

# BALKAN ANALYTIC FORUM



**Normativity**

**Normativity  
of Art**


*Edited by*

**Miroslava Trajkovski**  
*University of Belgrade*

**Emily C. McWilliams**  
*Duke Kunshan University*



UNIVERSITY OF BELGRADE  
FACULTY OF PHILOSOPHY



Welcome to the inaugural volume of the Balkan Analytic Forum's publication series. I am thrilled to be part of this effort, which seeks to nurture scholarly exchange and collaboration within the Balkan region and beyond. This volume comprises a selection of contributions around the themes of the Balkan Analytic Forum's inaugural conference series, including BAF1: Normativity, and BAF+: Normativity of Art. The essays in this volume encapsulate the depth and diversity of analytic philosophy's approaches to thinking about normativity, while also highlighting endeavors to bridge philosophical traditions, providing insights that resonate across philosophical landscapes. Working on this volume has been especially meaningful for me as a philosopher at a Sino-American joint venture institution with a global orientation. The orientation and mission of the Balkan Analytic Forum reflect the kind of rooted globalism that animates our work at Duke Kunshan University. The forum appreciates the historical and cultural embeddedness of intellectual traditions, while also striving to create knowledge and understanding through engagement among and across philosophical lineages, enriching the global discourse of philosophy. I am thrilled to be involved in this project. I extend my heartfelt gratitude to all those who supported the establishment of this forum and its inaugural events, and to Dr. Miroslava Trajkovski for her trailblazing and steadfast leadership in bringing it all together.

Emily C. McWilliams



*University of Belgrade – Faculty of Philosophy | 2024*



1838

Balkan Analytic Forum  
Normativity & Normativity of Art  
International Conference  
19–29. X 2023. Belgrade  
Proceedings

*Edited by*  
MIROSLAVA TRAJKOVSKI, University of Belgrade  
EMILY C. McWILLIAMS, Duke Kunshan University  
Belgrade 2024

*Publisher*  
Center for Contemporary Philosophy – Balkan Analytic Forum  
University of Belgrade – Faculty of Philosophy  
Čika Ljubina 18–20, Beograd 11000, Srbija  
www.f.bg.ac.rs

*For the publisher*  
Prof. Dr. Danijel Sinani  
Dean of the Faculty of Philosophy

*Scientific Organizing Committee*  
Marina Bakalova  
Miroslava Trajkovski  
Monika Jovanović  
Miloš Vuletić

*Reviewed by*  
Irina Deretić, University of Belgrade  
Amber Riaz, LUMS, Pakistan  
Ivana Simić, University of Florida, USA

*Scientific Programme Committee*  
Marina Bakalova, Bulgarian Academy of Sciences, Institute of Philosophy and Sociology, Bulgaria  
Mircea Dumitru, University of Bucharest, Romanian Academy, Romania  
Timothy Williamson, University of Oxford, New College, United Kingdom  
Stathis Psillos, University of Athens, Department of Philosophy & History of Science, Greece  
Vladimir Marko, Comenius University, Bratislava, Slovakia  
Lilia Gurova, New Bulgarian University, Department of Cognitive Science and Psychology, Bulgaria  
Miroslava Trajkovski, University of Belgrade, Department of Philosophy, Serbia

*Cover art and design by*  
Dragan Trajkovski

*Set by*  
Dosije studio, Belgrade

*Printed by*  
JP Službeni glasnik

*Print run*  
100

*Supported by:*

MINISTRY OF SCIENCE, TECHNOLOGICAL  
DEVELOPMENT AND INNOVATION  
of the Republic of Serbia



Република Србија  
МИНИСТАРСТВО НАУКЕ,  
ТЕХНОЛОШКОГ РАЗВОЈА И ИНОВАЦИЈА

Project number 451-03-66/2024-03/200163

ISBN 978-86-6427-286-5

<https://doi.org/10.18485/baf.2024.1>

Balkan Analytic Forum



Normativity  
&  
Normativity of Art

Conference Proceedings  
19–29. X 2023. Belgrade, Serbia

*Edited by*

MIROSLAVA TRAJKOVSKI  
University of Belgrade

EMILY C. McWILLIAMS  
Duke Kunshan University

Belgrade  
2024



# CONTENTS

- 7 | Contributors
- 9 | Editors' Preface:  
General Introduction to Balkan Analytic Forum
- 11 | Acknowledgements
- 13 | Roaring Balkans – *Vlada Stanković*
- 17 | Introduction: Normativity and Normativity of Art –  
*Miroslava Trajkovski*

## *BAF1: NORMATIVITY*

### 1. NORMATIVITY AND EXPLANATIONS

- 35 | *Damir Smiljanić* “On the Distinction between  
Descriptive and Prescriptive Metaphilosophy”
- 51 | *Thodoris Dimitrakos* “Historicizing Second Nature:  
The Consequences for the Is/Ought Gap”

### 2. NORMATIVITY AND KNOWLEDGE

- 73 | *João Carlos Salles* “Ernest Sosa’s Telic Virtue Epistemology”
- 87 | *Timur Cengiz Uçan* “Machines and Us: The Comparison  
of Machines and Humans at the Test of the Problematic  
of Solipsism”

### 3. NORMATIVITY AND COLLECTIVE INTENTIONALITY

- 129 | *Vasiliki Xiromeriti* “Collective Deliberation in Epistemic  
Groups: Lessons from Deliberative Democracy”
- 151 | *Ognjen Milivojević* “Searle and the Creation of Social Norm”

#### 4. **NORMATIVITY AND LOGIC**

- 167 | *Aleksandra Vučković* “Normativity and Truth in Naturalized Epistemology”
- 181 | *Miroslava Trajkovski* “Normativity, Validity and Semiotic Implication”

*BAF1+*: *NORMATIVITY OF ART*

#### 5. **ART WORKS AND COGNITION**

- 199 | *Ted Kinnaman* “Normativity in Art in Kant’s Aesthetics”
- 211 | *Isidora Novaković* “Philosophical Value of Literature: Machiavelli and Shakespeare”

#### 6. **ART WORKS AND COLLECTIVE INTENTIONALLITY**

- 231 | *Milan Popadić* “Can a Monument Be Bad? Normativity and Commemorative Values in Public Space”
- 247 | *Ivan Popov* “When is Art Interactive?”



## CONTRIBUTORS

- Emily C. McWilliams, Duke Kunshan University, China (0000-0002-8609-070X)
- Vlada Stanković, University of Belgrade, History Department, Center for Cypriot Studies, Serbia (0009-0004-2720-6583)
- Damir Smiljanić, University of Novi Sad, Department of Philosophy, Serbia (0000-0002-5791-620X)
- Thodoris Dimitrakos, University of Patras, Department of Philosophy, Greece (0000-0001-6990-1980)
- João Carlos Salles Pires da Silva, Federal University of Bahia, Philosophy Department, Brasil (0000-0002-4872-3465)
- Timur Cengiz Uçan, Bordeaux Montaigne University, *Sciences Philosophy Humanities* Mixed Research Unit Social Sciences of Contemporary Changes Department, France (0000-0001-7620-6601)
- Vasiliki Xiomeriti, graduate student, Université Jean Moulin Lyon 3, Philosophy Department, France (0009-0001-2669-304X)
- Ognjen Milivojević, graduate student, University of Belgrade, Department of Philosophy, Serbia (0009-0007-6696-8144)
- Aleksandra Vučković, University of Belgrade, Institute of Philosophy, Serbia (0000-0002-5960-2909)
- Miroslava Trajkovski, University of Belgrade, Department of Philosophy, Balkan Analytic Forum, Serbia (0000-0002-4927-2168)
- Ted Kinnaman, George Mason University, USA (0009-0008-7098-2982)
- Isidora Novaković, graduate student, University of Belgrade, Department of Philosophy, Serbia (0009-0007-2129-5580)
- Milan Popadić, University of Belgrade, Department of Art History, Serbia (0000-0002-9525-7575)
- Ivan Popov, St.-Kliment-Ochridski-University Sofia, Department for German and Scandinavian Studies, Bulgaria (0000-0001-6245-3329)



# EDITORS' PREFACE

## GENERAL INTRODUCTION

to

## BALKAN ANALYTIC FORUM

<https://www.f.bg.ac.rs/instituti/baf>

Welcome to the inaugural volume of the Balkan Analytic Forum's publication series. I am thrilled to be part of this effort, which seeks to nurture scholarly exchange and collaboration within the Balkan region and beyond. This volume comprises a selection of contributions around the themes of the Balkan Analytic Forum's inaugural conference series, including BAF1: Normativity, and BAF+: Normativity of Art. The essays in this volume encapsulate the depth and diversity of analytic philosophy's approaches to thinking about normativity, while also highlighting endeavors to bridge philosophical traditions, providing insights that resonate across philosophical landscapes. Working on this volume has been especially meaningful for me as a philosopher at a Sino-American joint venture institution with a global orientation. The orientation and mission of the Balkan Analytic Forum reflect the kind of rooted globalism that animates our work at Duke Kunshan University. The forum appreciates the historical and cultural embeddedness of intellectual traditions, while also striving to create knowledge and understanding through engagement among and across philosophical lineages, enriching the global discourse of philosophy. I am thrilled to be involved in this project. I extend my heartfelt gratitude to all those who supported the establishment of this forum and its inaugural events, and to Dr. Miroslava Trajkovski for her trailblazing and steadfast leadership in bringing it all together.

Emily C. McWilliams



The Balkan Analytic Forum aims to bring together experts in analytical philosophy from the Balkans to exchange their ideas, but it is also open to approaches that establish connections between analytical philosophy

and other philosophical traditions, as well as to interested experts from other parts of the world. The activity of the forum is to establish, through conferences and accompanying publications, a platform for discussion where scientists from the Balkans, and all those interested in analytical philosophy, can meet regularly and present the texts they are currently working on and their new publications.

The mission is to carry out basic, applied and developmental research in the domain of analytical philosophy; to publish the results of this scientific research and professional work; to include young researchers and doctoral students at the Faculty in the implementation in this through participation in the programs implemented by the Center; to participate in the organization of gatherings, symposiums, professional meetings and workshops for the purpose of training researchers in the field of analytical philosophy; as well as to cooperate with other institutions in the country and abroad, especially with countries from the Balkans. The work of the Balkan Analytic Forum has been institutionalized by receiving the support of the Department of Philosophy, the University of Belgrade, Faculty of Philosophy and the Ministry of Science, Technological Development and Innovation for organizing the first conference of the Balkan Analytic Forum: BAF1: Normativity, BAF+: Normativity of Art, which was held in three units in the period from October 19 to 29, 2023 (cf. <https://www.f.bg.ac.rs/instituti/baf/publikacije>). The organizing committee of the conference consisted of: Miroslava Trajkovski (Department of Philosophy, Faculty of Philosophy), Monika Jovanović (Department of Philosophy, Faculty of Philosophy), Miloš Vuletić (Department of Philosophy, Faculty of Philosophy) and Marina Bakalova (Institute of Philosophy and Sociology, Bulgarian Academy of Sciences). The Programme Committee included Marina Bakalova (Bulgarian Academy of Sciences, Institute of Philosophy and Sociology, Bulgaria), Mircea Dumitru (University of Bucharest, Romanian Academy, Romania), Timothy Williamson (University of Oxford, New College, United Kingdom), Stathis Psillos (University of Athens, Department of Philosophy & History of Science, Greece), Vladimir Marko (Comenius University, Bratislava, Slovakia), Lilia Gurova (New Bulgarian University, Department of Cognitive Science and Psychology, Bulgaria), Miroslava Trajkovski (University of Belgrade, Department of Philosophy, Serbia).

The conference has been supported by: Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and University of Belgrade – Faculty of Philosophy.

M. T.

## ACKNOWLEDGEMENTS

Grateful acknowledgments to the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and the University of Belgrade – Faculty of Philosophy who supported this event.

Grateful acknowledgments to Nenad Cekić, head of the Department of Philosophy who supported the BAF initiative, to Vlada Stanković, professor of history and head of the Center for Cypriot Studies, for his willingness to write an introductory text about the Balkans, to Stefan Mičić who offered his help in the finalization of the preparation of this collection. Thanks to Dragan for his design, his gift, to the BAF.



Vlada Stanković

## ROARING BALKANS

One thing for certain cannot be associated with the Balkans: boredom. Nobody can state, for better or worse, that the Balkans are boring. Whether we talk about history, politics, identities—national or regional—, sport, conspiracy theories and – why not – philosophy, the Balkans always stand out from the rest of Europe as still somewhat mystical, almost mythical and incomprehensible jumble of small nation-states, and regions of extremes and opposites.

What does a specialist in the history and culture of the Byzantine and post-Byzantine world, the Balkans included, have to say regarding the contemporary culture, intellectual and philosophical trends in the Balkans? Aside from personal intellectual curiosity, by studying the longest lasting European empire, its structure, changes, and impressive legacy of original thought, one could better understand contemporary Balkan complexities, a unique mixture of traditionalism and modernity buried deep in cultural roots of this always interesting and question provoking European region. The long and deep traditions of two European world empires, Roman-Byzantine and Ottoman, undoubtedly left their mark on the mentality of the collective Balkans. The traditions of an all-powerful empire, its emperor and the overwhelming centripetal power of the empire's center, one of the *par excellence* world's capital cities, Constantinople-Istanbul, are not only present and felt in the region, but are equally strongly emphasized, positively or negatively, in contemporary geopolitics and scholarly discourse. Often overstated or misused, this connection with a dominant regional power and its mighty center contributed to the creation in the Balkans of a slightly odd notion in contemporary Europe of the existence of, and even the need for a “higher authority”, which in turn somewhat denigrates all those below the highest authority/highest power, be it one's neighbors, ‘strangers’ from another village or region, or members of other nations.

For many a reason, therefore, the Balkans are much more than a peninsula – the Balkans are Europe's subcontinent ...

The bare notion of the Balkans provokes an image of the long-lasting, never-changing rivalries and enmities that, supposedly, stretch back at least to medieval times, even though it has been demonstrated persuasively that the ideological concept of the Balkans is actually the product of the 19th century European Orientalizing discourse (M. Mazower, *The Balkans. A Short History*, New York 2000/2nd ed. London 2005). The modern Balkans, which rose from the 19th century national struggles against the Ottoman empire, represented an enigma for European thinkers, policy-makers and adventurers in no small measure because the “young” Balkan nation-states built their aspirations for inclusion into modern Europe by looking back to the “glorious” medieval past, except in the case of the modern Greek state which based its ideology and identity predominantly on the legacy of Antiquity. A double-mirror effect had thus created a peculiar intellectual conundrum: the Balkan nations looked back to the past in order to develop modern political and cultural structures, with European ideologists and commentators accepting at face value the Balkan’s “fixation on the past”—categorized as incurable backwardness—which was then projected back to the Balkans as their unchangeable destiny.

This double trap, in itself a product of a lack of knowledge and intellectual laziness, had real-life consequences for the Balkans. Essentially disinterested and impatient with Balkan complexities and intricacies, the majority of Europeans did much more than just ‘imagining’ the Balkan subcontinent: they tried to differentiate and distance themselves from this incomprehensible mosaic of nations, states, cultures and beliefs, even though the Balkans’ obsession with the past was but a reflection of the broader European ideologization of national pasts (M. Todorova, *Imagining the Balkans*, New York–Oxford 1997; P. Geary, *The Myth of Nation. The Medieval Origins of Europe*, Princeton 2001). This vicious circle of non-communication and misunderstandings created an image of the Balkans as a second-rate Europe, and shaped preconceptions that were re-asserted every time when the seemingly unending flow of evidence of Balkan backwardness was broadcasted to the ‘world’. In other words, the judgments on the Balkans have been laid down, and only new examples could be piled up to additionally confirm it.

One aspect of the staying power of the misconceptions about the Balkans is best expressed by Mark Mazower: ‘On the lookout for evidence of Balkan bloodthirstiness, however, Western observers have often mistaken the myths spun by nineteenth-century romantic nationalists for eternal truths’ (Mazower, *The Balkans* 2005, p. 133). The double mirror duped the beholders from both its sides—the Balkans itself and the ‘Westerners’ which naively fell in the trap of believing in ‘centuries-old “ancestral



hatred”<sup>6</sup>, mistakenly surmising that their ‘world’ had become immune to any ‘historical’ disputes (T. G. Ash, *Free World. America, Europe and the Surprising Future of the West*, New York 2004, p. 133).

The story of the Balkans is, therefore, in many ways one of the lessons not learned, equally by the peoples from the Balkan subcontinent and the ‘others’. It is also the story of quickly forgetting honest people, positive examples, and valuable ideas, thoughts, and concepts, and focusing on the median, the negative and the bad –the story, in short, of easily giving up on the former and accepting the inevitability of the latter. But the Balkans are no more negative or backward than any other region, quite the opposite.

To use the specific, strong, *roaring* dynamics of the Balkans positively is one of the major aspirations of all intellectual endeavors. ‘For, the Balkans are the world in microcosm. All the elements of discord, all the elements of harmony, exist in the Balkans in a simplified and highly concentrated form. Solve the problem of the Balkans and you will have solved the problem of Europe.’ (L. White, *The Long Balkan Night*, New York 1944, p. 450).



Miroslava Trajkovski

## INTRODUCTION: NORMATIVITY AND NORMATIVITY OF ART

### NORMATIVITY

#### 1. Normativity and Explanations

Damir Smiljanić in “On the Distinction between Descriptive and Prescriptive Metaphilosophy” considers the issue of the norms related to philosophical explanations. In particular, the author examines the question of what attitude philosophy has in relation to itself. This topic occupies the attention of a significant number of contemporary philosophers. A detailed treatment of this problem is found, for example, in Timothy Williamson’s book *The Philosophy of Philosophy* (2007), as well as, Smiljanić notes, in Richard Raatzsch’s book of the same name *Philosophiephilosophie* (2000). However, Smiljanić does not take philosophy of philosophy as the subject of his discussion, but metaphilosophy. If we understand the qualification “meta” as Alfred Tarski understands the relationship between metatheory and object theory (such that the former is a theory about the latter), then the term “metaphilosophy” means “a theory about philosophy.” Although these two phrases can be taken as synonymous in many contexts, with regard to the question Smiljanić is considering, they are not. Let us stress that the phrase “philosophy of philosophy” already implies that reflection on philosophy is self-reflection within the discipline of philosophy. In his paper Smiljanić deals with the question of whether this reflection requires a special discipline, so, that the term metaphilosophy is more appropriate for it does not prejudice an answer. Smiljanić looks for the answer to the question in the conceptual framework that Nicholas Rescher sets up in his works. Within the analytical tradition, the examination of the nature, status and method of philosophy was particularly emphasized by philosophers of the Vienna Circle and the Berlin Group. Hence, it is relevant to

note that Rescher classified himself as part of the third generation of the Berlin Circle (cf. “The Berlin School of Logical Empiricism and its Legacy,” *Erkenntnis* 64 (3), 2006). Smiljanić analyzes Rescher’s views on metaphilosophy from a number of different angles. He especially considers the distinction Rescher draws between descriptive and prescriptive metaphilosophy. The former involves a historical presentation of philosophy, while the latter concerns the question of what a valid philosophy should be. Let us note that according to this distinction, the history of philosophy does not belong to philosophy, but to metaphilosophy, which should lead us to a wider reflection on the attitude of analytical philosophy towards historicity in general. Smiljanić defends the disciplinary autonomy of the metaphilosophy. Moving from here to Thodoris Dimitrakos’ paper where he defends the autonomy of normative explanations by historicizing the notion of second nature, we are reminded of Richard Rorty who says that historicist and metaphilosophical self-consciousness, is the best precaution against barren scholasticism (“Analytic and Conversational Philosophy”, *The Rorty Reader*, Wiley – Blackwell, p. 203).

**Thodoris Dimitrakos** in “Historicizing Second Nature: The Consequences for the Is/Ought Gap” deals with the issue of the status of normative explanations through the analysis of the dispute between scientific naturalists and normativists. The former take it that normative explanations are reducible to empirical-scientific explanations, while the latter claim that normative explanations are genuine. In his analyses, Dimitrakos takes into account the connection between McDowell’s understanding of autonomy and the content of this concept in the framework of German idealism.

The account given by the author “is naturalistic insofar as it leaves nothing ‘beyond the reach’ of scientific understanding and presents scientific explanations as constitutive of the space of reasons. It is also liberal in the sense that it rejects the eliminability of normative vocabulary and retains the genuineness of normative explanations.” (p. 64) Dimitrakos explains how rationality is not a power outside of nature but a capacity to take control of our lives by understanding how the causal goings-on work” (p. 64), he defends the autonomy of the space of reasons arguing that “the layout of the space of reasons is historically changeable.” (p. 65) The author questions the strength of John McDowell’s response, particularly aiming to “scrutinize John McDowell’s attempt to defend the genuineness of normative explanations by adopting the notion of second nature.” (p. 51) In general, second nature is an acquired nature, while first nature is the one we were born with. However, in different considerations, the term itself receives additional specifications. Dimitrakos carefully examines the func-

tion that the old Aristotelian notion of second nature has in McDowell's arguments in support of his thesis that "whatever is law-governed is part of first nature while the rest of nature consists of events belonging to the domain of second nature". The author underlines the problems of this thesis, but also those related to its revision, which McDowell later presents.

It is relevant to remind the readers that the concept of second nature explicitly enters into the definition of *knowledge how* given by Gilbert Ryle in his ground-breaking book *The Concept of Mind* (1949). It is therefore puzzling that McDowell in his famous book *Mind and World* (1994) not only overlooks this, but emphasizes in several places that this historically valuable concept is, in philosophy of his time, quite forgotten.

Finally, Dimitrakos considers Hume's law according to which one cannot derive ought-statements from is-statements. Contrary to that Dimitrakos believes that "the gap between is-statement and ought-statement is not completely unbridgeable". In this context, Anselm's credo that "something is true when it is as it ought to be" João Carlos Salles reminds us of, is a good introduction to Salles' paper.

## 2. Normativity and Knowledge

João Carlos Salles in "Ernest Sosa's Telic Virtue Epistemology" deals with telic normativity as something which is normally inherent to human performances, because they are goal oriented and valued accordingly. The specificity of Sosa's theory of general human performances, as Salles indicates, is that it "reiterates essential traits of a normative perspective when applied to knowledge." (p. 78)

The author analyzes Ernest Sosa's model (given in *Epistemic Explanations: A Theory of Telic Normativity and What It Explains*, 2021) for evaluating epistemic performances, based on epistemic modalities such as "sensitivity," "safety," and "security." The analysis is given in the context of Timothy Williamson's objection to Edmund Gettier's critique of the traditional definition of knowledge (K) as justified true belief (JTB). The author notes that the Gettier problem was what attracted Sosa to epistemology, and underlines how inspite "the strong influence of Williamson, with his formula of great rhetorical appeal (knowledge first!), the program remains alive, although now more for its stability than for its effervescence." (p. 77) However, Salles argues that "Williamson would only be right if our task were to look for definitions in the form of necessary biconditionals." (p. 77) Salles' opinion is that an interesting philosophical analysis of knowledge has a different form; one is given by Sosa in his unpublished manuscript which the author quotes:

“K is present, when it is, in virtue of, or grounded in, JTB. Or: Always, when JTB is present, then K is *thereby* present. And, moreover, when K is present, that is *because* JTB is present.”

So, as a response to Williamson’s claim (from *Knowledge and its Limits*) that knowledge is not analyzable, the difference between mere analysis and philosophical explanation is stressed.

Salles concludes that “Sosa’s current reflection, which seeks to analyze the components of a foundation in primary conditions for knowledge, reiterates the normative features of his all-encompassing perspective while taking a new step in his epistemology of virtues – now extended, improved, and ready to become a Dawning Light Epistemology, which will soon entirely deserve our attention.”

**Timur Cengiz Uçan** in “Machines and Us: The Comparison of Machines and Humans at the Test of the Problematic of Solipsism” does not address the issue of normativity directly, but rather approaches it through considerations of solipsism. I would call the approach to the question of normativity that Uçan takes Austinian, in his examination of excuses Austin observes that “so often, the abnormal will throw light on the normal” (“A Plea for Excuses”, *Proceedings of the Aristotelian Society*, New Series, Vol. 57 (1956 – 1957), p.6). And indeed, as Uçan quotes from the classic work of C. I. Lewis *Mind and the World Order* (1929), solipsism annihilates. There are two relevant quotes from Lewis’s “Experience and Meaning” (1934) that Uçan takes as a starting point. One is: “Descartes conceived that the lower animals are a kind of automata; and the monstrous supposition that other humans are merely robots would have meaning if there should ever be a consistent solipsist to make it.” The other is: “A robot could have a toothache, in the sense of having a swollen jaw and exhibiting all the appropriate behavior; but there would be no pain connected with it. The question of metaphysical solipsism is the question whether there is any pain connected with your observed behavior indicating toothache.”

Uçan explores, with the help of C. I. Lewis, Alan Turing and Ludwig Wittgenstein, the limits of the comparison between humans and machines. Specifically, he argues that “Turing achieved, under a description, in ‘Computing Machinery and Intelligence’, exactly that which Lewis argued against.” Uçan analyzes how Turing’s results lead to a change in understanding consciousness, which implied a different understanding of solipsism. He draws attention to the fact that even if it is accepted that machines can think and have emotions, the question can be raised whether they feel thoughts and emotions in the same way as we do. Finally, Uçan makes valuable comparisons between Wittgenstein’s thinking about

pain and Lewis's, especially in connection with the following paragraph of Wittgenstein's *Philosophical Investigations* (1953):

“But can't I imagine that the people around me are automata, lack consciousness, even though they behave in the same way as usual? — If I imagine it now—alone in my room—I see people with fixed looks (as in a trance) going about their business—the idea is perhaps a little uncanny. But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: 'The children over there are mere automata; all their liveliness is mere automatism.' And you will either find these words becoming quite meaningless, or you will produce in yourself some kind of uncanny feeling, or something of the sort.” (§420)

There is a famous observation made by Richard Gale in the book *Divided Self of William James* (Cambridge University Press, 2007), he says that one “gets the feeling that Wittgenstein wrote his *Philosophical Investigations* with an open copy of *The Principles of Psychology* before him, especially the chapter on “The Stream of Thought”” (p.165) The paper of Timur Cengiz Uçan makes one think that Wittgenstein was equally inspired by James's student C. I. Lewis.

### 3. Normativity and Collective Intentionality

Vasiliki Xiromeriti in her rich paper “Collective Deliberation in Epistemic Groups: Lessons from Deliberative Democracy” deals with the question how collective views are deliberately shaped by epistemic collaborations aiming at joint actions. The author relies on Michael Bratman's account of shared intentionality and shared agency, particularly exposed in his book *Shared Agency: A Planning Theory of Acting Together* (2014). Xiromeriti accepts that “shared intentions should not be understood as attitudes in individual minds (i.e., ‘I intend to do my part in our shared action’), nor as attitudes belonging to a collective super-mind – for such a mind does not exist. Rather, a shared intention is a network of appropriately interconnected individual intentions.” (p. 132) Deliberative democracy, as the author notes, is discussed as a theory of political justification, but, she thinks, the normative role of deliberation in genuinely epistemic contexts is underexplored. Xiromeriti rightly observes that “if epistemic questions can have more than one permissible solutions and cannot be tracked by appealing to some overriding principle (e.g., evidence), they need to be settled through argumentative interaction.” (p. 131)

The account the author proposes is described as dialectical and based on Bratman's analysis of shared intentional action which has the following

two advantages: “First, Bratman’s account makes the fewest possible normative assumptions on the ground level of explanation. For shared intentional action to be possible, it is not required that there is a strong institutional background or, more generally, a consensual evaluative or cognitive background among participants. Thus, it makes it possible to consider collaboration even in contexts where substantial disagreement prevails – in interdisciplinary groups, for example. Secondly, Bratman does not take shared intentions as given but addresses the process by which they come to be developed. Individuals may have a plan of acting together, but this plan needs to be ‘filled in’ through reasoning on the part of the members of the group.” (p. 132)

Xiromeriti considers the role of collective deliberation in epistemic collaborations focusing on the question of group epistemic justification and underlines that for successful epistemic collaboration, it is important to have well defined rules and processes that govern deliberation, since “the norms are vital for ensuring a fair, inclusive, and rational discourse among members.” (p. 142) The author concludes by explaining that her paper “contributes to a deeper understanding of how collaborative interactions shape knowledge production in interdisciplinary settings and enriches discussions on the intricate processes involved in collaborative knowledge creation.” (p. 146)

**Ognjen Milivojević** in his paper “Searle and the creation of social norm” deals with John Searle’s account (given in *The Making of Social World: The Structure of Human Civilization*, 2010) of institutional social reality which is characterized by collective intentionality. Instead of “institutional social reality” the author uses the phrase “normed social reality”. By normed social reality Milivojević understands “a network of assigned status functions connected in the two basic ways as described above, by stacking status functions on top of each other or by horizontally assigning multiple status functions to the same object, where the original nodes in the network are natural objects,”(p. 156) simply “a normed social reality is a network of social norms.” (p. 156) Searle argues that social norms are created through speech acts which are status function declarations (i.e. a system of rights and obligations). Through these acts the reality is both represented and changed. Namely, Searle claims that the speech acts constitutive of social norms have a double direction of fit. In the text, Milivojević discusses the argument that Arto Laitinen in “Against representations with two directions of fit” (2014) gives against this thesis.

