Alfred North. *Process and Reality*. An Essay in Cosmology. New York: The Free Press, 1978.

Licensed to Kill: Autonomous Weapons as Persons and Moral Agents

Alkis Gounaris
National and Kapodistrian University of Athens
E-mail address: alkisg@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0002-0494-6413

George Kosteletos
National and Kapodistrian University of Athens
E-mail address: gkosteletos@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0001-6797-8415

Abstract: The debate over the attribution of personhood to non-human entities is of an increasing concern to both academia and institutions. The intelligence, autonomy and efficiency exhibited by modern AI systems raise pressing questions regarding the moral responsibility issues their use entails. In our paper we focus our discussion on autonomous war machines, as their actions touch upon issues of life and death and their design, production and use cause philosophical controversies. Prompted by the classic position of Daniel Dennett defending the possibility that autonomous intelligent systems are responsible for their actions, we consider a) the argument of cognitive and / or functional equivalence of humans and machines, b) the argument of autonomy as such and c) the argument of excessive efficiency of the actions of intelligent machines. Our investigation upholds a skeptical stance towards the issue of recognition of moral personhood, while illuminating aspects such as the difference between cognition and intelligence, the necessary and sufficient conditions that imply moral responsibility and the differences between the ontological and the epistemological examination of the above problems. Finally, the contradiction between the demand for clear and solid decision-making criteria and the endless nature of a philosophical analysis that strives to be consistent is emphasized.

Keywords: artificial intelligence; autonomy; moral personhood; autonomous weapons; human - machine intelligence equivalence; AI excessive effectiveness.

doi: https://doi.org/10.12681/aprlp.49.925

In October 2017, the humanoid Sophia became the first artificial intelligence entity to become a citizen of Saudi Arabia.[1] Two years before, the European Parliament Committee on Legal Affairs had suggested the need to establish a legal framework for the recognition of the civil rights and obligations of intelligent "electronic persons" who make autonomous decisions.[2] This framework is still not outlined at the moment, as conflicting views are expressed on the subject, particularly regarding the issues of liability and moral responsibility resulting from the autonomous operations of intelligent systems.[3]

The issue of moral responsibility in AI systems concerns today, as we will see later, both the philosophical and the research community and is closely related to the concept of person. But can we "literally" attribute the term personhood[4] to artificially

[1] Chris Weller, "Meet the First-ever Robot Citizen - A Humanoid Named Sophia that Once Said It Would 'Destroy Humans,'" *Business Insider*, accessed July 30, 2020, https://www.businessinsider.com/meet-the-first-robot-citizen-sophia-animatronic-humanoid-2017-10?r=UK.

[2] About electronic persons see https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf.

[3] "For the purposes of liability, it is not necessary to give autonomous systems a legal personality." Further reading at https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608.

[4] It should be noted that the recognition of legal personhood in objects, animals, plants or artificial intelligence systems is an issue that has long preoccupied lawyers and philosophers, see Lawrence Solum, "Legal Personhood for Artificial Intelligences," *North Carolina Law Review* 70, no. 4 (1992): 1231-1287. Specifically, the attribution of the status of a person is discussed, corresponding to the attribution of the status of a legal entity to non-natural entities such as companies, institutions, municipalities, government agencies, etc. that carry out operations, contract, have rights and obligations, responsibilities and demands. At present, however, we will only be concerned with the attribution of moral status to intelligent machines, as such a perspective is directly related to personhood recognition in AI systems.

intelligent systems that behave like real people? According to the philosophical defense of the fictional character HAL 9000 by Daniel Dennett, provided these systems make intelligent and autonomous decisions and take effective actions, these actions can be evaluated "morally" just like the corresponding human ones.[5] That is, if a system thinks, acts and behaves like (or even better than) a human, it will be able to bear moral responsibility for its actions and be considered a moral person.

In this paper we will mainly consider the problem of the moral responsibility of machines, which leads to a number of issues concerning the moral person concept. We begin from the assumption that if someone or something can be characterized as a moral entity, then he/it can very well be considered as having the status of a person in general, while the opposite does not necessarily happen.

We have chosen to focus our discussion on 'autonomous' military machines, namely machines that purportedly decide autonomously on matters of life and death, due to the great urgency of the ethical issues caused by their design, construction and use, but we consider that our arguments can as well be valid regarding any other AI system.

Focusing our investigation on intelligent 'autonomous' military machines, we are faced with questions such as whether – and under what conditions – should intelligent systems make 'autonomous' life and death decisions. Also, whether intelligence, autonomy and efficiency are necessary as well as sufficient conditions for an agent to be considered a moral being. And if so, then should these war machines, in addition to being responsible for their actions, take up military positions and join the military hierarchy not as weapons but as soldiers? Would this possibly mean that they should enjoy the benefits of the Geneva Conventions regarding prisoners of war if arrested, or that they should be held accountable in military courts for their actions and omissions or insubordination?

---

[5] Daniel Dennett, "When Hal Kills, Who's to Blame? Computer Ethics," in *Hal's Legacy: 2001's Computer as Dream and Reality*, ed. David G. Stork, 351-365 (Cambridge, MA: MIT Press, 1997).

Today, the involvement of artificial intelligence systems in government, military, space and other operations is no longer the fictional content of films such as *2001, A Space Odyssey*. 'Autonomous' war machines, as well as other systems, already operate for defensive or offensive purposes and are tested in real war situations. States are already in an armaments race in order to gain a competitive edge, and the defense industry is paving the way in this research direction. This is one of the main reasons why we focus on AI war systems, considering whether the concept of moral person – and consequently of person – can be attributed to AI systems, as the severity of the consequences of their actions is proportional to the severity of the moral questions which the latter one raises.

The pressing context in which philosophers and AI researchers are called upon to deal with these new problems and the ethical issues that arise, is revealing. In 2015, Stuart Russell, Max Tegmark and other AI and Robotics Researchers, published an open letter, requesting a ban on the development of autonomous weapon systems and killer robots.[6] Among other things, they report that 'autonomous' weapon systems are today the third revolution in military operations (after gunpowder and nuclear weapons) and due to their relatively low cost and ease of manufacture, they are expected to be widely distributed and mass-produced, with the risk of being used for terrorist acts, ethnic cleansing, assassinations, destabilization of nations, enslavement of populations and selective extermination of national or social groups. For this reason, they call on AI researchers to refuse to participate in the research and construction of such weapon systems, the same way that biologists, chemists and physicists, respectively, widely support similar international agreements to ban chemical and biological or laser-equipped weapons.

---

[6] Stuart Russell, Max Tegmark, et al., "Autonomous Weapons: An Open Letter From AI & Robotics Researchers," *Future of Life Institute*, accessed July 25, 2020, https://futureoflife.org/open-letter-autonomous-weapons/.

In the opposite direction, the endorsers of these systems argue that war machines will only pursue their target in a legal and accurate manner and will strictly follow the provisions of the International Conventions for the wounded, civilians, prisoners of war etc., while in contrast to human soldiers they will never be under psychological pressure, they will not make mistakes due to fatigue and they will not commit revenge atrocities (as is often the case with soldiers, who may prove to be mentally and emotionally vulnerable). Therefore, intelligent machines can become in the future the ideal model of the moral soldier, as they will respect opponents, civilians, infrastructures etc.[7]

It is understood that the discussion around ethical problems raised by the design, production and use of autonomous weapon systems is related to:

> 1) Whether or not there should be such systems – a problem related to (1.1) their expediency and their possibly malicious use and (1.2) their ontological status, as formulated by their autonomy, intelligence and effectiveness, (1.2.1.) as well as whether their action is morally evaluable and (1.2.2) the systems themselves are, possibly morally responsible, and
> 2) How are we to determine if the system ultimately acted autonomously and as a moral person.

In the present investigation we focus on the epistemological question (2), the answer to which, however, is inextricably linked to (1.2), that is the ontological status and the criteria required to consider someone or something as a moral person.

Due to the conceptual vagueness as well as the differences in the use of the same terms between the philosophical and the technical vocabulary, we deem it appropriate to make some introductory clarifications.

---

[7] Ugo Pagallo, "Robots of Just War: A Legal Perspective," *Philosophy & Technology* 24, no. 3 (2011): 307-323; Wendel Wallach, and Collin Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford: Oxford University Press, 2009).

Speaking of intelligent autonomous military machines we refer mainly to Autonomous Weapon Systems (AWS), and Lethal Autonmous Weapon Systems (LAWS). These systems, as defined by the UK Department of Defense (2011), are capable of "understanding" instructions, intentions, environments etc. and after considering the alternatives, to decide autonomously and take actions that cannot be foreseen in advance.[8] Hereto, what is claimed to make war machines "perceive," "understand," decide and act alone, utilizing and evaluating complex information in order to achieve a specific mission, is Artificial Intelligence.[9]

Although philosophers disagree on the exact definition of intelligence, we could accept that by this term we mean the ability of an entity to achieve complex goals.[10] In other words, it is a computational process in which information is transformed through functions (op. cit.). According to Haugeland however, Artificial Intelligence researchers and developers aim to create a *genuine* intelligence, rather than an imitation of the human one.[11] In this sense, researchers are trying to build a non-biological intelligence that will have the characteristics of intelligent beings. In fact, they are trying to build machines with *cognition* that will be capable of intelligence.[12]

---

[8] "Joint Doctrine Note 2/11, The UK Approach To Unmanned Aircraft Systems," *Ministry of Defense*, accessed July 20, 2020, https://www.law.upenn.edu/live/files/3890-uk-ministry-of-defense-joint-doctrine-note-211-the.

[9] Peter Singer, *Wired for War* (New York: Penguin Press, 2009), 145.

[10] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (New York: Knopf, 2017), 73.

[11] John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge, Massachusetts: MIT Press. Chicago, 1985), 255.

[12] Regarding the metaphysical differences between Cognition and Intelligence, see Alkis Gounaris, "Human Cognition and Artificial Intelligence: Searching for the Fundamental Differences of Meaning in the Boundaries of Metaphysics," accessed January 14, 2019, https://alkisgounaris.gr/gr/research/human-cognition-artificial-intelligence/. This is a fundamental difference which, however, is not taken into account by the majority of AI researchers who equate the two concepts. According to our position,

This position stems from the assumption that the human brain is nothing more than a biological computing machine that produces human cognition and has the ability to achieve complex goals, that is, to have intelligence. The anthropomorphic view of artificial intelligence as well as the mechanistic view of the human cognition, enframes research and discussion within defined linguistic boundaries (psychological and mechanistic vocabulary) in which we perceive and define the abillities and functions of autonomous systems.

For example, we may say that the artificial intelligence system thinks, understands etc., or that the brain performs algorithmic calculations. In these cases we use language metaphorically, borrowing terms from different scientific vocabularies, and as a result this temporary loan from one language game is established with another meaning within a different language game. As the concepts of cognition, intelligence, consciousness etc. remain cloudy, indeterminate and are used in many different ways by both philosophers and AI specialists, their ontological clarification becomes particularly complicated.[13] As a result,

---

intelligence can be defined as the ability to achieve complex goals and is inextricably linked to computational ability, and cognition is defined as the ability of the cognitive being to learn, perceive and understand, to make value judgments and decisions, to give meaning to its world, etc., i.e. processes that are not necessarily related to computing capacity.

[13] Christian De Quincey, "Switched-on Consciousness: Clarifying What It Means," *Journal of Consciousness Studies* 13, no. 4 (2006): 6-10; David Levy, "The Ethical Treatment of Artificially Conscious Robots," *International Journal of Social Robotics* 1, no. 3 (2009): 209-216; Aaron Sloman, "A Systematic Approach to Consciousness (How to Avoid Talking Nonsense?)," accessed July 28, 2020, http://www.cs.bham.ac.uk/research/projects/cogaff/misc/consciousness.rsa.text. In fact, as Hoffmann and Hahn point out, this vagueness in the definition of intelligence leads respectively to an ambiguity as to the characterization of a machine as an AI system, see Cristian Hoffmann, and Benjamin Hahn, "Decentered Ethics in the Machine Era and Guidance for AI Regulation," *AI & Society* 35, no. 3 (2020): 635-644. Indeed, it seems practically impossible to know whether to classify a machine as an "Artificial Intelligence system" without first having a

most thinkers turn to the formulation of behavioral cues and ultimately behavioral criteria of intelligence.[14] Daniel Dennett seems to follow this shift towards behavioral criteria as well, albeit in part as we shall see,[15] defending the "human" behavior of HAL 9000.

The concept of autonomy has also had comparable linguistic adventures as we will see in more detail below, since it

clear definition of the term "intelligence." In this sense, the conceptual vagueness of the term "intelligence" also leads to a vagueness regarding the definition of the borders to the set of entities to which we attribute the term "Artificial Intelligence" – and it should be emphasized that as an already first serious consequence, we can't precisely define all the technological applications that fall within the field of analysis of AI ethics. The phrasing of the Turing-Red-Flag-Law, which essentially expresses a demand that all AI systems be indeed recognizable as such, is, after all, characteristic of the severity of the whole situation, see Toby Walsh, *It's Alive: Artificial Intelligence from the Logic Piano to Killer Robots* (Hamburg: Edition Körber, 2017).

[14] The first move towards finding behavioral criteria was made by Descartes, with his suggestion of the criterion of Language as well as the criterion of successful action-in-the-world, see Gerald J. Erion, "The Cartesian Test for Automatism," *Minds and Machines* 11, no. 2 (2001): 29-39; Keith Gunderson, "Descartes, La Mettrie, Language, and Machines," *Philosophy* 39, no. 149 (1964): 193-222; Virginia Savova, and Leonid Peshkin, "Is the Turing Test Good Enough? The Fallacy of Resource-Unbounded Intelligence," *International Joint Conferences on Artificial Intelligence Organization: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, IJCAI-07:545-550, accessed August 29, 2020, https://www.ijcai.org/Proceedings/07/Papers/086.pdf. In the 20th century, this shift to behavioral criteria was marked by Turing's introduction of the 'Imitation Game' - now known as the *Turing Test*, though Turing's intentions were diametrically opposed to those of Descartes, as the former turned to behavior in order to support an ontological equivalence of humans and machines, while the latter did so in order to support their ontological distinction, see Alan Turing, "Computing, Machinery and Intelligence," *Mind* LIX (1950): 433-660.

[15] See bellow on the criterion of excessive efficiency in this regard.

is used differently by moral philosophers and by the designers and engineers of Artificial Intelligence.[16] For Kantian moral philosophers, autonomy forms the basis of moral responsibility and the attribute of personhood[17] and is associated with free will and self-governance – namely, the possibility and the ability of the person to delimit his/her own actions. In order to have moral responsibility, a person must be autonomous or in any case free from coercion.

This means that the person should be free from external factors that can force one to act in a certain way (for example not to have a gun to their head) and not to be limited by uncontrollable internal factors that determine one's decision (for example not to be under the influence of a drug or in some uncontrollable mental state). The decision, that leads a person

[16]The 'technical' (i.e. the technological) use of the term "autonomy" usually refers to a long period of time between two consecutive energy charges, while in the case of weapon systems it means that the weapon has "fire and forget" ability, i.e. the ability to maintain focus and targeting upon the target chosen by the human operator, without the operator having to constantly intervene. On the contrary, the philosophical expression of the term "autonomy" is inextricably linked to moral responsibility and at the same time it is charged with a multitude of rich ontological contexts that, as we will see below, reach as far as the concept of cognition. It often happens that the researchers of AI start their reference to the "autonomy" of the machines in the 'technical' way but in the process, they forget about it and claim for these machines what a philosophical expression of this term would dictate. Thus, due to a misleading analogy, according to Wittgenstein, a similarity in the surface grammar of these two ways of delivering the term "autonomy," they come to support a similarity in depth grammar, that is, in meaning. We must, of course, say in advance that Dennett, whose argument we shall consider, does not make such a mistake and uses the term "autonomy" in the philosophical way. However, if his argument proves to be insufficient, the only way in which the use of this term in terms of AI systems may be possible will be in the end the 'technical' one.

[17] John Christman, "Autonomy in Moral and Political Philosophy," *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), ed. Edward N. Zalta, https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/.

to a specific act, should be determined by the person themselves in a reasonable manner.[18]

Autonomy is by definition a precondition for moral responsibility in such a way that moral responsibility entails autonomy. As Müller observes though, this relation is not inversely implied as well.[19] The term "autonomous systems" in a technical sense does not necessarily mean that these systems are morally responsible for their actions. According to this technical and weaker concept of autonomy, a mechanical system (intelligent or not) is considered autonomous in relation to its degree of control by the human factor. [20]

This weaker notion of autonomy leaves open the question of who ultimately controls the system and who bears the moral responsibility. This is the problem that in ethics is called Responsibility Gap[21] which we encounter in complex situations (e.g. in economics and business, in war, in international relations etc.) where the act in question, while it presupposes the participation of many people or bodies in an earlier stage of the act, ultimately cannot be accurately predicted or controlled in these previous stages. In autonomous AI, for example, questions are raised regarding the share of responsibility – if there is one – of programmers, developers, designers, research sponsors, the company that built the AI system etc., and even end users.

---

[18] Sarah Buss, and Andrea Westlund, "Personal Autonomy," *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), ed. Edward N. Zalta, https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/.

[19] Vincent C. Müller, "Ethics of Artificial Intelligence and Robotics," *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), ed. Edward N. Zalta, https://plato.stanford.edu/archives/fall2020/entries/ethics-ai/.

[20] Vincent C. Müller, "Autonomous Cognitive Systems in Real-World Environments: Less Control, More Flexibility and Better Interaction," *Cognitive Computation* 4, no. 3 (2012): 212-215.

[21] Regarding the Responsibility Gap in AI, see Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology* 6, no. 3 (2004): 175-183.