In particular, Laitinen claims that the notion of the double direction of fit is inconsistent since it leads to a vicious regress. As the historical roots of the idea of double fitting, Milivojević cites G.E.M. Anscombe’s



*Intention* (1957), while the term itself was coined by John Austin in “How to Talk-Some Simple Ways” (1953), but I would suggest that the origin of this idea can perhaps be found in John Dewey in his *Logic* (1939), in the chapters on the matrices of inquiry.

The author agrees with Laitinen’s criticism but notes that Laitinen gives only a negative reaction to Searle’s thesis and does not present an alternative thesis to Searle’s. Hence, Milivojević develops a concrete alternative and proposes the speech act without a double direction of fit. He starts with the question how our experience of normative life is established by speech acts. In a typical social activity, Milivojević distinguishes the following two features of normed social reality, which he terms Motivation and Expectation: “subjects are motivated to act in accordance with social norms, and subjects predict or expect that others will follow the same social rules as they do.” (p. 159)

Further, Milivojević introduces the *law of collective action* as follows: individuals associate with each other performing a collective action, if and only if everyone makes a net gain in their well-being from association. (p. 160)

Finally, the author gives his original definition of the speech act that creates social norms as “a collective statement of preferred normed collective action. Its form is: we claim that each of us derives the maximum possible welfare by counting X as Y in an environment C and we believe that this is common knowledge.” (p. 160) Given that social phenomena include collective intentionality, Milivojević concludes that “a normed social reality is a system of intentions to hold certain Xs, originally natural objects, as Ys in contexts C, a system based on a shared belief system about which status functions to assign to certain, originally natural objects, in order to achieve a maximally beneficial outcome for each agent.” (p. 162)

#### 4. Normativity and Logic

Aleksandra Vučković in “Normativity and Truth in Naturalized Epistemology” reflects on the notion of truth implied by W.O. Quine’s naturalized epistemology. The author argues that “there is a sense of normativity in Quine’s naturalized epistemology, but only insofar as we are willing to accept his imperfect notion of the truth.” (p. 167) This text is included in the section “Normativity and Logic” and not in the section “Normativity and Knowledge” for the following reasons. First, Quine’s naturalized epistemology is not a normative discipline, but an empirical one, so the author focuses on the issue of the notion of truth in this framework. Second, truth is supposed to be a normative notion. As to the former, Vučković notes

that “Quine was reluctant to accept the absence of normativity, possibly out of fear that it would leave his epistemology hollow,” (p. 169) in relation to this, she quotes Quine who in his “Reply to Morton White” (1986) says that normative epistemology is the technology of truth-seeking.

Hence the author poses the question: “Do we agree that there is normativity in naturalized epistemology? And if we do, where do we seek the source of this normativity?” (p. 169)

Vučković discusses some of the solutions to Quine’s problem of normativity. For example, one solution consists in accepting natural science as the source of epistemic norms. Vučković asks another question: “Can we agree with the Quinean notion of continuity between natural science and epistemology and *still* believe some *purely epistemological questions* exist and require its unique type of approach and source of normativity?” (p. 173)

The author considers the distinction between zero, medium and high normativity claims. The zero normativity claims assume no normativity. In this context Vučković cites Jaegwon Kim’s and Hilary Kornblith’s criticisms of Quine. Paul Roth argues against it, embracing instead *medium* normativity, i.e. the claim that there is a weak kind of normativity in naturalized epistemology, and the high normativity claim (defended by Richard Foley) that there *are* norms in naturalized epistemology. Vučković notes that the high normativity claim is “inseparable from the idea that the pursuit of truth is the main quest of the philosophical and scientific enterprise”. (p. 175) Hence, the author discusses Hilary Putnam’s evaluation (given in “Why Reason Can’t be Naturalized” (1982)) of Quine’s enterprise according to which “the main problem behind Quine’s loyalty to the pursuit of truth lies in the disbalance between how he interprets the notion of truth as opposed to the other concepts of a similar ontological background”. (p. 176) Summarizing Putnam’s argument Vučković concludes: “the truly naturalist notion of truth would entail nothing more than a semantic device that allows switching from one linguistic level to another.” (p. 176) Finally, such a notion of truth the author very explicitly qualifies as imperfect.

Miroslava Trajkovski in “Normativity, validity and semiotic implication” introduces a non-standard implication and terms it *semiotic implication*. The author argues that semiotic implication is an important interpretative tool and stresses some of its formal characteristics which differentiate it from standard implication, hence the sign “÷” for it is introduced. The implication is written as:  $(x)(P(x) \div Q(x))$  and is read as: P is an index of Q, for any P and Q, and Q is an icon of P, for Q different from P. The semiotic implication is linked to semiotic validity which calls

into question its compliance with the norm. In particular, there are at least three deviations in connection with semiotic implication: deductive validity is defined through an index; the enthymeme might be taken as a category of natural reasoning and abductive validity is defined through an icon. However, it is argued that these deviations are not necessary. Formal features of semiotic implication are non-reflexivity, transitivity and it contraposes. Peirce's idea that different arguments correspond to different signs is accepted but an error in Peirce's argumentation is pointed out. It is argued that the sign should be sought in the major premise.

Aristotle talked about signs in the context of enthymemes, so the relationship between semiotic implication and enthymeme is considered. Alan Ross Anderson & Nuel D. Belnap, in their paper "Enthymemes", claim that intuitionistic implication is enthymematic, it is presented that the semiotic implication is enthymematic in a way complementary to the intuitionistic one: when the major premise in the syllogism is an intuitionistic implication, the minor one is not necessary, while if the semiotic implication is the major premise, then the minor premise is not necessary.

The formal importance of the semiotic implication is reflected in the fact that it enables writing in the object language that, for example, milk is an index of pregnancy, or that a human is an index of an animal, while an animal is an icon for a human being. Qualities are semiotically related to other qualities, for example, smoke is a sign of fire. In the paper a starting point is a pragmatic understanding of attributing to a cognitive subject that s/he has the concept X. If this is the case then the subject must connect this concept with at least one other concept which either implies X or is implied by X. Hence, the subject who claims "X(t)" must at least, for some Y and Z, have an understanding of either " $(x)(Y(x) \supset X(x))$ " or " $(x)(X(x) \supset Z(x))$ ".

## NORMATIVITY OF ART

### 1. Art Works and Cognition

Ted Kinnaman in "Normativity in Art in Kant's Aesthetics" puts forward an intriguing thesis that Immanuel Kant "might accept the 'death of the artist,' because rational agency plays no role in explaining a work's success, that is, its beauty." (p. 207) The author defends this claim by juxtaposing the following theses from Kant: "beauty is a type of cognition", "beautiful art is an art to the extent that it seems at the same time to be nature" and "beautiful art is art of genius." Kinnaman places these in the

context of evaluating some works of art that marked the conceptual art in general of the 20th century. As examples Kinnaman takes Marcel Duchamp's *Fountain*, John Cage's *Music for Change*, surrealists' game *Exquisite Corpse*, Jasper Johns' *Flag*, and Jackson Pollock's drip paintings. The author demonstrates that Kant's theory of beauty given in the *Critique of the Power of Judgment*, if understood correctly, illuminates art in our time, while as a central problem of this theory, he sees the reconciliation of subjective and objective dimensions of taste. In particular, we have an apparently tolerant credo "De gustibus non est disputandum," on the one hand, but we have the expectation that others agree with our taste, on the other. The tension between the two is crucial for the issue of aesthetic normativity, and Kant's solution, as Kinnaman understands it, is to borrow the normativity of taste from the critical account of cognition, since in both kinds of judgement, cognitive and of taste, we expect others to agree with our judgement. (p. 200) Kinnaman reminds us of Kant's telling that nothing "can be universally communicated except cognition, and representation so far as it belongs to cognition." (p. 200)

In order to examine what *cognition in general* is for Kant, Kinnaman turns to the *Critique of Pure Reason*, claiming that "[t]he problem of the normativity of taste in the *Critique of Judgment* has its roots in the center of Kant's critical project." (p. 200) On Kinnaman's reading, "cognition in general" refers to "the goal of systematizing empirical cognition". (p. 202) Hence, the author takes it that for Kant the beauty in art as in nature is the suitability of its object for integration into a system of empirical cognition. (p. 205) Kinnaman underlines that "[t]he role of the beautiful object in Kant's account is important because it is what makes beauty *normative*: Kant assumes throughout his critical writings that our goal in cognition is to get the world right, that is, to represent it accurately." (p. 206)

In the end, let's ask ourselves if artistic archetypes are so deep in us that individual philosophical insights (like Kant's) centuries precede the realization of a work of art that we judge (see) as beautiful.

Isidora Novaković in "Philosophical Value of Literature: Machiavelli and Shakespeare" deals with the question of whether art has, in addition to an aesthetic dimension, a cognitive one. The author enters into the debate between aesthetic cognitivists and aesthetic non-cognitivists, aiming to show "that philosophy and literature can talk about the same significant truths, e.g. about human nature, motivation and ambition". (p. 211) Novaković analyses how William Shakespeare's work can help us to better understand the concept of virtue and in this context provides important information about the history and etymology of this term. The author compares the method of argumentation using thought experiments,

which she sees as indispensable means of philosophical analysis, with the content of literary works. One can only agree with her words “Literature is a thought experiment, and thus has a cognitive function. Truth finds its place in literature, at least in the sense that fiction takes some elements from the real world, so it can teach us, say, history and geography.” (p. 226)

The work of Iris Murdoch is cited in support of such an attitude. As an important theorist in this field Dorothy Walsh and Christoph Baumberger are mentioned. Walsh, in addition to *knowledge that* and *knowledge how*, considers *knowledge what it is like*. Baumberger takes that in the case of the knowledge we get from literature, it is a special kind that differs from what we normally call knowledge. The author compares Shakespeare and Niccolò Machiavelli asking “whether Shakespeare’s texts can offer us a basis for normative claims about Machiavelli’s ideas. Does Shakespeare provide us with a set of norms and rules for the (right) interpretation of Machiavelli’s texts?” (p. 211) Novaković focuses on the idea of fortune in both authors and finds similarities. She reminds us that the external and independent of force fortune is one of the key notions of Machiavelli’s political philosophy standing in the beginning of every political action. In her analyses, Novaković relies on the analysis of political philosophy in Shakespeare’s plays, given by Kosta Čavoški. The author concludes: “The influence of fortune and *virtù* together is best seen in the examples of conspiracies, which both Machiavelli and Shakespeare especially scrutinize as the greatest threat to the ruler. And aside from all of that, both writers are very humorous and sharp, although not at all radical, which is indicated by their use of irony. Irony in their works can also be interpreted in a strictly political sense when we try to discover their own personal stances. The interpretation of the nature of men lies in the centre of their political stances and here they finally diverge. All of these topics provide us with a framework in which we can interpret the words of one more easily through the words of the other, primarily Machiavelli’s through Shakespeare’s examples.” (p. 227)

## 2. Art Works and Collective Intentionality

**Milan Popadić** in “Can a Monument be Bad? Normativity and Commemorative Values in Public Space” considers the relationship between different aspects of evaluating monuments, particularly the transformations of their commemorative function in public space. The author starts from defining a monument as “an entity (sculpture, building, landmark...) erected (or recognized) as a sign of memory of a person or event.” (p. 231) Put simply “a monument is a physical structure that has the property of

commemoration; commemorativeness is a permanent property of a monument; commemorative values during the lifetime of the monument are variable. For example, a memorial dedicated to an army general can at one moment be a monument to a liberator, at another a monument to a conqueror.” (p. 234) The author relies on the differentiation and interweaving between private, public and cultural commemorative values that create the current modes of recognition of the monument.

Historical changes bring differences in the evaluation of monuments in public space. In this sense, the fall of the Berlin Wall brought changes in a large part of Europe. Popadić notes that the status of the Berlin Wall itself has changed, it was not erected as a monument, but it has been preserved as a monument, it is a non-intentional monument. On the other hand, monuments erected in memory of somebody or something are called intentional monuments. Popadić particularly compares two intentional Belgrade monuments: The Victor made by Ivan Meštrović and the monument to Stefan Nemanja made by Alexander Rukavishnikov.

The former is “a standing bronze male figure in the nude with symbols of peace and war (a falcon in the left hand and a lowered sword in the right), commemorating Serbia’s victory over the Ottoman and Austro-Hungarian empires during the Balkan Wars and the First World War.” (p. 237) The latter “is dedicated to the medieval ruler of Serbia, Stefan Nemanja (c. 1113–1199), the founder of the Nemanjić dynasty (who ruled between 1166 and 1371). The monument has a very complex pedestal: on the scepter of Saint Sava (the first Serbian archbishop and the youngest son of Stefan Nemanja), there is a broken Byzantine helmet; on the inside of the helmet are scenes from the life of Stefan Nemanja.” (p. 238) Popadić takes it that “compared to the monument to Stefan Nemanja, ‘The Victor’ seems as a textbook example how a monument should look like and how should be placed.” (p. 239) But of the monument to Stefan Nemanja the author gives a very insightful assessment, when he says that it is “literally in a *depression* (in terms of urban morphology), It seems that nothing testifies so clearly and directly as this depressed monument does to our confusion about values of the past mean in the present times, about the replacement of the idea of historical meaning by mere material grandeur, about the substitution of eloquence by the accumulation of content, about the swap of cultural and national development by political authoritarianism...” (p. 240)

Ivan Popov in “When is art interactive?” considers the contemporary Bulgarian art scene which is, according to him, dominated by discussions about interactivity. He states that after 1990 there was a change in the attitude towards art in Bulgarian society. It is clear that the fall of the Berlin

Wall was a turning point, after which it was possible for Bulgarian society as a whole to get involved in the trends of breaking with the tradition of placing the observer outside of the artwork. The author expresses the opinion that this process is related to “the reception of Martin Heidegger’s philosophical project of overturning the subject-object dualism.” (p. 249) Popov states that discussions about interactivity dominate the Bulgarian art scene. Hence, he analyzes the very concept of interactive art, providing at the same time “a critical reconstruction of the opposition between the eye and the body.” (p. 247)

Popov observes the problem of interactivity from the perspective of the history of ideas and philosophy and introduces us to the relevant literature and viewpoints. He refers particularly to Martin Heidegger’s “critique of metaphysics overcomes the subject-object divide, declaring untenable the claim that the self constitutes a kind of tabula rasa, which discovers the external world only in its encounter with it.” (pp. 248–249) The main question that Popov considers is whether interactive art constitutes a separate category.

The text draws attention to the viewpoints according to which the term “interactive art” is itself problematic, and demands that it be more precisely defined. For example, Dominic McIver Lopes proposes to define it as “the audience-induced change in the structure of the work’s so-called vehicle (be it a narrative, a visual image, a musical structure, etc.)” (p. 252) Shelby Moser, on the other hand, searches for the criteria “with the help of which artistic interactivity can become an operative category which could be implemented in our commerce with art.” (p. 252) However, Popov stresses that “both authors speak of interactivity in terms of the modification of the medium of the artwork, but not the effect this process has on the physical state of the viewer (i.e. the interactor). Provoking a bodily reaction on the side of the audience can certainly be an element of the overall conception of the particular work, but obviously does play a minor role in the philosophical analysis of the notion of artistic interactivity.” (p. 252)

Together with the author, we can ask what social reasons lead to the erasure of the border between the work of art and the viewer. Does interactivity give the artist a causal role and take away the teleological one? Given that for decades, curators have been signing themselves as the authors of art exhibitions, with the introduction of interactivity, we can wonder are we entering a post-author phase, a kind of a slippery slope, the end result of which is that the author is not an author?







BALKAN  
ANALYTIC  
FORUM

**BAF<sub>1</sub>**: Normativity



# 1. NORMATIVITY AND EXPLANATIONS



Damir Smiljanić

## ON THE DISTINCTION BETWEEN DESCRIPTIVE AND PRESCRIPTIVE METAPHILOSOPHY

**Abstract:** The author discusses the role of metaphilosophy, which is primarily concerned with the nature of philosophical thinking. When dealing with this topic, Nicholas Rescher's division of metaphilosophy into descriptive and prescriptive will be important, as it reinforces the self-sufficiency of this discipline's tasks. While descriptive metaphilosophy investigates what (historically) counts as philosophy, prescriptive metaphilosophy asks what valid philosophy should be. Thus, one studies the factual situation (e.g. in history of philosophy) and the other tries to determine the value criteria for specific philosophical positions (in which case a position is valid, what makes an argument better than the other, etc.). I will consider the question of whether it is possible for metaphilosophy to be guided by cognitive values without losing its descriptive character. Perhaps precisely "positional neutrality" (in the sense that the metaphilosopher does not favor a particular philosophical position) is the key value, although its acceptance makes the relationship between philosophy and metaphilosophy even more debatable: Is metaphilosophy just a part of philosophy or a discipline *sui generis* with a distance to philosophy? Is it a useful tool for the philosopher or is it redundant? The author will try to defend the disciplinary autonomy of the metaphilosophical approach.

**Keywords:** metaphilosophy, descriptive/prescriptive, cognitive values, positional neutrality, Nicholas Rescher.

### 1. Why Metaphilosophy?

Although the term 'metaphilosophy' has been used for eighty years, it has not yet established itself in contemporary philosophical discourse. What could be the main reason for its non-acceptance by members of

the philosophical community? Maybe the conviction of philosophers that they are capable enough to reflect on their own opinion and that it is not necessary to “delegate” this self-reflection to some artificial “meta-discipline”. This kind of self-thematization found its expression both in the speculative tradition (as “thinking about thinking”) and in the logical one (the use of “object-language” and “meta-language”). What is the point of a discipline solely dealing with philosophy? Is it not enough to rely on the history of philosophy? Does this established discipline not tell us clearly enough what philosophy is and in what forms it appears? The main problem is precisely whether philosophers can be *impartial* enough when speaking about their own theories as well as whether historians of philosophy can convincingly present the *aporetic* character of philosophy and explain the meaning of deep divisions and debates in its field. Perhaps metaphilosophy is still able to reveal some aspects of philosophy that usually escape its self-reflection.

If philosophers would see the essence of their work clearly enough, metaphilosophy would probably not be needed. What is it that prevents philosophers from adequately presenting what qualifies philosophy as a special way of thinking and speaking about the world? I would say that the main barrier is the limitation of their *worldview orientation*, the fact that in discussions they take and defend a specific *position*. Many philosophers are highly convinced of the correctness of their own arguments from the start. Thus, they are so occupied with defending their own position that they are not interested in the potential strengths of opposing positions (in the form of convincing arguments etc.). However, it is precisely this – the course of discussion in case of opposing positions, the way one’s own position is presented and the exchange of arguments – that should be of interest to a discipline such as metaphilosophy. For descriptive metaphilosophy, it is not primarily important which of the positions is “right” but how to properly establish communication between advocates of opposing positions as an issue of its own.<sup>1</sup> Obviously, the problem of differentiating between philosophy and metaphilosophy is quite complicated. At this point, we would already like to

---

1 However, if we regard metaphilosophy as a prescriptive discipline, it will adhere to the pretension of philosophy to identify the “stronger” argument in the discussion of a problem and offer one of its proposed solutions as “appropriate”. Such a metaphilosophy can “serve” a particular philosophical standpoint. (Exactly this is the case with Nicholas Rescher, who, although advocating the independence of metaphilosophical thinking, prioritizes the position of pragmatic idealism in the philosophical discussion.)

emphasize that no insurmountable gap should be created between the two – it is possible to argue for two completely different interpretations: both that metaphilosophy is part of philosophy and that it takes place on a completely different level of thought.

The distinction between philosophy and metaphilosophy seems quite different from the distinction between object-oriented disciplines *within* philosophy, for example ontology and epistemology or ethics and aesthetics. In this case, the difference does not only concern the nature of the subject being investigated, but also the researcher's attitude towards the subject. A philosopher asserts a certain thesis in order to solve a problem and defends it with arguments against the objections of other philosophers or (s)he disputes the thesis of his (her) opponent, again via arguments. On the other hand, the metaphilosopher does not intend to solve philosophical problems;<sup>2</sup> (s)he rather describes the way in which philosophers present and defend their theses in discussions, without prioritizing one philosophical position over the other – from a structural point of view, all philosophical positions are equal. A philosopher evaluates the answers to philosophical questions, while a metaphilosopher merely states them without evaluation. The key difference between philosophy and metaphilosophy seems to be the involvement of *values* in the theorizing process – philosophers are guided by certain values when researching their subject, which then has an impact on the structure of their theories as well as on their relationship to other positions. Most philosophers would agree with this, and probably consider the pursuit of truth a core value. An idealist and a materialist base their arguments on different evaluative assumptions – that is the reason why their approaches to problems and the theses they represent in discussions differ.<sup>3</sup> Therefore, metaphilosophy must devote its attention to the underlying values of philosophical positions, which usually remain unrecognized by philosophers.

## 2. The Perspective of Orientational Pluralism

When it comes to determining the difference between metaphilosophy and philosophy, Rescher's metaphilosophical studies are useful in many ways. Nicholas Rescher (1928–2024) is known for his numerous

---

2 At least the one dealing with his/her discipline in a descriptive sense.

3 For Johann Gottlieb Fichte, these differences even depend on the *character* of the thinker. (Cf. Fichte, 1984, p. 17)

works in the fields of logic, philosophy of science, epistemology and history of philosophy. There is almost no area of theoretical philosophy, in which he has not left some (written) mark. However, it seems to me that his works in metaphilosophy are being neglected – but if a history of metaphilosophical thought is ever to be written, I think it would have to mention Rescher's name.<sup>4</sup>

In the context of problematizing the difference between metaphilosophy and philosophy, I believe his “Essay on the Grounds and Implications of Philosophical Diversity”, called *The Strife of Systems* (Rescher, 1985), to be the most significant. In that essay, Rescher deals with the aporetic nature of philosophizing, which means that philosophical problems cannot be solved due to philosophers basing their thought on incommensurable assumptions. Contrary to common opinions, according to which the entanglement of philosophers in insoluble controversies speaks in favor of the infertility of philosophical discourse, Rescher claims exactly the opposite – namely, that insolvability of problems is an eminent feature of philosophy distinguishing it from other forms of rational thought. Far from resorting to some kind of relativism, he believes that the answers to philosophical questions are not given arbitrarily, but according to a certain logic or, more precisely, according to what he calls the “imperative of cognitive rationality”: in accordance with certain theoretical values, philosophers are forced to maintain the consistency of their claims by eliminating conflicting options from the repertoire of possible answers. For this purpose, Rescher draws up a unique methodological operation – the creation of the so-called *aporetic clusters*:

An *aporetic cluster* is a family of philosophically relevant contentions of such a sort that:

- (1) as far as the known facts go, there is good reason for accepting them all; the available evidence speaks well for each and every one of them, but
- (2) taken together, they are mutually incompatible; the entire family is inconsistent.<sup>5</sup>

Thus, it is necessary to omit at least one statement in order to obtain a meaningful whole.

Take the following set of statements as an example of an aporetic cluster:

---

4 The following are just some of his metaphilosophical books: *Metaphilosophical Inquiries* (Rescher, 1994), *Standardism* (Rescher, 2000), *Philosophical Dialectics* (Rescher, 2006), *Aporetics* (Rescher, 2008), *Philosophical Textuality* (Rescher, 2010), *Metaphilosophy* (Rescher, 2014), *Philosophy Examined* (Rescher, 2021).

5 Rescher, 1985, p. 21.



- (1) Philosophical questions cannot be solved by the mind.
- (2) There is nothing more useful than the mind to help solve philosophical questions.
- (3) Philosophical questions can be satisfactorily answered.

Each of these statements makes sense in isolation, but when observed jointly, they make an inconsistent whole. Meaning can be restored by discarding one of the statements. If the first statement is rejected, then we are dealing with a *rationalist* point of view; if the second one is removed from the cluster, the result is a *transcendental-philosophical* standpoint, and someone omitting the third statement represents either an *agnostic* variant of *skepticism* accepting the meaningfulness of philosophical questions but considering them unsolvable, or a more radical, *nihilistic* variant of *skepticism* (in this case, the questions are simply regarded as senseless). As the configuration of aporetic clusters may vary, it is clear that in any philosophical discussion multiple positions may be taken by different thinkers, each of which has a certain degree of plausibility. Therefore, Rescher favors the position of orientational pluralism:

Metaphilosophical pluralism maintains that distinct and conflicting positions are always in principle available with respect to philosophical issues. A specifically *orientational* pluralism goes beyond this in holding that there are different cognitive/value schemes, diverse probative perspectives, relative to which discordant alternatives can be validated vis-à-vis their competitors.<sup>6</sup>

It is important to note that this type of pluralism does not equal relativism, because, according to Rescher, it is necessary to *evaluate* positions and, thus, determine their truthfulness. We are only able to evaluate a philosophical position if we ourselves accept some values (e.g. objectivity). We cannot persuade someone to change his or her worldview merely by arguments, no matter how skillfully outlined and logically constructed they may be – a possible change of one's world view is a matter of value re-orientation, when someone literally undergoes a "conversion"<sup>7</sup> as a result of adopting a different normative perspective. Philosophers do not simply try out which position would suit them better, they rather adjust their choice according to their guiding values. We could say that philosophers do not arbitrarily choose a position, but that some value in the background directs them to their point of view. Rescher sees the advantage of orientational pluralism in the fact that it can shed light on the "anarchy

---

6 Ibid., p. 123.

7 In Rescher's words: "changes of heart".

of philosophical systems”, the complete disunity in terms of common assumptions and the philosophers’ lack of a readiness for consensus, which, according to many thinkers (such as Immanuel Kant), prevented philosophy from becoming a real science. From the perspective of Rescher’s pluralism, one can understand why a universal consensus is not possible in philosophical discourse. There is no consensus between representatives of philosophical positions precisely because they follow different values and ideals. A positivist starts from completely different values than a metaphysician (perhaps even denies the existence of values as separate entities unlike the latter). However, according to Rescher, the lack of agreement among philosophers is not an argument in favor of disputing philosophy’s scientific character, but to admit that its rationality is different from that of the empirical sciences. Orientational pluralism also has certain interpretive implications. Understanding a position means that we have succeeded in recognizing its normative background, although understanding it does not automatically imply accepting it as correct (Rescher thus distances himself from the quasi-hermeneutic equating of understanding and acceptance, or appropriation of a position). Anyway, what is crucial to Rescher’s concept of philosophy is that philosophical questions are answered on the basis of appropriate cognitive values, and those answers are further elaborated within philosophical systems. One of the main tasks of metatheoretical analysis is then to provide insight into the diversity of evaluative approaches to philosophical problems.

So what is Rescher’s perception of the task of metaphilosophy? Should it reconstruct the normative dimension of philosophizing? Is metaphilosophy also oriented towards some values, when it attempts to show the essence of philosophical thinking? Or does metaphilosophy, on the contrary, need to be value-neutral in order to fulfill its “mission”?

Rescher makes a distinction between *descriptive* and *prescriptive* metaphilosophy: while the former only describes which positions exist in the field of philosophical discourse without striving to determine to what extent they correspond with reality, the latter evaluates philosophical positions based on a certain (normative) assumption on what should be “real” philosophy. In his opinion, descriptive metaphilosophy is not a part of philosophy at all, since it deals exclusively with facts instead of normative criteria. In this respect, this type of metaphilosophy shares a lot of similarities with the *history of philosophy*,<sup>8</sup> although it might also

---

8 We mean the kind of history of philosophy that is closer to historical science than the history of philosophy of the Hegelian type, which is guided by teleological and normative presuppositions.

be seen as closely related to *sociology* or *psychology*. (Let us leave aside the complex question, whether the history of philosophy is guided by implicit value judgments and assumptions when studying its subject.) The key here is to study something as it *is*, not as it should be. Keeping to the facts, Rescher believes that all philosophers could agree on what history is telling us about the development of philosophical thought from antiquity to modern times. This does not mean that descriptive philosophers agree on every issue, but rather that they respect the historical development, as history<sup>9</sup> still provides the possibility to present a philosophical position without strictly determining its truth. On the other hand, prescriptive metaphilosophy remains a part of philosophy as it functions in the same manner as philosophical thinking – by distinguishing adequate (successful, effective, superior etc.) positions, theses and arguments from inadequate (unsuccessful, ineffective, inferior etc.) ones. While descriptive metaphilosophers share the same historical presuppositions,<sup>10</sup> prescriptive ones differ in their assessment of certain philosophical theories as they follow different normative directives – thus, orientational pluralism is based on different values than absolutism, skepticism, syncretism<sup>11</sup>. There can be no consensus between these forms of thought on a metaphilosophical level, just as there can be no consensus on a philosophical level.

Considering that he prefers a normative approach, it comes as little surprise that Rescher attempts to justify the standpoint of orientational pluralism based on advantages it allegedly has over other standpoints. This type of pluralism itself has a double meaning: (1) in a descriptive sense, orientational pluralism describes which criteria of cognitive evaluation are applied by representatives of philosophical positions, and (2) in a prescriptive sense it gives a judgment concerning the value orientation of philosophers by claiming that the stated value criteria are applied correctly, although they only have limited validity (in this context, there are no universal values shared by all contributors to the discourse). Rescher sees the main advantage of orientational pluralism in the fact that it allows us to recognize philosophy as a significant cognitive activity that is also compatible with the orientation towards humanistically relevant values. At the same time, this pluralism makes it possible to maintain the strictly scientific status of philosophizing, regardless of the fact that the pretensions to reaching an “absolute truth” are given up.

---

9 At least in the sense that is closer to scientific historiography.

10 Which does not mean that they represent the same point of view when dealing with content-related issues.

11 Rescher uses the term ‘conjunctionism’.

According to Rescher, what gives orientational pluralism an additional advantage over other positions is the complexity of its use – it can be used both at the level of philosophical (“doctrinal”) and metaphilosophical (“metadoctrinal”) thinking. As a philosophical standpoint, it incorporates the rather “dogmatic” attitude that from one perspective it is possible to accept only one “correct” position when solving philosophical problems; as a metaphilosophical approach it allows and acknowledges the existence of multiple plausible perspectives and these can be taken by other thinkers, although the adherent of orientational pluralism does not agree with them on a doctrinal-philosophical level.

If this interpretation of Rescher is accepted, then the challenging question is how these two seemingly contradictory functions of the named pluralism may be reconciled. Almost in a phenomenological manner, Rescher assumes the possibility of changing one’s attitude by distancing oneself from substantially dealing with philosophical questions, literally: “taking a step back”, by which we “parenthesize” our doctrinal convictions; then “from a distance” we can see all possible perspectives on these issues. In other words, by suspending the philosophical attitude, we arrive at the level of a metaphilosophical consideration of facts in the field of philosophy. This brings to mind Husserl’s suspension of the “natural attitude”, thus opening up the possibility of transcendental reduction as a precondition for the eidetic investigation of phenomena. Taking into account the postulation of the dual use of pluralism (as a philosophical view on the issue itself and a metaphilosophical overview of possible perspectives), we could say that Rescher clearly separates factual inquiry from the axiological approach. He transfers that difference to metaphilosophy itself, thus distinguishing its descriptive and prescriptive (normative) variant. Can we, however, offer an alternative interpretation of metaphilosophy, its purpose and nature?

### 3. On the Relationship between Philosophy and Metaphilosophy

By separating the field of factual inquiry from the axiological one, Rescher gets a basis for considering the essence of philosophy, but also the relationship between philosophy and metaphilosophy. Philosophy itself is defined as a type of research on problems that is determined by certain values. Facts are secondary here – Rescher says that philosophy is “overcommitted” in the factual sense. This means that the facts must be arranged, placed in a certain conceptual scheme and processed by certain

methods. Rescher describes this process of the philosophical “editing” of facts as follows:

Philosophizing always moves through two stages. At first there is a “pre-systemic” stage, where we confront a group of tentative commitments, all viewed as more or less acceptable, but which are collectively untenable because of their incompatibility. Subsequently there comes a “systematizing” phase of facing up to the inconsistency of the raw material represented by the “data”. And this becomes a matter of eliminative pruning and tidying up where our commitments have been curtailed to the point where consistency has been restored.<sup>12</sup>

Philosophy is not guided by facts like empirical sciences, but by forms and schemes of their arrangement. That is why those determinants leading a philosopher to the solution to a problem are important to make him see the matter from a certain perspective: the materialist assumes the non-existence of a separate mental or spiritual substrate, so in the dispute concerning the immortality of the soul he will take up the position that the soul is perishable (or even that it does not exist at all); an idealist, panpsychist or spiritualist, on the other hand, will claim something completely different as he is guided by a different perspective and other values.

Descriptive metaphilosophy abstracts from the value of adequacy, so it is not interested in which of these positions is “right” but describes their solution to the problem from a methodological point of view and reconstructs the very course and outcome of the debate between representatives of such conflicting positions. Prescriptive metaphilosophy, on the other hand, is not so neutral, but gets involved in the very course of the debate, considering that a certain position is more adequate or closer to reality, i.e. to the solution to the problem – meaning that the metaphilosopher makes decisions (i. e. descriptions imply the possibility of making a value judgement). Rescher solves the question of the relationship between philosophy and metaphilosophy by saying that prescriptive metaphilosophy is part of philosophy, while descriptive metaphilosophy is not. The pure form of descriptive metaphilosophy is the history of philosophy. Metaphilosophy stands, so to speak, with one foot in the field of philosophy, with the other in the field of its history. Does this threaten the independence of metaphilosophy as a separate discipline? Should not metaphilosophy, given its very name, be “above” or at least “next to” philosophy? In Rescher’s opinion, this is only the case with its descriptive variant. Rescher seems to ascribe greater dignity to prescriptive metaphilosophy as it is interested in substantial philosophical

---

12 Rescher, 1985, p. 20.

questions and debates among philosophers. Descriptive metaphilosophy is predestined for more “modest” tasks, which it shares with the history of philosophy.<sup>13</sup> Thus, the central problem is the demarcation between philosophy and metaphilosophy: are they strictly separated or do they interact with each other? It seems that Rescher remains undecided, and the reference to values should help him to keep philosophy and metaphilosophy in close contact.

Now, let us try to think more radically and assume some kind of hiatus between philosophy and metaphilosophy. The difference between these forms of thought would be the discrepancy between two levels of thought.<sup>14</sup> If we accept that discrepancy, it seems insurmountable. Philosophy and metaphilosophy would have no points of contact. Rescher suggests one option to maintain such a radical difference: understanding descriptive metaphilosophy as a *historical* discipline. The historian of philosophy should not be interested in which of the philosophical positions correspond to reality, but in pointing out which positions have generally been established over the course of history and how they related to each other (were they in conflict, complemented each other, cooperated, etc.). True, Rescher does not exclude the possibility that a historian of philosophy interprets history in a way supporting his personal philosophical preferences or even teleologically determines his own system as the end point of historical development (the latter was the case with Georg Wilhelm Friedrich Hegel). This kind of history of philosophy itself has a philosophical character. And yet, a descriptive philosopher, like a historian of philosophy, should present the course of historical events in the field of philosophy as impartially as possible.

However, the question is whether even the kind of metaphilosophy with systematic interests can fulfill the requirements of an exclusively descriptive approach. Rescher himself cites an example of the so-called *systematology*, a metaphilosophical discipline investigating the structure and forms of systems in philosophy, while relying on the system concept of Franz Kröner, who advocated the thesis of the necessity of pluralism of philosophical systems. Systematology does not deal with the question of which of the systems is the “best” or “most advanced”, but with their de-

---

13 Can we reverse the perspective and claim that the history of philosophy is a sort of metaphilosophy? A question for discussion would be whether the history of philosophy is a philosophical discipline or a metaphilosophical subdiscipline.

14 One will think of the difference between an object-language and a meta-language where the levels of speech are clearly demarcated. By the way, in his *Philosophical Investigations* (§121) Ludwig Wittgenstein denied the existence of such a type of metaphilosophy that would relate to philosophy, for example as grammar according to the ordinary use of the language.

sign, the way in which one system can develop from another or, in turn, lead to the formation of other systems. Of course, when Kröner opposes Hegel's idea of a comprehensive system, he seems to be forcing a philosophical thesis (pluralism versus monism). This is, however, rather the assumption, from which systematology emerges – this sort of metaphilosophy assumes the *fact* of the existence of many systems, which cannot be denied, even by a radical absolutist like Hegel. A philosopher may dispute the existence of multiple systems claiming that only one system is valid, while the rest are apparent forms of knowledge, but a metaphilosopher (systematologist) simply bases his perspective on the fact that there is a multitude of systems and sees nothing problematic in accepting this. On the other hand, when systematology deals with philosophical systems, it abstracts from their doctrinal differences, and sees the systems as components of a wider constellation, in which consideration of assertive claims (germ. *Geltungsansprüche* (J. Habermas)) is not in the foreground. It is more important to consider the complex dynamic of intersystem relations that develops according to a special logic. Therefore, descriptive metaphilosophy is not necessarily historical.

But there is another catch. Namely, Rescher assumes that descriptive metaphilosophy is limited to establishing the factual situation in philosophy without evaluating the content of theories. However, the question is whether it is possible for a metaphilosopher not to be – at least implicitly – guided by some values. Suppose someone is a supporter of Cartesian dualism in the philosophy of mind. If (s)he wants to investigate the development and outcomes of the debate about the nature of the human spirit, as it has been conducted since Descartes outlined his dualistic standpoint, (s)he will ignore his philosophical position (metaphysical dualism) and with equal studiousness approach the analysis of a monistic or pluralistic point of view (e.g. Spinoza's or Leibniz's metaphysics). (S)he will ignore the problematic arguments in support of these theories, and even abstract from them. This is possible because (s)he distances herself/himself from her/his own philosophical claims and research interests in the given context. We can say that (s)he is guided by the values of *cognitive impartiality* and *content disinterestedness*. We could unite these under the term *value neutrality*. Some might claim that value neutrality is self-contradictory: it is obviously a *value* in the back of a metaphilosopher's mind when he (she) wants to describe a position or a discussion, because it *urges* him (her) not to judge what is claimed or discussed there. Thus, one of the main objections to *radical skeptics* was that by allegedly refraining from any position, they were already taking a certain position. Yes, but both the radical skeptic and the metaphilosopher abstracting from his/her dualistic preferences *suspend*

the philosophical position, so that they can move to a *completely different* level of thinking and speaking. After all, when could a certain philosophical problem be solved by abstaining from some doxastic attitude? Descriptive metaphilosophy should refrain from sympathizing with a particular position – this is already an axiological recommendation. But it is needed in order to maintain the difference between the two levels of thematizing things (philosophical speech about objects and metaphilosophical speech about the attitude towards the object). *Positional neutrality* is an important assumption of descriptive metaphilosophy, because without a distance to one's own philosophical "taste" it is not possible to adequately deal with metaphilosophy. Rescher solves the possible self-contradiction by taking a complex position of orientational pluralism, which combines philosophical and metaphilosophical usage. There are other positions that are also capable of dealing with the danger of self-contradiction: *radical skepticism*, *perspectivism*, *syncretism*, etc. Obviously, these are positions that are able to reflect on their own dependence on a certain point of view.

Finally, the normative background of the relationship between philosophy and metaphilosophy remains to be considered. Assuming that values are equally important for both philosophers and metaphilosophers, the question is who, in particular, should investigate the issue. In philosophy, a separate discipline – *axiology* – is responsible for dealing with values. We have seen that prescriptive metaphilosophy is in charge of issues of normative determinism of philosophy. Who, then, investigates the values guiding metaphilosophers? If it is an axiologist, does philosophy still take primacy over metaphilosophy and even make it obsolete? If the metaphilosopher himself/herself is able to see the values influencing his/her approach to philosophy, how does he/she overcome the evaluative limitations of his/her view? Might someone with a broader perspective on these values understand them better? Are we then dealing with a "super-theory" that is able to encompass both philosophy and metaphilosophy? In order to avoid the "third man" problem, I would suggest that we "leave the ball" in the "court of philosophy" and simply assign that task to axiology. It should be noted that – similar to orientational pluralism – axiology can have two applications: both philosophical and metaphilosophical. The *taxonomy* of values and perspectives associated with them should be left to that discipline. Then one can enumerate and describe purely cognitive<sup>15</sup>, practical, aesthetic, culturally accepted, and other values. Thus, a new field of research for philosophy opens up here, especially as one gets the impression that axiology has been somewhat neglected since the days when the neo-Kantians dealt with it.

---

15 *Cognitive values* are especially important for metaphilosophers.



## 4. Plea for Philosophical Diversity

We should summarize the previous considerations on the relationship between philosophy and metaphilosophy. According to Rescher, metaphilosophy is a part of philosophy if it deals with the validity of philosophical theories from an axiological point of view; if it considers them from a purely historical point of view, then it occupies a place outside of philosophy. I have tried to revise that insight, which is itself metaphilosophical in nature: metaphilosophy can be positioned outside of philosophy if it accepts positional neutrality as the main guiding thread of its reasoning, and at the same time it can also have the character of systematic research and not be purely historically accentuated. In other words, metaphilosophy is a discipline dealing with philosophy, so *in that sense* it depends on the latter, but its approaches are independent and enrich our view of philosophy in any case.<sup>16</sup>

The point of metaphilosophy is not to overcome or even eliminate philosophy, but to *understand* it *better*. Therefore, metaphilosophy is not some kind of competitor to philosophy, but literally its assistant. However, even as such, it has its independence and it makes sense to deal with it. (I suppose Rescher would agree with this opinion.) Metaphilosophy gives expression to the desire for overcoming the limitation of the cognitive horizon when we reflect on our own position. After all, it has become clear that the position of the most comprehensive thinker is inserted into constellations that can only be seen if the multi-perspective character of philosophical thinking is taken into account. Philosophy necessarily has such a character and it is illusory to hope for the establishment of a comprehensive theory providing a satisfactory answer to the main philosophical problems at any time during its historical development. Should philosophers forget this, metaphilosophy is there to remind them that we benefit more from the debate of many different opinions than the domination of only one (dogmatic) opinion. In Rescher's words:

---

16 Examples of such metaphilosophical approaches can be found in the following publications: Karl Groos, *The Structure of Systems: A Formal Introduction to Philosophy* (Groos, 1924), Franz Kröner, *The Anarchy of Philosophical Systems* (Kröner, 1929), Eberhard Rogge, *Axiomatics of All Possible Philosophizing: The Fundamental Discourse of Logistics, Language Criticism and Metaphysics of Life* (Rogge, 1950), Hans Leisegang, *Forms of Thought* (Leisegang, 1951), and in recent times: Dimitrios Markis, *Protophilosophy: On the Reconstruction of Philosophical Language* (Markis, 1980), (of course) Nicholas Rescher, *The Strife of Systems: An Essay on the Grounds and Implications of Philosophical Diversity* (Rescher, 1985), finally two books with the same title: Richard Raatzsch, *The Philosophy of Philosophy* (Raatzsch, 2000), Timothy Williamson, *The Philosophy of Philosophy* (Williamson, 2021).

What counts as crucial from this overarching metaphilosophical standpoint is not the matter of ‘getting at the absolute truth’, but rather of enhancing the quality of the argumentation and gaining a deepened understanding of the structure of alternative positions.<sup>17</sup>

Therefore, the “anarchy of systems” is inevitable, but there is nothing wrong with that, on the contrary:

For centuries, most philosophers who have reflected on the matter have been intimidated by the strife of systems. But the time has come to put this behind us – not the strife, that is, which is ineliminable, but the felt need to somehow end it rather than simply to accept it and take it in stride. To reemphasize the salient point: it would be bizarre to think that philosophy is not of value because philosophical positions are bound to reflect the particular values we hold.<sup>18</sup>

In this sense: Let’s continue to fight for our values and enrich the philosophical discussion!

## References

- Fichte, J. G. (1984). *Versuch einer neuen Darstellung der Wissenschaftslehre* (1797/98). 2., verbesserte Auflage. Hamburg: Meiner.
- Groos, K. (1924). *Der Aufbau der Systeme: Eine formale Einführung in die Philosophie*. Leipzig: Meiner.
- Kröner, F. (1970). *Anarchie der philosophischen Systeme*. Graz: Akademische Druck- und Verlagsanstalt.
- Leisegang, H. (1951). *Denkformen*. 2., neu bearbeitete Auflage. Berlin: De Gruyter.
- Markis, D. (1980). *Protophilosophie: Zur Rekonstruktion der philosophischen Sprache*. Frankfurt am Main: Suhrkamp.
- Raatzsch, R. (2000). *Philosophiephilosophie*. Stuttgart: Reclam.
- Rescher, N. (1985). *The Strife of Systems: The Grounds and Implications of Philosophical Diversity*. Pittsburgh: University of Pittsburgh Press.
- Rescher, N. (1994). *A System of Pragmatic Idealism*. Volume III: *Metaphilosophical Inquiries*. Princeton: Princeton University Press.
- Rescher, N. (2000). *Philosophical Standardism: An Empiricist Approach to Philosophical Methodology*. Pittsburgh: University of Pittsburgh Press.
- Rescher, N. (2006). *Philosophical Dialectics: An Essay on Metaphilosophy*. Albany: State University of New York Press.
- Rescher, N. (2009). *Aporetics: Rational Deliberation in the Face of Inconsistency*. Pittsburgh: University of Pittsburgh Press.

---

17 Rescher, 1985, p. 273/274.

18 Ibid., p. 277.

- Rescher, N. (2010). *Philosophical Textuality: Studies on Issues of Discourse in Philosophy*. Frankfurt am Main et al.: Ontos Verlag.
- Rescher, N. (2014). *Metaphilosophy: Philosophy in Philosophical Perspective*. Lexington Books.
- Rescher, N. (2021). *Philosophy Examined: Metaphilosophy in Pragmatic Perspective*. Berlin, Boston: De Gruyter.
- Rogge, E. (1950). *Axiomatik alles möglichen Philosophierens: Das grundsätzliche Sprechen der Logistik, der Sprach-Kritik und der Lebens-Metaphysik*. Meisenheim am Glan: Westkulturverlag Anton Hain.
- Williamson, T. (2021). *The Philosophy of Philosophy*. Second Edition. Hoboken, NJ: Wiley-Blackwell.



Thodoris Dimitrakos

## HISTORICIZING SECOND NATURE: THE CONSEQUENCES FOR THE IS/UGHT GAP

**Abstract:** In this paper, I examine the normative problem from the standpoint of the following question: should normative explanations be regarded as a genuine form of explanation, or should we consider them as transient modes of intelligibility, requiring reduction to the valid explanatory frameworks of the empirical sciences? In particular, I scrutinize John McDowell's attempt to defend the genuineness of normative explanations by adopting the notion of second nature. I will show that, in light of the development of the social and behavioral sciences, McDowell's account lacks adequate defense against the skeptical arguments put forth by scientific naturalists. Thus, I will provide a further argument aimed at defending the autonomy of the domain of normative explanations without restricting the scope of the empirical sciences. I will conclude that the essential feature of second nature—or, more correctly, of rational second nature—is the human ability to cognitively grasp the causal goings-on that are part of both first and second nature. Finally, I will very briefly examine the consequences of my account for Hume's is/ought gap.

**Keywords:** Normativity, naturalism, John McDowell, Hume's law, social and behavioral sciences.

### 1. Introduction:

#### The normative problem as an explanatory problem

In this paper, my attention will be directed toward the normative problem, examining it through a distinct perspective—specifically, treating it as a matter of explanatory genuineness.<sup>1</sup> From this vantage point,

---

<sup>1</sup> The present paper is founded on the critical examination of John McDowell's 'naturalism of second nature,' as outlined in (Dimitrakos, 2020), and the logical distinction between normative and empirical-scientific explanations, as discussed in (Dimitrakos, 2021). However, here I also delve into the implications of my account for the is/ought gap.

a pivotal question arises: should normative explanations be regarded as a genuine form of explanation, or should we consider them as transient modes of intelligibility, requiring reduction to the valid explanatory frameworks of the empirical sciences? This, I take it, aligns with John McDowell's perspective on the normative problem: 'The problem posed by the contrast between the space of reasons and the realm of law, in the context of a naturalism that conceives nature as the realm of law, is not ontological but ideological' (McDowell, 1996, p. 78, fn 8).

From this perspective, we can frame the normative problem as a dilemma: either normative explanations lack genuineness, allowing them to be entirely assimilated into the realm of empirical-scientific explanations (as posited by the proponents of scientific naturalism), or normative explanations possess a unique status (they are *sui generis*), making their reduction into the explanatory patterns of empirical science impossible without sacrificing valuable information about the world (as asserted by normativists). Consequently, in the subsequent discussion, I will scrutinize McDowell's distinction between first and second nature primarily as an effort to uphold the authenticity of normative explanations. Additionally, I will try to show that, in light of the development of the social and behavioral sciences (human sciences for short), McDowell's account lacks adequate defense against the sceptical arguments put forth by scientific naturalists. Thus, I will provide a further argument aimed at defending the autonomy of the space of reasons without restricting the scope of the human sciences.

One final introductory remark is needed here. Normative concepts, explanations, etc. should not be considered as exclusively belonging to the domain of reason. The domain of reason is, of course, a normative domain but not necessarily *vice versa*. Many biological phenomena, for instance, are made intelligible in normative terms but not as rational phenomena. In what follows, for the sake of brevity, I will treat the normative realm as co-extensive with the domain of reason, even though, as I said, they are not. I acknowledge that a conceptual distinction between the rational-normative, the nonrational normative, and the non-normative domain is a philosophical necessity. However, in the present paper, I won't be concerned with the difference between the rational and the non-rational normative realm. I will focus on the contrast between the empirical-scientific understanding and the understanding which is proper to reason.<sup>2</sup> This is

---

2 By contrasting empirical-scientific explanations with the understanding proper to reason, I just follow McDowell's terminology. This terminology, by no means, implies that empirical-scientific explanations are not rational. The distinction lies in the presentation of the object of the explanation. Normative explanations portray the

why I am going to treat biological and other non-rational normative phenomena crudely as belonging to the non-normative domain. I will take the normative domain as co-extensive with the realm of reason, intentionally ignoring the part of this domain which belongs to the empirical-scientific understanding in order to examine the above-mentioned contrast in a simpler way.

My argumentation will take the following course. First, I will provide a distinction between the normative and the empirical-scientific explanations. Then, I will briefly present McDowell's both initial (*Mind and World*) and revised versions of second nature, focusing on the problems of these versions. In the fourth section I will list the possible normativist strategies for dealing with the problems in question. In the fifth section I will provide a further conceptual distinction between the reducibility and the eliminability of the normative realm. Next I will argue that the essential feature of second nature—or more correctly of rational second nature—is the human ability to grasp cognitively the causal goings-on that are part of both first and second nature. Finally, I will very briefly examine the consequences of my account in Hume's law, i.e. the logical gap between is-statements and ought-statements.

## 2. Normative vs Empirical-Scientific Explanations and the Logical Mapping of the Responses to the Normative Problem

Treating the normative problem as an explanatory issue requires distinguishing between normative and empirical-scientific explanations.<sup>3</sup> Normative explanations render actions or belief modifications intelligible by subsuming them under a norm or a set of norms. For example, if someone questions why I believe *q*, I could respond that I believe *p* and also hold the belief that if *p*, then *q*, i.e., I can explain my belief modification by showing how my belief modification conforms to the epistemic norm of *modus ponens*. Similarly, if asked why the person in front of me stopped her car and assisted a stranger experiencing a heart attack, I

---

object of their understanding as rational, which is why they are termed 'proper to reason,' while empirical-scientific explanations do not present their object (such as pendulums, tides, molecules, cells, etc.) as rational. I am indebted to an anonymous reviewer for this clarification.

3 My presentation of the distinction between normative and empirical-scientific explanations is inevitably sketchy here. For a detailed presentation, see (Dimitrakos, 2021).

could explain that she expressed solidarity with a person in need. In other words, her action can be explained by showing how it conforms to a practical norm. Conversely, empirical-scientific explanations make things or events in the world intelligible by showing how they conform to the causal order outlined by one or more empirical sciences. This is the case in which we explain someone's belief or action by appealing to the social milieu or the psychological mechanisms that dictate their thoughts and their behavior. For instance, the case where we explain someone's adherence to an ideology based on the interests of the social class to which they belong, or the case where we explain repetitive actions as an expression of obsessive-compulsive disorder.

Based on this logical distinction, we can provide the fundamental logical mapping of the possible replies to the normative problem. Scientific naturalists claim that every normative explanation is, in principle, reducible to a set of empirical-scientific explanations.<sup>4</sup> This entails that the domain of normative explanations is altogether eliminable. Normativists, on the other hand, argue that the domain in question is ineliminable because normative explanations (or at least some of them) are genuine, i.e., incapable of being reduced to empirical-scientific explanations. We cannot reduce genuine normative explanations to empirical-scientific explanations without losing crucial informational content about the world of human beings. However, it is crucial to discern two strands of thought within the normative camp. Idealistic views ground the ineliminability of the normative domain in some sort of non-natural entities (such as values or the soul as supersensible entities) or non-natural cognitive faculties (e.g., the Cartesian mental intuition). On the other hand, liberal naturalists defend the ineliminability of the domain of normative explanations within a naturalist framework that leaves nothing outside the realm of nature (De Caro & Macarthur, 2004, 2010, 2022). John McDowell's (1995, 1996, 2008, 2009, 2010, 2018) conception of second nature is one of the most well-discussed instantiations of the liberal naturalist perspective in contemporary analytic philosophy.

### 3. Second Nature as a Reply to the Normative Problem

McDowell claims that the normative problem is the result of a mistaken idea which takes the ontological territory of nature to be co-extensive

---

4 They have to be placed into the 'ordinary stream of explanation' (Turner, 2010, p. 11). The 'ordinary stream of explanation' is the network of explanations employed by the empirical sciences.



with the subject matter of empirical-scientific knowledge. He rejects this view by suggesting that 'nature includes second nature' (McDowell, 1996, p. xx). Second nature is an Aristotelian notion (*Nichomahean Ethics*) which refers to the acquired or cultivated aspects of human behavior and character that arise through conscious or unconscious effort and habituation. Unlike "first nature", which encompasses innate qualities and biological predispositions, second nature involves the development of virtues, skills, and ethical dispositions through practice and societal influence. McDowell adopts the concept of second nature in order to assert that individuals are born as mere biological creatures and are transformed into thinkers and agents, that is, rational animals, by engaging in the process of language initiation. Language does not serve only as a means for communication but also 'as a repository of tradition, a store of historically accumulated wisdom about what is a reason for what' (McDowell, 1996, p. 126).

The notion of second nature plays a dual role. On the one hand, it helps to sever the conceptual tie between the ontological notion of nature and the subject matter of the empirical sciences, thereby avoiding the eliminability of the normative domain proposed by scientific naturalists. On the other hand, since second nature is still nature, it helps to avoid idealism or supernaturalism (or 'rampant platonism' in McDowell's terms), which would make our capacity to respond to reasons 'look like an occult power' (McDowell, 1996, p. 83). However, to effectively play this dual role, the concept of second nature demands several philosophical explications that confront various puzzles. These puzzles forced McDowell to revise the notion that he initially presented in *Mind and World*.