For Dennett, however, as we shall see below, an Artificial Intelligence system that operates autonomously and effectively can be evaluated morally like any other moral person, as long as it demonstrates intelligent behavior similar to human behavior (in specific objectives). Dennett in his now classic article entitled "When Hal Kills, Who's to Blame? Computer Ethics," which according to Sparrow, is the most serious modern philosophical defense of the position that machines could be held responsible for their actions, builds his argument by first citing the iconic chess victory of the first IBM computer, Deep Blue, over world champion Gary Kasparov in 1996.[22]

In particular, he claims that we recognize and admire the ability of the computer to win in chess and congratulate its developers for the achievement, but this victory belongs to the computer and not to the developers. If the developers faced the world champion, they would obviously lose to him in a few minutes. The responsibility of the developers for the victory of Deep Blue is equivalent to the responsibility of Kasparov's coach or teacher, but ultimately the "responsibility" for the result of the match is born by the players themselves and specifically Kasparov and Deep Blue.

Dennett's argument is extremely relevant if one considers two important AI achievements that essentially signal a future that concerns us. The first has to do with the consecutive victories in 2016 of the AI system called AlphaGo built by Google's DeepMind against Lee Sedol, world champion and one of the most important players of all time in the GO game. Sedol quit after his defeats, admitting that AI is now invincible.[23] The peculiarity of GO is that unlike chess, it relies not only on the computing ability of the players but also

---

[22] Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24, no. 1 (2007): 62-77.

[23] James Vincent, "Former Go Champion Beaten by DeepMind Retires after Declaring AI Invincible," *The Verge*, accessed August 1, 2020, https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat.

on more complex cognitive skills, with many claiming that it is actually a kind of art.[24] The second achievement is the total dominance of DeepMind's AI system in virtual air combat, in 2020, over top pilots of the United States Air Force with F16 Viper fighters.[25] The significance of this victory lies in the fact that in addition to computing skills, perception of three-dimensional space, physical skills and deceptive movements are required.

Dennett extends the reasoning for accountability proportionally, by moving from Deep Blue to HAL 9000, a heuristically programmed Algorithmic Computer[26] who is the main character of Stanley Kubrick's film *2001, A Space Odyssey*.[27] HAL has infinitely greater computing power than Deep Blue, operates "autonomously" and carries out life and death operations, since in order to ensure the success of its mission when it realizes that it is in danger, HAL decides to kill the spacecraft crew in which it was installed, and gain full control. Dennett attributes moral personality traits to HAL because this autonomous intelligent machine exhibits human

---

[24] As Tegmark points out there are far more possible positions in GO than there are atoms in the universe, which means that no computer system can analyze all the interesting sequences of future movements, see Tegmark, 114.

[25] Fabienne Lang, "AI Flawlessly Beats US Air Force F-16 Pilot in Simulated Dogfight," *Interesting Engineering*, accessed August 21, 2020, https://interestingengineering.com/ai-flawlessly-beats-us-air-force-f-16-pilot-in-simulated-dogfight.

[26] Heuristic mechanisms are computer problem-solving techniques which evaluate and select intermediate situations by rejecting the rest, in order to save time. In AI, although these techniques are algorithmically coded, they are not considered "exactly" algorithms, as algorithms always lead to accurate results, while these mechanisms more closely resemble human "intuitive" thinking and educated guess.

[27] The script of the film was based on the novel of the same name by Arthur Clarke; see Arthur C. Clarke, *2001: A Space Odyssey* (New York: New American Library, 1968).

behavior, regardless of whether it repents, feels remorse, feels, or understands what it means to be a moral person.

In our view, however, the arguments put forward by Dennett do not sufficiently prove the position that HAL can be characterized as a moral person.

I. The Argument of Equivalence

Initially, the supposed equivalence of Kasparov's relationship with his coach and Deep Blue with its developers is not logically obvious. In particular, this equivalence can be supported in two ways:

> a) The computer is ontologically equivalent to the human athlete or
> b) The computer is not necessarily ontologically equivalent to the human athlete but the 'developer – computer' relationship is functionally equivalent to the 'coach – athlete' relationship, i.e. these two relationships can both be described in common functional terms. In other words, the study of both of these relations at a functional level can lead to an identical description: the two relations are reduced to the same set of functions performed.

In the case of a), that is, in the case where one argues that the computer is ontologically equivalent to a human athlete, the logical fallacy of a circular argument is being committed, as in the end we come to take for granted what we are trying to prove.

Related to this, to say "The responsibility of the developers for the victory of Deep Blue is equivalent to that of the coach or teacher of Kasparov," based on the assumption that the computer is ontologically equivalent to Kasparov, takes for granted what needs to be proven – i.e. this equivalence. One would expect that we would provide sufficient evidence to demonstrate this human - computer ontological equivalence, instead of simply making an affirmative statement that ends

up being essentially a tautology, hence a sentence without real "epistemological value."[28] Specifically, to say that "The victory belongs to the computer because the 'developer – computer' relationship is the same as the 'coach – athlete' relationship," and that "the 'developer – computer' relationship is the same as the 'coach – athlete' relationship, because the computer and the athlete are ontologically equivalent," is like saying "The computer and the athlete are ontologically equivalent because they are ontologically equivalent." The only way to escape this tautology is:

> a1) To finally face the problem head on, trying to answer the question: Under what criteria can we establish an ontological human – machine equivalence or distinction? This is the most central, timeless and persistent philosophical question of AI.
> a2) To try to disengage the discussion of accountability and (ultimately) moral status from the issue of the ontological human – machine equivalence or distinction. But how easy is it to separate these two in our thought? What else could provide a sufficient criterion for assigning moral status to an entity other than the ontology of the latter? Are there any examples of acceptable human thought in which the rendering of moral status and ontology were not correlated in one way or another? All moral status queries soon lead to ontology status queries.

In the case of b), that is in case we would attempt to attribute the same moral status to both Kasparov and Deep Blue on the

[28] In addition, one should explain the terms under which two entities are considered ontologically equivalent and adequately justify these terms. For example, we could suggest functionalist terms, but then we would have to justify our choice to make a functionalist description. In addition, the functionalist description will make us confront the problems discussed below in relation to b).

basis of a functionalist equivalence of the 'coach – athlete' and the 'developer – computer' relationships, we are called upon to demonstrate this very functionalist equivalence of these two relationships either

> b1) through a 'coach – developer' and 'athlete – computer' functionalist equivalence (in this case, the equivalence of the relationship of the 'coach – athlete' and the 'developer – computer' pairs is established by demonstrating the relations of equivalence of the respective members of these pairs)[29] or
> b2) because the respective members of the pairs are not functionally equivalent but the pairs that these members form, happen to be (in this case the equivalence does not lie in the members, but in the relationships they enter into with each other).[30]

Moreover, in the face of the prospect of self-programmed and self-reproducing machines, the argument based on the parallelism of developers to coaches and machine to athletes is invalidated, as the role of the human programmer becomes unnecessary.[31]

But let us look in more detail at the problems that arise from trying to prove a functionalist equivalence. Regarding b1) we must emphasize that proving a functionalist 'computer –

---

[29] For example: a-b = g-d, because a = g and b = d.

[30] For example: a ≠ g, and b ≠ d, but a-b = g-d.

[31] For an interesting analysis of the philosophical implications of the possible development of self-programming and self-reproducing machines, see John Von Neumann, *Theory of Self-Reproducing Automata* (Urbana: University of Illinois Press, 1966); Rudy Rucker, *Infinity and the Mind: The Science and Philosophy of the Infinite* (Princeton, N.J.: Princeton University Press, 1982), 157-188. The idea of self-replicating machines is not new. For one of the first technical analyses of the possibility of self-reproducing machines, see Edward F. Moore, "Artificial Living Plants," *Scientific American* 195, no. 4 (1956): 118-126.

athlete' equivalence implies proving a functionalist 'human – machine' equivalence which in turn has not been possible so far. The most well-known and organized attempt to establish an ontological equivalence of 'human – machine,' the theory of Functionalism and especially of *Machine Functionalism*, has presented serious problems, some of which are already found in the fundamental assumption of this theory, i.e. in the position that thought equals computation.

This is a position whose proof has not been reached yet, as in addition to the cloudy image we have regarding the ontology of the Mind, there are significant and well-established obstacles in the nature of any computation in which the notion of infinity is after all involved.

This very problem of the possibility of an infinite computation was pointed out by Turing himself (on whose theoretical *Machine Functionalism* is largely based) who proved that a general algorithm to solve the Halting Problem cannot exist.[32]

---

[32] Alan Turing, "On Computable Numbers With an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* 42, Series 2 (1937): 544-546; Alan Turing, "On Computable Numbers With an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* 43, Series 2 (1938): 544-546. For a comprehensible and detailed presentation of the issue of non-computability as well as for its implications regarding AI, see John L. Casti, and Werner De Pauli, *Gödel, A Life of Logic* (Cambridge: Basic Books, 2000). Also see Hubert L. Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason* (New York: MIT Press, 1992), chapters 5 and 10; Rucker, 157-188; John R. Lucas, "Minds, Machines and Gödel," *Philosophy* XXXVI (1961): 112-127. For the optimistic and ultimately opposite to Dreyfus and Lucas approach, see Paul Benacerraf, "God, the Devil and Gödel," *The Monist* 51 (1967): 9-32. Note that one functionalists' gambit in order to escape the impasse of non-computability, is to support the position that intelligence could be fully reproduced by a suitably "complex" Turing Machine. However, this position, apart from being analogous to the Church-Turing thesis and an unproven position, creates a new problem, as Jaegwon Kim notes, since

Still, beyond the problem of non-computability, Functionalism inevitably falls into a circular argument, as it fails to define any functions without referring to mental terms, thus it ends up trying to establish the possibility of cognition in machines while actually presupposing it.[33]

In addition to these specific logical fallacies, a functionalist attempt to prove the above equivalence, faces two major ontological problems that functionalists are called upon to solve in general. The first one is that the functionalist description ignores or fails to describe the qualitative and subjective appearances (the phenomenal aspect) of mental states, that we call qualia.

Focusing solely on the input-output relationship of a system (human, animal, machine, etc.) Functionalism leaves open a rather paradoxical possibility: Two systems may have exactly matching inputs (stimuli) and outputs (behavioral manifestations), but completely different or even inverted qualia – that is, to experience completely different or even inverted 'internal states.' It is also possible that qualia can be completely absent from one of the two systems.[34] The paradox, here, is that according to *Machine Functionalism*, these two systems are considered functionally equivalent, despite their differentiation in the level of qualia.[35]

---

functionalists are now called upon to determine what complexity is and what the appropriate complexity threshold is, beyond which a Turing Machine succeeds in demonstrating intelligence, see Jaegwon Kim, *Philosophy of Mind* (USA: Westview Press, 1998), 151-156.

[33] Kim, *Philosophy of Mind*, 153, 154.

[34] Ned Block, "Troubles with Functionalism," in *Readings in Philosophy of Psychology*, vol.1, ed. Ned Block, 268-305 (Cambridge: Harvard University Press, 1980). For the two opposing views on the possibility or non-existence of inverted or absent qualia, see David Shoemaker, "Caring, Identification, and Agency," *Ethics* 114, no. 1, (2003): 88-118; Ned Block. "Are Absent Qualia Impossible?" *Philosophical Review* 89 (1980): 257-274.

[35] At this point there have been objections from some philosophers who deny the existence or the epistemological validity of qualia during the ef-

fort of knowledge (inspection) of the Mind [for example see Paul M. Churchland, *Matter and Consciousness* (Cambridge, Massachusetts: MIT Press, 1988); Keith Frankish, "Illusionism as a Theory of Consciousness," *Journal of Consciousness Studies* 23, nos. 11-12 (2016): 11-39; Keith Frankish, *Illusionism: As a Theory of Consciousness* (Exeter: Imprint Academic Publishing, 2017); Georges Rey, "A Reason for Doubting the Existence of Consciousness," in *Consciousness and Self-Regulation*, vol. 3, eds. Richard J. Davidson, Gary E. Schwartz, and David Shapiro, 1-39 (New York: Plenum, 1983); Georges Rey, "A Question About Consciousness," in *Perspectives on Mind*, eds. Herbert R. Otto, and James A. Tuedio, 5-24 (Dordrecht: D. Reidel Publishing, 1988); Kathleen Wilkes, "Yishi, Duh, Um and Consciousness," in *Consciousness in Contemporary Science*, eds. Antony Marcel, and Edoardo Bisiach, 16-41 (Oxford: Oxford University Press, 1988)]. However, it is difficult to imagine that in the absence of qualia we could talk about the experiences of taste, smell, color, touch, etc. or even illusory experiences. Finally, it is difficult to see how we could categorize our stimuli, recognizing for example the taste or the aroma of a fruit we have eaten before [for the opposite position, see Daniel Dennett, "Quining Qualia," in *Consciousness in Contemporary Science*, eds. Antony J. Marcel, and E. Bisiach, 42-77 (New York: Oxford University Press, 1988); Valerie Hardcastle, *The Myth of Pain* (Cambridge, MA: MIT Press, 1999)]. Moreover, when it comes to the knowledge of consciousness itself, the distinction between illusion and reality collapses and therefore any critique of the epistemological validity of qualia regarding the knowledge (inspection) of the Mind becomes problematic at the very least: "Where consciousness is concerned the existence of the appearance is the reality", see John Searle, *The Mystery of Consciousness* (New York: The New York Review of Books, 1997), 122. In any case, we see here, on the occasion of the present as well as the immediately preceding footnote on qualia, that a functionalist approach to the question of attributing moral personhood in machines, such as the one attempted by Dennett, may open up many more issues than those it is coming to close. Even if Dennett opposes the existence of qualia, the issue remains open and one of the most debatable in modern philosophy, see Dennet, "Quining Qualia," 42-77. Therefore, invoking a functionalist analogy between the 'coach – athlete' and 'developer – machine' relationships would bring us face to face with this serious and still-pending philosophical ontological problem, leading to an endless

The second ontological problem focuses on whether – and if so, to what degree – does an intelligent machine or intelligent operating system *understand* or realize the meaning of the computational process and the result it produces. The most popular description of this problem has been made by John Searle in 'The Chinese Room Argument.' With this argument Searle showed that the successful syntax of physical symbols by the machine does not require the machine to understand these symbols.

Therefore, machines do not understand and in the end their implementation of a successful syntax as it takes place during the execution of an algorithm is not a demonstration of cognitive ability.[36]

The above two ontological problems that functionalists have to solve, prove that it is not self-evident that a machine that simulates human behavior is intelligent merely because it demonstrates an input-output mapping that matches that of a human in a given task. Dennett, however, a priori rejects the

---

discussion that would gravely deviate from the clarity that a criterion used to attribute moral status must have within the context of a branch of Applied Ethics such as AI Ethics.

[36] John Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3, no. 3 (1980): 414-457; John Sealre, *Minds, Brains, and Science* (Cambridge Massachusetts: Harvard University Press, 1984). Another famous argument against *Machine Functionalism* is the Multiple Realization Argument. This response is extremely interesting, as the Multiple Realization Argument was originally articulated to support Functionalism. However, understanding the way in which the Multiple Realization argument affects Functionalism and the response to it, requires an extensive reference to the structure and operation mode of the Turing Machine as well as an extensive bibliographic reference, that go beyond the main purpose and the allocated length of this article. For an overview of how Multiple Realization affects *Machine Functionalism*, see Hilary Putnam, *Representation and Reality* (Cambridge: MIT Press, 1992). For a more comprehensive analysis of the relationship between Functionalism and Multiple Realization, see Jaegwon Kim, *Supervenience and the Mind* (Cambridge: Cambridge University Press, 1993).

useful role of qualia in cognitive science and disagrees with Searle, arguing that the Chinese Room "understands" as a comprehensive system the meaning of the result it outputs, and ultimately adopts the attitude of a rational behaviorist towards the Kasparov - Deep Blue (Human - Machine) equivalence, content with the end result and the behavior of the compared entities.

This disagreement demonstrates that the functionalist approach is characterized by ontological issues that remain pending to this day. Therefore, for the time being, it does not seem to be the most appropriate for the consolidation of an easy-to-use and robust criterion in order to attribute moral status to machines. In any case, as we have seen, the substantiation of the functionalist equivalence fails already at a logical level. Therefore, we should probably go back to the need of directly addressing the basic question of AI referred to above, namely the question of the ontological equivalence or human - machine distinction and eventually to a).

Finally, regarding b2), that is, the functionalist comparison not of the members that make up the pairs 'coach – athlete' and 'developer – computer' but of the relationships that these pairs form, we must observe that already the 'coach – athlete' relationship seems to be characterized by a much higher level of freedom than the 'developer – computer' relationship. The computer's actions seem to be much more dependent on the developer's commands, than the athlete's actions bound by the commands of his coach.

In fact, in the functionalist definition of the 'developer – computer' relationship there is the program factor, which does not seem to have a functional analogy in the case of the 'coach – athlete' relationship. In addition, one could argue that the developer and the machine are involved in an endless loop of dynamic interaction and in an ongoing dialogue that simultaneously determines the actions of both.

In any case, this discussion regarding the laxity or not of the 'coach – athlete' relationship versus the 'developer – computer'

relationship, brings forth the terms of environmental programming (the environment as a developer) and ultimately of autonomous agency. These are terms that have a timeless presence in the effort to address the fundamental philosophical question of AI as to the ontological identification or distinction of human – machine.[37]

II. The Argument of Autonomy

The fact that Dennett, among other things, invokes the autonomy of HAL in order to attribute moral responsibility to HAL, thus moral personhood, is not something new in the field of AI Ethics. Other thinkers and researchers have also linked the attribution of moral personhood to the machines with the issue of autonomy.[38] Moreover, empirical studies in the Psychology of Human-Computer Interaction indicate that the majority of people consider the ability of a machine to make choices as being one of the basic criteria for attributing moral responsibility to this machine.[39] At first glance, this connection of the attribution of moral personhood to the machines with the concept of autonomy seems quite reasonable, especially under a Kantian approach.

---

[37] At this point an intersection – or rather a common conclusion – of a) and b) is found again. It seems, therefore, that even under a functionalist attempt to bypass the direct confrontation of the ontological question of the human – machine identification or distinction – that is, even with the gambit of reducing an ontological question to functionalist terms, the basic features of the question and their impasses remain fully valid.

[38] David Calverley, "Toward a Method for Determining the Legal Status of a Conscious Machine," in *Proceedings of the AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment*, eds. R. Chrisley, R. Clowes, and S. Torrance, 75-84 (Hatfield: University of Hertfordshire, 2005); Sparrow, "Killer Robots," 62-77.

[39] Andrew E. Monroe, Kyle D. Dillon, and Bertram F. Malle, "Bringing Free Will Down to Earth: People's Psychological Concept of Free Will and its Role in Moral Judgment," *Consciousness and Cognition* 27 (2014): 100-108.

In the case of HAL, Dennett attempts to overcome the issue of the possible heteronomy of a programmed computer by comparing HAL with the case of a genetically or 'environmentally programmed' moral agent. If genetic programming and human experiences exempt humans from their moral responsibilities, then they should do so for HAL. At this point, we practically have the articulation of the argument that the environment is for the humans what the programmers are for the machines. According to this line of thought, one could say that even if – as shown above – it would be quite difficult if not impossible to establish a coach – programmer analogy in detailed functionalist terms, there could be at least some trainer – environment parallelism that could possibly prepare the grounds for the support of an ontological equivalence between humans and machines.

Here one could object this view by stressing the fact that Dennett overlooks an important aspect which makes the use of the term "autonomy" a metaphorical one. Specifically, it could be supported that contrary to the case of humans, each autonomous AI system integrates a certain given goal, a predefined task. For instance, it is not possible for such a machine to temporarily postpone the execution of its task in order to take a break and have a cup of coffee or read a book. Every task of an AI system is predefined, given, inescapable and extraneously determined (determined 'from the outside') in such a way that any notion of autonomy is negated. This is due not just to the fact that the machine is programmed in a certain way, but because the goal of its existence is integrated in its essence. Every machine is a 'machine for…,' namely it is built to perform a certain function in order to achieve certain goals regardless of their complexity. The conception – not to mention the construction – of an intelligent machine with no particular goals seems a real challenge for AI research.[40]

[40] It has to be noted that although one of the basic visions of the researchers in the first years of the AI scientific program was the creation of general purpose machines (the concept of the Universal Turing Machine, projects like Allen Newel's and Herbert Simon's The General Problem

Nevertheless, if we want to be really fair to Dennett, we have to ask ourselves how different are humans compared to machines, regarding the issue of a goal integrated into their existence. Do humans really come to life and grow up free from goals not chosen by them but chosen by their environment? Often, humans are nurtured, bred more or less explicitly to be given a certain purpose in their lives. The extreme cases of the 'tightly closed' priories and religious orders, the more usual cases of political youth clubs, the church and people's introduction to a system of religious faith, the training of the priests, military training and finally the more imperceptible and intangible ways of training from the environment, such as family members leading by example, gender-specific role-taking are examples of the wide variety of environmental influences over humans. But even before we consider all these, the very fact of a human's birth integrates a goal extraneous to this human, namely the choice of one's parents to bring him/her to life (in order to achieve a continuation of their name or to satisfy their parental or sexual instincts or even to satisfy the social role models, the wishes of their families etc.). Therefore, one arrives at the following question: Up to which level of environmental influence could an entity be thought of as being autonomous? In other words, which is the threshold of intervention beyond which the environmental influences are considered as programming, as a mechanism of reaching to a heteronomy of an entity's will? Which is the threshold of the extraneous intervention beyond which the entity is considered to have integrated to the essence of its existence a goal extraneous to it? At this point we seem to be asking for a quantitative criterion (specifically a threshold), since the issue of attribution of moral status is also usually dealt with a quantitative manner (we usually attribute different levels of

Solver and the cognitive architecture SOAR, are examples of this vision), such a development has not yet taken place - possibly due to ontological restrictions in the very the nature of a machine.

moral status to different entities).[41] Thus, it seems that before we are able to identify this threshold, we can't totally reject Dennett's argument of a parallelism between the environment and the programmers. Given that the humans undergo a kind of programming by their environment, absolute autonomy might not even exist for humans either. Therefore, for now, it seems that we don't have the right to support a distinction between humans and machines on the basis of an argument of goals being imposed to the machines by their human creators and programmers.

If we really want to identify a problem in the use of the criterion of autonomy we will have to shift the focus of the discussion from the machine – programmer relation to the definition and the determination of the limits of the philosophical concept of autonomy and to the way in which this concept is related to the attribution of moral status. Moreover, we will also have to focus on our ability to identify the presence of autonomy in an entity.

Previously in this text, we saw that for one to be acknowledged as an autonomous agent one must not be in a status of internal or external coercion, namely not to have a gun pointed to his/her head or not to be in a mental state that is not controlled by him/her.

However, if we want to be precise with the definition of the concept of autonomy we need to be in the position to answer the four following questions:

> 1) Which are the presuppositions of autonomous agency? Which are the features and the properties that an entity has to have in order to act as an autonomous agent? In other words, how is the concept of autonomous agency delimitated?

[41] Regarding particularly for the different levels of the attribution of moral responsibility for acts of war and specifically for the distinction between adult and children soldiers as well as for a parallelism between the latter and AI weapons see Sparrow, "Killer Robots," 62-77.

> 2) How can we identify autonomy? Which are the indications that we need to have in order to regard an entity as being an autonomous agent?[42]
> 3) Is autonomous action – especially the action of a moral agent – necessarily linked to the property of the cognitive being? In other words, is an agent's mental state (and cognition in a broader sense) a necessary condition for autonomous action?
> 4) Is the issue of autonomous agency attribution totally symmetrical to the issue of moral status attribution? Does the characterization of an entity as an autonomous agent necessarily entail that this entity can also be characterized as a moral agent?

First, we have to see that questions 1 and 2 are linked, since some of the features and the properties required for reaching autonomous agency can inform the criteria for the autonomous agency identification. For instance, if the feature F is demanded so that an entity E is truly autonomous, then a safe criterion for the identification of autonomous agency in an examined entity E would be the identification of F as a feature of E. Question 1 is an ontological question (What is the autonomous agency?) while question 2 is an epistemological question (How can we know the existence of autonomous agency?). However, frequently the answer to the epistemological question is strongly defined by the answer to the ontological question.[43]

---

[42] The determination of the features and therefore the safe indications of autonomous agency is crucial since these indications will form the basis of the ontological evaluation and classification of the entities under the question of moral status attribution. See right below, in the main text.

[43] A typical example of the connection between an ontological and an epistemological question is Thomas Reid's introduction of the 'Other Minds Problem' as a critique in the way in which Berkeley approached the concept of mind; see Anita Avramides, *Other Minds* (London, New York: Routledge, 2001), 139-180. Here it must be pointed out that apart from the concept of autonomy, this connection between the ontological and

Nevertheless, with regard to question 1, a plurality of definitions of – and finally presuppositions for – autonomous agency exists.[44] Which of all these views is the correct one? Thus, which of all these views should be the basis of the discussion regarding the attribution of moral personhood to AI systems? It seems that until now most of the researchers in the field of AI Ethics have adopted internalist approaches (in the sense that they refer to the concept of consciousness and to mental states like intentions, beliefs, emotions etc.), and therefore they approximate or they are even in complete alignment with what in the traditional field of autonomous agency analysis is known as the Coherentist View.[45] According to the Coherentist View

the epistemological question exists also with regard to any other concept that has been related to the attribution of moral personhood. It is reasonable that concepts like consciousness, cognition and intelligence have also an ontological and an epistemological question with the answer to the first affecting the answer to the latter which in its turn affects the feasibility of the ontological classification of the examined entities.

[44] Returning to the above analysis with regard to the 'environment as a programmer' argument, we have to see that this plurality of autonomous agency definitions and presuppositions affects also in a negative way our ability to identify the threshold of extraneous intervention beyond which an entity has to be considered as integrating an extraneous goal to the essence of its existence. For a review of the way in which the problem of defining the limits of the concept of autonomous agency is connected to the problem of defining the limits of the extraneous interventions see Buss, and Westlund, "Personal Autonomy."

[45] Calverley, "Toward a Method," 75-84; Manuel De Landa, *War in the Age of Intelligent Machines* (New York: Swerve Editions, 1991); David Levy, *Intimate Relationships with Artificial Partners* (Ph.D. Diss., Maastricht University, 2007); Steven Pinker, "Can a Computer Ever Be Conscious?," *US News & World Report* 123, no. 7 (1997), accessed July 28, 2020. https://stevenpinker.com/files/pinker/files/computer.pdf; Solum, "Legal Personhood," 1231-1287; Sparrow, "Killer Robots," 62-77; Steve Torrance, "Could We, Should We, Create Conscious Robots?" *Journal of Health Social and Environmental Issues* 4, no. 2 (2004): 43-46. For a detailed presentation of all the views that are until now proposed regarding autonomous agency see Buss,

an agent has control over his/her action if and only if the motive of his/her action is in coherence with some mental state representing the agent's point of view.[46] Nevertheless, different advocates of the Coherentist View propose respectively different mental states as being the proper ones for an autonomous agency. Specifically, these mental states can either be related to some long-term goals, motives and plans[47] or to emotions and mainly emotions of 'caring.'[48] This raises again the issue of the

---

and Westlund, "Personal Autonomy." Given the reasonable space limit in this text, we have decided to focus only on the Coherentist View since up to now this is the one characterizing the discussion in the field AI Ethics. The analysis of the problems or solutions that could possibly come up by examining the rest of the traditional philosophical views regarding autonomous agency could be part of a new fruitful reflection presented in a new article in the future. For the time being and for the needs of the present article, we will be confined in just mentioning that the existence of these other views increases the 'noise' in the analysis of the issue regarding the attribution of moral status to the machines.

[46] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," in *The Importance of What We Care About*, ed. Harry Frankfurt, 11-25 (Cambridge: Cambridge University Press, 1988a).

[47] Gary Watson, "Free Agency," *Journal of Philosophy* 72, no. 8 (1975): 205-220; Michael Bratman, "Practical Reasoning and Weakness of the Will," *Noûs* 13, no. 2 (1979): 153-171; Michael Bratman, *Structures of Agency: Essays* (Oxford: Oxford University Press, 2007).

[48] Harry Frankfurt, "The Importance of What We Care About," in *The Importance of What We Care About*, ed. H. Frankfurt, 80-94 (Cambridge: Cambridge University Press, 1988b); Harry Frankfurt, "On Caring," in *Necessity, Volition and Love*, ed. Harry Frankfurt, 155-180 (Cambridge: Cambridge University Press, 1999); Agnieszka Jaworska, "Caring and Full Moral Standing," *Ethics* 117, no. 3 (2007a): 155-180; Agnieszka Jaworska, "Caring and Internality," *Philosophy and Phenomenological Research* 74, no. 3 (2007b): 529-568; Agnieszka Jaworska, "Caring, Minimal Autonomy, and the Limits of Liberalism," in *Naturalized Bioethics: Toward Responsible Knowing and Practice*, eds. Hilde Lindemann, Marian Verkerk, and Margaret Walker, 80-105 (Cambridge: Cambridge University Press, 2008); Shoemaker, "Caring, Identification, and Agency," 88-118.

plurality of definitions which leads to a reasonable question: Which of all these criteria is the right one? Based on which of all these proposals should we judge the autonomy of humans, animals and machines? The problem of conceptual vagueness makes its appearance once again.

Moreover, the Coherentist View is a good example to return back to the link between questions 1 and 2, since we see here the way in which our inability to come up with a definite and universally accepted answer to question 1, leads also to an inability to provide a definite answer to question 2. Specifically, the plurality of the mental states proposed under the Coherentist View as being the decisive features of autonomous agency – hence the plurality of answers to question 1 – delivers a fatal strike to our chances of reaching to an unambiguous and final answer regarding question 2: How can we know which mental states should we seek to identify in an entity under examination in order to consider this entity as autonomous and therefore qualified for an attribution of moral personhood?

Besides its conceptual vagueness, the problem of the identification of mental states in other entities brings us directly against one of the most central problems in the Philosophy of Mind: The Other Minds Problem. How can we verify the existence of mental states in the entities that surround us? In fact, this question is actually divided into the following two questions:

> a) How can we know whether other beings around us have any mental states at all?, and
> b) If they do have mental states, how can we know the content of these mental states?[49]

In trying to approach the issues of attribution of moral personhood to machines through the Coherentist View of autonomous agency we are faced with the following appearances of the Other Minds Problem: How can we know whether a

[49] Avramides, *Other Minds*, 1.

machine has any mental states and especially mental states of the kind that is related to a point of view of the machine itself? How can we know if a machine has motives and plans and if these motives and plans are for the long-term? How can we know whether a machine has emotions and whether these emotions are related to 'caring?'[50]

[50] Of course, we have to mention that, apart from the Coherentist View, the Other Minds Problem is also an obstacle for any other internalist approach of the issue of moral personhood attribution, even for the approaches that do not refer to the criterion of autonomous agency. We can briefly refer here to a trend within the AI Ethics field that examines the issue of moral personhood attribution to the machines through the concept of patiency, see Tom Regan, *The Case for Animal Rights* (Berkeley & Los Angeles: University of California Press, 1983); Mane Hajdin, *The Boundaries of Moral Discourse* (Chicago: Loyola University Press, 1994); Hoffmann, and Hahn, "Decentered Ethics," 635-644; Luciano Floridi, and J.W. Sanders, "On the Morality of Artificial Agents," *Minds and Machines* 14 (2004): 349-379; Levy, "The Ethical Treatment," 209-216; Wallach, and Allen, *Moral Machines*. This is a concept which in its turn is usually linked to the concept of sentience. The latter was introduced for the first time as a criterion for the attribution of moral status in non-human entities by Peter Singer and with reference to the animals [see Peter Singer, *Animal Liberation: A New Ethics for Our Treatment of Animals* (New York: New York Review of Books, 1975); Peter Singer, *Practical Ethics* (Cambridge: Cambridge University Press, 1993)], but now it has been also introduced to the discussion regarding the AI systems, see Levy, "The Ethical Treatment," 209-216; Jonathan Owen, and Richard Osley, "Bill of Rights for Abused Robots: Experts Draw up an Ethical Charter to Prevent Humans Exploiting Machines," *The Independent*, last modified April 1, 2007, https://www.independent.co.uk/news/science/bill-of-rights-for-abused-robots-5332596.html. The basic line of thought regarding the concept of moral patiency supports the view that if AI systems and especially robots are sentient-thus capable of suffering-they should possibly be thought of as victims. However, a question arises of whether we could ever be able to know if machines actually suffer. Indeed, some AI Ethics researchers have started to note the obstacle of the 'Other Minds Problem,' see David Gunkel, *The Machine Question: Critical Perspectives on AI, Robots and Ethics*

At this point, we would like to stress how crucial the answer to question 2 is, when we work in the context of Applied Ethics where sound and practical ontological criteria are required, which, in turn, will lead to sound and handy criteria of moral status attribution in all the grades and shades of the latter. Thus, we would say that until now the treatment of question 1 has not yet led to results really useful for the treatment of question 2. In other words, the question 1 is until now approached in a way that is non-productive for the demands and the needs of Applied Ethics (in this case of AI Ethics).

Due to the dead end in which one is led when confronting the Other Minds Problem, a possible strategy could be an attempt to bypass this problem and examine the autonomous agency criterion irrespectively of any reference to mental states. Such a strategy though would bring forth question 3 ('Is the autonomous action – especially the action of a moral agent – *necessarily* linked to the property of the cognitive being?').

Let us think, for instance, a vehicle with a damaged navigation system, a conventional car with a failing steering rack or with broken brakes. Can we support the view that this vehicle exhibits a kind of autonomy in the sense that its action is not controlled by the driver?[51] It is true that usually we are

(Cambridge, MA: MIT Press, 2012); Hoffmann and Hahn; Levy, "The Ethical Treatment," 209-216.

[51] Regarding this example, one could say that this car is indeed not controlled any more by its human-driver but is now fully under the deterministic laws of nature that totally define its movement. Therefore, not being controlled by its human-driver does not necessarily mean an autonomous agency. Under an extreme naturalistic approach one could support the view that this is also the case with the human-driver. The driver is also under the deterministic laws of nature. Thus, a denial of an entity's autonomous agency on the basis of a reference to the laws of nature could be also applied to the case of humans thus striking the idea of human autonomy too. On the other hand, this would be a maneuver fatal for the whole project of Ethics, thus a maneuver that would violently interrupt and end once and for all the whole present discussion (and search for solutions

not tempted to think that the uncontrolled movement of the car is similar to the autonomy that we think that characterizes the humans. This is due to the fact that what we are looking for here is a *certain type* of autonomy; an autonomy linked to a *certain type* of agency.[52] Which is the essential characteristic of this agency? Why don't we even think to raise the question of attributing this type of autonomous agency to an uncontrolled conventional car that moves with its brakes broken but we do so in the case of a 'smart' vehicle, a computer and above all a human?

Possibly because contrary to the case of the uncontrolled conventional car, in the case of the human we have a priori accepted the property of the cognitive being and in the case of the computer or the 'clever' car there is at least a *suspicion* thus a still open possibility of cognition.[53]

---

in the field of AI Ethics and Applied Ethics in general) not by providing answers to the questions raised but by negating the whole context within which these questions are born and raised. Nevertheless, in the preset analysis we adopt a compatibilist view supporting that the natural laws and the criteria of moral status attribution belong to discrete conceptual fields (namely the ontological and the evaluative).

[52] At this point, recall the described above difference between the philosophical and the 'technical' (technological) use of the term "autonomy."