### 3.1. The Mind and World version

In *Mind and World*, McDowell rejects the assumption that the domain of nature is co-extensive with the subject matter of the empirical sciences by distinguishing between first and second nature. In the terminology that I use in the present paper, second nature is the subject matter of normative explanations while first nature is the subject matter of empirical-scientific explanations. In this version of second nature, emphasis is placed on a nomological conception concerning scientific explanations. In essence, McDowell assumes that the empirical sciences render phenomena intelligible by subsuming them under natural laws. Law-governedness is the essential feature of first nature. Consequently, he associates empirical-scientific explanations with nomological explanations, adopting implicitly a neo-positivist (Hempel & Oppenheim, 1948) perspective on scientific

explanations.<sup>5</sup> The specific kind of intelligibility attributed to second nature involves situating an event within the space of reasons. Thus, the differentiation between first and second nature was grounded on the logical distinction between the realm of (natural) law and the space of reasons (McDowell, 1996, p. 78). In short, in *Mind and World's* version of liberal naturalism, whatever is law-governed is part of first nature while the rest of nature consists of events belonging to the domain of second nature.

### 3.2. The revised version

McDowell revised his account by taking into consideration a broader conception of natural sciences, which encompasses more than just mathematical physics and the relevant explanations based on law-governedness. He included biology in his philosophical framework, leading him to the conclusion that second nature is not exclusive to human beings<sup>6</sup>; it is only a necessary, though not sufficient, condition for rationality. The acquisition of rationality presupposes second nature, but the existence of second-natural phenomena does not necessarily guarantee the presence of rationality. Some second-natural phenomena require an intelligibility that is not significantly different from the intelligibility needed to understand first-natural phenomena (McDowell, 2008, p. 220).<sup>7</sup> Only a subset of second-natural phenomena becomes intelligible by placement in the space of reasons.

This revision makes McDowell's account more refined and science-informed. It incorporates a more nuanced understanding of the explanatory patterns used in the natural sciences, without altering its fundamental idea: the intelligibility associated with placement in the space of reasons is *sui generis*, that is, 'beyond the reach of the natural-scientific understanding' (McDowell, 2008, p. 217). Importantly, this kind of intelligibility is concerned with phenomena belonging to nature; they are not spooky or occult. In the *Mind and World* version of the account, these phenomena were identified with second-natural phenomena, whereas in the revised version they are only associated with a subset of second-natural phenomena. However, this nuanced version leaves room for problems of another sort.

### 3.3. The problem: The naturalistic threat of human sciences

I claim that McDowell's account stops taking into consideration the lessons from the empirical sciences exactly at the point where the threat

5 McDowell (2000) himself recognized this conception as Russellian.

6 For instance, Pavlovian conditioning is a part of second nature not only of human beings but also of other mammals.

7 Take for instance Pavlovian conditioning.

is less acute. If after biology we try to incorporate human sciences into our philosophical image, the eliminativist naturalistic threat returns. The subject matter of the human sciences encompasses the entire spectrum of human actions and beliefs. Therefore, nothing within second nature can, in principle, be ruled out from the understanding offered by these sciences. It appears that the emergence of the human sciences introduces a more pressing philosophical worry than the challenge posed by the natural sciences. This challenge seems to directly threaten the autonomy of the space of reasons, attempting to render human second nature fully intelligible through explanatory modes foreign to any placement in the space of reasons. Scientific naturalists can claim, putting forth a history-informed argument, that the scientific revolution of the 16th and 17th centuries and the relevant emergence of mathematical physics set us free from the ancient and medieval superstition that considered first nature as the realm of final ends. Next, the emergence of biology and modern medicine exempted a part of second nature from the kind of intelligibility which is proper to reason. Now, with the inclusion of the human sciences, the rest of second nature is exempted from this supposedly unique understanding. One could conclude that the more we scrutinize reality, and while the empirical sciences become gradually more mature, the more apparent it becomes that the only legitimate form of intelligibility arises from the perspective of the empirical sciences. Ultimately, the scientific naturalist might posit that the only philosophy informed by science is the kind of naturalism that identifies nature with the subject matter of the empirical sciences.

Despite its monolithic conception of scientific explanations and, in one sense because of it, the *Mind and World* version of McDowell's account could provide a philosophical ground for the genuineness of normative explanations. In this version, the notion of second nature is co-extensive with the domain of application of the normative explanations, and hence the genuineness of the latter is secured by the existence of the former. The domain of normative explanations could easily be determined by the logical contradistinction to the notion of law-governedness. In the revised version, on the other hand, there is no such possibility. Second nature and the domain of application of normative explanations cease to be co-extensive notions. Second nature is divided into one rational and one non-rational part: one part consists of phenomena like Pavlovian conditioning, which are part of empirical-scientific understanding, and another part consists of phenomena made intelligible by placing them into the logical space of reasons. But then second nature cannot provide a philosophical ground for the genuineness of the normative explanations. We need a further criterion for demarcating the second-natural phenomena which have to be explained by the empirical sciences (e.g., biology)

from the second-natural phenomena which have to be made intelligible by placing them in the space of reasons, that is, by providing a normative explanation. In short, we need a criterion for demarcating rational from non-rational second nature. The logical contradistinction to the law-governedness can provide no help here.

#### 4. Responding to the Threat

It seems that if we want to defend the genuineness of the domain of normative explanations, we are left with the following options.

- (a) To endorse some version of *anti-naturalism* in the philosophy of human sciences, asserting that human sciences do not exclude normative explanations.

Anti-naturalism of this sort claims that there is a logical gap between the explanatory patterns of the human and the natural sciences. The most traditional expression of this view is Hermeneutics.<sup>8</sup> But this option suffers from two major problems. First, the evolution of the empirical sciences does not appear to support the stark differentiation between natural and human sciences. By highlighting the distinctions within both categories, the unity of natural and human sciences becomes plausible. The rejection of the monolithic view that the nomological model in natural-scientific explanations entailed has given rise to a more comprehensive understanding of scientific explanations. In this broader perspective, there is no clear-cut division between the explanatory patterns employed by the human and natural sciences. As Sandra Mitchell (2009, p. 131) points out, '[t]he types of knowledge gained of the social world are much like the types of knowledge we can claim of the biological world' (Mitchell, 2009, p. 131). Second, and more important for my argument, opting for the anti-naturalist approach in the philosophy of human sciences doesn't dissolve the philosophical problem concerning the relationship between nature and reason; rather, it relocates this issue to a different philosophical subarea. Instead of addressing why normative explanations are genuine and cannot be entirely reduced to empirical explanations offered by the

---

<sup>8</sup> Hermeneutics in philosophy of history and in philosophy of social sciences in general involves a distinction between Naturwissenschaften (natural sciences) and Geisteswissenschaften (human sciences), not just in terms of the scientific subject matter but also concerning the distinctive logical forms of explanation. Sciences of nature "involve Erklären (explanation by way of laws) while [sciences of spirit] involve Verstehen (hermeneutic understanding from the "inside")" (Macarthur, 2010, p. 134).

human sciences, the normativist now needs to answer why normative explanations within the human sciences domain are genuine and resistant to reduction to explanations of a different logical kind. The fundamental philosophical requirement to present an argument against the eliminability of the domain of normative explanations persists.

- (b) To establish a demarcation criterion, distinguishing cases that should be understood through empirical-scientific understanding from those that should be understood by placing them in the space of reasons. The Kantian categorical imperative is a paradigmatic expression of formulating such a criterion. It distinguishes between cases of heteronomy (i.e., those requiring understanding through empirical factors) and cases of autonomy (i.e., those that are genuine expressions of reason).

The fundamental flaw of this demarcationist view is that it is vulnerable to sceptical historicist arguments. Anti-normativists consider the diversity of normative contexts throughout history and the world to raise doubts about the genuineness of the domain of normative explanations. We can reconstruct their argumentation as follows: given that ‘most of the people in history and in the present were and are living in normative error’ (Turner, 2010, p. 181)<sup>9</sup>, the validity of appealing to binding rules of reason and of their corresponding explanatory power is called into question. This argument follows the structure of the pessimistic meta-induction against scientific realism, which underlines that all past scientific theories once deemed successful are now regarded as false. ‘Therefore, the pessimist concludes, current successful theories will turn out to be false as well’ (Mizrahi, 2013, p. 3210). Similarly, Turner (2010) argues that appealing to some normative force that binds rational beings, along with the demarcation of cases representing genuine expressions of rationality, is chimerical. This is because our historical record reveals that, by any universal criterion, most people in the past have lived in normative error. The error becomes evident as we can explain people’s actions and beliefs not by invoking the truth or goodness of their beliefs and actions, respectively, but by taking into account various biological, psychological, or sociological causal factors. Therefore, the pessimist concludes, no demarcation criterion is capable of singling out cases of genuine rational expression. Consequently, every aspect of human thinking and behavior should be considered as the potential subject matter of empirical scientific research.

---

9 The normative error here is nothing more than a deviation from normative standards. The historical evolution of the normative frameworks reveals that all previous human beings used to live contrary to the present normative standards.

McDowell correctly rejects both options for defending the ineliminability of the normative domain. He never refers to any kind of anti-naturalism in the philosophy of the social sciences and he explicitly rejects the demarcationist option in saying:

[...] there is no criterion, if by that we mean some general formula that it might be possible to apply to mark off genuine reasons from impostors. It is incumbent on one to reflect about whether what seem to one to be reasons really are reasons, doing one's best not to be taken in by impostors. And there is no straightforward, as it were mechanical, way to guard against the risk. Any general formula one came up with would itself come within the scope of the obligation to reflect (McDowell, 2010, p. 12).

McDowell doesn't seek an infallible criterion to distinguish genuine expressions of rationality. He emphasizes that while autonomy<sup>10</sup> is a capacity which can fail, this does not prove that it doesn't exist at all. Rational subjects may sometimes act or think based on what just appears to be a reason rather than responding to genuine reasons, a point acknowledged by the anti-normativist argument from normative error. This recognition leads to the distinction between genuinely normative facts and seemingly normative facts. However, accepting the existence of genuinely normative facts doesn't necessitate committing to an infallible criterion for their demarcation. The sceptical argument from normative error is only effective against normativist accounts striving to establish an infallible criterion for identifying authentic expressions of rationality. Normative foundationalism (establishing an infallible criterion) and normative scepticism (rejecting the domain of normative explanations due to their fallibility) are not our exclusive options. One can argue that humans possess the capacity to respond to reasons, even if occasional failures occur. In short, the sceptical threat of the argument from normative error applies only to normative foundationalism (i.e., demarcationism). But, as McDowell shows, the defense of the autonomy of the normative domain can also be articulated on the ground of normative fallibilism, thereby sidestepping the aforementioned sceptical threat.

Nevertheless, normative fallibilism alone cannot provide a decisive argument against the eliminative version of naturalism. It can only show

---

10 McDowell embraces the concept of autonomy within the framework of German idealism. Here, autonomy signifies the ability for self-governance or self-legislation, where individuals act in accordance with principles they determine through rational deliberation. This stands in opposition to heteronomy, where moral principles are imposed externally. In this context, autonomy is closely associated with the concepts of reason and freedom. As Sebastian Rödl (2007, p. 105) points out, '[i]t is the principle thought of German Idealism that self-consciousness, freedom, and reason are one.'

that it is a less vulnerable option than foundationalism. Scientific naturalists can still ask for philosophical grounds to support the idea that certain natural phenomena lie beyond the reach of empirical-scientific understanding. Given my preceding argumentation, these grounds can be provided neither by just invoking the concept of second nature nor by proposing a demarcation criterion to identify genuine expressions of rationality. McDowell's quietist approach, though, does not recognize the need for further arguments. He merely claims that liberal rather than scientific naturalism should be our default position. In McDowell's (2006, p. 237) words, the default view should be that 'human beings are unique among living things – outside the reach of the sort of understanding achievable by a scientific biology – in virtue of the freedom that belongs with our responsiveness to reasons as such, [...] unless it can be shown to be wrong'. But taking into consideration the human sciences, I think, makes McDowell's 'naturalism of second nature' more vulnerable to the crude naturalistic threats and his quietist strategy less convincing.

## 5. A Further Argument: Reducibility and eliminability

In my view, the defense of normativism requires a further argument which can rely on a conceptual distinction that is necessary for rejecting scientific naturalism. In particular, I claim that we have to distinguish between the concept of explanatory reducibility and the concept of the eliminability of the domain of normative explanations. Only lacking this distinction can scientific naturalism be plausible. I am going to provide the distinction and the relevant argument against scientific naturalism in two steps.

### 5.1. Reduction and Epistemic Normativity

First of all, in order to reduce normative explanations to scientific explanations we need the normative vocabulary at our disposal. The reduction of a normative explanation to a scientific explanation presupposes that the latter is true and hence the explanatory reduction justified. But this also presupposes the genuineness of the domain of normative explanations. The reduction should be normatively explained, otherwise its validity is at stake. The elimination of the normative vocabulary makes us philosophically blind to the distinction between simply being taken-to-be-justified and actually being-justified. Empirical-scientific explanations can only reveal why something is taken to be justified by a person

or a community and not why something is actually justified. For instance, when we emphasize the role of the social milieu of Weimar Democracy in the emergence of the Quantum mechanics,<sup>11</sup> we explain why specific scientists considered their theoretical choices rational. However, to determine whether those choices were genuinely rational, we also require the application of normative criteria. In this sense, empirical-scientific explanations cannot sustain the distinction between ‘taken to be justified’ and ‘actually justified.’

Nonetheless, every act of reduction of the normative to empirical-scientific explanations cannot but be presented as genuinely justified. It presupposes the distinction between being taken-to-be-justified and being-justified and hence preserves implicitly the autonomy of the domain of normativity. Therefore, while each normative explanation can be reduced to a set of empirical-scientific explanations, the domain of normative explanations is not altogether eliminable. To put it in a slogan, the reducibility of each and every normative explanation does not entail the eliminability of the domain of normative explanations. We cannot deny the autonomy of normativity and claim that cognitive acts of reduction are correct at the same time.

## 5.2. Rejecting God’s point of view

What makes the scientific naturalist image plausible is an implicit assumption that should be revealed and rejected. The assumption is that the act of explanatory reduction is performed from a standpoint external to the person or persons to whom the normative explanation is attributed. I have to be extremely sketchy here<sup>12</sup> but consider the following case: I perform an action because I am convinced that it is justified. Someone else, a sociologist for instance, explains my conviction in terms of a causal order entailed by sociology. This is a paradigmatic case of reduction for scientific naturalists. My action and the relevant beliefs have been taken out of the space of reason and have been placed in a social causal chain. In this sense, it is presumed that the domain of normative explanations has been impoverished while the domain of scientific explanations has been enriched. Scientific naturalists think that if we follow this logical path to its end, then we have to conclude that the domain of normative explanations is eliminable through the gradual reduction of its various parts. Now let’s take into consideration a slightly different version of this scenario. I

---

11 See (Forman, 2011).

12 See the examples in (Dimitrakos, 2020) for more details.



myself acquire access to the sociological knowledge that explains my action. Then I cease to believe that my action is actually justified.<sup>13</sup> I understand that my behavior was the result of social causes that forced me to act like this. In this version of the scenario, my space of reasons has not been impoverished. On the contrary, it has been enriched in an important sense. If the act of reduction is correct, my new space of reasons contains a few more justifications that prevent me from making a mistake on the issue. During reflective scrutiny about what is a reason for acting in a particular way, taking into consideration the sociological knowledge about my condition can prevent me from believing uncritically that my action is justified. Thus, embracing the cognitive content of the act of reduction enriches the space of reason because it leads to what we may call the *restriction of the possibility of normative error*.

Scientific naturalists are eliminativist with regard to normative content because they take for granted the first scenario. But the first scenario presupposes God's point of view: an external vantage point overarching thought and world. Envisaging the space of reasons as shrinking after every act of reduction requires the putative standpoint of a reason that is not affected by this cognitive act. Only from this standpoint can the normative vocabulary be considered eliminable. But the presupposition of this sort of vantage point is very problematic, especially for accounts which aim to be naturalistic. Therefore, the reduction of a normative explanation to a scientific explanation is not an episode toward the necessary gradual shrinking of the space of reasons, but an episode toward its potential expansion. Equating reduction with elimination presupposes that the subject of reduction is necessarily different from the object of reduction, and this entails that there is something like a super-reason that accomplishes the task of reduction, a super-reason that is not and could not be affected by this very cognitive act of reduction.<sup>14</sup>

---

13 Sometimes, this kind of knowledge doesn't automatically lead to the decision that my action is not actually justified. This occurs only if I realize that the rational ground of my action has disappeared, and additional conditions are necessary for this recognition. The process can be intricate, but its description is not the focus here. The key point is to draw attention to the fact that the cognitive act of reduction is not necessarily supplied by an external point of view to the agent. This suffices for my argumentation here. I owe this clarification to the comment of an anonymous reviewer.

14 Of course, there can always be cases where the human beings that are the objects of the empirical inquiry will never have at their disposal the result of this inquiry. But this is just a contingent issue. My point is that human beings can, in principle, be aware of the act of reduction of a normative explanation to a set of empirical-scientific explanations.

## 6. Rational Second Nature as Knowledge of the First Nature

Scientific explanations are constitutive for our freedom<sup>15</sup> rather than a threat to it. They help us become freer. They contribute to the restriction of the possibility of normative error and consequently prevent us from being mere slaves of the various causal goings-on, permitting us to take control of our lives. My main argument is that by rejecting the idea that explanatory reducibility entails the eliminability of the domain of normative explanations, we can create a suitable logical space between scientific naturalism and idealism. The perspective I am proposing is naturalistic insofar as it leaves nothing ‘beyond the reach’ of scientific understanding and presents scientific explanations as constitutive of the space of reasons. It is also liberal in the sense that it rejects the eliminability of normative vocabulary and retains the genuineness of normative explanations. Rationality is not a mysterious power outside of nature but a capacity to take control of our lives by understanding how the causal goings-on work, that is, by gradually knowing more about the cases in which we do not have control of our lives. Furthermore, it is impossible to get rid of the kind of intelligibility which is proper to reason without appealing to the extremely questionable idea of God’s point of view. Therefore, we can say that the essential feature of second nature—or more correctly, of rational second nature—is the human ability to grasp cognitively the causal goings-on that are part of both first and second nature. Second nature is the purely natural ability to adjust our beliefs and actions to the knowledge of the causalgoings-on which dictate our lives. In short, rational<sup>16</sup> second nature is the knowledge of the first.

As I have argued elsewhere (Dimitrakos, 2020), my account has two main philosophical consequences that diverge from McDowell’s liberal naturalist perspective. The first is that freedom, understood as a condition coextensive with the subject matter of normative explanations, is not an ‘either you have it or you don’t capacity’ (Pippin, 2008, p. 214). There are various degrees of freedom that depend on our knowledge of the causal

---

15 As I mentioned above (see fn 10), freedom is conceptually associated with autonomy and reason. Therefore, this proposition implies that scientific explanations are constitutive of our capacity for autonomy, i.e., for acting in accordance with right reason.

16 I want to stress that only the rational part of second nature can be identified as knowledge of the first nature. As we have seen, second nature can be divided into two domains: one that requires a kind of intelligibility not essentially different from that which takes first nature as its subject matter, and one that requires a kind of intelligibility that is proper to reason.

goings-on that affect us. The more we know the physical, biological, psychological, or sociological factors that affect our lives, the more we can endorse beliefs and undertake actions which can be considered as expressions of our rationality, and consequently, our freedom. The second is that the layout of the space of reasons is historically changeable. This kind of historicism does not entail normative relativism. The idea of the historical changeability of the layout of the space of reasons does not necessarily entail the idea that moral or epistemic judgements can genuinely change truth-value or moral value when they change historical context. It just implies the idea that the space of reasons is reorganized when new judgements become candidates for truth-value or moral-value, as the result of the ongoing expansion of our empirical-scientific knowledge of the causal factors that dictate our lives.

Now, I would like to briefly explore the philosophical consequences of my account for the putative logical gap between ‘is’ and ‘ought’ statements.

## 7. Blunting the Guillotine

The is/ought problem or gap is the central issue in contemporary metaethical conversations. It has its roots in what is often called Hume’s guillotine (Black, 1964), which challenges the logical possibility of deriving an evaluative or ought-statement from premises that are exclusively descriptive or is-statements. In a famous passage in *Treatise*, Hume writes:

In every system of morality, which I have hitherto met with, I have always remark’d, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz’d to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, ‘tis necessary that it should be observ’d and explain’d; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it (Hume, 2007, 3.1.27).

As the passage makes clear, Hume considers the derivation of ought-statements from is-statements inconceivable and he argues that the philosophers who make such derivations need to provide an argument which shows the sustainability of this kind of inference. In other words, he takes it that the default position should be that we cannot deduce normative conclusions from descriptive premises. The ‘dominant historical interpre-

tation' (Bělohrad, 2011, p. 263) attributed to Hume the idea that no further argument of this sort can be found and hence that the is/ought gap is unbridgeable. The 'informal philosophical thesis that one "can't get an ought from an is" or, less snappily, that purely descriptive premises never entail normative conclusions' (Russell, 2022), is often called 'Hume's law'.

I would like to argue that Hume's law seems to disregard the logical asymmetry between confirmation and refutation. Even if it is impossible to deduce an ought-conclusion from factual premises, it isn't equally impossible to deduce the refutation of an ought-statement from is-statements. Those are the cases where a normative error is revealed by the reduction of a normative explanation to a set of empirical-scientific explanations. Let me use an example. Someone, say X, believes that gender-based division of labor is right because men and women are substantially different. Therefore, they argue, this kind of division of labor takes advantage of the different skills and inclinations between people of different genders, and hence, maximizes the efficiency of the labor force in society. Let's suppose now that X becomes familiar with sociology, history, and gender studies. X now realizes that his conviction that men and women have substantially different inclinations and skills was the result of his immersion in the dominant patriarchal ideology. He may now think that it was a mistake to believe that gender-based division is right. It's not the case that we ought to follow a gender-based division of labor. The example shows, I think, that the gap between is-statement and ought-statement is not completely unbridgeable. Even if it is a fallacy to make a deduction of the following sort:

Is-statement1  
 Is-statement2  
 -----  
 Ought-statement.

It does not follow that it is equally fallacious to make one of the following sort:

Is-statement1  
 Is-statement2  
 -----

It is not the case that [ought-statement].<sup>17</sup>

The question now is whether the proposition 'it is not the case that [ought-statement]' is a factual or a normative one. If my previous argu-

---

17 For instance, it is not the case that gender-based division is right.

mentation about the epistemic normativity and God's point of view is sound, the propositions of the sort 'it is not the case that [ought-statement]' are not mere factual propositions. Only the propositions of the sort 'it is not the case that we think/believe/suppose that [ought statement]' are mere factual propositions. But, as I attempted to show above, the rejection of the philosophical presuppositions of scientific naturalism entails securing the distinction between being taken-to-be-justified and being-justified, and thus the distinction between the propositions of the sort 'it is not the case that [ought-statement]' and the propositions of the sort 'it is not the case that we think/believe/suppose that [ought statement]'. Hence, there are at least some propositions of the kind 'it is not the case that [ought-statement]' which are normative and they can be deduced by factual premises.<sup>18</sup> This means that the is/ought gap is not totally unbridgeable. By combining (a) the logical asymmetry of confirmation and refutation, (b) the rejection of the philosophical presuppositions of scientific naturalism, and (c) the possibility of reduction of (at least some of) the normative explanations to a set of empirical scientific explanations, we can reach the conclusion that we can infer normative statements from factual premises.

Two remarks are necessary at this point. First, the epistemology of the refutation of ought-statements appears way more complicated than the epistemology of the refutation of is-statements. For instance, refuting the conviction that gender-based division of labor is right appears extremely more complicated than refuting the conviction that all swans are black. However, as the Duhem-Quine thesis and other lessons from the philosophy of science teach, the refutation of factual propositions can be equally complex. Moreover, complexity is not the main point here. The point is whether or not there can be, in principle, an inference of normative statements from descriptive statements. Propositions of the sort 'it is not the case that [ought-statement]' show that there can be.

Second, one may object that this kind of proposition offers only a negative determination. They inform us about what actions we should avoid rather than providing positive guidance about what we ought to do. But statements, the objection could continue, need to have positive content in order to be normative. I wouldn't disagree with that. I am not suggesting that propositions of the sort 'it is not the case that [ought-statement]' can exclusively provide the normative content that we need. Nonetheless, I argue that even if they don't provide the entire content, it is difficult to deny that they do provide some content. As Hegel says in *Logic*, complet-

---

18 For instance, the sociological, historical, and other descriptive propositions about the dominant patriarchal ideology.

ing Spinoza's famous motto that determination is negation (*determination negation est*), every negation is also determination.

The one thing needed to *achieve scientific progress* – and it is essential to make an effort at gaining this quite *simple* insight into it – is the recognition of the logical principle that negation is equally positive, or that what is self-contradictory does not resolve itself into a nullity, into abstract nothingness, but essentially only into the negation of its particular content; or that such a negation is not just negation, but is the negation of the *determined fact* which is resolved, and is therefore determinate negation; that in the result there is therefore contained in essence that from which the result derives – a tautology indeed, since the result would otherwise be something immediate and not a result. Because the result, the negation, is a *determinate* negation, it has a *content* (Hegel, 2010, 21.38, emphasis in the original).

I think that this kind of relation between the normative and the factual creates a middle ground between scientific naturalism and idealism. This middle ground is necessary for liberal naturalism. The middle ground is created through a rejection of both the view that the domain of ought-statements needs to be reduced to the domain of is-statements (crude naturalism), and the view that ought-statements are absolutely independent of any factual information about the world (idealism). Liberal naturalism requires an intermediate position according to which factual statements about the world play a significant role in the determination of the normative content. However, they don't exclusively dictate this content.

## 8. Conclusions

The account that I proposed here is necessarily incomplete. I am aware that a lot of philosophical questions concerning the relationship between factual and normative content need to be answered. For instance, apart from the factual content provided by the cases of reduction of a normative explanation to a set of empirical-scientific explanations, what are the other sources of determination of normative content for an account that aspires to be liberal naturalist? These questions disclose difficulties since they can reinstate the oscillation between crude naturalism and idealism. Despite the difficulties though, I think that the way to articulate a coherent liberal naturalist account goes through the three theses I attempted to defend here:

- (a) Rejecting the idea that nature is coextensive with the realm of law is not sufficient in order to cope with the threat of scientific naturalism. We also need to reject the implicit idea that the re-

ducibility<sup>19</sup> of each and every normative explanation to a set of empirical-scientific explanations entails the eliminability of the domain of normative explanations.

- (b) In the light of the abovementioned rejection, the human sciences should be considered as constitutive for our freedom rather than a threat to it. Rational second nature should be first and foremost understood as knowledge of the first nature.
- (c) The is/ought gap is not absolute. Even if we cannot derive ought-conclusions (i.e. conclusions that consists in ought-statements) from factual premises we can derive refutations of ought-conclusions from factual premises.

## References

- Bělohrad, R. (2011). The Is-Ought Problem, the Open Question Argument, and the new science of morality. *Human Affairs*, 21(3), 262–271.
- Black, M. (1964). The Gap Between “Is” and “Should.” *The Philosophical Review*, 73(2), 165.
- De Caro, M., & Macarthur, D. (2004). Introduction: The Nature of Naturalism. In *Naturalism in Question* (pp. 1–20). New York: Columbia University Press.
- De Caro, M., & Macarthur, D. (2010). Introduction: Science, Naturalism, and the Problem of Normativity. In M. De Caro & D. Macarthur (eds.), *Naturalism and Normativity* (pp. 1–19). New York: Columbia University Press.
- De Caro, M., & Macarthur, D. (2022). Introduction. In *The Routledge Handbook of Liberal Naturalism* (pp. 1–4). London: Routledge.
- Dimitrakos, T. (2020). Integrating first and second nature: Rethinking John McDowell’s liberal naturalism. *Philosophical Inquiries*, 8(1).
- Dimitrakos, T. (2021). The Source of Epistemic Normativity: Scientific Change as an Explanatory Problem. *Philosophy of the Social Sciences*, 51(5).
- Forman, P. (2011). Scientific Internationalism and the Weimar Physicists: The Ideology and Its Manipulation in Germany after World War I. In *Weimar Culture and Quantum Mechanics* (pp. 27–56). London: ICP.
- Hegel, G. W. F. (2010). *Georg Wilhelm Friedrich Hegel: The Science of Logic* (G. Di Giovanni, Ed.). Cambridge: Cambridge University Press.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15(2), 135–175.
- Hume, D. (2007). *A Treatise of Human Nature* (D. F. Norton & M. J. Norton (eds.); Vol. 1). Oxford: Clarendon Press.
- Macarthur, D. (2010). Taking the Human Sciences Seriously. In M. de Caro & D. Macarthur (eds.), *Naturalism and Normativity* (pp. 123–141). New York: Columbia University Press.

---

19 I want to stress that reducibility here means only the possibility of reduction.

- McDowell, J. (1995). Two Sorts of Naturalism. In *Virtues and Reasons* (pp. 149–179). Oxford: Oxford University.
- McDowell, J. (1996). *Mind and world: with a new Introduction*. Harvard: Harvard University Press.
- McDowell, J. (2006). Response to Graham Macdonald. In C. Macdonald & G. Macdonald (eds.), *McDowell and His Critics*. Oxford: Blackwell.
- McDowell, J. (2008). Responses. In J. Lindgaard (ed.), *John McDowell: Experience, Norm, and Nature* (pp. 200–267). Oxford: Blackwell Publishing Ltd.
- McDowell, J. (2009). Naturalism in the Philosophy of Mind. In *The Engaged Intellect: Philosophical Essays* (pp. 257–278). Harvard: Harvard University Press.
- McDowell, J. (2010). Autonomy and Its Burdens. *The Harvard Review of Philosophy*, 17(1), 4–15.
- McDowell, J. (2018). Responses. In A. Abath & F. Sanguinetti (eds.), *McDowell and Hegel*. Berlin: Springer Verlag.
- Mitchell, S. (2009). Complexity and explanation in the social sciences. In C. Mantzavinos (ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice* (pp. 130–143). Cambridge: Cambridge University Press.
- Pippin, R. B. (2008). *Hegel's Practical Philosophy*. Cambridge: Cambridge University Press.
- Rödl, S. (2007). *Self-Consciousness*. Harvard: Harvard University Press.
- Russell, G. (2022). How to Prove Hume's Law. *Journal of Philosophical Logic*, 51(3), 603–632.
- Turner, S. (2010). *Explaining the Normative*. Cambridge: Polity Press.



## 2. NORMATIVITY AND KNOWLEDGE



João Carlos Salles

## ERNEST SOSA'S TELIC VIRTUE EPISTEMOLOGY<sup>1</sup>

**Abstract:** Ernest Sosa's work recently took the form of a theory of telic normativity. As he presents, telic normativity is inherent to actions and attempts that characterize human performances, being telic because they are aimed at ends and often normative because we say they are better if successful and, therefore, if they reach their objective. The specific aims of this paper are, at first, to show the core of his new theory, notably his model for evaluating epistemic performances, associated with the shifting of position in his work on epistemic modalities such as "sensitivity," "safety," and "security." In a second moment, we suggest how Sosa is going forward just after achieving the current stage of his reflection – that is, formulating a "dawning light epistemology."

**Keywords:** Ernest Sosa, normativity, epistemic modalities, telic virtue epistemology, dawning light epistemology.

1. Ernest Sosa is one of the most important contemporary philosophers. His intellectual contribution has been remarkable on the philosophical scene for around 60 years. Furthermore, it is a unique contribution. After all, since 1974, his work constitutes its proper field of reflection, the virtue epistemology, recently taking, in 2021, the form of a theory of telic normativity, which now, as a process in full swing, deepens as a Dawning Light Epistemology.

In this text, we will seek to present the unique model of normative evaluation of human performances, particularly epistemic performances, taking into account his 2021 book *Epistemic Explanations: A Theory of Telic Normativity and What It Explains*. This model shows us how Sosa, on the one hand, always faced the challenges of the research program sparked by Edmund Gettier, not accepting *tout court* the reprimand to the program made by Timothy Williamson.

---

1 This research is supported by the Brazilian agencies CAPES and CNPq.

To this end, we will briefly present (1) the characteristic features of the traditional definition of knowledge as justified true belief, (2) Gettier's famous critique, and (3) Williamson's objection. Such a succinct presentation highlights Sosa's path, offering him only a context. Thus, on the other hand, we will show how Sosa developed a complex explanation for a multifaceted phenomenon such as knowledge without limiting himself to a definition that would supposedly resist successive counterexamples. He successfully established the normative frameworks of a taxonomy.

Given the scope of this text, we will not be able to address specific points we nevertheless mentioned. Sosa's reflection is, after all, a constant work in progress. Moreover, even the recent publication of this normative model only propels him to the new steps that he has already presented at several conferences.

In this way, the specific aim of this presentation is to show a quintessential aspect of his 2021 theory (notably, the normative model and the shifting of position in his work on epistemic modalities such as "sensitivity," "safety," and "security"). However, it is also essential to indicate that Sosa is going forward just after achieving the current stage of his reflection – that is, now formulating a "dawning light epistemology," with which he benefits a lot from his more recent and peculiar dialogue with Wittgenstein's *On Certainty*.

2. Knowing is not as trivial as it seems. The term has different meanings, and it is often crucial to reduce it so that the epistemic aspect stands out. Let us take knowing here as an action focused on facts, whose truth would be presented through declarative sentences. This is the meaning that interests Sosa throughout his career. As a result, pragmatic components, however important they may be and despite the centrality they increasingly acquire in Sosa's work, are subordinated to semantic aspects, being decisive, for example, the relationship between the notion of knowledge and notions such as evidence, belief, and justification. Knowing would be, then, for Sosa, equivalent to getting it right through statements.

Some conditions are primarily accepted, as stated by the traditional view of knowledge as Justified True Belief (JTB). (i) For something to be known, it must occur and can then be represented in a true proposition. So, for something to be known, it must be true. There may even be knowledge of what it is not, but precisely that it is not. The reflection is extensive

in the history of philosophy, as in the Aristotelian lesson that a proposition tells the truth if it says things are what they are or if it says they are not what they are not – a lesson, moreover, reiterated by Sosa: “What is not so cannot be known to be so: if anyone is to know that  $p$  then at a minimum it must be so, it must be true, that  $p$ .” (Sosa, 1994, p. xi)

Of course, the truth of  $p$  is thus a necessary but insufficient condition for knowledge. After all (ii), much of what is true goes wholly unnoticed and, therefore, unknown. We must take this into account and believe that something is (or is not) in order to know about it. A subjective act is required, manifesting as a belief or a judgment, whether or not it can be proven. Therefore, to believe that  $p$  is also a condition for us to know  $p$ , but obviously, it is still not enough.

However, (iii) belief alone does not bring the truth. Furthermore, even the conjunction between belief and truth can be a coincidence, as when we get a multiple-choice question correct from a mere guess. In addition to believing that something is accurate, we need a justification for our belief – with which knowledge would be a true and justified opinion. As said in *Theaetetus*, in a classic version of the traditional definition, the possibility of presenting a path to truth is also an indispensable feature of knowledge: “(...) when someone gets hold of the true judgment of something without an account, his mind is in a state of truth about it but doesn't know it.” (Plato, 2014, pp. 95–96 (202c))

3. Suppose we were satisfied with this view of knowledge as Justified True Belief. In that case, we should only ask for more perfect justifications since merely to believe being justified is not objectively being justified – as believing in following a rule is not following that rule, to remember a Wittgensteinian dictum.

However, even that favorable JTB conjunction, although required, would be insufficient. With the famous Edmund Gettier's 1963 paper, “Is Justified True Belief Knowledge?”, we learned better: the conjunction between belief, truth, and justification does not guarantee that, finally, we know. Gettier's problem, with precision and clarity, would reveal the scandalous inadequacy of the standard definition of knowledge.

The two counterexamples of his paper became immediate classics, one conveying logical consequence in a calculus structure of analyzed propositions and the other in a calculus structure of unanalyzed propositions. In both cases, long story short, the epistemic consequence does not follow the logical consequence, and so we may arrive at the truth by chance. Therefore, the subjective search for more objective justifications would not satisfy our epistemic claims, as even a justification according to logical inferential guidelines may fall short of knowledge.

Taking the simple Gettieresque example of inference by addition, we have that, from the truth of 'p', we can infer the truth of the disjunction between this proposition and any other since the truth of 'p' is enough to guarantee the truth of 'p v q'. The logical consequence is perfect. It is indifferent whether 'q' is true or false, being sufficient the truth of 'p'. However, the proposition 'p' being false, the truth of the disjunction is reached by mere chance, thus depending on the truth of 'q', which was simply unknown before. Our belief about the truth of 'p v q' would be accurate and justified, but it would not be a case of knowledge since it does not follow the pattern of the intended normativity that should surround our getting it right.<sup>2</sup>

4. Because of such undesirable possibilities raised by Gettieresque situations, a rich and specific research program has been developed with great technical sophistication, trying to answer this central question: What are the necessary and jointly sufficient conditions for us to have any knowledge and, therefore, for our eventual access to truth not to be a work of chance?

Indeed, a logical justification only sometimes carries with it epistemic correctness. Whoever seems to arrive reasonably at the truth may still be threatened by illicit favors of fortune. So, from the 1960s onwards, we have a community prepared to use heavy analytical artillery to operate the minor technical aspects – an extensive community equipped with the appropriate tools to get down to the nitty-gritty of what previously did not even seem like a problem.

Much ink was thrown into papers, re-editing (we must admit) a kind of scholasticism, now with analytical philosophy's rigor and scientific flavor. So, it seemed to some to be a relief to imagine that such a research program could cease.

After all, despite the results and valuable side products, the program awakened by Gettier came to be seen as perhaps idle or stimulated by an illusion. When someone thinks they have reached a safe place, they likely will face the truth that there are many ways of being wrong.

If someone, on the contrary, is justified, it would be expected that the consequences drawn from initial beliefs do not reach the truth by mere

---

2 Sosa claimed that the Gettier problem was what first gripped him in epistemology. Indeed, he published one of the first reactions to the Gettier paper in 1964 and summarized its two examples: "Suppose S has good evidence for his belief that p, from which in turn he deduces that p v q. But, unknown to S, (~p) & q. So, all three conditions for knowledge specified in the view under examination are fulfilled; but we still do not want to say that *Sknows* that pvq." (Sosa, 1991, p. 15)

chance. In the case of luck, the beliefs being well founded, some inferential standard was disrespected. If so, the definition of knowledge would be in order. Although vague, it would be sufficient, even without indicating under what conditions it would be satisfied.

Nevertheless, being so, the definition itself would not rule out the possibility of a false justification, nor would it finally indicate what a reasonable justification would be, just as the competence of archers would manifest itself in how they can hit the target, even though they may miss.

5. It is beyond our intention and almost impossible to recompose the research program's very twists of examples and counterexamples. We only intend to highlight one milestone in the program's evolution, Ernest Sosa's virtue epistemology,<sup>3</sup> which offers, for instance, a sufficient response to one of the strongest objections to the program itself, the criticism of Timothy Williamson,<sup>4</sup> who believes that it is not possible to submit the notion of knowledge to an analysis, seeming to return the notion to the condition of a truism, without which would not even be possible to think and which, therefore, we fail to analyze.

Despite the strong influence of Williamson, with his formula of great rhetorical appeal (knowledge first!), the program remains alive, although now more for its stability than for its effervescence. After all, Williamson would only be right if our task were to look for definitions in the form of necessary biconditionals.

Moreover, the race would be vain and idle if we were hunting for counterexamples in which the equivalence between a justified true belief and knowledge would be false. In this case, above all, the notion of knowledge would be present in some way, even in a non-obvious way, in the examples of justified true belief, whose description would supposedly rule out such a possibility:

---

3 Facing the challenges posed by the analysis of knowledge, Sosa developed a path all his own with the proposition of a virtue epistemology, which is markedly normative. This truly original elaboration has its inception, we believe, in 1974, with the text "How do you know?". The famous 1980 text, "The Raft and the Pyramid," is also recognized as a seminal document of its genesis, so Sosa's epistemology of virtues can have two beginnings, both deserving of wide celebration. Regarding these two origins and the characteristic features of Sosa's epistemology of virtues, see our paper Salles, J. C. "The Amazing Mr. Magoo: To celebrate Virtue Epistemology's 50th anniversary," published in Spanish, with the translation by Carlos Caorsi, as "El increíble Sr. Magoo: Para celebrarel 50 aniversario de la Epistemología de la virtud," *Revista Elenkhos*, Vol. 6, N. 2, December 2023.

4 Timothy Williamson's *Knowledge and Its Limits* is deservedly one of the most influential contemporary epistemology books.

So, the interesting philosophical analysis of knowledge would go beyond the mere correlation of K with JTB, even beyond the proposed *necessary* correlation of K with JTB. The interesting philosophical proposal would be rather of the following form:

K is present, when it is, in virtue of, or grounded in, JTB. Or: Always, when JTB is present, then K is *thereby* present. And, moreover, when K is present, that is *because* JTB is present. (Sosa, 2023a, p. 2)

Even admitting that Williamson was correct in many aspects, Sosa then states: “epistemologists interested in Gettierology were addressing interesting philosophical explanations of how knowledge comes to be, of how it is metaphysically grounded.” (Sosa, 2023a, p. 2) In short, we are not in the game of mere analysis, but in that of philosophical explanation – a game that shall never cease and always needs justification. In other words, the normativity of the epistemic endeavor sustains its proper relevance and continuity – something that Sosa’s theory of normativity beautifully exemplifies.

#### 6. Let us show a few traits of Sosa’s theory of telic normativity.

First, his defense of such an explanatory exercise (never dissociated from analysis) implies preserving one specific aspect of the philosophical activity of analyzing knowledge, namely, its normative dimension tantamount to understanding knowledge as a unique example of human action aimed at purposes – which is, moreover, better exemplified by alethic purposes. Therefore, Sosa’s theory of general human performances reiterates essential traits of a normative perspective when applied to knowledge.

Of course, normativity has several meanings, but, in any case, if we think about epistemic normativity, it is appropriate to emphasize that truth has a normative tint. Remembering Anselm’s lesson, something is true when it is as it ought to be. Thus, in the end, the truth is a matter of correctness, the correctness appropriate in each instance. A proposition normatively must be able to mean what it intends to express with it, just as it must be able to express, when declarative, how things are; that is, it must be able to saywhathappens correctly. (Cf. King, 2006, pp. 214–219)

So, we can use the notion of normativity in epistemology’s field in a somewhat limited but quite reasonable way, indicating with it the select set of procedures adopted when we aim to achieve the truth competently and thus obtain knowledge. Thus, “epistemic norms” are simply the procedures we follow when thinking and reasoning competently.

Second, Sosa’s model for evaluating performances makes evident, being normative, the central phenomena of a telic theory, such as attempt, success, competence, aptitude, and achievement:



If an archer shoots at a certain target, we can assess that shot in various respects. First, does it *succeed*? Does it hit the target? Second, how competent is the shot? The arrow may exit the bow with an orientation and speed that would normally take it straight to the bullseye. Even if a gust diverts it, the shot might still be competent. It can be *adroit* without being *accurate*. And it can be accurate by luck, without being adroit. But even a shot that is both accurate and adroit might still underperform. An arrow adroitly released from a bow may be headed straight to the bullseye when a gust diverts it so that it would now miss the target narrowly, except that a second gust eases it back on course. The archer succeeds in that attempt to hit the target, and the shot is also competent, as the arrow leaves the bow perfectly directed and with the right speed. But the shot is accurate because of the lucky second gust, with a distinctive luck that repels competence. (Sosa, 2021, p. 18)

However, this is more than a simple example among many others. Sosa offers us a general model for the normative evaluation of human performances, a model that becomes even clearer when the telic horizon of performance is also alethic – a normative model for evaluating human performances aimed at goals and, in particular, an evaluation of the action of the epistemic agent. Triple AAA, therefore, is a normative model, an almost grammatical criterion on identifying a result with the agent's merit, a measure of attribution of responsibility and the meaning of the act itself.

7. Telic normativity is inherent to actions, such as attempts that characterize human performances.<sup>5</sup> It is telic because they aim at ends and often normative because we say they are better if successful and achieve their objective. It is also better for attempts to manifest competence and achieve success through competence, not by chance.

That is why we prefer persuasion to the use of force, an excellent diagnosis to mere guessing, the expert's advice to the charlatan's opinion. Also, we attribute merit to regular athletic performances rather than casual successes. After all, as Sosa reminds us, "to reach Larissa through ignorant luck is not to flourish." (Sosa, 2015, p. 142)

In light of a theory of telic normativity, the model accounts for performances in general and, in particular (as in the case of competing air gusts), of Gettieresque situations. The approach is, therefore, entirely normative, with prescriptions such as

1. It is better to be competent than inept;

---

<sup>5</sup> "Telic normativity is inherent in action, in attempts. An attempt is "better" (as an attempt), a better attempt, if it succeeds than if it fails; better if competent than if incompetent; and better if its success is apt, through competence rather than luck. This all seems trivial for instrumental, end-relative normativity." (Sosa, 2023b, p. 189)

2. It is also better to be successful than not to be successful;
3. Even better is to succeed based on our merit, our competence, and how appropriate our action is.

A telic and epistemic gesture following this model can then be defined as:<sup>6</sup>

- a) **not being knowledge (or not being an act creditable to the agent), if:**
  - 1)  $\sim$ Accurate,  $\sim$ Adroit and  $\sim$ Apt (like a blind shot that fails);
  - 2) Accurate but  $\sim$ Adroit and  $\sim$ Apt (like a blind shot that hits the target);
  - 3) Adroit but  $\sim$ Accurate and  $\sim$ Apt (like a shot made competently but which deviates from the target);
  - 4)  $\sim$ [Adroit and  $\sim$ Accurate, but Apt] (The negation of the conjunction indicates that it is a grammatical limit of the model (to use a Wittgensteinian expression), as it indicates something that, after all, can never occur. For it to be apt, it needs to be accurate and adroit as it is accurate because it is adroit.);
  - 5)  $\sim$ [Accurate and  $\sim$ Adroit, but Apt] (Another grammatical limit of the description);
  - 6) Accurate and Adroit, but  $\sim$ Apt (like a shot that reaches the target, but due to the effect of double bursts that compensate each other, this being the truly Gettieresque situation of a starting competence that is not maintained upon arrival, as the accuracy does not express competence, and it is not open to debate whether the presence of luck, in any case, suppresses telic credit);

A telic gesture can be defined as:

- b) **being knowledge (or an act creditable to the agent) at various levels or gradations if:**
  - 1) Accurate, Adroit, and Apt (thus, basic level of knowledge, animal knowledge, either because it does not involve reflection or because it is unsafe);

---

6 With minor differences (perhaps improvements), we recover below our interpretation of the telic model as a normative evaluation criterion, as presented before in our text "A Gnoseologiassegundo Ernest Sosa" (Revista *Trans/Form/Ação*, v. 44, p. 63–96, 2021), from which we also took some of the following considerations about modal conditions for knowledge.

- 2) Accurate, Adroit, and completely Apt (that is, reflective knowledge, which can be safe knowledge at a higher level, with safety no longer being a necessary condition for knowledge but rather a distinctive note of the higher level of knowledge, a security knowledge).

We then have a triple standard for evaluating performances (AAA), to which, by the way, is added a triple component of evaluating adroitness, as the reliability of the competence depends on a more internal ability (skill, preserved even when we sleep), but which can be compromised by our effective state (shape) and by the concrete conditions in which the performance develops (situation).

Such a telic evaluation model also has practical applications. They are complex and far from automatic. However, the model serves as a nuclear formula for what can be charged to all human activity and what can be used for its evaluation, with the specifications and additions that prove necessary – sometimes, case by case. Such a model offers a broad perspective for evaluating performances, opening up an arc of questions that crosses every purpose-driven human endeavor. Therefore, Sosa states: “Our aim is a theory of knowledge that will fit a unified and broad-scope epistemology.” (Sosa, 2017, p. 210)

In this case, pragmatic aspects previously subordinated to epistemic aspects acquire unusual relevance to fulfill this broader purpose. It is no coincidence that Sosa's work, in its current progress, benefits from a more intense dialogue like Wittgenstein's *On Certainty*. As we will show in another paper, the analysis of human performances makes Sosa's reflection, together with Wittgenstein's, move to the concrete ground of actions involving language games amidst life forms. Then, it will come to the fore notions like “background conditions,” “hinge propositions,” and “bracketed domains.”

8. As presented in *Epistemic Explanations*, the normative model argues the merit or responsibility of the epistemic agent and the agent in general. Reliability indicates the probability of an attempt's success, given the model's components. That would be, let us say, a “safety condition” so that an attempt results in a hit, being harmless to luck. Let us quickly recover the reasons for introducing the modal notion of ‘safety’ for knowledge evaluation.

One way of analyzing knowledge was to propose a modal addition by which belief would acquire an uncontroversial characteristic note that would elevate it to the rank of knowledge. For this, for example, the be-

lief should be sensitive so that, if it were not true, we would not have it ( $\sim p \rightarrow \sim Bp$ ) – if not  $p$ , then we would not believe that  $p$ . (Nozick, 1981)

Now, the condition of sensitivity is hardly sustainable and cannot resist skeptical threats. If there is *sensitivity*, if the facts were different, our senses (our gauges) would show something different. If what happens, if reality were different, the appearance would be different, and, therefore, our beliefs would be different.

However, the epistemic agent would be powerless in the face of the skeptic's claim that our experience can be illusory and misleading. In this case, our instruments would not indicate any change. They do not in themselves indicate, in the skeptical scenario, that they are disconnected, and we do not, after all, have access to the fact of access itself, as we have no connection with the connection – in short, we are not sensitive to the fact presupposed by the condition of sensitivity.

To avoid such an undesirable upshot, Sosa replaced sensitivity with a distinct epistemic modality, safety ( $Bp \rightarrow p$ ). As sensitivity and safety are not equivalent, it seemed advantageous that novelty<sup>7</sup> – namely, a belief capable of postulating the knowledge condition would be one such that we would have it only if it were true. Later, Sosa admitted that a belief can yet constitute knowledge, even if unsafe because it is overdetermined, but this applies only to the circumstance of so-called “animal knowledge.”

For instance, imagine that a baseball player's performance would be at risk because of the imminence of a power outage during a game. Can this risk deny the player's credit for hitting the ball? Can mere danger similarly suppress our epistemic capacity and our telic credit? Not at all. This confusion between knowledge and the presupposition of knowledge would lead us to suppress the possibility of human knowledge, its consequence being even more absurd, namely, the pure and straightforward assumption of no possible knowledge.

If we are lucky enough to escape this most devastating danger, this luck does not compromise our performance; in this case, we are not being reckless. For instance, the imminent risk of an electrical blackout, which was not confirmed due to extreme luck, does not suspend the credit of the victorious athlete hitting a ball.

As a result, Sosa has recently insisted that *not all luck corrupts a performance* – which, moreover, modifies the entire characteristic of the fa-

7 As Sosa explains in one of his most quoted texts, “[a] belief is sensitive iff had it been false, S would not have held it, whereas a belief is *safe* iff S would not have held it without it being true. For short: S's belief  $B(p)$  is sensitive iff  $\sim p \rightarrow \sim B(p)$ , whereas S's belief is safe iff  $B(p) \rightarrow p$ . These are not equivalent, since subjunctive conditionals do not contrapose.” (Sosa, 1999, p. 146)

mous Gettier problem. So, we can see that the possible abandonment of safety as a necessary condition for knowledge affects only the b1 clause (related to animal knowledge), being still a normative condition for full knowledge (i.e., our b2 condition of an epistemic gesture Accurate, Adroit, and completely Apt).

9. Let us summarily present some final considerations, either derivative of what we presented or signaling future steps we could not present in this paper.
  - (1) Normativity is at the basis of epistemic justification, which, as such, can both explain what it is to know and *ipso facto* justify the relevance of explanatory undertakings.
  - (2) Normativity touches on modal issues essential to judging reflexively in a chain of knowledge classification. The hierarchy of knowledge does not have strict philosophical interest; nevertheless, being normative is far different from a mere taxonomy.
  - (3) The reflection ends by asking for more answers about the agent's perspective to identify whether justified, for example, in suspending an action – and suspending a judgment is an exercise that can never be reduced to a mere causal impulse.

Thus, Sosa's current reflection, which seeks to analyze the components of a foundation in primary conditions for knowledge, reiterates the normative features of his all-encompassing perspective while taking a new step in his epistemology of virtues – now extended, improved, and ready to become a Dawning Light Epistemology, which will soon entirely deserve our attention.

A highly respected scholar, Tim Crane, announced to the Four Winds in a recent tweet: "I'm rather surprised to say this, but I think Ernest Sosa may have solved epistemology." If Crane is correct (and we hypothesize that he is) and Sosa's proposal has the intended scope, the limits of possible experience will have been redefined from an epistemic point of view as if Sosa carried out a kind of new Cartesian revolution.

To carry out such a "revolution," Sosa does not disregard the fruits generated by the Gettiresque race in search of an analysis of sufficient and necessary conditions for what constitutes knowledge. However, by presenting a normative model, his theory goes beyond a definition capable of resisting the invention of any counterexample. On the contrary, it benefits from the very counterexamples that challenge it since it leads us to explanations of how what we call knowledge occurs. In its complexity, the

model situates different cases within its normative frameworks without imprisoning the different situations in the same straitjacket. The model allows us, on the contrary, a more flexible taxonomy through which the complexity of phenomena that involve knowledge can find its place within the framework of a well-established normativity.

*Knowledge* can be technically defined as an apt belief. In that case, this does not transform the work of analysis into the mere search for biconditional statements but instead allows the explanation of diverse manifestations and, therefore, of distinct classifications, which even incorporate different modal gradations. Thus, being clear about what knowledge is not, we are also better informed about what, for instance, animal knowledge, reflective knowledge, and full reflective knowledge can be.

In any case, true or false, Crane's statement is powerful. It is equivalent to claiming the proof of a theorem that has resisted for centuries or that a consolidated theory has been refuted. However, in this case, the statement is more comprehensive. It challenges all of us (or, at least, the science community) to answer what knowledge ultimately is and to draw from it the consequences for the valuation of human activity in general. Crane's claim is not self-proving; it does not follow from his academic authority, but it certainly deserves some attention, as we have tried to grant here.

## References

- Gettier, E. (1963). Is Justified True Belief Knowledge?. *Analysis*23: 121–123.
- King, P. (2006). Anselm. In D. Borchert (ed.), *Encyclopedia of Philosophy*. Farmington Hills: Thomson Gale, vol. 1.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Plato. (2014). *Theaetetus*. Translated by John McDowell. Oxford: Oxford University Press.
- Sosa, E. (1991). *Knowledge in Perspective*. Cambridge: Cambridge University Press.
- Sosa, E. (ed.) (1994). *Knowledge and Justification*. London: Dartmouth Publishing Co.
- Sosa, E. (1999). How to Defeat Opposition to Moore. *Philosophical Perspectives*, 13.
- Sosa, E. (2015). *Judgment and Agency*. Oxford: Oxford University Press.
- Sosa, E. (2017). *Epistemology*. Princeton: Princeton University Press.
- Sosa, E. (2021) *Epistemic Explanations: A Theory of Telic Normativity and What It Explains*. Oxford: Oxford University Press.

- Sosa, E. (2023a). *Philosophical Methodology*. unpublished manuscript.
- Sosa, E. (2023b). Default Assumptions and Pure Thought. In L. R. G. Oliveira, (ed.) *Externalism about Knowledge*. Oxford: Oxford University Press.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Wittgenstein, L. (1975). *On Certainty*. Oxford: Blackwell.