[53] Although regarding the humans we have definitely accepted the property of the cognitive being which, of course, also implies intelligence, this is not the case with the 'smart' machines. For them the question of cognition remains open even though we answer positively regarding their ability to present intelligence (even in various levels). On the contrary, in the case of a heteronomous machine like the conventional car we a priori answer negatively both for the property of cognitive being and the ability of intelligence. Therefore, it seems that we have three levels in the attribution of the property of the cognitive being and the AI systems are placed in a middle ground (some prefer to call it a 'grey area') somewhere in between the full attribution of the property of the cognitive being (the case of humans) and the total rejection of this possibility (the case of conventional machines). We would like to stress here that the AI systems are not placed towards the negative end together with the rest of the machines due to

It seems, then, that in general we accept that the autonomous agency can't be but a *cognition-related* agency. Therefore, with regard to question 3 ('Is the autonomous action – especially the action of a moral agent – *necessarily* linked to the property of the cognitive being?'), we would answer that based on the dominant views in the fields of Ethics and AI Ethics (but also on the dominant views in our everyday life) the autonomous action *is indeed necessarily linked* to the concept of cognition. However, this concept is not treated in a uniform and unambiguous way, as an 'all or nothing' feature but rather as something that presents quantitative and qualitative variations.[54] Hence, there are still cases of human beings to which we deny the attribution of autonomous agency and therefore the attribution of a full-fledged or at least a partial moral status. Infants, certain categories of mental patients, humans in a comatose or vegetative state are only some of the cases of human beings for which we find it difficult to reach universally accepted and final answers regarding the attribution of cognitive agency and finally of moral status. Consequently, although we think of the autonomous agency as necessarily linked to cognition, the latter seems to be characterized by many different levels and instances which finally lead to speculation on and questioning of the need for adopting different levels in the attribution of moral personhood via the criterion of autonomy. Thus, we would like to complete our answer to question 3 ('Is the autonomous action – especially the action of a moral agent *necessarily* linked to the property of the cognitive being?') as follows: Based on the currently dominant views in the field of AI Ethics and Applied Ethics in general, the autonomous agency is necessarily linked to the property of cognition, but given the quantitative and qualitative differences that we acknowledge in the latter, this

---

the fact that they 'behave' (or behave?) in an intelligent way which creates a suspicion that this could be something more: a cognitive way. Could we ever manage to have something more than just a simple suspicion?

[54] The problem of conceptual vagueness comes forth again here with regard to the concept of cognition.

necessary link leads to a *non-unequivocal* correlation between the autonomous agency and the attribution of moral status. In the end, considering all the above analysis with regard to the Other Minds Problem, this unbreakable link between the autonomous agency and the cognition bequeaths to the first with all the conceptual, ontological and epistemological problems of the latter. As a result of this, the autonomous agency becomes a criterion quite difficult to use for the attribution of moral personhood.

Here, it is also worth mentioning that – at least under the Coherentist View – the autonomous agency becomes difficult and problematic to use as a criterion due to its connection with some other concepts. Specifically, the coherentist account constitutes a point of intersection between the discussion for autonomous agency and the traditional and arduous reflections regarding the concept of the person. This happens in three ways: (i) The demand for the existence of goals and mental states under the point of view of an agent is equivalent to the demand for a delimitation of a *personal* point of view (ii) The existence of long-term goals, plans and motives as the essential features of autonomous agency presupposes the "diachronic unity" of this personal point of view. Thus, it presupposes the continuity, the survival through time of the agent's identity, therefore the survival of *the same* person.[55] (iii) The acknowledgement of the

[55] At this point it becomes obvious that especially the version of the Coherentist View which proposes the long-term intentional mental states as essential for the autonomous agency asks for a "psychological continuity" which is equivalent to the psychological consistency needed for the preservation of the 'sense of the self' and finally of the person's identity. Here, the discussion for the delimitation of the concept of the autonomous agency overlaps with the problems of the preservation in time of the property and of the identity of the person. In other words, this coherentist account of the autonomous agency brings us against what is known as the 'persistence' and the 'characterization question' of personhood. For a detailed analysis of these two problems see Eric T. Olson, "Personal Identity," *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed.

long-term or short-term intentional mental states as being the essential ones for the attribution of autonomous agency might raise a question regarding the delimitation of the agent's will . Which manifestations of intention are thought of as extrinsic to the will, (i.e. as extraneous, as coming from outside the will and imposed on it) and which as intrinsic, namely as pure products of the will? Are there any completely intrinsic intentions? Which is the limit of distinction between the intrinsic and the extrinsic intentions? In other words, what is the border that distinguishes a person from the surrounding world? Moreover, can our impulses, our short-lived and very short-term strong desires be thought of as products of our will? Finally, are our personality traits endogenous or exogenous factors with regard to our will? This conjunction of the issue of autonomous agency with the question regarding the concept of the person is an example of the way in which the philosophical analysis and the conjunction of different concepts leads to an increase rather than a decrease of the philosophical problems, since any new concept (e.g. "person") that is introduced to help us clarify a previous concept (e.g. "autonomous agency") brings with it its own problems of delimitation.

The issue of the delimitation of the will and the inclusion (or not) of the impulses and the very short-term strong desires, brings forth question 4 as well ('Is the issue of autonomous agency attribution totally symmetrical to the issue of moral status attribution? Does the characterization of an entity as an autonomous agent necessarily entail that this entity can be also characterized as a moral agent?').

As one can easily see by looking to the relevant bibliography as well as from our everyday practice, different views regarding the limits of the will lead to respectively different answers to the above question. For example, we usually don't attribute full autonomy to drug addicts. As a consequence of this, we also don't attribute to them full moral status. The discussion

Edward N. Zalta, https://plato.stanford.edu/archives/fall2019/entries/identity-personal/.

over the limits of their moral responsibility has proven to be quite long and arduous. Thus, in the case of drug addicts, the autonomy - moral status relation seems to be symmetrical, namely the negation of full autonomy leads to a negation of a full moral status. Contrary to this, in other cases, for instance in the case of people that have undergone brainwashing or indoctrination, we usually don't attribute autonomy (the traditional bibliography on the issue of autonomous agency is quite clear with this) but we usually do attribute a moral status (for instance moral responsibility for their actions even if those were dictated by their indoctrination). Namely, while according to most of the philosophical accounts of autonomous agency these people are not considered to be fully autonomous agents, they are nevertheless acknowledged to have a full moral status. In this case the autonomy-moral status relation seems non symmetrical, since the negation of autonomy has not led to a respective negation of moral status. We see then that the symmetry of the relation between autonomy and moral status changes on a case-by-case basis; a fact that makes the use of the autonomy criterion even more problematic.

In conclusion, we would say that Dennett's invocation of autonomy does not provide his argument with robustness and clarity. Autonomy is a criterion that for now is characterized by conceptual vagueness – thus by ontological ambiguity – but also by epistemological difficulties due to its correlation (at least under the most popular in the field of AI Ethics trend of the coherentist approach) with the concepts of cognition and personhood.

III. The Argument of Excessive Effectiveness

In his attempt to ground even more convincingly his argument in favor of attributing moral responsibilities to AI systems, Dennett supports the view that we recognize and admire the skill and the ability of the computer (i.e. Deep Blue) to win in chess and we congratulate its programmers for the achievement, but the victory belongs to the computer itself and not to its

programmers. *If the latter faced the world champion in chess (i.e. Kasparov), they would obviously lose within minutes*. At this point, Dennett seems to articulate an argument based on the excessive effectiveness of Deep Blue. This computer has indeed proven to be extremely effective in chess and of course it has been proven much more effective than its programmers (and most of the humans). According to Dennett, this effectiveness superiority of the computer over its human-programmers constitutes a sufficient reason for attributing the victory to the first and not to the latter. Could this specific argumentation by Dennett open the path for a successful answer to the responsibility gap question? Namely, could excessive effectiveness constitute a sound, sufficient and universally accepted criterion for the attribution of moral status – in this case, moral rights – to AI entities and even more generally to acting entities around us (humans, animals, machines etc.)? This possibility calls for an examination of the following question: Has, until today, existed any successful application of the excessive effectiveness criterion to humans, to animals, or to machines?

As seen at the beginning of the present article, Max Tegmark and Stuart Russell also refer to the criterion of excessive effectiveness, in this case in order to appeal for a limitation or even a prohibition of AI weapons. They do so by comparing AI weapons with weapons of mass destruction and stressing on their similarity in terms of their excessive effectiveness to kill. With the occasion of this appeal a question comes up: How come we don't attribute moral responsibility to nuclear or chemical weapons on the basis of their excessive effectiveness like Dennett suggests us to do in the case of Deep Blue? Both this supercomputer and the weapons of mass destruction present excessive effectiveness. Confining the discussion only to the level of effectiveness, we see that if Deep Blue is much more effective than its human-creators in winning a game of chess, the nuclear and the chemical weapons are similarly much more effective than their human-creators in killing. So, why hasn't until now any argument been articulated in favor of

a moral responsibility attribution to these weapons like it has been for Deep Blue? The above appeal by Tegmark and Russell equates the weapons of mass destruction with the AI systems (in this case AI weapons) on the basis of an analogous risk which in its turn implies an analogous excessive effectiveness. If the effectiveness superiority over the human-creators is analogous in the cases of Deep Blue and the weapons of mass destruction, why are we not ready to open a similar discussion for the attribution of moral status to the weapons of mass destruction like Dennett does with regard to the attribution of moral status to Deep Blue? It seems that the excessive effectiveness criterion is not applied in a consistent way to the machines.

At this point, one could answer that contrary to Deep Blue and most AI systems, weapons of mass destruction do not perform in an intelligent – or at least an intelligent-like way – and therefore there is not any issue of attributing moral responsibility to the latter.[56] However, we must point out that with such an argument: A) One has to define what does one mean with the term "intelligent" (or "intelligent-like") and thus one will again need to directly face the problem of defining the limits of the concept

[56] Of course, here we need to stress that nowadays most weapons of mass destruction are navigated and controlled by AI systems. Therefore, AI is now an integral part of weapons of mass destruction to the point that the latter can be classified as AI weapons. So the distinction between AI systems and weapons of mass destruction is no longer standing in practice. However, for the sake of the above discussion, let us assume that the weapons of mass destruction do not have AI features and belong to another class of machines. The very reference by Tegmark and Russell treats them in exactly this way in order to achieve the wanted comparison – and finally correlation – with AI weapons, not in the basis that the weapons of mass destruction employ AI but on the basis of an analogous risk. Moreover, if we prefer, we can confine our analysis and refer only to the older generation of weapons of mass destruction, for instance, to the first atomic bombs that were conventional bombs not having even the simplest system of guidance.

"intelligence" as well as "cognition."[57, 58] Thus, we seem to have here a behavioral criterion that not only fails to relieve us of the arduous problem of the delimitation of cognition, but it actually throws us back to it.[59] B) One diverts the whole analysis from the

[57] Already the distinction between an "intelligent" and an "intelligent-like" way again brings forth the Chinese Room Argument and the possibility of a simple imitation of intelligent behavior. In other words, it brings us against traditional questions of the Philosophy of Mind that we tried to bypass by introducing the excessive effectiveness criterion.

[58] For the differences between the terms "cognition" and "intelligence" see footnote 12.

[59] This can be easily seen form the fact that the above line of arguments and counter-arguments leads the advocate of Dennett's position to a circular argument and finally to a tautology. Specifically, Dennett's initial argument can be expressed with the following abstract statement: "We must attribute the moral responsibility of an action A to an entity E, if E performs A with excessive effectiveness." In order to face the counter-argument that entities to which we usually don't attribute moral responsibility also present an analogous excessive effectiveness, the above argument was rephrased as follows: "We must attribute the moral responsibility of an action A to an entity E, if E performs A with an intelligent (or intelligent-like) way [and with an excessive effectiveness]." However, in any case, even on the level of an everyday naïve psychology, the attribution of moral responsibility to an entity implies that this entity is intelligent (e.g. it is characterized by mental states of an intentional character). Thus, we have to ask: What more is added here (compared to the naïve psychology approach) with the criterion of excessive effectiveness? The above last version of Dennett's argument could be finally expressed as follows: "An entity E is intelligent if it acts in an intelligent (or intelligent-like) way." At this point we have to see that if we choose the version with the term "intelligent" we end up with a tautology (even if we distinguish between the terms "intelligence" and "cognition," the epistemological value of the above sentence can't surpass that of a tautology, since intelligence is a sub-set of cognition). On the other hand, if we choose the version with the term "intelligent-like," we avoid expressing a tautology, but we are confronted with the Chinese Room Argument. Note also that in the last abstracted version of Dennett's argument any reference to the excessive

criterion of the excessive effectiveness by introducing one more term – that of the intelligent (or the intelligent-like) way which after all seems to be finally more sufficient and decisive from the one that we initially tried to uphold (i.e. excessive effectiveness). In the end, if what distinguishes the AI systems from mass destruction weapons is the intelligent (or intelligent-like) way of their action, then what reasons do we have to refer to the criterion of the excessive effectiveness? The focus of our analysis has now been definitely shifted towards another criterion and any reference to the excessive effectiveness now seems redundant. In fact, if we carefully examine the way in which the above arguments were juxtaposed, the excessive effectiveness seems to be more of an element of similarity rather than of distinction between the AI systems and the conventional weapons of mass destruction. After all, it was this justified remark regarding the similar effectiveness that led to the adoption of the criterion of the intelligent (or intelligent-like) way in the first place.

Carrying on with our analysis regarding the application of the excessive effectiveness criterion, let us now, for the sake of the conversation, bypass the problem of defining what 'an intelligent way' is. Let us see that the inconsistent use of the excessive effectiveness criterion can be revealed even if our focus is confined only to the set of machines that we call "AI systems." Specifically, although an attribution of moral status is proposed for a super-computer like HAL 9000, this is not also the case with AI weapons. Excessive effectiveness is a feature that characterizes both the first and the latter. So, why do we start up a discussion of moral status attribution only for HAL? Which is the distinctive difference, the *differentia specifica* between them? Is it that HAL participates in a predominantly human activity as a member of a space expedition, while the weapons of mass destruction are not (killing is not an activity

---

effectiveness is completely missing, so it seems that the criterion for the attribution of moral status has been shifted from the concept of excessive effectiveness to the concepts of intelligence and cognition (see in the main text).

characteristic only of humans)? Nevertheless, we would answer that with this argument:

> A. One substitutes again the excessive effectiveness criterion with another criterion, namely the criterion of the field of human action.
> B. One accepts a delimitation of the term "cognition" that coincides exclusively with the delimitation of the term "human action." Therefore, one denies tacitly the attribution of the property of cognitive being to animals, an issue that is still debated and for which many of those who would like to deny the moral status of AI weapons answer positively supporting the possibility of animal moral rights.
> C. Therefore, we see that the inconsistency in the use of the excessive effectiveness criterion remains, even if we confine the discussion within the set of the AI systems.

Things are no better concerning the application of this criterion to humans. It is widely accepted and verified in practice that the human kind presents a remarkable diversity of skills which in any case are not distributed in a uniform way. People vary regarding their special abilities, their 'talents' as well as their weaknesses. However, we usually try not to have a similarly diverse attribution of moral status to them, although we don't always succeed in this task. Quite often people are considered morally responsible for their actions in fields in which they don't present an excessive effectiveness, whereas there are cases in which a mitigation of moral responsibility is attempted for actions in which people do present such effectiveness. A typical example of this is the case of people who have suffered damage in brain areas related to the triggering and the control of the so-called pro-social emotions. Usually, such individuals end up becoming serial killers and mass murderers since they combine a high level capacity to plan a murder – therefore an excessive

effectiveness of executing it – with a lack of moral restraints.[60] These people are frequently addressed as mental patients, thus as individuals having a reduced autonomy due to their mental illness. Eventually, we see that not only the attribution of moral status is not symmetrical to the attribution of excessive effectiveness (namely, the delimitation of moral status is not univocally related with the delimitation of any excessive effectiveness), but also it is rather based on other criteria like the criterion of autonomy (which brings us back to the previous discussion regarding the problems of the autonomous agency). So, if in the case of humans we avoid linking effectiveness to the attribution of moral status, why should we do so in the case of the machines?

In fact, in the case of machines – as well as animals – excessive effectiveness has been used sometimes as an indication of a non-cognitive, 'automatic' nature and, therefore, of a nature inferior to that of humans, and some other times as a proof of these entities' moral or cognitive superiority over humans. Respectively, René Descartes was the first who supported the view that an exhibition of an excessive effectiveness in certain actions on behalf of an entity is a safe indication – and thus a sound behavioral criterion – of the automatic nature of this entity.[61] For Descartes, the "automata" (animals and machines) function not based on rational mind but completely based on the specificities in the structure of their body. Therefore, they present an excessive effectiveness in certain areas of action due to the specific structure of what is nowadays called "hardware.""[62] This is a position that has also been adopted by

[60] Clare Allely et. al., "Neurodevelopmental and Psychosocial Risk Factors in Serial Killers and Mass Murderers," *Aggression and Violent Behavior* 19 (2014): 288-301.

[61] René Descartes, "Letter to the Marquess of Newcastle," in *The Philosophical Writings of Descartes: Volume 3*, trans. John Cottingham, Robert Stoothoff, and Murdoch Dugald, 302-304 (New York: Cambridge University Press, 2004), 304.

[62] René Descartes, "Discourse of the Method of Rightly Conducting One's

some modern researchers in the field of AI, in their effort to come up with a solid behavioral criterion for a safe judgment regarding cognition in the Turing Test context.[63] Here, the 'hardware' specificities have been substituted by what we could refer to as 'software' specificities, namely the specialization of the machine's program. A completely opposite use of the excessive effectiveness criterion has been made by a philosopher almost contemporary to Descartes, specifically by Michel de Montaigne in his support of the doctrine of *Theriophily*. According to Montaigne, the fact that the animals exhibit a remarkable and quite higher than humans effectiveness in certain actions, constitutes a sufficient proof of the animals' superiority over the humans and finally of the animals' right to have a moral status fully respected by the humans. [64]

Therefore, it seems that the philosophical analysis has not yet settled down with regard to the relation between excessive effectiveness and the attribution of cognitive abilities or finally the attribution of an ontological status that would be also related to a moral personhood attribution. On the contrary, up to now, the discussion is characterized from completely opposite ways of using the excessive effectiveness criterion. To the extend that moral responsibility is related to cognition, we could say that Dennett's view that excessive effectiveness constitutes a sufficient reason for attributing the win to Deep Blue and moral responsibility to HAL is diametrically opposite to the view of Descartes and to

---

Reason and of Seeking Truth in the Sciences," in *The Philosophical Writings of Descartes: Volume 1*, trans. by John Cottingham, Robert Stoothoff, and Murdoch Dugald, 11-151 (New York: Cambridge University Press, 2004), 141.

[63] Donald Michie, "Turing's Test and Conscious Thought," in *Machines and Thought. The Legacy of Alan Turing*, vol. 1, eds. Peter Millican, and Andy Clark, 27-51 (Oxford, New York: Oxford University Press, 2002).

[64] Michel de Montaigne, "An Apology for Raymond Sebond," in *Michel de Montaigne: The Complete Essays*, trans. Michael A. Screech, 489-683 (London, New York: Penguin Books, 2003).

the view of those modern AI researchers who try to ground the Turing Test on a correlation between excessive effectiveness and the total absence of cognitive abilities. According to the approach made by Descartes and all those who treat excessive effectiveness as an indication of an entity's 'automatic nature,' Deep Blue should never have the right to be attributed with the victory in a chess game. On the other hand, according to Dennett, excessive effectiveness constitutes a sufficient reason for attributing the victory to Deep Blue and moral responsibility to HAL. One could possibly support the view that Dennett's position seems more compatible with that by Montaigne. However, opposite to what Montaigne supports regarding the animals, Dennett does not use the excessive effectiveness criterion to support a superiority of Deep Blue and HAL over the humans. He rather argues for an equivalence between these super-computers and the humans. Under a rough description, we could say that until now we have three different uses of the excessive effectiveness criterion on behalf of the philosophers:

> 1) Descartes' use of excessive effectiveness as an evidence of other beings' (animals and machines) inferiority compared to the humans
> 2) Montaigne's use of excessive effectiveness as an evidence of the superiority of other beings (animals) over the humans
> 3) Dennett's use of excessive effectiveness as an evidence of equity between other beings (machines) and the humans.

Which of these uses is the correct one? For now, the only safe claim we can make is that the excessive effectiveness criterion is not being used with a constant, univocal and thus consistent way for the ontological comparison of humans with other entities. This inconsistency leads logically to a respective non-univocal and non-consistent use of the excessive effectiveness criterion for the attribution of moral status to these entities.

IV. Conclusions

In this article we set out to examine the possibility of attributing moral personhood to AI systems. Our analysis focused exclusively on AI weapons, and this due to the severity of the consequences their use may result in; this severity is proportional to the sharpness and the intensity of the ethical issues that this use raises. In other words, we referred specifically to the case of AI weapons because it constitutes the most pressing of all the contexts in which the philosophers and the AI researchers find themselves confronted with the problem of moral status attribution to AI entities. Nevertheless, we think that the arguments and the conclusions that we have presented in the above text have a rather general validity – namely, they can be applied to any machine characterized as an "AI system" – since they are not based on aspects that are specific only to AI weapons. On the contrary, they can also apply to any other machine. Moreover, we have chosen to base our analysis on a scepticist response to the arguments supported by Daniel Dennett in his text *When HAL Kills, Who's to Blame? Computer Ethics*, which is considered to be a milestone of contemporary philosophical analysis in favor of the attribution of moral status to the machines. After all, the reference to HAL and to the murder that this system commits in the famous film *2001: A Space Odyssey* makes Dennett's analysis relevant to the ethical issues raised regarding AI weapons.

Specifically, we supported that Dennett's analysis is mainly based on three basic arguments: The analogy between the programmer – machine and the coach – athlete relations, the machine autonomy argument and the argument of excessive effectiveness (the last two as sufficient criteria for the attribution of moral status to an AI system).

With regard to the first argument, we showed that the support of an analogy in the programmer – machine and trainer – athlete relations as an argument in favor of the machine moral status is already a logical fallacy. First, because it constitutes a

circular argument given that it assumes the conclusion or in other words it presupposes what is to be proven, namely the ontological equivalence between the human and the machine. Second, because in the case that one considers this analogy as a functionalist one, one is confronted with the logical problems inherent in the foundations of *Machine Functionalism* as well. In addition, this view of a functionalist analogy faces also the ontological problems of *Machine Functionalism*.

Concerning the argument of machine autonomy, we initially observed that Dennett's programmers-environment parallelism, thus challenging the unconditional, absolute human autonomy, is not totally groundless. However, we showed that Dennett's appeal to the criterion of autonomy faces the problem of conceptual vagueness which is raised by a plurality of autonomy definitions. Moreover, according to the most popular – at least in the field of AI Ethics – autonomy account, namely, according to the internalist view, one is inevitably confronted with the Other Minds Problem and also with certain well-known and traditional problems regarding the property of personhood like the 'persistence' and the 'characterization problem.' Moreover, we showed that the appeal to the criterion of autonomy pits one's analysis against the ambiguity of the delimitation of the will. It is this ambiguity that leads to a non-symmetry in the relation between the attribution of autonomy and the attribution of moral personhood, namely to the inconsistent use of the criterion of autonomy.

Until now, the use of the excessive effectiveness criterion has been proven to be similarly inconsistent, both regarding the human – machine and the human – animal distinction. After all, the application of this criterion seems to take place with an arbitrarily selective way not only regarding AI weapons, but also other machines like the weapons of mass destruction.

We think that our counter-arguments presented above respond to a large part of the contemporary discussion regarding the literal attribution of moral status – and thus of the property of moral personhood – to AI systems and especially

to AI war machines. We support the view that for now and until the ontological and epistemological issues related to human cognition and artificial intelligence are resolved in a satisfying way, any discussion towards this direction can be made only with a metaphorical use of the words "autonomy," "personhood," and "moral status." Besides, we should not overlook the fact that the exhibition of morally relevant actions (actions that can be morally evaluated) on behalf of the machines is something completely different from the attribution of moral responsibility to the machines for their actions.[65]

Dennett's view is in favor of the attribution of moral status to the AI systems. Our present analysis did not aim at supporting the opposite view, namely a view against the attribution of moral status to these systems. It rather aimed at demonstrating the fact that based on the dominant current argumentation in the field of AI Ethics, the question regarding the attribution of moral status to the machines can only remain *undecidable*. Thus, we are once again confronted with a contradiction well known to anyone working in the field of Applied Ethics, specifically with the contradiction between the demand for clear and sound moral decision criteria and the interminable nature of a philosophical contemplation that tries to be consistent.

References

Allely, Clare, Helen Minnis, Lucy Thompson, Philip Wilson, and Christopher Gilberg. "Neurodevelopmental and Psychosocial Risk Factors in Serial Killers and Mass Murderers." *Aggression and Violent Behavior* 19 (2014): 288-301.

Anderson, Michael, and Suzan L. Anderson. "Machine Ethics: Creating an Ethical Intelligent Agent." *AI Magazine* 28, no. 4 (2007): 15-26.

Avramides, Anita. *Other Minds*. London, New York: Routledge, 2001.

[65] Michael Anderson, and Susan L. Anderson, "Machine Ethics: Creating an Ethical Intelligent Agent," *AI Magazine* 28, no. 4 (2007): 19.

Benacerraf, Paul. "God, the Devil and Gödel." *The Monist* 51 (1967): 9-32.

Block, Ned. "Are Absent Qualia Impossible?" *Philosophical Review* 89 (1980): 257-274.

Block, Ned. "Troubles with Functionalism." In *Readings in Philosophy of Psychology*, vol. 1, edited by Ned Block, 268-305. Cambridge: Harvard University Press, 1980.

Bratman, Michael. "Practical Reasoning and Weakness of the Will." *Noûs* 13, no. 2 (1979): 153-171.

Bratman, Michael. *Structures of Agency: Essays*. Oxford: Oxford University Press, 2007.

Buss, Sarah, and Andrea Westlund. "Personal Autonomy." *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). Edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/.

Calverley, David. "Toward a Method for Determining the Legal Status of a Conscious Machine." In *Proceedings of the AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment*, edited by R. Chrisley, R. Clowes, and S. Torrance, 75-84. Hatfield: University of Hertfordshire, 2005.

Casti, John L., and Werner De Pauli. *Gödel, a Life of Logic*. Cambridge: Basic Books, 2000.

Christman, John, "Autonomy in Moral and Political Philosophy." *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). Edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/.

Churchland, Paul M. *Matter and Consciousness*. Cambridge, Massachusetts: MIT Press, 1988.

Clarke, Arthur C. *2001: A Space Odyssey*. New York: New American Library, 1968.

De Landa, Manuel. *War in the Age of Intelligent Machines*. New York: Swerve Editions, 1991.

De Quincey, Christian. "Switched-on Consciousness: Clarifying What it Means." *Journal of Consciousness Studies* 13, no. 4 (2006): 6-10.

Dennett, Daniel. "Quining Qualia." In *Consciousness in Contemporary Science*, edited by Antony J. Marcel, and E. Bisiach, 42-77. New York: Oxford University Press, 1988.

Dennett, Daniel. "When Hal Kills, Who's to Blame? Computer Ethics." In *Hal's Legacy: 2001's Computer as Dream and Reality*, edited by David G. Stork, 351-365. Cambridge, MA: MIT Press, 1997.

Descartes, René. "Discourse of the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences." In *The Philosophical Writings of Descartes: Volume 1*, translated by John Cottingham, Robert Stoothoff, and Murdoch Dugald, 111-151. New York: Cambridge University Press, 2004.

Descartes, René. "Letter to the Marquess of Newcastle." In *The Philosophical Writings of Descartes: Volume 3*, translated by John Cottingham, Robert Stoothoff, and Murdoch Dugald, 302-304. New York: Cambridge University Press, 2004.

Dreyfus, Hubert L. *What Computers Still Can't Do: A Critique of Artificial Reason*. New York: MIT Press, 1992.

Erion, Gerald J. "The Cartesian test for Automatism." *Minds and Machines* 11, no. 2 (2001): 29-39.

Floridi, Luciano, and J.W. Sanders. "On the Morality of Artificial Agents." *Minds and Machines* 14 (2004): 349-379.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." In *The Importance of What We Care About*, edited by Harry Frankfurt, 11-25. Cambridge: Cambridge University Press, 1988a.

Frankfurt, Harry. "On Caring." In *Necessity, Volition and Love*, edited by Harry Frankfurt, 155-180. Cambridge: Cambridge University Press, 1999.

Frankfurt, Harry. "The Importance of What We Care About." In *The Importance of What We Care About*, edited by Harry Frankfurt, 80-94. Cambridge: Cambridge University Press, 1988b.

Frankish, Keith. "Illusionism as a Theory of Consciousness." *Journal of Consciousness Studies* 23, nos. 11-12 (2016): 11-39.

Frankish, Keith. *Illusionism: As a Theory of Consciousness*. Exeter: Imprint Academic Publishing, 2017.

Gounaris, Alkis. "Human Cognition and Artificial Intelligence: Searching for the Fundamental Differences of Meaning in the Boundaries of Metaphysics." Accessed January 14, 2019. https://alkisgounaris.gr/gr/research/human-cognition-artificial-intelligence/.

Gunderson, Keith. "Descartes, La Mettrie, Language, and Machines." *Philosophy* 39, no. 149 (1964): 193-222.

Gunkel, David. *The Machine Question: Critical Perspectives on AI, Robots and Ethics*. Cambridge, Massachusetts: MIT Press, 2012.

Hajdin, Mane. *The Boundaries of Moral Discourse*. Chicago: Loyola University Press, 1994.

Hardcastle, Valerie. *The Myth of Pain*. Cambridge, MA. Massachusetts: MIT Press, 1999.

Haugeland, John. *Artificial Intelligence: The Very Idea*. Cambridge, Massachusetts: MIT Press. Chicago, 1985.

Hoffmann, Christian, and Benjamin Hahn. "Decentered Ethics in the Machine Era and Guidance for AI Regulation." *AI & Society* 35, no. 3 (2020): 635-644.

Jaworska, Agnieszka. "Caring and Full Moral Standing." *Ethics* 117, no. 3 (2007): 460-497.

Jaworska, Agnieszka. "Caring and Internality." *Philosophy and Phenomenological Research* 74, no. 3 (2007): 529-568.

Jaworska, Agnieszka. "Caring, Minimal Autonomy, and the Limits of Liberalism." In *Naturalized Bioethics: Toward Responsible Knowing and Practice*, edited by Hilde Lindemann, Marian Verkerk, and Margaret Walker, 80-105. Cambridge: Cambridge University Press, 2008.

Kim, Jaegwon. *Philosophy of Mind*. USA: Westview Press, 1998.

Kim, Jaegwon. *Supervenience and the Mind*. Cambridge: Cambridge University Press, 1993.

Lang, Fabienne. "AI Flawlessly Beats US Air Force F-16 Pilot in Simulated Dogfight." *Interesting Engineering*. Accessed August 21, 2020. https://interestingengineering.com/ai-flawlessly-beats-us-air-force-f-16-pilot-in-simulated-dogfight.

Levy, David. *Intimate Relationships with Artificial Partners*. Ph.D. Diss., Maastricht University, 2007.

Levy, David. "The Ethical Treatment of Artificially Conscious Robots." *International Journal of Social Robotics* 1, no. 3 (2009): 209-216.

Lucas, John R. "Minds, Machines and Gödel." *Philosophy* XXX-VI (1961): 112-127.

Matthias, Andreas. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology* 6, no. 3 (2004): 175-183.

Michie, Donald. "Turing's Test and Conscious Thought." In *Machines and Thought. The Legacy of Alan Turing*, vol. 1, edited by Peter Millican, and Andy Clark, 27-51. Oxford, New York: Oxford University Press, 2002.

Ministry of Defense. "Joint Doctrine Note 2/11, The UK Approach To Unmanned Aircraft Systems." Accessed July 20, 2020. https://www.law.upenn.edu/live/files/3890-uk-ministry-of-defense-joint-doctrine-note-211-the.

Monroe, Andrew E., Kyle D. Dillon, and Bertram F. Malle. "Bringing Free Will Down to Earth: People's Psychological Concept of Free Will and its Role in Moral Judgment." *Consciousness and Cognition* 27 (2014): 100-108.

Montaigne, Michel de. "An Apology for Raymond Sebond." In *Michel de Montaigne: The Complete Essays*, translated by Michael A. Screech. Penguin Books, 2003.

Moore, Edward F. "Artificial Living Plants." *Scientific American* 195, no. 4 (1956): 118-126.

Müller, Vincent C. "Autonomous Cognitive Systems in Real-World Environments: Less Control, More Flexibility and Better Interaction." *Cognitive Computation* 4, no. 3 (2012): 212-215.

Müller, Vincent C. "Ethics of Artificial Intelligence and Robotics." *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). Edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2020/entries/ethics-ai/.

Olson, Eric T. "Personal Identity." *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Edition by Edward N. Zalta. https://plato.stanford.edu/archives/fall2019/entries/identity-personal/.

Owen, Jonathan, and Richard Osley. "Bill of Rights for Abused Robots: Experts Draw up an Ethical Charter to Prevent Humans Exploiting Machines." *The Independent*. Last modified April 1, 2007. https://www.independent.co.uk/news/science/bill-of-rights-for-abused-robots-5332596.html.

Pagallo, Ugo. "Robots of Just War: A Legal Perspective." *Philosophy & Technology* 24, no. 3 (2011): 307-323.

Pinker, Steven. "Can a Computer Ever be Conscious?" *US News & World Report* 123, no. 7 (1997). Accessed July 28, 2020. https://stevenpinker.com/files/pinker/files/computer.pdf.

Putnam, Hilary. *Representation and Reality*. Cambridge: MIT Press, 1992.

Regan, Tom. *The Case for Animal Rights*. Berkeley & Los Angeles: University of California Press, 1983.

Rey, Georges. "A Question About Consciousness." In *Perspectives on Mind*, edited by Herbert R. Otto, and James A. Tuedio, 5-24. Dordrecht: D. Reidel Publishing, 1988.

Rey, Georges. "A Reason for Doubting the Existence of Consciousness." In *Consciousness and Self-Regulation*, vol. 3, edited by Richard J. Davidson, Gary E. Schwartz, and David Shapiro, 1-39. New York: Plenum, 1983.

Rucker, Rudy. *Infinity and the Mind: The Science and Philosophy of the Infinite*. Princeton, N.J.: Princeton University Press, 1982.

Russell, Stuart, Max Tegmark, et al. "Autonomous Weapons: An Open Letter from AI & Robotics Researchers." *Future of Life Institute*. Accessed July 25, 2020. https://futureoflife.org/open-letter-autonomous-weapons/.

Savova, Viginia, and Leonid Peshkin. "Is the Turing Test Good Enough? The Fallacy of Resource-Unbounded Intelligence." *International Joint Conferences on Artificial Intelligence Organization: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, IJCAI-07: 545-550. Accessed August 29, 2020. https://www.ijcai.org/Proceedings/07/Papers/086.pdf.

Sealre, John. *Minds, Brains, and Science*. Cambridge Massachusetts: Harvard University Press, 1984.

Searle, John. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 414-457.

Searle, John. *The Mystery of Consciousness*. New York: The New York Review of Books, 1997.

Shoemaker, David. "Caring, Identification, and Agency." *Ethics* 114, no. 1 (2003): 88-118.

Shoemaker, Sydney. *Identity, Cause, and Mind*. Cambridge: Cambridge University Press, 1984.

Singer, Peter. *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York: New York Review of Books, 1975.

Singer, Peter. *Practical Ethics*. Cambridge: Cambridge University Press, 1993.