Timur Cengiz Uçan

## MACHINES AND US: THE COMPARISON OF MACHINES AND HUMANS AT THE TEST OF THE PROBLEMATIC OF SOLIPSISM

**Abstract:** The first objective of this article is to propose a reflexion about the limits of the comparison or analogy or metaphor between humans and machines. This comparison which runs through the history of European philosophy (Aristotle, 1995, 1253b23; Descartes, 2006, pp. 157–159; Onfray de la Mettrie, 1996, 3–39; Kant, 2007, §65; Lewis, 1934, p. 144; Sartre, 2003, p. 248; Wittgenstein, 1947, Ts-229, 448), is basic for functionalism, and central for the development of medical sciences. For the distinction between parts of living bodies, in particular, between organs, involves the consideration of distinct and mutually compatible biological ends, whose coordinated functioning together renders satisfaction possible. However, although the affirmation of the comparability of these two types of cases is not problematic as such, the affirmation of the identity or indistinctness of these relations is not without posing problems, whether conceptual or practical. If humans are under some aspects *like* machines and inversely, as some tasks are realizable by humans or machines, another thing is to suppose affirming that humans *are* machines, or that machines *are* humans (see C. I. Lewis, 1934). The stake of this point is considerable, for its range is not only the literality of the personification involved by the humanization or biologization of machines as robots (for we are not surprised by saying that such robot sweeps, achieves actions, smiles), but also that the depersonification involved by the machinization or metaphorical dehumanization of humans (whether to express an appreciation of the realization of a task by a person or to express the horror and the inhumanity, the absence of emotions involved by the realization of an action by a person). But its range also concerns: the extension of our concept of autonomy, the asymmetry of our relations to rules, principles, laws, of humans and machines, and in fact to a stronger extent our concept of *relation*. The question is thus whether this comparison, pertinent under some aspects in some contexts for certain ends, could have been adequate, turned out not be a comparison at all, such that the metaphorical could have become in such cases, literal. This affirmation could have seemed entirely incompatible with new possibilities of liberation rendered possible by technological innovations. In reality that is not the case since these possibilities are understood

as such against the background of precedent possibilities. The problem we then shall pose is the following: to which extent does the comparison or metaphor or analogy of human machine render possible the necessarily nonrestrictive limits of intelligibility? What are the limits of this comparison? To which extent does the recourse to this comparison turn out beneficial? To contribute to the resolution of this problem, I shall propose to put the comparison between machines and us and of us with machines at the test of the problematic of solipsism. To achieve this task, I present the criticism made by Lewis of solipsism (1934), and then present Turing's critical reconception of solipsism (1950). I then attempt to establish the way in which Wittgenstein, with his criticism of solipsism (1953), functionalism, and reductionism, solves the problems encountered by the conceptions of solipsism of Turing and Lewis.

**Keywords:** artificial intelligence, consciousness, C. I. Lewis, machines, solipsism, A. Turing, L. Wittgenstein.

## Introduction<sup>1</sup>

This first objective of this article is to propose a reflexion about the limits of the comparison or analogy or metaphor between humans and machines. This comparison, which runs through the history of European philosophy (Aristotle, 1995, 1253b23; Descartes, 2006, pp. 157–159; Onfray de la Mettrie, 1996, 3–39; Kant, 2007, §65; Lewis, 1934, p. 144; Sartre, 2003, p. 248; Wittgenstein, 1947, Ts-229, 448), has been studied (cf. Kennedy, 2022), and is basic for functionalism and central to the development of medical sciences. The distinction between parts of living bodies, in particular, between organs, involves the consideration of distinct and mutually compatible biological ends, whose coordinated functioning together renders satisfaction possible.

However, although the affirmation of the comparability of these two types of cases is not problematic as such, as the comparability and eventually the similarity of relations between wholes and ends is involved by scientific and engineering practices (for example, the wing of the plane is *like* the wing of the bird and inversely), the affirmation of the identity or indistinctness of these relations is not without posing problems, whether conceptual or practical. If humans are under some aspects *like* machines and inversely, as some tasks are realizable by humans or machines, another thing is to suppose affirming that humans *are* machines, or that ma-

---

1 Many thanks to Donald Cornell, to the reviewers, and to the editors of this volume for their helpful remarks and criticisms about this text.

chines *are* humans. And it is uncertain that whoever achieved or even, strictly speaking, *tried* or could have tried such affirmation.<sup>2</sup>

The stake of this point is considerable, for its range is not only the literality of the personification involved by the humanization or biologization of machines as robots (for we are not surprised anymore by saying that such robot achieves actions as sweeping, smiling, etc.), but also that the depersonification involved by the mechanization or metaphorical dehumanization of humans (whether to express an appreciation of the realization of a task by a person or to express the horror and the inhumanity, the absence of emotions involved by the realization of an action by a person). Its range also concerns: the extension of our concept of autonomy, the radical asymmetry of our relations to rules, principles, and laws, of humans and machines, and in fact to a stronger extent our concept of *relation*. Another way to formulate the conceptual difficulty (as is ordinarily, frequently, commonly “verified” that we are “humans” in our ordinary internet uses), is that of the indeterminacy of what we do when we *lend* to machines what we know of other humans, and of what we do when we lend to humans what we could *not*, strictly speaking, have ignored of machines, conceived to render possible either the better execution of some tasks, or the simple execution of some tasks (strictly speaking unrealizable by humans without their intermediacy).

The question is thus whether this comparison, pertinent under some aspects in some contexts for certain ends could have been adequate, turned out not to be a comparison at all, and the metaphorical could have become in such cases, literal. Surely, numerous technological innovations (biological computers, interfaces, and tools adjunctive to human bodies) render, for some conceptions, to some extent porous (cf. Kennedy, 2022) conceptual distinctions that could have seemed sealed, and mutually uncommunicative. Yet, if the open-endedness or intrinsic evolutivity of language is undeniable, it is uncertain that in the case of the comparison of humans and machines, we could have had to grant that this comparison could have ceased to be one, and became a unique literal means of expression. This affirmation could have seemed entirely incompatible with new possibilities of liberation rendered possible by technological innovations. In reality that is not the case since these possibilities are understood as such against the background of precedent possibilities: the intelligibility of history as social and objective science is tied to this point. The problem

---

2 This negation might seem incompatible with some uses of the metaphor between humans and machines, as that, for example, of Wittgenstein (1947, Ts-229, 448), but one central purpose of this article is to propose the epistemological elucidation that it is not. On this see also Bouveresse (2022, pp. 259–260).

we shall then pose is the following: to what extent does the comparison or metaphor or analogy of humans and machines render possible the necessarily unrestrictive limits of intelligibility? What are the limits of this comparison? To which extent does the recourse to this comparison turn out beneficial?

The response to this question is also important to think about some structural similarities of debates about ecological or climatic catastrophisms in relation to the development of artificial intelligence: similarly to ways in which catastrophistic narrations about climate provide occasions to think of the reality of the ecological emergency, catastrophistic narrations about artificial intelligence provide occasions to think the reality of the possibility of conceptions and detrimental uses of artificial intelligence. This is not unrelated to the fact that environmental or technological misuses are too often causes of environmental or technological catastrophes. But equally important is to remark that such catastrophisms should not be held as the presentation of some paralyzing aspect of reality in any sense whatsoever. Not only the transformations of facts (by contrast notably with the analyses or the explanations of facts) but also the misleading presentations of false facts as true, exaggerations (as under-evaluations) neither substitute nor could have substituted for the conception of artificial intelligences or for ways in which artificial intelligences can contribute to the resolution of environmental problems.

To contribute to the realization of this task we shall propose to put the comparison between machines and us and of us with machines at the test of the problematic of solipsism. To achieve this task, I shall first present the criticism made by Lewis of solipsism in "Experience and Meaning" (1934). I shall then present Turing's critical reconception of solipsism in "Computing Machinery and Intelligence" (1950) and propose a critical assessment of this conception against the background of philosophical results achieved earlier by Lewis. I will then attempt to establish the way in which Wittgenstein's criticism of solipsism, functionalism, and reductionism in the *Philosophical Investigations* solves problems centrally encountered by the critical conceptions of solipsism of Turing and Lewis.<sup>3</sup>

---

3 The notion of machine is not used in any theatrical sense throughout the text. The proposed approach is both critical of the very reductive criticisms of theatricality involved, for example, by Fried's notion of theatricality, and of the very inflationist conceptions of theatricality involved in some conceptions that allegedly would continue or have achieved the criticism of the Enlightenment. Theatricality is neither a problem nor a solution per se. But, as I shall attempt to render clearer in the third part of this text, in which I shall present a study of Wittgenstein's criticism of solipsism, self-estranged theatricality is delusory.

## 1. The critical conception of solipsism of C. I. Lewis

### 1.1. The problem of the solipsistic supposition according to which we '*are*' machines

“To repudiate all such transcendence is to confine reality to the given, to land in solipsism, and in a solipsism which annihilates both past and future, and removes the distinction between real and unreal, by removing all distinction of veridical and illusory” (Lewis, 1929, p. 183)

“Descartes conceived that the lower animals are a kind of automata; and the monstrous supposition that other humans are merely robots would have meaning if there should ever be a consistent solipsist to make it. The logical positivist does not deny that other humans have feelings; he circumvents the issue by a behaviouristic interpretation of “having feelings.” He points out that your toothache is a verifiable object of my knowledge; it is a construction put upon certain empirical items which are data for me – your tooth and your behaviour. My own toothache is equally a construction.” (Lewis, 1934, p.144)

C. I. Lewis expressed in 1929 a critical diagnosis of solipsism: Solipsism would be a position where one would arrive as the result of a repudiation – of a refusal – of “transcendence”. Such a would-be position, solipsism, *thusly* reached (inasmuch as Lewis presupposes that we can distinguish solipsisms), would involve: the annihilation of past and future, and the removal of the distinction between the real and the unreal as the outcome of the removal of any distinction between the veridical and the illusory. Among conceivable and eventually conceived solipsisms, such solipsism would be peculiarly unbeneficial, and delusory. For the rejection of every distinction between the real and the illusory, incompatible with the reality of past and future, can seem to leave as our only option a self-contradictory assumption according to which only the present and whatever is presently and sensorially available could exist (metaphorically “given” to mind). But if whatever is sensorially available to us is all that is real, then whatever is not sensorially available to us is not real. So according to the conception of solipsism, devised and critically diagnosed by Lewis, we either would have to acknowledge “transcendence”, that is, that the real could not possibly reduce to the sensorially available, or could not think a distinction between the veridical and the illusory, the real and the unreal, the past and the future. Conceptual distinctions of relevant opposites could but should not be entirely suppressed. There would be the possibility of delusory entrapment within a possibility that is not a possibility.

Lewis attempted to this extent to account for a distinction between a conception of solipsism which is inherently delusory, from a philosophi-

cal acknowledgement of (the reality of) reality: the idealism presented by “the world is my idea” could not ultimately but turn out to be acknowledgement of the fundamental and natural similarity between the idea of the world of an individual person, and the world whose idea is that of an individual person.<sup>4</sup> But Lewis did not render explicit in 1929 the motives of his critical conception and diagnosis of solipsism. He proceeds to such clarification in 1934, in the above quoted passage. He there argues that among solipsisms, a much more problematic solipsism would consist in the supposition that humans could not and would not be anything but robots. Thereby Lewis leaves aside the traditional characterization of solipsism, which would consist in the claim that a single person could be the only reality there is, and of which Schopenhauer had earlier argued that it would be claimed only in psychiatric institutions. According to the conception criticized by Lewis, any attempt to identify another human would necessarily fail and amount to an attempt to misidentify a robot – and not the opposite.<sup>5</sup> Any attempt to distinguish other humans from robots and robots from other humans would necessarily fail. Lewis does not unfold his diagnosis, but the difficulty is easily expressed: unlike humans, robots are tools conceived and produced to achieve the automatic achievement of tasks according to human desires, some of which are unachievable otherwise. The result of the negation of the conceivability of a distinction between robots and humans, the affirmation of the reducibility of humans to robots cannot but raise multiple problems concerning our relations. For, although some humans have engendered some other humans, no human has engendered every other human. No robot has engendered or could engender a human. Every robot has been produced by humans or by other robots, themselves produced by humans. Humans could not be reducible to tools, may have their own conceptions of which they are more or less conscious, and have their own desires and ends. Conceptions according to which humans could be produced for nothing but the satisfaction of the desires of other humans, and peculiarly, of their genitors, are abnormal: human procreation could not be reducible to slave production.

Two important aspects of the criticism made by Lewis of solipsism are to be considered. Lewis not only argues in favour of a distinction among solipsisms, but also among solipsists, according to the eventual consistency of claims and actions. The mere affirmation of the reducibility of

---

4 An approach which is relevantly comparable with that of Wittgenstein in the *Tractatus* (2003, 5.62) which inspired Lewis.

5 Considering the direction of the use of the comparison of machines and humans to explain the criticism of Lewis matters – as remarked by Bouveresse about Wittgenstein’s approach (2022, p. 259).

the only reality there is to oneself by a person is not coherent, as earlier brought out, and such incoherence is *prima facie* manifest: such solipsism would involve simultaneous negation of the previously considered solipsistic affirmation and inversely. But Lewis (as Sartre in 1943 with *Being and Nothingness*, Part 3) also considers the eventuality of solipsistic maintaining of (solipsistic) inconsistency. Indeed, Lewis considers a difficulty with respect to the activity in which “supposing”, and its results – “suppositions” – consist. A supposition results from an eventually expressed and eventually collective activity of thinking a truth. And in many ordinary, unproblematic and desirable cases, the truth of a fact is not and could not be dependent upon the decision of someone else. Thus, at first sight, Lewis can seem to be claiming that, as the achievement of a supposition by someone is directly dependent upon the action of only one person and no one else, the supposition that humans are merely robots can successfully be achieved by whoever thusly supposes. All cases considered: either a person supposes that humans are robots, or a person does not suppose that humans are robots. If we grant that the negation of the conceivability of a distinction between humans and robots is monstrous, in the sense of problematically abnormal, then its achievement cannot be really successful. But then the true answer of the question “Can one relevantly and successfully achieve the negation of the distinction between robots and humans?” could seem to remain indeterminate, as could seem relevant to negate the relevance of the previously expressed conditional for practical purposes. This is the difficulty addressed by C. I. Lewis just after the quoted passage.<sup>6</sup> The question of the determinacy of the true answer to the question “Can the distinction between robots and humans be negated?” could be, according to logical positivism, circumvented by means of a behaviouristic interpretation:

“The logical positivist does not deny that other humans have feelings; he circumvents the issue by a behaviouristic interpretation of ‘having feelings.’ He points out that your toothache is a verifiable object of my knowledge; it is a construction put upon certain empirical items which are data for me – your tooth and your behaviour. My own toothache is equally a construction.” (Lewis, 1934, p. 144)

According to such a picture, human relations could be reducible to partially communicative behaviours of humans which would consist in the sensorially accessible part of otherwise inaccessible data of humans about

---

6 This difficulty is also addressed by Sartre who explicitly presents behaviourism as solipsism put into practice (2003, 253) and also, as we shall see, by Wittgenstein (2009, §420).

each other. The difficulty brought out by the circumventing pointed out by Lewis is that the affirmation that the feelings of others can be accessed only indirectly – through behaviours – cannot but have consequences with respect to the evaluation of a human’s own feelings by oneself:<sup>7</sup> if the feelings of others are mental objects, constructions which are forever only partially accessible to an individual person, then one’s own feelings are also constructions which are forever only partially accessible to others, and eventually to that individual person oneself. Multiple difficulties arise from such an unreflexive “strategy”: among which notably mutual alienation, devaluation of knowledge, and possibly destructions.<sup>8</sup>

## 1.2. Is the moralistic rejection of the comparison of humans and machines philosophically receivable?

Humans-to-machines reductionism is, on Lewis’ terms, “monstruous” in that strictly carried out, such conception involves for practical purposes the self-contradictory negation of the conceivability of any distinction whatsoever between (other) humans and machines. The neglect of this problem has consequences with respect, notably, to our understandings of our experiences (as shared common experiences would be unintelligible as such), to our respective knowledges of others (which also would be unintelligible as such). “Reduction”, in this sense, ultimately leads to mutual alienation, devaluation of knowledges, and eventually to destructions. To this extent, Lewis raised the question of the identity of methodological solipsism with solipsism, a question to which Putnam, Sartre, Descombes, and Wittgenstein also provided positive answers (Sartre, *Being and nothingness*, 2003, p. 253; Putnam, “Why reason can’t be naturalized”, pp. 236–7, 1996; Descombes, *La denrée mentale*, 1995, p. 289; Wittgenstein *Philosophical Investigations*, §420): *Methodological solipsism, mere internalism, reductionism with respect to mind* (other minds) is not ultimately distinct from solipsism.<sup>9</sup>

---

7 See Uçan (2016) on this.

8 These problems, which are related to *problematic* skepticism, contrarianism, and denialism are further considered and exemplified in the second part of the present article.

9 This way of expressing their common criticisms could not conceivably reduce to a would-be “argument of authority”, and rather involves acknowledgment of the fact that distinct philosophers from different philosophical traditions have at diverse times and places reached independently the same result in diverse ways with respect to the would-be distinction between “methodological solipsism” and “solipsism”: ultimately there is no such distinction. As shall be rendered clear in the third part of the present text, the sort of possibility of verification that would allegedly be necessary for the



Remarkably, Turing, with the attempt to render clearer that machines think, has done in 1950 under one description exactly that which Lewis argued against. That is to say, one thing is to compare humans and machines for the achievements of some ends, goals, tasks, and finalities, and another thing is to reject that distinctions between humans and machines can be achieved whichever are the considered ends, goals, tasks, finalities. Especially against the background of a tacit agreement to the traditional verticalist *scala naturae* conception of a hierarchy of lives according to degrees of complexity and soulfulness, the comparison of humans and machines may have seemed to unavoidably involve the breaking of a taboo. But does the approach for scientific purposes of relations of parts and wholes of organisms as mutually coordinated involve the negation of the receivability of the moralistic criticism of the comparison of humans and machines?

To reply exhaustively to this question, the precision of the sense of the question, and the consideration of distinct cases will prove beneficial. The comparison of humans and machines is basic to functionalist achievements whose results are undeniable – notably in medical sciences. Inasmuch as we can compare parts of wholes of human organisms with parts of wholes of mechanisms constructed for definite ends or aims, we can distinguish functions and ends or coordinated parts of wholes.<sup>10</sup> Such comparisons contribute to render conceivable the resolution of theoretical problems, required for practical resolutions of health problems, and the conception of preventive and curative practices, which can be institutionalized. Whether such achievements do involve “metaphysics” can be asked. For as we shall see, although Turing rightly called into question moralistic ways of criticizing the achievability of the analogy of humans and machines for scientific purposes, this criticism was achieved by Turing with a misleading and distorted picture of other cultures, and especially of Islamic cultures, while Lewis had earlier argued that the resolution of the problem raised by the solipsistic supposition – “metaphysical solipsism” – required very limited, and more integrative, dependence to “metaphysics”:

“A robot could have a toothache, in the sense of having a swollen jaw and exhibiting all the appropriate behavior; but there would be no pain connected with it. The question of metaphysical solipsism is the question whether there is any pain connected with your observed behavior indicating toothache.” (Lewis, 1934, p. 145)

---

establishment of the truth of solipsism cannot be possibly be verified and is not, could not be, a possibility of verification at all.

10 On the compatibility of the criticism of the sufficiency of at least some “mechanistic world-view” see Putnam (1975a, pp. 364, 366, 385).

Lewis grants the conceivability of “metaphysical solipsism”, which would consist in the question of whether there is and could be any pain connected with an observed behaviour (for example, a behaviour indicating toothache). “Metaphysical solipsism”, as a solipsism, implies wrongly calling into question the existence of a connexion between an observed behaviour and pain. Such connexion could be unverifiable and unknowable. From the outset, the conception of “metaphysics” involved by the “metaphysical solipsism” envisaged Lewis is very minimal. Unmoralistically, such conception involves just the acknowledgment of the commonality of the veridicality of the expressions of their pains by humans.<sup>11</sup> Such a conception is compatible with any moralistic conception of the veridicality of the expressions of their pains by humans, that is, any conception according to which one must only veridically express that pain is felt by oneself *because* of some prescription, rule, law internal to a world-conception. Any such conception is indeed compatible with the existence of connexions between behaviours expressive of pains and experiences of pains (by contrast with the cases of machines and robots) and incompatible with fake expressions of pains by persons while no pain is felt by them.

The receivability of the moralistic criticism of the comparison of humans and machines is to this extent debatable: the mere rejection of the relevance of such comparison by appeal to a principle, religious or not, is not receivable since functional achievements (by contrast with functionalism) are not only conceivable but achieved and further will be achieved. The use of such comparison has a central place in the development of medicine, for the autonomous development of persons, individual or collective (institutional). But moralistic criticisms of the rejection of any conceivable distinction whatsoever between machines and humans because of a prescription, rule, or law internal to a world conception present some truth, as the negation of the distinction between robots or machines and humans does not result, could not have resulted, in the indistinctness or abolishment of the distinction between machines and humans. Such criticisms seldom are satisfactory, at least, if the appeal to a prescription, rule, or law, is meant to coerce the acknowledgment of the expression of pain as such, of the existence of a connexion between a behaviour expressive of the experience of pain, and the experience of pain.

---

11 “Any metaphysics which portrays reality as something strangely unfamiliar or beyond the ordinary grasp, stamps itself as thaumaturgy, and is false upon the face of it.” (Lewis, 1929, p. 10).

## 2. The critical conception of solipsism of A. Turing

### 2.1. “Can machines think?”

As mentioned, Turing achieved, under a description, in “Computing Machinery and Intelligence”, exactly that which Lewis argued against. In this part, I will propose a philosophical and epistemological study of Turing’s conception and criticism of solipsism in that article. I will attempt to render clear that although Turing there established that machines can somehow be unproblematically said to think, that thoughts and actions can relevantly be ascribed to machines, the conception of solipsism there put forward is, to express the point in Lewis’ terms, “thin” (1929, p. 30). The reduction of the problematic of solipsism to one and only one of its aspects, socially regrettably enough contributed to the replacement of a philosophical conception of solipsism by an unphilosophical one, whose consequences are yet to be brought out, studied, and criticized. Turing indeed introduces a conception of solipsism, to carry out a criticism of solipsism, in one of the counter-objections to the objections to the argument proposed with *Computing Machinery and Intelligence*, namely, the would-be objection that is called by Turing, “the argument from consciousness”. To critically assess this conception, let us first recall the problem posed by Turing and the replacement strategy proposed as an indirect means to achieve the resolution of the problem.<sup>12</sup>

After having proposed a consideration of the question “Can machines think?”, Turing considers a difficulty concerning an answer to this question (Turing 1950, p. 433). Uncritical adherence to an understanding of the question employing definitions that somehow “reflect so far as possible the normal use of the words” would be scientifically and philosophically problematic. Sciences and knowledge do progress with linguistic uses – uses of words – which are neither necessarily incompatible nor necessarily compatible with, independent from uses that are normal or considered as normal within a community, a society, of linguistic practitioners. Were we to restrict ourselves only to available “normal use of the words”, novelty,

---

12 This problem is deeply related to the relations of our conceptions of common sense with the one of Turing, inasmuch as (quasi-)paradoxically, common sense is necessarily debatable, open both to philosophical and unphilosophical contestations and acknowledgements. In that, Turing’s approach faces difficulties similar to the one of Sartre (2003, pp. 481-489) as their conceptions of common sense are not, at least, *that* common. Yet uncommon claims of common sense can desirably become common. For a historical and philosophical account of the development of Turing’s conception of common sense in relation to Wittgenstein see Floyd (2021).

improvement, discoveries, and creations, would almost be impossible, creativity could be reducible to exhaustion of combinations of allowed moves predetermined by social norms, and social norms would be unquestionable, whichever these are. But, if we would merely reject available “normal use of the words”, similarly, novelty, improvement, discoveries, and creations would almost be impossible, as novelties, improvements, discoveries, and creations could not be expressed within, and eventually understood, by a community, a society of linguistic practitioners. Thusly posed, everything can seem as if we are unavoidably entrapped in a predicament:

Either we accept that machines can think, reject “the normal use of the words”, the relevance of the examination of meanings involved by common uses of words. But then we might be led to assume that we must to rely on a statistical evaluation of the meanings of “meanings”. But then the justification of the answer could not be provided in any community anyway, and then both the meaning of the question and the end achieved by the asking of the question are lost.

Or we reject that machines can think, accept “the normal use of the words”, the examination of meanings involved by common uses of words as both relevant and sufficient. But then we cannot justify our answer except by reiterating appeals to “the normal use of the words”.

Turing thusly presents a dilemma which could not be resolved and which would result from opposite demands: that of the uncritical adherence to the common meanings of words for the sake of communication and critical rejection of the common meanings of words for the sake of novelty, discovery, and progress. Whether the phrase “machines can think” is true or false is a question that cannot, as such, be directly and satisfactorily answered. As a means for an indirect resolution of the problem raised by the question “Can machines think?”, Turing presents a replacement strategy with “the imitation game” (Turing, 1950, p. 433). In this “game” an interrogator has the objective to identify out of two persons with whom communication is achieved from a distance and without visual contact, a woman and a man, who is a woman and who is a man, provided that the man will attempt to make the identification fail. Such a game should be considered as a correct replacement to the initial question of whether the man is replaced by a machine.<sup>13</sup> Such replacement of the man by a machine in the game can indeed result in a different outcome, which can justify a reassessment of the relative positions of the humans playing the game, and also the way in which both “the imita-

---

13 For historical and philosophical accounts of “Turing machines” see (Kennedy, 2021; Floyd, 2021, Mundici and Sieg, 2021).

tion game” and the concept of game are to be conceived and understood. Drawing a conclusion from the previously mentioned difficulty related to the use of common definitions of words, Turing replaces the question “Can machines think?” by another “which is closely related to it and is expressed in relatively unambiguous words.” The questions, in fact, the allegedly equivalent questions, are the following:

“We now ask the question, ‘What will happen when a machine takes the part of A in this game?’ Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, ‘Can machines think?’” (Turing, 1950, p. 434)

Turing proposes in this way to reconceive the relations between concepts and applications. A satisfactory answer for the question “Can machines think?” could involve a reconception of our concepts both of humans and machines. That the interrogator is not in the vicinity of both the machine and the woman, rules out a sexist misunderstanding of the expression “who is a human”. The remarkable point to which Turing draws attention to is that “A machine can be constructed to play the imitation game satisfactorily” (Turing, 1950, p. 435), that is to say, a machine can be conceived and constructed to lure an interrogator into thinking that a woman is a man (no essentialism involved). Turing’s objective is indeed to render clear that automated and closely approximate replications of human actions by machines can be achieved (that is, indirectly by humans) (Turing, 1950, p. 438). Turing’s argument involves the acknowledgment that a machine that can replicate the behaviour of any discrete-state machine can be produced: “Provided it could be carried out sufficiently quickly the digital computer could mimic the behaviour of any discrete-state machine” (Turing, 1950, p. 441).

A few conclusions can thus be drawn if Turing’s clarification that machines can necessarily rightly be ascribed thoughts and actions is accepted: it would be a mistake to suppose the possibility of beneficially reducing Turing’s problem to itself without considerations of application. The problem raised by the question “Can machines think?” does not, and could not reduce to the conceivability of the affirmation of the indistinction of machines and humans, or to the negation of the distinction of machines and humans. That there are games at which humans and machines can play together and which can both be won and lost by humans and machines does not imply that the distinction or difference between humans and machines can (relevantly or without loss) be rejected. On the contrary, the commonality of such situations implies that the personification involved

by the humanization or biologization of machines as robots (as when we say of a machine or robot that such machine achieves actions, as sweeping or similarly) could not imply its own literality.

It is relevant to say of a machine or robot that such machine or robot achieves actions (which could be achieved by humans as well), that actions can relevantly be ascribed to robots or machines, since there is no relevant doubt with respect to the availability of a distinction between machines or robots and humans. The ascription of an action to a machine or robot is derivative in the sense that when a human person ascribes an action to a machine or a robot, that person does not ascribe an action (or expression) to a machine or robot which could eventually transform into or turn out to be a human. For then, there would not be any test of concepts in their relation to their uses or applications.<sup>14</sup> Actions in such cases are ascribed to a machine which has been constructed to render possible the automated (and eventually) better execution of a task which otherwise would eschew to one or several humans, or of a task which otherwise would remain unachieved by humans (as some human actions necessarily involve the mediation of the actions of machines to be achieved). The obviousness of such a point is probably more easily and better understood if one considers that: mechanisation or metaphorical dehumanization of humans, which involves the depersonification of humans at the occasion of the comparison of one or several machines with one or several humans, also has contraries, or “opposite poles”. Ordinary language uses do indeed involve distinguishing between: desirable cases in which, humans are appreciated for their mode of realization or achievement of a task as, or even better than machines (as conceiving an artificial intelligence or winning a game of go), and undesirable cases in which, lived horror or inhumanity of humans is expressed due to their realization of a task or action whose realization by a human necessarily implies the rejection of felt or observed shared human emotions. To this extent, Turing might have, on this point, involuntarily underestimated the resources of our (common, ordinary, everyday) linguistic means, our “natural” languages.

Turing did, since the 50s, envisage the evolution of logical space, the space of possibilities, our possibilities, with respect to the fact that the ascription or attribution of thoughts and actions to machines by humans is unproblematic (Turing, 1950, p. 442). But less obvious is that Turing’s evaluation of one’s own question can be agreed with under one’s own terms. Obviously, the “original” question “Can machines think?” should be discussed – for example, regrettably enough, many people could lose

---

14 The analysis of the test proposed with this article is different and independent from the one proposed by Gonçalves (2024).

their work if such a question is not publicly addressed.<sup>15</sup> Most probably, it is not senseless to consider that, in the 1950s, the question was too remote from most persons' lives, interests and concerns, to be considered as somehow linked and eventually determinative of their own conceptions of their lives. In that sense, Turing's writing about one's own question that this question is "too meaningless" could eventually be understood. Nevertheless, such evaluation does not, and could not imply that there are, or could be, *degrees* of logic, logicity, logicality or logicalness.<sup>16</sup> In that, Turing's evaluation of one's question is arguably in tension with Turing's own achievements in "Computing Machinery and Intelligence". No unacceptable fact was involved in Turing's expressions of one's own conception of computers. No one would in any way deny jointly that computers have been constructed and do not exist. And remarkably enough, even contrarianist conceptions implicitly addressed by Turing in the objections section, would involve as a step of the conception of their destructive efforts the acknowledgment of the existence of a targeted existence (As the buddhas of Bamiyan). In that, Turing's conception of consciousness and solipsism, I will attempt to render clear, is not, and could not be successful, turn out adequate.<sup>17</sup>

Let's consider the "*The Argument from Consciousness*" that Turing wants to contest and of which Professor Jefferson is presented by Turing as a notable defender (Turing, 1950, pp. 445–447).<sup>18</sup> The argument is that if a machine could write a poem or compose a musical piece *because of*

15 That liberatory possibilities involved by the conception and the use of artificial intelligences (as the execution of some tasks can be automated and dispensed with) should not make us forget that the challenges thereby raised present social significance: the realization of the antic dream of the liberation from repetitive work is no more than it was, a wish whose realization would be, as such, relevantly available to every one (Aristotle, 1995, 1253b23).

16 On this, Lewis' criticism of the alogical is to be reminded: "Sometimes we are asked to tremble before the specter of the "alogical" in order that we may thereafter rejoice that we are saved from this by the dependence of reality upon mind. But the "alogical" is pure bogey, a word without a meaning." (Lewis, 1929, p. 246).

17 The first, narrowly theological objection considered by Turing, consists in denying that animals or machines can think on the basis of the affirmation that only humans (by contrast with animals and machines) have souls or are soulful, and that thinking is a function of the soul (Turing, 1950, p. 443). This objection is of little interest for the problem posed and addressed in this article, and Turing's question: "How do Christians regard the Muslim view that women have no souls?" at best is expressive of a distorted picture of Islamic cultures. As a clue of a conceivable reply, the falsity of the question can be established by the true affirmation by a person, whether a Muslim believer or not, to know someone who is Muslim and believes that women have souls, that women are soulful.

18 For a study of the context of the debate between Turing and Jefferson see Gonçalves (2024, Sections 4.6 and 5.5).

*thoughts and emotions felt*, then at least one machine could think or be considered conscious, and therefore machines could think, or be considered conscious. But, and the following was right when expressed by Jefferson quoted by Turing, machines have not achieved such artistic productions. Thus, machines do not think, are not to be considered conscious. Maybe Professor Jefferson would have liked to add: “because machines cannot experience, feel and act as we – humans – do” (underlining mine), but such addition would arguably render clearer a tension internal to Professor Jefferson’s conception. Turing’s interpretation is that such argument consists in a rejection of the validity of the test. Turing achieves to render clear a difficulty in simultaneously attempting to maintain that machines can fail humans into determinate misidentifications (that is to say, machines not only can make someone believe that someone is anyone else, but also make someone believe that someone is someone else), and that machines cannot think: only if machines think can these achieve an action which is inconceivable without previous reflexions. The objectivation of the realization of such failing of a human by a machine can be successfully achieved and verified, the loss of a human face to a machine noted both by machines and humans. In that, what was later to be called the “Turing test” is formally valid, as its falsification is conceivable, and the criteria of the test are public and publicly acknowledgeable by relevant expert practitioners.

Nevertheless, less clear is that “the argument from consciousness” consists in a rejection of the formal validity of the test, except maybe, the last move involved by Professor Jefferson’s reply, which involves presenting something undone as something that cannot be done, the presenting of a limit as a restrictive limit. For, the reflexions involved by the (derivatively) intentional aspect of the failure of humans by machines, are, strictly speaking those of the (eventually) other humans who conceived and constructed the considered machine, rather than only or merely the achievement of the machine considered in isolation from its conceivers and producers.

Otherwise put, the transition from the question “Can humans fail other humans into thinking that machines are humans, by conceiving and constructing machines which can lure humans into thinking that machines are humans *as good as* humans who can lure other humans into thinking that someone is another?” to the question “can machines fail humans into thinking that machines are humans?” is at least unclear, not to say undue or illegitimate. For, even in the intricate case in which the machines (which can lure humans into thinking that machines are humans) have been conceived and constructed by other machines (conceived and constructed by humans to lure other humans into thinking that machines are humans), it is not rendered true that machines self-conceived themselves by themselves – that is, autonomously in an underivative sense – to lure humans into thinking



they are humans rather than machines.<sup>19</sup> Quite the contrary, *only inasmuch as humans conceived machines*, which can conceive other machines, which can lure humans into thinking that machines are humans, can it be rendered true, and not only in a narrow experimental sense but in a historically accurate way, that humans can be *lured* into thinking that the machines conceived and produced by the machines they conceived and produced are humans rather than machines. To this extent, we should probably reject, not that the test is valid, but rather that the imitation game does consist in a test at all. That is to say, if any test is involved by the “imitation game”, this test is different from the presented test (that of the testing of the thinking of machines),<sup>20</sup> and strictly irreducible to the “verification” of the humanity or humaneness of humans (as in “tests” which we are, as internauts, frequently asked to achieve). What is at stake is rather whether the production of a luring situation by a source-of-language, of source-of-language conception could be acknowledged (conception(s) according to which (a) “private language” could be conceived).

To reject the validity of the test, according to Turing would be equivalent – under its most extreme form – to defending solipsism, which *could be*, Turing grants, “the most logical view” (Turing, 195, p. 446). So not only that there could and would be degrees of logic, of logicity, of logicality, of logicalness, but there could also be *consistent solipsism*, with solipsism defined as the thesis according to which the *only* way to know *that* – the fact that – someone thinks is to *be* that (particular or individual) person and feel oneself thinking. A parallelism, an analogy, could be made with the case of machines: the only way to know *that* a machine thinks is to *be* the machine and feel oneself thinking.

The motives of Turing’s partial objection to Professor Jefferson’s eventual objection (as Turing agrees with Professor Jefferson against solipsism) can then be brought out: verification of whether machines are humans is impossible. Thus, it would be sufficient to lure a human into thinking that a machine is a human to establish that machines can think. And among

---

19 Putnam considers a similar intricate case in which the question whether robots are conscious is posed about robots produced by other robots, and argues that its answer involves a decision concerning the treatment of robots within one’s linguistic community (Putnam, 1975b, pp. 406-407) and not a discovery. This article is fully compatible and in agreement with Putnam’s rejection that considerations with respect to the applicability of the concept of consciousness are meant to be decided on the basis of a discovery. But this article does not argue in favour of the conception according to which consciousness-ascriptions to robots could have been without truth value *until* a decision is taken with respect to the question whether robots are conscious.

20 On this see Davidson (2004, p. 83). And for a criticism of misleading uses of the argument see Descombes (1995, p.156).

many reasons that can be provided, some of which have been previously explained, such would both be too much and not enough, especially since the premise according to which verification of whether machines are humans is impossible is left uninterrogated. Prof. Jefferson's defence of the "argument from consciousness", according to Turing, then amounts to verifying whether machines are humans is impossible. But suppose such verification could consist in an artistic production by – literally – a machine. Until such production is achieved, that machines can think will not have been established. If Prof. Jefferson were right at that time, then today, Turing would be right and Prof. Jefferson would be wrong since artistic productions produced by machines have failed even expert juries. But it is remarkable that the victory has nevertheless been attributed to a human (by contrast with cases in which victories were attributed to an artificial intelligence as in, for example, the games of chess or go). Nevertheless, could the production of a luring situation of humans by machines be conceivably determinative as Turing argued for? This is, at best, unclear. A reappraisal of the conception of solipsism presented by Turing will prove important, necessary, and beneficial. For Turing both grants that solipsism could have been "the most logical view to hold", and that the only problem involved by such a "view" would be that communication would be rendered difficult. Not only that Turing does not address the question of the logicity, or logicity, or logicalness, of solipsism, or of whether solipsism could be logic or logical,<sup>21</sup> but also, and more importantly Turing neglects both the initially non-philosophical and philosophical conceptions and criticisms of solipsism.

The problems raised by solipsism were indeed not reducible to difficulties of communication. Even only according to the analyses of Lewis, that the difficulties raised by solipsism are very concrete, as concrete as the negation or denegation of the reality of pain involved by contrarianisms and denialisms, among which behaviourism, has been shown. Not only that the problem posed by non-philosophical and philosophical solipsism(s) is not reducible to Turing's conception, but also undue belief in such reductive conception of solipsism can lead to the neglect of solipsism, even *to* solipsism, and this, even despite Turing's achievements.

"In short then, I think that most of those who support the argument from consciousness could be persuaded to abandon it rather than be forced into the solipsist position. They will then probably be willing to accept our test." (Turing, 1950, p. 447)

---

21 The negative answer is involved by the negative answer to the conceivability of an exclusively private language, a philosophical result that is not explained in this paper (On this, see Uçan, 2016; 2023).

Turing indeed assumes that there would be an exhaustive alternative between two opposite possibilities. Either we abandon the argument from consciousness – as we would need to be persuaded to abandon the argument from consciousness to integrate the results from the Turing procedure, or to be persuaded that we cannot integrate the results from the Turing procedure if we maintain the argument from consciousness. Or we are forced into the solipsist position.

And indeed, if we grant both: that the argument from consciousness and the results of the Turing procedure are not compatible, and, that we need to reject solipsism even if that involves rejecting the argument from consciousness, then quasi-unavoidably, the conclusion seems to follow from the premises: we probably will accept Turing's test, allegedly "our test". But an undue dichotomism, or at least, an undue use of a dichotomy in a non-dichotomic case is involved by Turing's conception both of consciousness and solipsism. Indeed, the whole "pressure" exerted on defences of consciousness turns around the ambiguity involved in the would-be claim according to which "Machines cannot *feel* thoughts and emotions". That is to say, the phrase can be used both to express that machines do not feel thoughts and emotions as we do (and how could we be surprised about that?), or to remind ourselves that expressions of feelings of thoughts and emotions authored by machines are really produced by machines (as we can be astonished by the similarity of expressions authored by machines and humans). Could we have really meant that machines *lack* the sensibility of humans? This is, I shall try to render clearer, at best unclear.

## 2.2. Commensurability and incommensurability of the facts of humans and machines

Let us remark that the central range of cases integrated by the Turing procedure, the replacement strategy, is *the range of commensurable actions of humans and machines (automated and eventually automatically)*. That is to say, the Turing test is formally valid in an unproblematic sense, as some actions can be realized by both humans and machines, even if its philosophical relevance can and should be criticized. A machine can perform, realize, and achieve for you exactly that which you could perform, realize, and achieve by yourself (for example, cleaning the floor of a room). But there is a range of cases in which the actions of humans and machines are not commensurable. You can and cannot fly at 900 km per hour at 11 km of height under different descriptions. You can thusly fly with a plane, even if you are not the pilot. You obviously could not thusly fly without such

a plane. But the impossibility involved is not ‘real’ (and even could not be such), and could even less be determinative of a restrictive limitation internal to humans.<sup>22</sup> Turing is, to an extent, clear about this distinction:

“We do not wish to penalise the machine for its inability to shine in beauty competitions, nor to penalise a man for losing in a race against an aeroplane.” (Turing, 1950, p. 435)

To be relevantly assessed as successful or failed, won or lost, the actions, performances, and achievements of humans and machines need to be relevantly compared. It would nowadays not belong to our expectations, shared human expectations, for a human to fly without a plane at 900 km per hour at 11km of height. For a human to fly involves the use of a tool or a machine which renders possible the achievement of a flight. Imagining the contrary is not impossible and eventually rather comical. But the missing of the comical in such a case could be tragic. After all, cannot we conceive that the conception and production of planes by humans imply the past acknowledgment of the existence of a restrictive limitation internal to humans? This line of argument, is, I argue, truly addressed by Turing, although relatively indirectly, in “Computing Machinery and Intelligence”.

Remarking the range of cases of incommensurable action to humans and machines does not imply that the ends attained by machines (in the sense of the tasks achieved by machines that cannot be done by humans, e.g. exhaustively surveying the data of a mega-database) are *not* the ends of some humans. Such ends can be and are attained by humans only since the attainment of such ends has been envisaged, and conceived, and machines and robots constructed along the lines of such conceptions to render possible their attainment by some humans. The realization of such ends could not be possible otherwise, that is, without the mediation of the past conception and construction of the machines which rendered possible the attainment of ends whose attaining was previously impossible. Machines are practically necessary for the attainment of some ends in this respect. Although some actions achieved by machines are incommensurable with actions achieved by humans, with respect to their realization, as the former can do what the latter could not, the same does not apply to the ends of these actions, which are commensurable. We considered that the ascription of thoughts and actions to machines by humans is derivative (of human ascriptions of thoughts and actions among themselves, rather than from past

---

22 On the distinction between the criticism of mechanistic conceptions of the human mind and the usability of a Turing machine as a model for some realizations of the human mind see Putnam (1975a, p. 366; 372).

conceptions and productions of machines). The ascription of ends to machines by humans, the self-ascriptions of ends by machines, the ascriptions of ends by machines to other machines, or even to humans, are likewise derivative of humans' ascription of ends to themselves. The ends of the machines could not, as such, be alien to those of humans. To this extent, the evaluations of the commensurability or incommensurability of the facts of machines and humans is circumstantial. Humans, as such, would neither be limited nor unlimited without machines. And similarly, without humans, as such, machines would neither be limited nor unlimited.<sup>23</sup>

### 2.3. The limits of metaphorical expression

We considered that the affirmation that machines (sometimes) (metaphorically) think is unproblematic. The important pivotal point is the distinction between the metaphorical and the literal senses of our comparative claims about machines and humans, eventually via the mediation of a comparison of some of their aspects. To say that someone is a machine is neither necessarily problematic (consider the case of the use of a metaphor to express the appreciation of a modality of an action's achievement) nor necessarily unproblematic (case of would-be unmetaphorical use for expression of depreciation, eventually expressive of lack of expectable emotion). To say that a machine is someone is neither necessarily problematic (case of the use of a metaphor to express the appreciation of the similarity, of the accurateness of the replications by a machine, of someone's human behaviours) nor necessarily unproblematic (case of would-be unmetaphorical uses involving identity confusion, a case different from the cases considered by Turing). Similarly, to say that a machine thinks is neither necessarily problematic (case of the acknowledgment of the machine-mediated realization of a task or action, whose realization without one or several machines sometimes is and sometimes is not conceivable), nor necessarily unproblematic (cases of the depreciation of a human by others, and of the solipsistic others-as-tools conception).

To this extent, although Turing was right about the unproblematicity of the affirmation that machines can think and (sometimes) think, the philosophical acknowledgedability of Turing's reconception and

---

23 These remarks are entirely compatible and in agreement with Putnam's rejection of the unavoidability of a trilemma concerning the application of the concept of consciousness to robots: it is at best unclear that we could have been bound either to affirm that robots are conscious, or deny that robots are conscious, or express our unavoidable ignorance with respect to the eventual truth of the question whether robots are conscious (Putnam, 1975b, p. 407).

displacement of the problematic of solipsism can and is to be contested. The problem with solipsism never was and could not have been merely reducible either to the correct identification of the thoughts of a person or to the correct observation of the achievement of the activity of thinking by a person.

The least that can be said is that the central aspects of solipsism, brought out by notably by Wittgenstein (also by Sartre and Putnam, but it is unsurprising that Turing did not discuss their works) are neglected.<sup>24</sup> The aspect that is centrally neglected, and which is elucidated by Lewis' critical conception of solipsism (and as we shall study, by the one of Wittgenstein as well) is the consideration of the eventuality of the experience of pain. Too much (or not even anything) is done by Turing about solipsism by granting that solipsism could be "the most logical view". Although, under one's own terms, Turing's focus on an aspect of solipsism is understandable and relatively beneficial, such focus and such reconception of solipsism has arguably contributed to the substitution of a thin non-philosophical conception of solipsism to previous philosophical and critical conceptions of solipsism, if we consider the influence of "Computing Machinery and Intelligence", and of Turing's works and achievements. In this sense, the reconception and the displacement of the problematic of solipsism proposed by Turing is not philosophically receivable, or acceptable. For rejection of asymmetrical pain ascriptions resulting in delusive false impossibilities does not imply, could not imply the rejection of the relevance of the acknowledgment of asymmetries between humans and machines with respect to ascriptions of pain. The phrases "machines cannot feel" and "machines do not feel" could not conceivably be reduced to each other and attempts to reject such irreducibility, I shall try to render clearer in the third part of this paper, cannot but turn out delusory.

To answer the question of the limits of the comparison of machines and humans thus involves considering the dimension of successfulness of the achievement of the comparison – its "performative dimension" – so to speak. Comparisons can be successfully achieved. Reflexion concerning circumstances in which the realization of comparisons of humans and machines turns out to be successful can obviously also be achieved. This point matters to remark and address several important difficulties in Rorty's account of metaphor explained and used by Kennedy, in a won-

---

24 No would-be "argument of authority" is involved by such expression of the issue. Putnam himself achieves the criticism of the intelligibility of the argumentative dimension of the would-be argument of authority, which nevertheless does not, could not consist in a mere rejection of specialization, authorship, authority, truth, and history (Putnam, 1996, p. 233).

derful article entitled “Gödel, Turing and the Iconic/Performative Axis”, difficulties which have been only partially addressed so far, to evaluate the place of the machine metaphor in our languages, cultures, societies, forms of life:

Rorty’s elaborate account of metaphor, of the way metaphor operates in language, is useful here. Metaphors, for Rorty, are “private acts of redescription” originating “outside” of language – “outside”, metaphorically, in the sense of unintelligibility; and his account turns on the idea of the literalized metaphor, literalization being what happens when a metaphor breaks into sensibility; when a phrase like, for example, “point of view” comes to mean something like an attitude toward something—becomes, in other words, literalized:

Between . . . [between living and dead metaphor] we cross the fuzzy and fluctuating line between natural and non-natural meaning, between stimulus and cognition, between a noise having a place in a pattern of justification of belief. Or, more precisely, we begin to cross this line if and when these unfamiliar noises acquire familiarity and lose vitality through being not just mentioned . . . but used: used in arguments, cited to justify beliefs, treated as counters within a social practice, employed correctly or incorrectly.

Rorty sees the creation and literalization of metaphors as the “fuel of liberalism”, and “a call to change one’s language and one’s life”. As such, metaphors are a sign of the viability of a shared social practice; evidence of the ability of that practice to continually transform itself, to produce new meaning, through the creation of metaphors. (Kennedy, 2022, pp. 3–4, underlining mine)

Rorty’s conception, as explained by Kennedy, involves several assumptions concerning the place, the origin, and the integration of metaphors in ordinary linguistic practices. Metaphors would be acts of redescription, new and different acts of description of whatever has been described utilizing a nonmetaphorical expression: the imaged or metaphorical expression would be a new description. Such redescriptions would somehow be “private”, and correlatively, the origin of such acts could be mysterious as such. However, what would such redescriptions be redescriptions of? Of a non-imaged, non-visual, non-metaphorical expression, of a possible or actual literal use of language? On Rorty’s account explained by Kennedy, the reply to such questions is relatively ambiguous and yet would somehow turn out successful (as metaphors would be signs of the viability of a shared social practice).

But, as earlier raised by Sartre (2003, pp. 536–538), the question whether language could be an author, by itself, ought to be raised anew: Could the personification of language be as such relevant at all? Even more problematically, a difficulty, earlier evoked, that I will not attempt to

address in this paper, but of which an aspect is relevant our consideration of the difficulty raised by such account of metaphor, is the problematic of private language. For distinct ways of considering the relevance of the use of metaphors are, for explanatory purposes, distinguishable.

One thing is to acknowledge metaphorical expressions as direct or indirect expressive means, another thing is to suppose considering metaphorical expressions means as unavoidably indirect expression means (one could not but use a metaphor in some would-be noncontrastive sense). Then, to present metaphors as “private acts” could seem to call into question the availability of our medium of expression, language, if any such distinction is supposed. And for a very simple reason, in fact, as one thing is to affirm that we indirectly use a metaphor to express whatever we could not have affirmed, cannot affirm otherwise (for we have not found a non-metaphorical expression to our metaphorical expression, although we can), and in such case there is no such thing as an implicit would-be exclusion of a possibility that is not possibility involved.

But another thing would be to suppose ourselves able to affirm that we use a metaphor to express whatever we could not have conceivably affirmed otherwise. The concept of the literalization of metaphors, of literal metaphoricality is to this extent “a double-edged sword”, that is to say, certainly not a risky weapon, due, allegedly, to the sharpness of both of its edges. But a concept that is similar to a weapon, a sword, whose edges both cut, and which has relevant ways of effective handling.

To suppose ourselves to be granting that we unavoidably have to use metaphors because we cannot express – ourselves – whatever we suppose ourselves able to be willing to express, otherwise than by employing metaphors, amounts to underestimating both our possibilities of expression and our (eventual) successfulness in our searches for new or better means and ways of expressing ourselves. The Rortian approach advocated for, to an extent, by Kennedy, is thus not satisfactory; not that such a conception is ‘risky’ or ‘dangerous’ as such (the criticism I make is one of intelligibility, but not moralistic), but that such a conception of the evolution of language (acknowledged in a way that is partially congruous with Wittgenstein’s philosophically pragmatic remarks in *Philosophical Investigations* concerning the fact that language, language-uses, change), may lead to confusions, sometimes somehow involved by aspects of Turing’s article, such as that of the confusion of humans with machines and inversely.

Although the limits of intelligibility indeed extend with our expressions, our actions, our doings, such a remark could not have implied that any linguistic use, any action, or any doing necessarily consists in an extension of the limits of intelligibility by itself in a relevant sense. Remark-



ing that a fact is historical (its historicity, so to speak), by means of the remark of the compatibility of the expression of a fact with relevant, accurate, more comprehensive and extensive historical narrations could not relevantly be equated with the creation, the conception, or the production of a new way of understanding, doing, explaining, or achieving. The difficulty with the assumption or the supposition of “degrees” of logic, logicity, logicality, logicalness is not ‘after all’ a difficulty related to our incapacity to distinguish between conceivable or actual courses of actions which are more or less relevant, or even adequate for the attainment of some ends. Quite the contrary, the difficulty is rather that there could not be such a difficulty and arises from an eventual tension between the acknowledgment of the existence of diverse systems of logic, world-conceptions, and the uniqueness of a way that is our own to understand.

### 3. The critical conception of solipsism of Wittgenstein in the *Philosophical Investigations*

#### 3.1. Dissolution of the problem raised by functionalism and reductionism: A ‘thought experiment’ by Wittgenstein.

“420. But can’t I imagine that the people around me are automata, lack consciousness, even though they behave in the same way as usual? — If I imagine it now—alone in my room—I see people with fixed looks (as in a trance) going about their business—the idea is perhaps a little uncanny. But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: “The children over there are mere automata; all their liveliness is mere automatism.” And you will either find these words becoming quite meaningless, or you will produce in yourself some kind of uncanny feeling, or something of the sort.

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.”

Wittgenstein expresses in §420 that assuming some symmetry between pain ascriptions to humans and machines, between consciousness ascriptions to humans and automata, results in false and eventually delusory impossibilities. Before pressing this point, let us recall that:

- (1) Automata are machines which have been built to achieve some actions by themselves, once somehow activated. Once built and activated, the realization of foreseeable and foreseen actions of

automata does not depend anymore on (although it is eventually controllable by) their conceivers, producers, and activators. This contrast is involved by the very intelligibility of our eventually successful ascriptions of failures to automata. In such cases whatever was to be relevantly considerable was considered at some stage to render possible the achievement of an action by an automaton, and yet the predicted outcome, the successful achievement of an action by an automaton, did not result from its conception, production, and activation. In this sense the failure of the achievement of an action by an automaton is ultimately intelligible and understood by us derivatively. Nowadays automata, robots, come with and under warranty. We would not take responsibility for each conceivable failure of the functioning of an automaton, of a robot, even if under some description we are the one or ones who have failed to make the automaton function.

- (2) “Consciousness” as used by Wittgenstein in this paragraph of the *Investigations* both is and is not used as in phenomenological conceptions under different descriptions. If by consciousness we mean, as in many phenomenological conceptions and accounts, a moment of mental life eventually correlated to irreducibly lived moments (by us or others), as in expressions such as “consciousness of happiness”, “consciousness of joy”, “consciousness of sadness”, that we could express also otherwise, then Wittgenstein’s use of the notion of consciousness in this passage is not phenomenological in the sense previously defined. But if by consciousness we mean, the fact that we can gain consciousness, take consciousness, that at such time and place, I, you, us, them is happy, joyful, or sad, rather than allegedly remarking from within ‘isolated’ or ‘separated’ ourselves that happiness, joyfulness, sadness is somehow ‘happening’ in ways in which not only are remote but cut from ourselves, unavailable to ourselves, separated from ourselves, then surely Wittgenstein’s use of the notion of consciousness is at least compatible with such phenomenological conceptions of consciousness.
- (3) According to the traditional conceptions of soulfulness or consciousness, to have a soul or to be conscious *is* to be a soul or to have (a) consciousness (See Sartre, 2003, pp. 127–129; pp. 310–315, p. 619). The derivative and metaphorical “property” of a soul would have eschewed to each of us as a result of some attribution about which nothing could conceivably have been done by us – humans, an attribution about which several narratives exist.

And in any case, as a result of such an endowment, we necessarily would *have* in ourselves, and ourselves *be what* necessarily could *not* be *had* in themselves by such existents which are not human, and could anyway not have *been* such existents. Consciousness thusly conceived could be some sort of additional ingredient or substance presented by some existents eventually encountered within visual space, and which could and would be in itself *lacking* from other existents eventually encountered within visual space. To render the point clearer: no essentialization of consciousness is involved by such an expression: in fact, quite the contrary.<sup>25</sup> Such lack both can and cannot be observed by us humans who are soulful or conscious, as we could understand that we are provided, endowed, or gifted with exactly the soul or consciousness that could not have been provided, endowed, or gifted to other living existents. Correlatively, we could not have provided the soul or consciousness that we were – as humans – to tools, or objects we construe, as automata, as machines, as robots. This would be an impossibility we could not but acknowledge were we to understand our ‘natural’ place. But we can nevertheless imagine how wonderful would be for such existents to be provided with – like us – a soul or a consciousness.