Singer, Peter. *Wired for War*. New York: Penguin Press, 2009.

Sloman, Aaron. "A Systematic Approach to Consciousness (How to Avoid Talking Nonsense?)." Accessed July 28, 2020. http://www.cs.bham.ac.uk/research/projects/cogaff/misc/consciousness.rsa.text.

Solum, Lawrence. "Legal Personhood for Artificial Intelligences." *North Carolina Law Review* 70, no. 4 (1992): 1231-1287.

Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy* 24, no. 1 (2007): 62-77.

Tegmark, Max. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf, 2017.

Torrance, Steve. "Could We, Should We, Create Conscious Robots?" *Journal of Health Social and Environmental Issues* 4, no. 2 (2004): 43-46.

Turing, Alan. "Computing, Machinery and Intelligence." *Mind* LIX (1950): 433-660.

Turing, Alan. "On Computable Numbers with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42, Series 2 (1937): 230-265.

Turing, Alan. "On Computable Numbers with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 43, Series 2 (1938): 544-546.

Vincent, James. "Former Go Champion Beaten by DeepMind Retires after Declaring AI Invincible." *The Verge*. Accessed August 1, 2020. https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat.

Von Neumann, John. *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press, 1966.

Wallach, Wendel, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press, 2009.

Walsh, Toby. *It's Alive: Artificial Intelligence from the Logic Piano to Killer Robots*. Hamburg: Edition Körber, 2017.

Watson, Gary. "Free Agency." *Journal of Philosophy* 72, no. 8 (1975): 205-220.

Weller, Chris. "Meet the First-ever Robot Citizen – A Humanoid Named Sophia that once Said it Would 'Destroy Humans.'" *Business Insider*. Accessed July 30, 2020. https://www.businessinsider.com/meet-the-first-robot-citizen-sophia-animatronic-humanoid-2017-10?r=UK.

Wilkes, Kathleen. "Yishi, Duh, Um and Consciousness." In *Consciousness in Contemporary Science*, edited by Antony Marcel, and Edoardo Bisiach, 16-41. Oxford: Oxford University Press, 1988.

Sartre and Personhood

Marica Rajković
University of Novi Sad
E-mail address: marica.rajkovic@ff.uns.ac.rs
ORCID iD: https://orcid.org/0000-0003-0901-4102

Abstract: The author's intention is to explain Sartre's concept of personhood through three important contexts within his philosophy. First one is Sartre's philosophy of existence, the second one is Sartre's idea of engagement, and the third one is Sartre's concept of humanism. Some of the most important points for understanding Sartre's concept of personhood and man in general are ideas of choice and possibility, that place human existence on a whole new ontological foundation, different to both natural and artificial necessities and objects. Human being is a being of possibility that carries constant responsibility for their actions. The idea of personhood in Sartre's philosophy is not founded on psychological or anthropological theories, but sets up as an ontological, political and practical concept.
Keywords: engagement; existentialism; Hegel; humanism; personhood; possibility; responsibility; Sartre.

Il y a toujours une possibilité pour le lâche de ne plus être lâche, et pour le héros de cesser d'être un héros.[1]

I. Personhood and existentialism

The idea of personhood in Jean Paul Sartre's philosophy contains numerous differences to the ideas of personhood in other philosophies of existence. The main reason for these differences lies in Sartre's

[1] Jean-Paul Sartre, "L'existentialisme est un humanisme." See Jean-Paul Sartre, "Existentialism is a Humanism," 14, http://web2.slc.qc.ca/sbeaudoin/z-Fall_19/K_S_F19/Sartre.pdf: "[...] and that there is always a possibility for the coward to give up cowardice and for the hero to stop being a hero." Also quoted in *Existentialism from Dostoyevsky to Sartre*, ed. Walter Kaufman (New York: Meridian Books, 1989).

doi: https://doi.org/10.12681/aprlp.49.926

mutual opposing of essence and existence, through which his understanding of personhood will be conceptualized. The idea that, when it comes to human being, existence precedes the essence – shows not only Sartre's unique approach to the idea of personhood, but also his daring to break down the fundamental structure of Western philosophical thought. In *Being and Nothingness*, Sartre argues that *the dualism of being and phenomenon no longer has a legitimate status in philosophy, nor does the dualism of phenomenon and essence*.[2] This, however, does not mean that the abolition of all dualisms has been achieved, because there remains one dualism that cannot be abolished: dualism between *finite and infinite*.

The finite phenomenon requires to be transcended towards infinity,[3] but it is no more the matter of usual philosophical hiatus but requires the whole new status and relation to infinity. Finite and infinite are not divided as opposites, but different as pervaded moments of the same entirety that synthesizes them. Infinite shows itself as *the infinite in final form,* and phenomenon represents only the meaning of itself and does not refer to something "behind" or "beyond," as it was represented in Kant's philosophy, for example.

Sartre sharply opposes the idea that essence is contained in the subject itself, thus actually contradicting the entire metaphysical tradition starting with Aristotle. The essence is all that in a human being that can be labeled with words "it is," therefore, the essence is not contained in the object but rather represents its meaning and reason. Likewise, being is not merely one of the qualities of object, nor does the existence of objects in any way participate in being. The object simply *is*, and it is the only possible determinant of its existence, claims Sartre.

In Hegel's philosophy, the idea of existence represents first of three categories of phenomena, and to step into existence means *to change and remain the same*. He explains that the

[2] Ž.-P. Sartr, *Biće i ništavilo: ogled iz fenomenološke ontologije* (Beograd: Nolit, 1983), 7-9.

[3] Ibid., 9.

phenomenon (as phenomenon)[4] does not represent *something else* but shows its own reality and meaning.[5] Sartre stands at a similar viewpoint, adding that every conscious existence exists as a *consciousness of existence*,[6] i.e., that the object cannot be separated from the consciousness of the object. Consciousness does not appear *before* its susceptibility to a given object, nor does the object *precede* the consciousness: there is only a being that is one and undivided *as an existence*. Existence comprises the essence because consciousness is not possible *before being* but represents the unity of phenomenon and existence. The only way to make the phenomenon dependent on consciousness is to place the object as *opposed* to consciousness – not by its presence but by its absence, therefore by nothingness.

Sartre claims that the *consciousness is a being* whose *existence sets the essence*, but it is also the consciousness of a being whose *essence implies the existence*. In addition to the aforementioned closeness to Hegel's understanding of existence, Sartre demonstrates a willingness to affirm another thesis made by the philosophy of German idealism – that a human being does not belong to the *domain of being,* but to the *domain of "ought to,"* i.e. to the *domain of possibility*. The subject is not yet a personality. Personality is created through the realization of possibilities of the subject, and this realization is made by making choices. Personhood is, therefore, a purposeful realization of possibilities conducted through concrete actions made by human being.

Being-for-itself is a being of possibility, while *being-in-itself* already is all that it is and can be. Human being is, therefore, a being of possibility, of "not yet achieved," of "pursued" and insufficient. "Man is condemned to be free,"[7] because once thrown into the world, he is responsible for everything he does, claims Sartre, adding that those possibilities are nothing else than freedom – the meaning and reason of human existence.

---

[4] G. V. F. Hegel, *Fenomenologija duha* (Beograd: BIGZ, 1986), 88.

[5] H.-G. Gadamer, *Hegelova dijalektika* (Beograd: Plato, 2003), 45.

[6] Sartr, *Biće i ništavilo,* 14.

[7] Jean-Paul Sartre, *Existentialism is a Humanism*, ed. John Kulka, trans. Carol Macomber (New Haven: Yale University Press, 2007), 29.

Although Sartre disagrees with Hegel's "logical and ontological equation"[8] of being and nothingness, he agrees with Hegel that freedom must have concrete and historical content. He believes that Hegel is closer to existentialism than it firstly appears,[9] because of his claim that the existences are associated with each other in the history that are creating and which, as a concrete universality, is what judges and what transcends them.

The subject of existentialism is a human individual in the social field, among other individuals, a personality that is created by division of labor and exploitation, but who fights against alienation and patiently conquers this field piece by piece.[10] Sartre does not claim that all manifestations of personality are equally important: hierarchy should exist, personality needs to be viewed from different angles – however, always as a whole. Man continually streams beyond his own position – transcending his own situation, in order to objectify himself. This transcendence is a project – not in the form of a will, because will is abstract, but it *can* have a willing *character*. In Sartre's opinion, existence is a perpetual disturbance of equilibrium, which separates from itself and is thrown through the field of possibility, determining its choices and freedom. Existence is not an existent substance, but it is not irrational either. It can be explained by a dialectical method, which is transcending but containing at the same time, and is explained as a form of choice, because personality is conditioned precisely by its process or act. Such existentialism is completely opposed by positivism, which apparently rejects the a priori, but actually decides in advance following its own scheme.

II. Personhood and engagement

In the wake of Hegel's thought, Sartre explains that the idea of freedom must be historically and socially contextualized, because otherwise it is merely an empty concept. Being condemned to

[8] Sartr, *Biće i ništavilo,* 39.
[9] Ž.-P. Sartr, *Egzistencijalizam i marksizam* (Beograd: Nolit, 1970), 14.
[10] Ž.-P. Sartr, *Kritika dijalektičkog uma* (Beograd: Nolit, 1983), 80.

be free, human being is actually being responsible for each and every action and choice. If freedom had no concrete realization, being-for-itself would be reduced to being-in-itself, and Sartre notes that this is exactly what happened to Heidegger's concept of existence, since it didn't get its historical content. Heidegger does confront with the notion of possibility through the idea of *a project*, but that idea has not been historically fulfilled through concrete actions and therefore remains incomplete.

Sartre wonders: if a man is *zoon politikon*, how is it possible that his fate is not resolved once and for all by gaining political freedom?[11] By answering this question we can also answer the question of the meaning of human existence and purpose of personhood. The answer is close to Kant's practical ideas: our goal is not something that is given, but rather something that is assigned! Our roles are always in the future,[12] each of us is assigned the tasks to accomplish and the possibilities to actualize. By reaching it, our aim is not achieved once and for all, on the contrary, reaching it and pursuing it is a process that doesn't terminate as long as we live. And if we thought there were areas where we could avoid that pursuit – e.g. culture or art, Sartre will show us that it is not so: art (above all literature) also involves constant practical and political choices.

In "engaged literature" Sartre explains that responsibility cannot be avoided by ignoring it. Ignoring the choice is also a choice, and not acting expresses the action and attitude as strongly as any other act. During the challenging political and historical times, silence can be very loud, claims Sartre, and abstract freedom can never compensate for lack of concrete acts. The most responsible among the artists are prose writers: engaged writer know *that words and acts are equal*.[13] The writer must plead the times and circumstances in which he lives and creates – if he does not do so, Sartre claims that he did express his attitude anyway.

---

[11] Ž.-P. Sartr, "Angažovana književnost," in Ž.-P. Sartr, *Šta je književnost* (Beograd: Nolit, 1984), 13.

[12] Sartr, *Kritika dijalektičkog uma*, 67.

[13] Sartr, *Šta je književnost*, 28.

Concrete action is a necessary part of the realization of freedom, just as the relation with others is a necessary segment of each subject. Subject cannot be accomplished without *the other*, without intersubjectivity. With this argument, Sartre approaches both Hegel's and Husserl's philosophy, pointing out that only through *difference* can one see the real state of one phenomenon. Personhood could not be achieved without *others*, literature is empty without engagement and freedom is nothing without *action*. Although famous for his sentence "hell is other people,"[14] Sartre clearly demonstrates that without others, man cannot exist.

Others are the condition of my existence, claims Sartre, in relation to them the whole world is being established, and that world is called intersubjectivity.

Sartre understands praxis as an inwardness of objectivity, because subjectivity is actually also the part of the objective process. Marxism overlooked the idea that there is no history without actual, individual and living people, and by its progressive method, it predetermined what was yet to happen. Unlike Marxist (progressive) – existentialistic method is heuristic: both progressive and regressive. Existentialism does not regard deviations as coincidences but sees them as concrete realities. The result of an existentialistic pursuit will not be a general personality trait but a personhood in its full objectification.

Marxism ignores the concrete determinants of human life and through historical totalization it retains only an abstract framework and "universal patterns." Sartre cites the example of Napoleon Bonaparte, to point out the absurdity of such abstractions and generalizations: Bonaparte was not some kind of "a man in general," determined by his role in the historical moment, but on the contrary – a concrete personality that made a certain historical situation possible! Sartre points out the paradox that Marxism actually stops the dialectic stream, totalizes human activities within a homogeneous flow, and does

[14] Ž.-P. Sartr, "Iza zatvorenih vrata," in Ž.-P. Sartr, *Drame*: *Izabrana dela*, vol. 5 (Nolit: Beograd, 1981).

not really move beyond the Cartesian understanding of time. And dialectics cannot be sustained unless the time is understood as dialectical. This means that neither man nor his actions are *in time,* but that the time is *provided by man's action*.

Atheist-based existentialism is further exacerbated by the fact that man no longer shares responsibility with any other being or force, but becomes abandoned and all responsibility falls on him. This is the point of Sartre's thesis that man is *condemned to be free*: since there is no general morality, man is free and obliged to choose according to his own conscience.

If a man has made himself something he does not want to be, e.g. a coward, it is not only that he cannot claim responsibility *anywhere outside himself,* but with every future failure to correct it, his responsibility increases. Man's obligations and responsibilities never stop, and his fate is in his hands, concludes Sartre.

III. Personhood and humanism

Existentialism should affirm the uniqueness of historical events, refusing to understand them as a mere sequence of a priori moments. Consequently, there is a need for a dialectic that will be able to follow the historical flow in its truthfulness, without insisting that the contradictions should be rationally resolved and neutralized. Sartre points out that the contradictions are the real source and basis that make ideas. The contradiction is what brings tension to every process, but also what gives the frame to the idea and event that is being clarified. Also, contradictions in ideas allow ambiguity, which determine the historical event itself and make it possible and concrete. What rationalism dismisses as coincidence – is in fact what makes all human life, concludes Sartre.

What is most frightening about existentialism, explains Sartre, is the fact that it leaves one not only with the *possibility* but also with *the necessity* of *choice*. Also, existentialism provides us with clarification of the idea of man – a being in which *existence* precedes *essence* and which exists before it can be defined by any term. At first, man *exists*, then he meets himself in the world and

finally, he defines himself: if man in existential terms cannot be defined, it is because he is *nothing at first*. Only later he will start *to become*, and he will *be* what he is willing *to do* of himself. This means that there is no human nature: man is not only what he *sees* in himself, but also what he *wants* from himself. This is why man is nothing but *what he does of himself*.

This is the first principle of existentialism and is called subjectivity, explains Sartre. This allows man to have dignity that is not given to inanimate objects, since man primarily exists, ie. he throws/projects himself towards the future. Man is a *project* that lives in a subjective way. The project is not a mere volition, but a human responsibility for what he is. The first step of existentialism is to put every person in the possession of what he is and to place full responsibility on his existence. By being responsible for oneself means that one is responsible for humankind in general. Every act by which an individual creates the person he wants to be, at the same time creates an image of what man in general should be, because with every choice, man establishes values that apply not only to him but to all people.

Man is at all times forced to perform acts of choice, because there is no one else who could do it for him. Certainly, this must result in the feeling that all humanity has directed its gaze to every act that a single man makes, because all humanity is treated by that act. Any assumption of responsibility necessarily carries with it the anxiety of that act, Sartre points out. In doing so, however, each act is confirmed as a possibility that has gained its value by making a choice.

Existentialism must return to the essence of humanism, whose traditional values it rejected, claims Sartre: wanting your freedom also means wanting the freedom of others, so this basis even provides the possibility of reconciliation of conscience. It should be acknowledged, however, that earlier humanism was absurd, because it attributed the merits of extraordinary individuals to *people in general*, and the point is on the contrary: to present *responsibility* as something that can transcend from an individual to such a generality.

Humanism has two very different meanings: in one sense it is a conception that sets man as its goal and highest value, while in another sense it means that man makes himself by continually transcending and finding himself. Sartre understands this second sense as the essence of *existentialist humanism,* which represents not only philosophical discourse but also an activism to live by.

IV. Concluding remarks

Sartre explains that the idea of freedom must be historically and socially contextualized, because otherwise it is merely an empty concept.

Our roles are always in the future, each of us is assigned the tasks to accomplish and the possibilities to actualize. By reaching it, our aim is not achieved once and for all, on the contrary, reaching it and pursuing it is a process that doesn't terminate as long as we live. Concrete action is a necessary part of the realization of freedom, just as the relation with others is a necessary segment of each subject. Subject cannot be accomplished without the other, without intersubjectivity.

Not only that – authentic philosophy should never avoid those contradictions and coincidences, on the contrary: it is the duty and the main purpose of philosophy to clarify the areas that other disciplines avoid or consider unexplainable.

Personhood could not be achieved without others, philosophy should embrace contradictions, literature is empty without engagement and freedom is nothing without action. Sartre demonstrates that the most feared path is usually the one we should choose, because by confronting the most unpleasant ideas often opens the doors to wider picture and the solutions of the most difficult tasks. Although famous for his sentence "hell is other people," Sartre clearly demonstrates that without those other people, man cannot exist. It is not the first time that this verse from Friedrich Hölderlin's poem is being quoted in the context of existential philosophy: "But where the danger is, also grows the saving power."

Being condemned to be free, human being is actually being responsible for each and every act. Every act of a man is the result of his choice, based on a freedom that cannot be rejected. This is the ultimate commandment of freedom, out of which all obligations derive.

In doing so, however, each act is confirmed as a possibility that has gained its value by making a choice. This further means that neither man nor his actions are *in time,* but that the time is *provided by man's action*. By making unpleasant choices man is condemned to be responsible, but that also means that his fate is in his hands and that entire world and its history are nothing but his creations.

The idea of personhood in Sartre's philosophy is not founded on psychological or anthropological theories, but sets up as an ontological, political and practical concept.

The others, the difference, the contradictions are what brings tension to every process, but also what gives the frame to the idea and event that is being clarified. Also, contradictions in ideas allow ambiguity, which determine the historical event itself and make it possible and concrete. What rationalism dismisses as a coincidence – is in fact what makes all human life, concludes Sartre.

References

Gadamer, Hans-Georg. *Hegelova dijalektika*. Beograd: Plato, 2003.

Hegel, Georg Vilhelm Fridrih. *Fenomenologija duha*. Beograd: BIGZ, 1986.

Kaufman, Walter, ed. *Existentialism from Dostoyevsky to Sartre*. New York: Meridian Books, 1989.

Sartr, Žan-Pol. *Biće i ništavilo: ogled iz fenomenološke ontologije*. Beograd: Nolit, 1983.

Sartr, Žan-Pol. *Drame*: *Izabrana dela*, vol. 5. Nolit: Beograd, 1981.

Sartr, Žan-Pol. *Egzistencijalizam i marksizam*. Beograd: Nolit, 1970.

Sartr, Žan-Pol. *Kritika dijalektičkog uma*. Beograd: Nolit, 1983.

Sartr, Žan-Pol. *Šta je književnost*. Beograd: Nolit, 1984.

Sartre, Jean-Paul. *Existentialism is a Humanism*, edited by John Kulka, translated by Carol Macomber. New Haven: Yale University Press, 2007.

Bioethics and the Person

Ioannis Ladas
National and Kapodistrian University of Athens
E-mail address: ioaladas@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0002-7767-026X

Abstract: In this short essay I will discuss the concept of the person. More specifically, I intend to focus on identifying the perspectives that the scientific discussion has created for the relationship of person and bioethics. I will argue that humans as persons, therefore as moral agents, are not identified by their biological existence, and that the respect of human dignity is a fundamental individual right, due to which we share an obligation to protect human persons against all forms of coercion. I will conclude with the view that all issues that fall within the field of bioethics would vanish into thin air, in case real, selfless love was the dominant feature as far as our inter-personal relations are concerned. If love pervades bioethics throughout as its starting point, its 'body' and its final end, it would be certain that the value of humanity and personhood would be safeguarded, moral agents' rights would not be infringed, and humans wouldn't ever be used merely as means. Hence, my concluding thesis is that, in order to overcome the deadlocks bioethics deals with, we should be oriented towards the Bioethics of Love.
Keywords: Bioethics; person; autonomy; dignity; love.

I. Introduction

The term *person* has a *technical meaning* within philosophy, and especially within ethics or bioethics. At the core of the personhood debate are two fundamentally different approaches: 1. on the one hand there is a belief in the inalienable and intrinsic value of human life, 2. and from the other hand is dependent on the existence of one or more attributes or abilities.[1] For this reason, the determination of the

[1] Dónal P. O'Mathúna, "Personhood in Bioethics and Biomedical Research," *Research Practitioner* 7, no. 5 (2006): 167.

doi: https://doi.org/10.12681/aprlp.49.927

nature of the person is one of the key issues of bioethics, especially nowadays, when many words have lost their meaning and are arbitrarily used to denote other things and states. A conceptual clarification of the authentic concept of the person and their properties becomes very pertinent, useful and enlightening in order to dispel the confusion of meanings of our era.

The English term "person" is ambiguous. We often use it as a synonym for "human being." However, the Greek term *πρόσωπο* (person)[2] is an etymologically composite word, originating from the phrase "προς ὦπα" (in front of the eyes), denoting the part of the head located where the eyes are.[3] The human face, however, is a concept with theological, philosophical, legal and aesthetic charge. It is also a referential concept, intertwined with the concept of personality.[4] It denotes one's relationship with other human being(s). Specifically, it denotes that one is open to other person(s). This relationship is what distinguishes a person from "the concept of static individuality."[5]

As Hans Urs von Balthasar (1905-1988) points out,

> […] few words have as many layers of meaning as person. On the surface it means just any human being,

[2] In Greek "πρόσωπο" also means "face." In ancient Rome, the word 'persona' (Latin) originally referred to the masks worn by actors on stage. The various masks represented the various "personae" during the play. Leonard William Geddes, "Person," in *Catholic Encyclopedia*, ed. Charles Herbermann, Edward Pace, Conde Fallen, Thomas Shahan, and John Wynne (New York: The Encyclopedia Press, 1911),

[3] Γεώργιος Μπαμπινιώτης, *Λεξικό της Νέας Ελληνικής Γλώσσας* (Αθήνα: Κέντρο Λεξικολογίας, 1998), 1518.

[4] Σταυρούλα Τσινόρεμα, "Το Πρόσωπο και η Αρχή της Προσωπικότητας," στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επ. Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή (Αθήνα: Παπαζήσης, 2016), 86.

[5] Ελένη Καλοκαιρινού, "Το Ανθρώπινο Πρόσωπο και η Φιλοσοφία: Για μία Ηθική του Προσώπου," στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επ. Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή (Αθήνα: Παπαζήσης, 2016), 68.

> any countable individual. Its deeper senses, however, point to the individual's uniqueness which cannot be interchanged and therefore cannot be counted. The complexity of the word's history, almost impossible to unravel, corresponds to this multiplicity of meanings, and almost from the beginning this history reflects the word's various aspects of meaning that cannot be synthesized.[6]

What does a person consist of? The answers given in relation to this question can be classified in two main groups: 1. The category of teachings of dualism, according to which a person is the sum of two independent hypostases, the body and the soul, that directly or indirectly affect one another,[7] and 2. the category of monism, according to which a person is defined as an inseparable uniform, whose body and soul are properties or predicates. Furthermore, the distinction between persons in the strict sense and "social persons," in case of fetuses and infants is very interesting.[8]

Regardless of the determination of its nature, the person has been treated by philosophers as a being to which moral characteristics are ascribed. Philosophy of the Human Person examine trans-empirical concepts like human nature, human dignity, fundamental human rights, the human soul, and human destiny[9]. Despite all aforementioned matters being deemed particularly interesting, the sole objective of the current study is to examine historically the concept of person and, alongside, to identify the perspectives that the scientific discussion has hitherto created for the concept of person and bioethics.

---

[6] Hans Balthasar, "On the Concept of Person," *Communio* 13 (1986): 18.

[7] For an exhaustive discussion of the notion of personhood see Evangelos D. Protopapadakis, *From Dawn till Dusk: Bioethical Insights into the Beginning and the End of Life* (Berlin: Logos Verlag, 2019), 24ff.

[8] Herman Tristram Engelhardt, *The Foundations of Bioethics* (New York: Oxford University Press, 1996), 135.

[9] Oswald Mascarenhas, "The Ethics of Dignity of the Human Person," in *Corporate Ethics for Turbulent Markets: The Market Context of Executive Decisions*, ed. O. Mascarenhas (Bingley: Emerald Publishing Limited, 2018), 11.

II. Historical view of the concept

"The word 'person' receives its special dignity in history when it is illuminated by the unique theological meaning."[10]

If we review the concept of person historically, we will see that it is first cited in the patristic philosophical theological tradition.[11] As aptly noted by Metropolitan of Pergamon John Zizioulas, ancient Greek philosophy is "non-personal in its substance,"[12] since it is incapable of composing the person with something permanent, thereby failing to create a philosophy of the person.[13] In this context, he underlines the significance of identifying the hypostasis with the person, because the person is no longer what perches on the specific being, but it is the being's hypostasis, and also because the hypostasis of the being no longer amounts to substance, but to the person.[14]

The person in its philosophical meaning, i.e. moral personality, is mainly associated with theology and Trinitarian disputes.[15] The Holy Trinity is defined as three persons with one and the same substance and the distinction of the persons is not due to the particularity of substance, but to the way of existing. The theological schools of Alexandria and Antioch have identified the concept of person with the concept of nature, while the ambiguous and ambivalent term "hypostasis," taken to mean person or the substance, has become a source of misunderstanding because of the different interpretations of linguistic symbols and denotations, signifier and signified, between eastern and western theology.[16]

---

[10] Balthasar, 18.

[11] Καλοκαιρινού, 69.

[12] Ιωάννης Ζηζιούλας, "Ἀπό τό Προσωπεῖον εἰς τό Πρόσωπον. Ἡ Συμβολή τῆς Πατερικῆς Θεολογίας εἰς τήν Ἔννοιαν τοῦ Προσώπου," στο *Χαριστήρια εἰς Τιμήν τοῦ Μητροπολίτου Γέροντος Χαλκηδόνος Μελίτωνος* (Θεσσαλονίκη: Πατριαρχικό Ἵδρυμα Πατερικών Μελετών, 1977), 287.

[13] Καλοκαιρινού, 70.

[14] Ζηζιούλας, 297.

[15] Μυρτώ Δραγώνα-Μονάχου, "Εισαγωγή," στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επ. Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή (Αθήνα: Παπαζήσης, 2016), 21.

[16] Μιχαήλ Μαντζανάς, "Βιοηθική και Πρόσωπο: Αρχαία, Βυζαντινή και

In a strictly philosophical spirit, the concept of person is associated with modern-time philosophy. Especially with Immanuel Kant, the moral person has shouldered the weight of dignity and autonomy where, in the context of his teachings on categorical imperative, people are persons who should always be treated as ends and never as means.[17] On the contrary, according to Jeremy Bentham, John Stuart Mill, and utilitarianism in general, the person becomes a means to advance another person's welfare and benefit and the concept of person is degraded or eradicated.[18]

It is natural that, throughout the centuries that followed, many versions of the philosophy of person have been developed both in Europe and elsewhere, and we can now talk about the philosophies of person (in plural).[19] For example Locke defines "person" as "a thinking intelligent being that has reason and reflection and can consider itself as itself."[20] They all place the person in the center, free, unique, ready to open up and relate to other persons or, alternatively, to offer himself/herself as a "gift" to others, unlike totalitarianism and national socialism that fully subordinate persons and unlike individualism that renders a person a "wolf" to others.[21]

III. The concept of Person in Bioethics

Bioethics is commonly understood to refer to the ethical implications and applications of the health-related life sciences. "Personhood is the focus of all ethical debates in biomedicine but there are two opposite approaches to the definition of per-

---

Σύγχρονη Οπτική," στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επ. Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή (Αθήνα: Παπαζήσης, 2016), 143.

[17] Δραγώνα-Μονάχου, 22.

[18] Καλοκαιρινού, 79.

[19] Ibid, 78.

[20] Locke, *Essay Concerning Human Understanding* (London: Dent, 1961), 260.

[21] T. D. Williams, J. O. Bengtsson, "Personalism," in *The Stanford Encyclopedia of Philosophy*, ed. N. Z. Edward (Spring 2014 Edition), http://plato.stanford.edu/archives /spr2014/entries/personalism.

sonhood."[22] As professor Antonio G. Spagnolo mention "in the reductionist perspective, the moral status of the person is attributed to the subject capable of a moral life or a 'valuable life;' in the personalistic approach, all human beings are considered persons from the beginning of life to the time of natural death, which is all human beings are persons in ontological sense."[23] Hence, in bioethical debates, the concept of the person plays a major role, because it is intimately connected with questions about the value of life.[24]

The rapid development of biosciences and biotechnology, which expand our potential of interfering with human nature, raise questions about the moral boundaries of such interference with respect to the value of the person and the freedoms and rights associated with personality.[25]

Bioethics[26] invests in the unlimited value of humans as rational and self-determined beings and this is why there are concerns about human persons in almost all bioproblems.[27] They are raised before the creation of human life, follow its course (mapping of human genomes, selection of gender, transplants, cloning, ageing delay) and are relevant even after it comes to an

---

[22] Antonio Spagnolo, "Personhood: Order and Border of Bioethics," *Journal of Medicine and the Person* 10 (2012), 99.

[23] Ibid.

[24] John Harris, "The Concept of the Person and the Value of Life," *Kennedy Institute of Ethics Journal* 9, no. 4 (1999): 293-308.

[25] Engelhardt, without getting involved in the labyrinth of theonomic and philosophical analysis, fully covers the concept of person, as it operates in the field of bioethics. He claims that "a person is self-conscious, rational, free to choose and in possession of a sense of moral concern." Engelhardt, 105.

[26] In modern bioethics, the concept of person is significant and many arguments in favor of one or the other "method" are based and/or rely thereon.

[27] From a bioethical viewpoint, the distinction between actual and potential persons is also significant, in reference to fetuses as potential persons and to humans in the face of euthanasia.

end (burial/cremation).[28] Furthermore, debates on the correlation between biological existence, personality and moral capacity have intensified - in addition to bioethics in environmental ethics. Of course, since animal ethics is part of bioethics, it would be an omission not to mention that some researchers supported that nonhuman animals can be persons. For example Daniel A. Dombrowski, who relies on the thought of neoclassical like Alfred North Whitehead, and Charles Hartshorne.[29] In any case the questions raised on bio-problems are many and complex. For instance:

1. Regarding fetal experiments, the key question is what a fetus is. Is it a person or just research material? Those who agree that fetuses are tissues of the female body are ready to accept medical experiments. Among those who believe that fetuses are potential or actual human beings, some do and some do not accept the challenge when protection of fetuses is ensured and their benefit is pursued. Both, however, ask whether a person who has dignity can be used in various ways as a guinea pig.
2. Regarding unused fertilized ovaries: If the fetus is not a person, then why not allow it to be sold? Ethically speaking, the fetus cannot be treated as a "simple means", as laboratory waste, in case of unused fetuses following medically assisted reproduction. Its treatment is ethically evaluated on the basis of care befitting the value associated with human persons.[30]
3. Regarding the legitimacy of (i) abortion, (ii) certain new practices of assisted reproduction (e.g. prenatal biomedical screening and embryo selection), (iii) genomic intervention. All these are associated with the normative issue of whether human persons are affected, and whether due respect and

[28] The concept of person is very important in the prenatal and the before-the-end-of-life painful state of humans in reference to the so-called borderline conditions. Δραγώνα-Μονάχου, 23.

[29] Daniel A. Dombrowski, "Are Nonhuman Animals Persons? A Process Theistic Response," *Journal of Animal Ethics* 5, no. 2 (2015): 135.

[30] Τσινόρεμα, 86, 109-110.

protection of persons includes future or potential persons. For instance, a pregnant woman does not hold power or ownership over the fetus as if it were a simple thing. When she decides to end a pregnancy or not, she takes action. She acts as a person capable of assuming the responsibility of becoming a mother. Because of its inherent characteristics, this decision is subject to moral accountability.[31]

4. Regarding cloning. This method threatens the sanctity, diversity and uniqueness of a person, since the original loses its uniqueness due to its substitution with the copy and the copy loses its uniqueness because it is deprived of originality and self-determination. Furthermore, as Hubert Doucet mentions "in the recent debates on human cloning, the respect and dignity of the person have influenced the concerns of those who are demanding an international moratorium on the possibility of cloning a human being."[32]

5. Regarding transplants. A moral issue is raised about the purchase and sale of organs. Any relevant commercial activity is an entirely immoral act because it shows lack of respect to human persons and life and offends human dignity.

6. Regarding treatment methods. The fact that the possibility of treatment, the method of treatment, the length of treatment and the method of treatment depend on the patient's financial status offends human dignity. Because is unfair to the financially weaker and also turns human persons into tools for wealth.

7. Regarding issues relating to the end of life. Moral dilemmas are raised concerning decisions relating to the end of people's lives, particularly with the development of new medical technologies that enable artificial prolongation of key biological functions of the body using mechanical means, even

[31] Ibid., 86.

[32] Hubert Doucet, "The Concept of Person in Bioethics Impasse and Beyond," in *Personhood and Health Care*, ed. David C. Thomasma, David N. Weisstub, and Christian Hervé, (Dordrecht: Springer, 1999), 121.

without expecting any medical benefit.[33] Euthanasia, in the context of Kant's approach, seems to distort the autonomy of the moral person. As a result, it is destroying its morality and is brutally offending its dignity because (i) it is an inherently contradictory moral choice and, therefore, it cannot become universal law, (ii) in the context thereof, the moral person ceases to be an end in itself and is demoted to a means. A gravely ill person lacking consciousness, a mentally retarded person or a person in a coma does not transform into something else as soon as such person loses their consciousness or mental powers.[34]

8. Regarding Neurological Science and Technology: "Modern advances in neurological science and technology pose profound challenges for our traditional concepts of the human person: they generate metaphysical and moral questions about beings at the edges of human life, from embryos that are not yet conscious, to persons who have lost their capacity for rational thought or have become permanently unconscious."[35]

At this point it should be noted that the major role of the concept of person in bio-problems has been widely disputed. For

[33] Kirsten Rabe Smolensky, "The Rights of the Dead," *Hofstra Law Review* 39 (2009): 764; Ευάγγελος Πρωτοπαπαδάκης, Κλωνοποίηση και Βιοηθική: Κλωνοποίηση Ανθρώπων και Δικαιώματα (Αθήνα: Παπαζήσης, 2013), 30; Stanley Benn, "Abortion – Infanticide and Respect for Persons," in *The Problem of Abortion*, ed. Joel Feinberg (Belmont, CA: Wadsworth Publishing Company, 1973), 99-100; Τσινόρεμα, 108-109.

[34] Ευάγγελος Πρωτοπαπαδάκης, "Η Ευθανασία και το Διακύβευμα της Αυτονομίας," στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο,* επ. Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή (Αθήνα: Παπαζήσης, 2016), 128.

[35] David Perry, "Some Issues in Contemporary Neurological Science and Technology," adapted from a presentation at a "Works in Progress" forum sponsored by the Center for Biomedical Ethics at Stanford University on December 11, 2001, https://www.scu.edu/ethics/focus-areas/bioethics/resources/ethics-and-personhood/.

example, Professor Bert Gordijn argues that "the concept of the person is unsuited to be a central concept in bioethical debates, because its use is connected with serious problems."[36] According to him "first, the concept is superfluous. Secondly, it is a confusing concept and it lacks pragmatic use. Thirdly, its use leads to simplifications". For this reason, he supports that "relinquishing the concept of the person could enhance the clarity and quality of bioethical debate."[37]

IV. Conclusion

The sacred character of human life gives meaning to human dignity, which science treats with respect. Of course, regarding the matter of founding the principle of dignity, positions differ. Some answers are atheistic and some are theistic. In any case, however, we would not now be talking about the morality of human rights and human dignity without the historical contribution of Christianity to the enhancement of human persons, since the concept of person, both historically and existentially, is integrally linked to theology. Human dignity is linked with the creation of humans in the image of God and the concept of human person that signals his/her relationship with God and fellow humans. Humans as persons, therefore as moral subjects, are not identified with their biological existence.[38] The demand for respect for the value of humans does not simply amount to nor is it exhausted in biological existence. Dignity is associated with the concept of person and self-determination and is defined by the person's ability to maintain moral autonomy so as to assess and make moral decisions on matters concerning them. Respect for human dignity is one of the fundamental individual rights and an obligation to protect human persons against all forms of power.

---

[36] Bert Gordijn, "The Troublesome Concept of the Person," *Theoretical Medicine and Bioethics* 20 (1999): 347.

[37] Ibid.

[38] Τσινόρεμα, 112.

However, some of us ask ourselves whether the lack of respect for human persons and human dignity is due to the lack of love. Whether the bio-problems described above would not exist if there was real and selfless love. If love was the beginning, the middle and the end of bioethics, it is certain that the value of human persons would be respected, their rights would not be infringed and humans would never be used as means. Therefore, to overcome all dead-ends, we must be oriented towards the Bioethics of Love. Without it, the respect for human life and human person has no future.

References

Balthasar, H. "On the Concept of Person." *Communio* 13 (1986): 18-26.

Beauchamp, Tom, and James Childress. *Principles of Biomedical Ethics*. Oxford: Oxford University Press, 1979.

Becher, Gerhold, ed. *The Moral Status of Persons. Perspectives to Bioethics*. Amsterdam: Rodopi, 2000.

Benn, Stanley. "Abortion – Infanticide and Respect for Persons." In *The Problem of Abortion*, edited by Joel Feinberg, 92-104. Belmont, CA: Wadsworth Publishing Company, 1973.

Chan, S., and Harris, J. "Human Animals and Nonhuman Persons." In *The Oxford Handbook of Animal Ethics*, edited by T. Beauchamp & R. Frey, eds., 304-331. Oxford: Oxford University Press, 2011.

DeGrazia, D. "On the Question of Personhood Beyond Homo Sapiens." In *Defense of Animals: The Second Wave*, edited by P. Singer, 40-53. Oxford: Blackwell 2007.

DeGrazia, David. *Human Identity and Bioethics*. Cambridge: Cambridge University Press, 2005.

Dombrowski, Daniel A. "Are Nonhuman Animals Persons? A Process Theistic Response." *Journal of Animal Ethics* 5, no. 2 (2015): 135-43.

Doucet, Hubert. "The Concept of Person in Bioethics Impasse and Beyond." In *Personhood and Health Care*, edited by David C. Thomasma, David N. Weisstub, and Christian Hervé, 121-128. Dordrecht: Springer 1999.

Engelhardt, Herman Tristram. *The Foundations of Bioethics*. New York: Oxford University Press, 1996.

Etheredge, Francis. *The Human Person: A Bioethical Word*. St. Louis, MO: En Route Books and Media, 2017.

Gordijn, Bert. "The Troublesome Concept of the Person." *Theoretical Medicine and Bioethics* 20 (1999): 347-359.

Harris, John. "The Concept of the Person and the Value of Life." *Kennedy Institute of Ethics Journal* 9, no. 4 (1999): 293-308.

Kant, Immanuel. *The Metaphysics of Morals*, edited by Mary Gregor. Cambridge: Cambridge University Press, 1991.

Mascarenhas, Oswald. "The Ethics of Dignity of the Human Person." In *Corporate Ethics for Turbulent Markets: The Market Context of Executive Decisions*, edited by O. Mascarenhas, 11-41. Bingley: Emerald Publishing Limited, 2018.

Mathúna, Dónal P. "Personhood in Bioethics and Biomedical Research." *Research Practitioner* 7, no. 5: (2006): 167-174

O'Neill, Onora. *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press, 2002.

Protopapadakis, Evangelos D. *From Dawn till Dusk: Bioethical Insights into the Beginning and the End of Life*. Berlin: Logos Verlag, 2019.

Smolensky, Kirsten Rabe. "The Rights of the Dead." *Hofstra Law Review* 39 (2009): 763-803.

Spagnolo, Antonio. "Personhood: Order and Border of Bioethics." *Journal of Medicine and the Person* 10 (2012): 99-102.

Teichman, Jenny. "The Definition of Person." *Philosophy* 60, no. 232 (1985): 175-85.

Williams, T. D., and J. O. Bengtsson. "Personalism." In *The Stanford Encyclopedia of Philosophy*, edited by N. Z. Edward (Spring 2014 Edition), http://plato.stanford.edu/archives /spr2014/entries/personalism.

Γιανναράς, Χρήστος. *Το Πρόσωπο και ο Έρως*. Αθήνα: Δόμος, 2001.

Γρινιεζάκης, Μακάριος. *Εισαγωγή στη Βιοηθική*. Αθήνα: Γρηγόρης, 2014.

Δραγώνα-Μονάχου, Μυρτώ. "Εισαγωγή." Στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επιμέλεια Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή, 13-52. Αθήνα: Παπαζήσης, 2016.

Ζηζιούλας, Ιωάννης. "Ἀπό τό Προσωπεῖον εἰς τὸ Πρόσωπον. Ἡ Συμβολὴ τῆς Πατερικῆς Θεολογίας εἰς τὴν Ἔννοιαν τοῦ Προσώπου". Στο *Χαριστήρια εἰς Τιμὴν τοῦ Μητροπολίτου Γέροντος Χαλκηδόνος Μελίτωνος*, 287-323. Θεσσαλονίκη: Πατριαρχικό Ίδρυμα Πατερικών Μελετών, 1977.

Καλοκαιρινού, Ελένη. "Το Ανθρώπινο Πρόσωπο και η Φιλοσοφία: Για μία Ηθική του Προσώπου." Στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επιμέλεια Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή, 67-84. Αθήνα: Παπαζήσης, 2016.

Μαντζανάς, Μιχαήλ. "Βιοηθική και Πρόσωπο: Αρχαία, Βυζαντινή και Σύγχρονη Οπτική." Στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επιμέλεια Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή, 137-160. Αθήνα: Παπαζήσης, 2016.

Μπαμπινιώτης, Γεώργιος. *Λεξικό της Νέας Ελληνικής Γλώσσας*. Αθήνα: Κέντρο Λεξικολογίας, 1998.

Πελεγρίνης, Θεοδόσιος. *Οι Πέντε Εποχές της Φιλοσοφίας*. Αθήνα: Ελληνικά Γράμματα, 1997.

Πρωτοπαπαδάκης, Ευάγγελος. "Η Ευθανασία και το Διακύβευμα της Αυτονομίας." Στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επιμέλεια Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή, 115-136. Αθήνα: Παπαζήσης, 2016.

Πρωτοπαπαδάκης, Ευάγγελος. *Κλωνοποίηση και Βιοηθική: Κλωνοποίηση Ανθρώπων και Δικαιώματα*. Αθήνα: Παπαζήσης, 2013.

Τσινόρεμα, Σταυρούλα. "Το Πρόσωπο και η Αρχή της Προσωπικότητας." Στο *Βιοηθικοί Προβληματισμοί ΙΙ. Το Πρόσωπο*, επιμέλεια Μαρία Κανελλοπούλου-Μπότη, και Φερενίκη Παναγοπούλου-Κουτνατζή, 85-114. Αθήνα: Παπαζήσης, 2016.

Contributors

Georgios Arabatzis

George Arabatzis is Associate Professor of Philosophy at the Naional and Kapodistrian University of Athens. He studied philosophy in Paris, France. A Scholarship holder of the French Government, he obtained his PhD with honors from the École de Hautes Études en Sciences Sociales and made a post-doctoral research in Torino. He has been a Visiting Scholar-Research Fellow at Princeton University, the University of Texas at Austin, the University of Helsinki, the University Charles in Prague, the University of Jassy, Romania, etc. He has given lectures at the above universities and other ones, and has taught philosophy at the Aristotle University of Thessaloniki, at the Postgraduate programme in Philosophy of the University of Athens, the Greek Open University, and the École de Hautes Études en Sciences Sociales, Paris. His research is focused on Byzantine philosophy, its Ancient Sources and its modern perception. He works also on iconology, medieval and modern (cinema, and digital art).

George Boutlas

Georgios Boutlas obtained his Bachelor's Degree from the School of Medicine of the National and Kapodistrian University of Athens. He also holds a Master's Degree in Philosophy by the Department of Methodology, History and Theory of Science, NKUA, and a PhD in Bioethics by the same department. He is Research Fellow and Head of the Liaison Department of the NKUA Applied Philosophy Research Laboratory. His research interests are focused on Bioethics, Kantian ethics, and Applied Ethics. He is Adjunct Lecturer at the University of Ioannina, the National and Kapodistrian University of Athens, and the Open University of Cyprus. He has published several articles in domestic and international philosophical journals.

Alikis Gounaris

Alkis Gounaris is Research Fellow at the NKUA Applied Philosophy Research Laboratory, and the Head of its Scientific Organizations Department. He has been teaching as an Adjunct Lecturer at the Department of Philosophy of the National and Kapodistrian University of Athens since 2012. His areas of expertise include Philosophy of Human Cognition and Artificial Intelligence, Philosophy of Cognitive Sciences, Ethics, and Applied Ethics. He holds two Bachelor's Degrees, one in Sport Sciences, and another in Philosophy; he has also earned a Master's Degree in Philosophy, and a PhD Degree in Ethics and the Philosophy of Mind, all four by the National and Kapodistrian University of Athens. Alkis Gounaris has contributed more than fifty papers and book chapters in domestic and international publications.

Panagiotis Kormas

Panagiotis Kormas is a Medical Doctor specialised in Orthopaedic Surgery. He received his PhD in Philosophy from the Department of Philosophy of the National and Kapodistrian University of Athens; currently he is a Research Fellow at the Applied Philosophy Research Laboratory, and Adjunct Lecturer at the University of Athens. He has served as a member of the Central Board of Health and as the Vice-President of the National Emergency Center. Currently, he serves as a Special Administration Counselor on health issues and a Medical Services Coordinator at major insurance companies. Dr. Kormas is also a member of the Medical Advisory Board in The Bioinformatics and Human Electrophysiology Lab of the Ionian University, as well as a member of the International Advisory Committee of the Hellenic Initiative Against Alzheimer's (HIAA). Several works by him have been published in various international medical journals, and conference proceedings.

George Kosteletos

George Kosteletos is Research Fellow at The NKUA Applied Philosophy Research Laboratory, and co-Head of its Research Department. He is also an Adjunct Lecturer at the Department of Philosophy of the National and Kapodistrian University of

Athens and the Open University of Cyprus. His areas of expertise include Philosophy of Mind and Artificial Intelligence, Philosophy of Cognitive Sciences and Neuro-ethics. George Kosteletos holds an academic scholarship by the NKUA Medical School as Research Fellow at the Psycho-physics Laboratory conducting research in the field of Moral Neuroscience/Psychology. He has contributed papers and book chapters in domestic and international publications.

Ioannis Ladas

Ioannis Ladas studied Theology at the National and Kapodistrian University of Athens, and Philosophy at the University of Patras. He holds two Master's Degrees: one in Ethics from the University of Athens, and one in Dogmatic and Symbolic Theology from the Aristotle University of Thessaloniki. He earned his a PhD in Ethics by the National and Kapodistrian University of Athens for a Thesis entitled *The Problem of the Philosophical Foundation of Bioethics and H. Tristram Engelhardt, Jr.* He is also the Projects Manager of *Conatus - Journal of Philosophy*, and Research Fellow at the NKUA Applied Philosophy Research Laboratory. Several articles of his have been included in domestic and international publications; he has also been the Guest Editor of *Conatus - Journal of Philosophy* 3, no. 2 (2018) Spcial Issue "On H. Tristram Engelhardt, Jr.: Ethics on the Edge of Faith."

Antonia Moutzouri

Antonia Moutzouri graduated from the University of Patras with an M.S. in Chemical Engineering and a PhD in Biomedical Engineering. Her research has been in the fields of Cell Mechanical Behaviour and Cell-Material Interactions; she is the author or the co-author of seven articles that have been published in peer-reviewed international journals, while she has also made several presentations in conferences related to her fields of expertise. She has worked as a Biomedical Liaison and a Business Associate for multinational companies mostly focusing on the development and application of cutting-edge autologous cell therapies, as well as the development of novel bioengineering solutions for soft tissue injury

prevention. Her research interests extend to the areas of philosophy and Bioethics in particular, with particular emphasis on the field of technology-assisted biomedical applications. Antonia Moutzouri is a Research Fellow at the Applied Philosophy Research Laboratory of the National and Kapodistrian University of Athens.

Una Popović

Una Popović is Associate Professor of Philosophy at the University of Novi Sad. She completed her BA studies in Philosophy at the Faculty of Philosophy, University of Belgrade, in 2008. She has a PhD in Philosophy (Aesthetics & Ontology) by the same University (2014); her PhD Thesis is entitled *Heidegger's Philosophy of Language*. Una Popović has published two books, *Language and Being: Fundamental Ontology of Language* (Belgrade, 2019), and *The Truth of Sensibility: Baumgarten and the Founding of Aesthetics* (Belgrade, 2019), as well as numerous scientific articles. Her research interests include aesthetics, philosophy of language, phenomenology, and history of philosophy. She is a member of the Scientific Board of the Aesthetic Society of Serbia, and the Serbian Philosophical Society.

Dragan Prole

Dragan D. Prole is Professor of Philosophy (Ontology, Contemporary Philosophy) at the University of Novi Sad, and guest lecturer at the Academy of Arts (Aesthetics, Contemporary Aesthetics). He is member of the Editorial Board of several philosophical journals (Conatus – Journal of Philosophy, Greece; Phainomena, Slovenia; Društveni pregled, Bosnia and Herzegovina; Theologos, Serbia). He has published eight books in Serbian as the single author, and six anthologies as the editor or the co-editor. Dragan Prole has translated six books from German into Serbian (Schelling, Husserl, Waldenfels); he has also been on study visits to the University of Berlin, Weimar, Leuven, Graz, Heidelberg and Vienna. He has been invited as a visiting lecturer to several Universities (Leuven Husserl Memorial Lecture 2019, Uppsala, Krakow, Athens, Ljubljana, Vienna, Oßmannstedt, Weimar, Skopje, Prague). Dragan Prole has received four awards for his books (best

book in philosophy, theory of literature and art in 2011, best book of the year 2013, award "Stevan Pešić" 2013, and best essays award "Sreten Marić" in 2016).

Marica Rajković

Marica Rajković is Assistant Professor of Philosophy at the University of Novi Sad. Her research focuses on aesthetics, philosophy of culture, philosophy of art, and philosophy of technology. The wider field of her scientific research encopass aesthetics, philosophy of German idealism, and poetical philosophy. In 2016 she defended her PhD thesis entitled *Aesthetics in the Philosophy of German Idealism*. She is a member of the Aesthetical Society of Serbia, and a member of the research project "Teaching philosophy in Serbian high schools in Vojvodina in the interwar period." She is the author of thirty scientific papers that focus mostly on poetical philosophy, and Western philosophy in general.

Goran Rujević

Goran Rujević is Assistant Professor at the Department of Philosophy in Novi Sad. He received his PhD from the University of Novi Sad for a thesis that establishes the notion of predictive gap, which can be used to describe discrepancies between predicted and real outcomes of scientific theories. His main areas of interest are philosophy of natural sciences, philosophy of mathematics and methodology of teaching and popularising philosophy. He has also translated several books from English and German into Serbian, including Immanuel Kant's "Universal Natural History and Theory of the Heavens," and Nick Bostrom's "Superintelligence."

Damir Smiljanić

Damir Smiljanić is Professor of Philosophy at the University of Novi Sad. He obtained his Master's Degree on Philosophy, Theatre and Media Studies, Sociology from Friedrich-Alexander-University of Erlangen-Nuremberg in 1998. He obtained his PhD Degree from the same University in 2005. Since 2007 he is employed by the Department of Philosophy at the Faculty

of Philosophy in Novi Sad. His main areas of interest are Metaphilosophy, Methodology, Epistemology, Philosophy of Language, and Media Philosophy. He has published the following books: *Philosophical Positionality in the Light of Perspectivism* (in German, Marburg, 2006), *Synesthetics* (Novi Sad, 2011), *Irritations* (Novi Sad, 2014), *Aporetics* (Novi Sad, 2016), and *Atmosphere of Death* (Novi Sad, 2018). He won the first prize in the Mainländer Essay Competition by the City of Offenbach/Main in April 2005. His book *Atmosphere of Death* was awarded by the Association of Writers of Vojvodina as the best book in 2018.



THIS VOLUME WAS DESIGNED BY ACHILLEAS KLEISOURAS, TYPESETTED IN GARAMOND TYPEFACE AND PRODUCED BY THE NKUA APPLIED PHILOSOPHY RESEARCH LAB PRESS. IT WAS PRINTED IN NOVI SAD, REPUBLIC OF SERBIA, BY FILOZOFSKI FAKULTET U NOVOM SADU IN ACID FREE 80GR PAPER. THE ELECTRONIC VERSION OF THIS VOLUME IS HOSTED BY THE NATIONAL DOCUMENTATION CENTER OF THE HELLENIC REPUBLIC.



HELLENIC-SERBIAN **PHILOSOPHICAL DIALOGUE** SERIES

p-ISBN 978-618-83729-2-4 e-ISBN 978-618-83729-3-1