This is the sense of ‘lack’ involved by Wittgenstein’s ‘thought experiment’ at the beginning of §420. That we can analogically or metaphorically envisage that artificial existents, as automata, robots, or machines, *could* have been provided a soul or consciousness, if these existents *had been* humans involves our acknowledgement that, in fact, these existents could not have been provided a soul or consciousness, as these existents are not humans. Even if we can imagine that these existents could have wished to be provided a soul or consciousness, although these could not have had a soul or consciousness, a soul or consciousness could not have remained unwished-for by these existents if these existents could have imagined consciousness or soulfulness. To this extent, such existents would lack precisely the soul or consciousness each of us is or has. Our assessment of these existents would have but to remain oscillating, once and for all

---

25 On this, Sartre was and is right against Heidegger: if anything is metaphorically ‘essential’ to consciousness, that is non-coincidence with “itself”. Expressed otherwise: according to the traditional picture, animals lack a soul or conscience; their reality is intelligible and accessible to us only negatively and privatively. But although the consciousness of animals might arguably be firstly intelligible and accessible to us negatively, it is at best unclear that such consciousnesses could need to be rendered intelligible and accessible to us privatively.

(we could be condemned to idle so to speak). Wittgenstein invites us not to remain constrained by such exercises of our imagination. Yes, we also can imagine that other humans are automata, “lack consciousness”. That is to say, we can imagine that people around us, at an occasion, are wrongly assumed by us, not to be automata, machines, or robots. It is sufficient to imagine that the substitution or replacement of humans by automata, machines, or robots would have been achieved with automata, machines, or robots whose actions would replicate, mimic, or reproduce the behaviours – actions – of persons whose behaviours – actions – are replicated so well, so accurately, that the substitution or replacement would remain undetectable by us. However, if we try, such automata would be ‘logically’ indistinguishable from humans and inversely (at least according to the traditional conception of logic addressed by Frege, Russell and Wittgenstein). Importantly, whether we are imagining to be with others (who in fact are not others but automata) while we are not, or are with others (who in fact are not others but automata) while we are, does not change the ‘thought experiment’ and its outcome. For in neither case, could be rendered true that humans are automata, or automata are humans, in ways in which we so far, until now, could have failed to notice, to discover.

Yes, we can *imagine* that we wrongly assume that machines *are* humans, but imagining such a case involves reconceiving what the holding of such a case, the happening of such a fact, would consist in. The distinctions between humans and automata would not thereby be rendered unavailable. The availability of such distinctions would remain implied by the intelligibility of the situation as such (one’s hesitation with respect of the identity of the existents in the surroundings, one’s discovery of a failure to identify a human or a robot). That we do not know could not imply in such cases that the truth about the eventual identification mistake could not conceivably be known by us – the realization of the replacement itself would involve the concerted action of several persons. To this extent, the ‘merely direct’ reading of this passage remains superficial. If §420 only addressed the ‘risk’ involved by a superficial conception of solipsism and functionalism, then §420 could have been ended with its first question. But this is not the case. Wittgenstein envisages ‘in the first person’, or invites us to envisage by ourselves, one way in which we could conceive the result of the imaginative exercise of our imagination, in determinate circumstances, the first range of cases considered above.

Let us imagine that we are not with others who in fact are not others but automata, and that we are alone in our rooms, in one’s room, and *imagine* that we are with others who in fact are not others but automata. Wittgenstein then expresses a conceivable result of such an imaginative exercise of our imagination, which can eventually be considered as quite

deceptive: "I see people with fixed looks (as in a trance) going about their business". Such a description of an imagined situation could be either very similar or very dissimilar from our ordinary experiences (not experiments): after all, one might or might not have experienced cases in which the focus of persons with whom one works seems very irrelevant or very relevant. But, more centrally, could not one have expected the outcome of the imaginative exercise of one's imagination to be seeing-automata-and-not-people? Was not the case envisaged, the case in which one is wrongly assuming that humans are automata 'after all'? But importantly enough, such an imagined case does not involve such a conclusion – another case could, but one independent from the former and that we would have to imagine.

Wittgenstein does call attention to the openness and necessarily public conceivability of the result of the 'thought experiment'. If there is no conceivable way of discovering – and especially as we are considering imagined cases – that humans – conscious existents – are automata – existents which supposedly lack consciousness – then there also is no conceivable way of discovering that automata – existents which supposedly lack consciousness – are in fact humans. Then the realization of the delusiveness of the would-be result of the would-be attempt to distinguish people and automata by presenting asymmetries with respect to attributions of consciousness to humans and to machines as involving reciprocal (and necessarily restrictive) impossibilities, is rendered conceivable: the reductive and ingredientist conception of consciousness, the conception according to which consciousness could exist as an ingredient of some bodies, is necessarily misleading.

Wittgenstein then invites us to interrogate ourselves with the eventual feelings we could experience, if we would imagine such a result, and notably the feeling of uncanniness. One might 'after all' remain unconvinced by one's own realization that the imagination of a case of delusory confusion of humans with automata, or machines or robots, does not, and could not amount to the establishment of the eventuality of the relevance of such confusion as such. Could not, and should not some feelings constitute (metaphorical) grounds on the basis of which we could and should reject that humans could be machines or that machines could be humans?

That moralistic resolution of the problem is rejected by Wittgenstein. For we can realize the meaninglessness of the feeling of uncanniness produced by means of the meaningless use of some of our words (the case of would-be attempt of reduction of children to mere automata by *consideration* of their reducibility to mere automata, turns out 'ineffective' in would-be 'optimal' circumstances, that is, in the vicinity of children), and thus the non-conclusiveness of the delusory outcome of the 'thought experiment' can be realized. That is to say, the words by means of which we

supposed an understanding of the reality of the situation to be rendered available to us, lose their sense as we understand that such use of words were not rendering anything available except a misunderstanding of the reality of the situation to ourselves. And we can also realize the correlative meaningless effectivity of the meaningless use of some of our words in the production by us in ourselves of the feeling of uncanniness.

The liberation from the would-be disjunctive entrapment within a dilemma between the meaninglessness of the feeling of uncanniness and its meaningless production by us in ourselves, does not consist in a conclusion, could not be drawn on the basis of premises, and is nevertheless not unargumentative. Rather, appropriation or reappropriation is realized by us by exhaustive consideration or reconsideration.

This realization renders available a non-psychologistic or non-psychological and philosophical achievement with respect to seeing: to use Wittgenstein's examples, though we could use other examples as well, we can see the cross-pieces of a window as a swastika, see that another figure (necessarily imagined) could be obtained by subtraction of some of its elements to a figure (necessarily perceived) in some cases. Importantly, the example put forward by Wittgenstein is a case in which the figure from which another figure can be obtained presents the dimensions and the elements from which the other figure could somehow be obtained. The figure which can be obtained yet is not, and could not, be reducible to an ingredient of the figure from which such figure can be obtained. For the figures and the ends, if any, achieved by production are not necessarily dependent on each other. Not every figure could be, or is meant to be, obtained from every other figure anyway. Some figures could be obtained by us by using some other figures. But some figures could anyway be obtained from each other.

This remark does not, could not imply restriction, or acknowledgement of restrictive limitation. We can also imagine the figure of the swastika to be completed so as to form the figure of the cross-pieces of a window. To this extent, we can see the figure of a swastika as the variant of the cross-pieces of a window and inversely. But would we consider the realization of the completion – not its eventuality – of a figure, to produce one of its variants, or, the subtraction of the elements of a figure to produce one of its variants, then each figure is seen by us or imagined by us as a limiting case of the other, *for the operations which are to be achieved to produce one from the other could not conceivably be the same.*

We neither unavoidably could have had to construe the figure of a swastika to construe the figure of the cross-pieces of a window or the opposite, contrary to the assumption of the ingredientist conception. Con-

straints about figure productions could not have had to be unavoidably thought of as signs of restrictive limitations, and can be thought as unrestrictive limits of modes of conception, production, in cases in which figures are conceived by us, and of constraints – unrestrictive constraints – concerning the modes of conception, production of figures from each other, in cases in which is envisaged the obtaining of a figure from another. The relation between seeing the cross-pieces of a window as a swastika and seeing a living human being as an automaton are similar.

We can imagine, to an extent, the obtaining of the later from the former: the conception and construction of machines, robots, automata have been rendered possible by the subtraction of aspects and dimensions of the lives of humans. It is possible to produce a robot, an artificial intelligence, an automata that replicates aspects and dimensions of the lives of humans. But it is also possible to produce a robot, an artificial intelligence, an automata that does not replicate aspects and dimensions of the lives of humans, for the life of a human, or for the lives of humans. There is not and could not be a common ingredient – consciousness – that would need to be added to some and not others so as to render possible the reversion of the relation: such concept of consciousness is delusive.

To this extent, §420 not only addresses the risk of solipsism involved in the reductionist and functionalist conception, but also the would-be attractiveness of a contrarianist form of reductionism and functionalism, namely “methodological solipsism”. That is to say, if one difficulty is that of the credulity related to a naive form and conception of solipsism, another one is that of the incredulity related to a sophisticated form and conception of solipsism which is “methodological solipsism”, whose distinction from solipsism needs, as earlier remarked, needs to be criticized.

The determinacy of Wittgenstein’s concern with solipsism has been in some sense unhelpfully neglected. The recent publication of the Whewell’s Court Lectures (Wittgenstein and Smythies, 2017) provided us with important passages in which Wittgenstein expresses one’s critical stance concerning solipsism, and the relation between the criticism of solipsism and the problematic of the philosophical relevance of pain:

“Suppose someone said: ‘I am having pain: the other person hasn’t got real pain’ – Solipsism, solipsistically speaking.

We are up against one definite use of language. If I say, ‘Lewy hasn’t got real pain’, he’ll be offended. I’m belittling his sufferings. This I don’t want to do.

The answer would be: ‘Sometimes yes, sometimes no.’ It would be a distinguishing property of language as we know it.” (Wittgenstein and Smythies, 2017, p. 115)

Solipsism is definable under its own terms as the denegation of the reality of the pain of others. Such denegation is according to Wittgenstein one “absolutely definite use of language”. The first important aspect of the case of pain, the reason for which this case constitutes a hard case, is that the case of pain is a (unrestrictively) limiting case of paradigmaticity and verification. An important paradigmatic aspect of pain is that pain has *degrees*, but the objectivation of (the experience of) pain does not necessarily involve reliance on quantification. This could not mean that a quantificational system cannot be used in order to render objective or objectivate the reality of pain, but that there is no such thing as an unavoidable use of a quantificational system to objectivate and objectively agree about the reality of the eventually high degree of the pain of someone (and for example, to evaluate the need for the use of some drugs to attenuate someone’s pain). Pains and degrees of pain can be expressed and measured in diverse ways, and whatever the used measure system is, provided public criteria, the results of the measure will be translatable into other measure systems, eventually with some little loss in accuracy, but negligible loss (and eventually undefinitive) with respect to the ends in which the measurement activities are carried out.

However, the objectivation and the eventual measures of pain imply the acknowledgment of the necessary secondariness of the denegation of the reality of pain. That is to say, we can well imagine or observe that someone fakes feeling or resenting some pain. But such cases are understandable as such against our having internalized the available intelligibility of a primary range of cases, in which, pain is felt and is somehow expressed by someone. It is as pain is felt that pain is expressed and not as pain is expressed that pain is felt. With respect to verification this might seem to cause, generate, induce, or raise a problem: by contrast with other cases of measurement activities, not only that someone’s pain is not necessarily perceived, but it also is not always relevantly expectable to be observed or objectivated, except by the mediation of our acknowledgment of the words of others. Verification of pain thus can at least sometimes be assumed to be impossible.

That was the position of the problem addressed, as we earlier studied, by Lewis, and involved by the criticisms made both by Lewis and Wittgenstein of verificationism. For, if it is acknowledged that sometimes verification of the feeling of pain by someone is impossible, then it is not inconceivable that such verification could always be lacking. If words could conceivably be used by others as by oneself to affirm that pain is felt although that is not the case (for some ends, whichever are these), maybe conceivable doubt concerning the expressions of pains of others could



always be relevant. If such doubt can at least seem to be always relevant, then a verification could always be missing in the case of pain. But then, even in one's own case, pains could eventually be unverifiable, always probable only, although, one does not see the way in which one could be wrong in expressing one's pains, which are not, strictly speaking, ascribed to oneself by oneself, but expressed by ourselves.<sup>26</sup>

## 2. The resolution of the problem posed by C. I. Lewis

Wittgenstein's analysis of the philosophical relevance of pain, and aspects of his dissolution of the problematic of private language responds to a central aspect of the problem addressed by Lewis: the negation of the reality of pain, involved by the solipsistic claim as earlier defined.

Let us recall that Lewis does grant the conceivability of "metaphysical solipsism" and argues in favor of the relevance of a minimal sense and conception of "metaphysics" which is meant to provide some grounding to the false rejection, the wrong calling into question of the existence of a connexion between observed behaviour and pain. Such connexion could be unverifiable and unknowable in the absence of the acknowledgment of the existence of a "metaphysical" connexion which would provide the ground, ensure the existence, of a connexion between an observed behaviour and pain.

We considered that one centrally beneficial aspect of Lewis' conception, which is congruent with Wittgenstein's criticism of solipsism and methodological solipsism, is that his minimally "metaphysical" conception is compatible with any moralistic conception of the veridicity of the expressions of their pains by humans, any conception according to which one must only veridically express that pain is felt by oneself because of some prescription, rule, law internal to a world-conception (or form of life, in Wittgenstein's terms). Indeed, any such conception is compatible with the existence of connexions between behaviours expressive of pains and experiences of pains (by contrast with the cases of machines, robots) and incompatible with fake expressions of pains by persons while no pain is felt by them.

But the force of this conception is also in some sense a weakness. For every connexion between a behaviour expressive of pain and the experience of pain should arguably be *grounded*, inasmuch as if such groundings did not exist, then the claim of which the grounding constitutes the

---

26 On this see Putnam (1975a, p. 362).

basis would not be grounded. The force and contextual relevance of such conception stems from the establishment of the necessary compatibility of each true conception with each other with respect to shared human needs and interests. The weakness of this conception is related both to the modes of the conception and to the reply which would arguably be required to be made to “metaphysical solipsism”, to be refuted under its own terms.

For we already considered that in some sense an exhaustive generalization would be, according to Lewis, involved by the legitimate acknowledgeability of a relevant doubt of the existence of a connexion between a behaviour and a pain. That is to say, if such connexion can relevantly sometimes be assessed to be lacking, then nothing precludes that such connexion could always be lacking. A dichotomous approach should nevertheless, according to Lewis, enable us to settle the question: we should be able to assess that: either a pain is connected to a behaviour and reciprocally, or a pain is not connected to a behaviour and reciprocally. A pain cannot be connected and not be connected with a behaviour in the same sense and reciprocally, a behaviour cannot be connected and not connected with a pain in the same sense. A pain can sometimes be connected to a behaviour (for example, one sometimes tells others about one’s headache; others sometimes tell us about their headaches). A behaviour can sometimes be connected to a pain (for example, someone might consider that such and such behaviours and actions are done by a person when that person feels pains in one’s knees, which are not similarly achieved by each one else in such case). Nevertheless a pain is not each time connected to a behaviour (sometimes one does not tell others about one’s headaches; sometimes others do not tell us about their headaches). And neither is a behaviour each time connected to a pain (for example, one can truly consider that another person faked *again* being in pain).

“Metaphysical” anti-solipsism is meant to provide an infallible response to “metaphysical” solipsism: the false denial of the existence or the inexistence of a connexion between a behaviour and a pain and reciprocally must always be wrong. And as a result of the consideration of the comparison between robots and humans, we studied that according to Lewis there is no such thing as relevantly rejecting each conceivable distinction between, or affirming the indistinctness of, humans and robots. That is to say, according to Lewis there should never be a “consistent solipsist” who could make “the monstrous supposition that other humans are merely robots”, as also this could have for outcome or result the provision of meaning to solipsism although solipsism should not be provided any meaning at all.

In a sense, Wittgenstein invited us in §420 to make exactly the supposition that Lewis invited us to reject – but in way which is different from that of Turing – and which has for first result to liberate us, if required, from the tacit acknowledgement of the eventuality of an event, which all things considered, could not have happened anyway: the transformation of humans into robots and reciprocally as the result of our ‘thought experiment’.<sup>27</sup> Imagination is not meant to be restricted in any sense if the issue raised by solipsism can be addressed at all. But Wittgenstein’s conception enables us to solve the problem posed by Lewis, with a radically different account of generality, a different account of relations between solipsism and skepticism, and a different account of the requirements internal to the intelligibility of the metaphor of humans as machines.

First, on Wittgenstein’s approach, the possibility for a person to fake an expression of pain does not, could not invalidate or disprove, that we express our pains. Quite the contrary, in fact. As mentioned, Lewis’s conception does not imply that the first range of cases we need to consider when observing the expression of pain by someone are cases of persons who are faking being in pain. And ultimately, Lewis also rejects methodological solipsism. Methodological solipsism is also as considered by Wittgenstein a sophisticated form of solipsism according to Lewis, but a “metaphysical response” should nevertheless be provided according to him to “metaphysical solipsism”.

That is the sense in which “metaphysical” solipsism should be at least in principle always be established to be wrong by “metaphysical” anti-solipsism which necessarily is common to every conception of “metaphysics” compatible with human needs and interests. A relevant contrast between the approaches of Lewis and Wittgenstein can then be spelled out: if on Wittgenstein’s approach, it is unclear that solipsism, understood under one’s own terms, could be right, in the way solipsism requires to possibly be, according to Lewis, to be right *at all*, then it is no more clear is that the wrongness of solipsism, understood under one’s own terms, should be establishable, according to Lewis, for the wrongness of solipsism to be established at all.

In other terms, while Lewis grants the possibility of the truthfulness of a “metaphysical” sort of solipsism to render explicit that any coherent “metaphysical” anti-solipsism can establish its falsity – except if the solipsist is coherent enough in the self-production of one’s solipsism, Wittgenstein does not grant the possibility of the truthfulness of a “metaphysical” sort of solipsism which would await its refutation. Solipsism does never,

---

27 And obviously although the cyborg is neither a mere human nor a mere robot, the consideration of the cyborg case is not as such a sufficient answer to the problem which is not that of the lack of an intermediate case between robots and humans.

could not start to present the relevance which would justify the acknowledgement of the existence of its false grounds. The negative replication of the difficulties generated, produced, raised, and posed by solipsism could not be relevant at all, and even less, philosophically.

### 3. The resolution of the problem posed by Turing

We have thus reached another crossroad. Both Turing and Wittgenstein invite us to do what Lewis invites us to reject, but not in the same senses. Wittgenstein invites us to reject metaphysical anti-solipsism and methodological solipsism with solipsism, but not the notion of consciousness.<sup>28</sup> While Turing does not invite us to reject metaphysical anti-solipsism and methodological solipsism with solipsism, but does neither preserve the notion of consciousness.

From the outset it can be remarked that some counter-objections envisaged by Turing to one's own argument idle. The contrast between machines and humans could not be blurred or rendered less accurate by the acknowledgment that machines (also) think. Our concept of consciousness does not necessarily, could not have necessarily implied reliance on the presenting of unrestrictive limits brought out at the occasion of the comparison of machines and humans as (restrictive) impossibilities. No one is or should be considered as eventually *forced* or coerced into a solipsistic position, and especially not for the sake of the establishment of the truth of an argument. The criticism of solipsism can be more direct and should be more direct to be addressable at all.

Further, Wittgenstein's evolutive conception of language renders conceivable to think the possibility of compatibility or agreement with respect to the ascription of actions to machines (and artificial intelligences) without calling into question the relevance of the notion of consciousness which is central in world-conceptions (by contrast with the notion of subjectivity), and to account for the distinction between humans and machines whenever required. The production of a luring situation is not, could not be conclusive in the way Turing presented, and Turing's achievements are (hopefully!) independent from an argument that is no more, and could not have been conclusive anyway. One way to express the point made by Wittgenstein is to remark that natural history is both *natural* and *historical*, that our history is not the history of men, but of humans.

---

28 Wittgenstein used the comparison of humans and machines when he defined "Turing's 'machines'" as "humans who calculate" (Wittgenstein, 1947, Ts-229, 448). On this passage see Floyd (2012a, p. 40; 2012b) and Shanker (1987, pp. 615-623), and on the relations of Turing and Wittgenstein see Floyd (2021, pp. 123-126).

## Conclusion: Independences and Forms of Life

This article proposed a reflexion about the limits of the comparison, analogy or metaphor between humans and machines. As such, the comparison could not be problematic: humans and machines present common aspects, and instances of such comparison are implicit in ordinary, engineering, scientific, and medical practices. That many progresses have been rendered possible also with, or in ways compatible with the use of this comparison could not have had to be established again. But the extent to which the comparison can be metaphorically literally understood, if it can be metaphorically literally be understood at all, is, as studied, a question whose stakes are of primary philosophical importance. Thematizations of this comparison of the XXth century, whether entirely philosophical – such as those of Lewis and Wittgenstein, or presenting philosophical significance – such as that of Turing, are indeed intertwined with the problematic of solipsism. As much as linguistic practices are concerned, we considered that there exist appreciative and depreciative ordinary linguistic uses which do involve this comparison, and testify of the available intelligibility of distinct ranges of cases which do not involve, and are not compatible with the confusion of machines with humans.

The first part of this paper presented Lewis' critical conception of solipsism against this background. The affirmation of the indistinction, or the negation of any distinction between machines and humans is necessarily problematic, necessarily misleading or delusive. According to Lewis, a minimally "metaphysical" conception is required so as to disprove "metaphysical" (and methodological) solipsism which consists in the negation of the existence of connexions between pains and behaviours expressive of pain.

However, we considered in the second part of the article, that the successful establishment by Turing that machines can be unproblematically be said to think – notably by means of the introduction of the "imitation game" – involves the assumption of a disjunctive entrapment between either defending that machines think, or, defending both consciousness and (a reconceived and unphilosophical conception of) solipsism. Fully acknowledging Turing's criticism of a traditional conception of consciousness according to which machines would be *deprived* of thoughts and emotions, we nevertheless considered that Turing's reconception of solipsism contributed to the substitution of an unphilosophical conception of solipsism to a philosophical conception of solipsism in an undue way.

Indeed, philosophically accounting for consciousness could neither necessarily involve to grant that machines *could* be *deprived* of thoughts

and emotions, nor to defend solipsism. The criticism of such unavoidable disjunctive entrapment is achieved by Wittgenstein in the *Philosophical Investigations*, as studied in the third part of this article. Indeed, reciprocal asymmetries with respect to attributions of pains and consciousnesses to humans and machines are inconceivable when presumed as involving reciprocal and necessary restrictive impossibilities. Infallible response to “metaphysical” solipsism then is no more than “metaphysical” solipsism, required or relevant to address the problematic of solipsism. At stake is no less than our conceptions of science, of diversity and of forms of life: scientism could not substitute for science, exclusion could not be compatible with diversity, forms of life could not be compatible with solipsism. The relatedness of some forms of life could not imply the mutual dependence of each form of life with each other. This can seem to be incompatible with ecology, but, on the contrary, is not: wholistic reflexion is not and could not be based upon mechanistic reductionism.

## References

- Aristotle (1995). *Politics*. Edited by R. F. Stalley. Translated by Ernest Barker. Oxford & New York: Oxford University Press.
- Bouveresse, J. (2022). *Les vagues du langage*. Lonrai: Seuil.
- Copeland, J. (2004). *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*. Oxford: Oxford University Press.
- Descartes, R. (2006) *A Discourse on the Method*. Translated by Ian Maclean. Oxford & New York: Oxford University Press.
- Davidson, Donald. (2004). “Turing’s Test.” In *Problems of Rationality* (pp. 77–86). Oxford: Clarendon Press.
- Descombes, V. (1995). *La Denrée Mentale*. Minuit: Paris.
- Floyd, J. (2012a). Wittgenstein, Carnap, and Turing: Contrasting Notions of Analysis. In P. Wagner (ed.), *Carnap’s Ideal of Explication and Naturalism* (pp. 34–46). London: Palgrave Macmillan. [https://doi.org/10.1057/9780230379749\\_4](https://doi.org/10.1057/9780230379749_4).
- . (2012b). Wittgenstein’s Diagonal Argument: A Variation on Cantor and Turing. In P. Dybjer, S. Lindström, E. Palmgren, and G. Sundholm (eds.), *Epistemology versus Ontology* (pp. 25–44). New York & London: Springer. <https://link.springer.com/book/10.1007/978-94-007-4435-6>.
- . (2021) Turing on ‘Common Sense’: Cambridge Resonances. In J. Floyd and A. Bokulich (eds.), *Philosophical Explorations of the Legacy of Alan Turing* (pp. 103–149). Cham: Springer International Publishing. <https://link.springer.com/book/10.1007/978-3-319-53280-6>.
- Gonçalves, Bernardo. (2024). *The Turing Test Argument*. New York & London: Routledge.

- Kant, Immanuel. (2007). *Critique of Judgment*. Edited by Nicholas Walker. Translated by Martin Meredith. Oxford & New York: Oxford University Press.
- Kennedy, J. C. (2021). Turing, Gödel and the 'Bright Abyss.' In J. Floyd A. Bokulich (eds.), *Philosophical Explorations of the Legacy of Alan Turing* (pp. 63–92). Cham, Switzerland: Springer International Publishing, 2021. doi.org/10.1007/978-3-319-53280-6.
- . (2022). Gödel, Turing and the Iconic/Performative Axis. *Philosophies* 7 (6), 141. https://doi.org/10.3390/philosophies7060141.
- Lewis, C. I. (1923). A Pragmatic Conception of the A Priori. *The Journal of Philosophy* 20 (7) (pp. 169–177). https://doi.org/10.2307/2939833.
- . (1929). *Mind and the World Order: Outline of a Theory of Knowledge*. New York: Dover Publications.
- . (1934) Experience and Meaning. *The Philosophical Review* 43 (2), 125–46. https://doi.org/10.2307/2179891.
- . (1970). *Collected Papers*. Edited by J. D. Goheen and J. L. Mothershead. Stanford: Stanford University Press.
- Onfray de la Mettrie, J. (1996). *Machine Man and Other Writings*. Translated by Ann Thomson. Cambridge, New York, Melbourne: Cambridge University Press.
- Mundici, D., and Sieg, W. (2021). Turing, the Mathematician. In J. Floyd and A. Bokulich (eds.), *Philosophical Explorations of the Legacy of Alan Turing* (pp. 39–62). Cham: Springer International Publishing.
- Putnam, H. (1975a) Minds and Machines. In *Mind, Language and Reality* (pp. 342–61). Cambridge, New York, Melbourne: Cambridge University Press.
- . (1975b) Robots: Machines or Artificially Created Life?. In *Mind, Language and Reality* (pp. 386–407). Cambridge, New York, Melbourne: Cambridge University Press, 1975.
- . (1996). Why Reason Can't Be Naturalized. In *Realism and Reason*. (pp. 229–47). Cambridge: Cambridge University Press, 1996.
- Sartre, J.-P. (2003). *Being and Nothingness*. Translated by H. Barnes. London: Routledge.
- Shanker, S. G. (1987). Wittgenstein versus Turing on the Nature of Church's Thesis. *Notre Dame Journal of Formal Logic* 28 (4), 615–649.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind* LIX (236), 433–60. https://doi.org/10.1093/mind/LIX.236.433.
- Uçan, T. (2023). Autonomy, Constitutivity, Exemplars, Paradigms. *Conversations: The Journal of Cavellian Studies*, (10), 52–79. https://doi.org/10.18192/cjcs.vi10.6613
- . (2016) The Issue of Solipsism in the Early Works of Sartre and Wittgenstein 2016. *University of East Anglia Digital Repository*. https://ueaeprints.uea.ac.uk/id/eprint/62314/1/2016UcanTUPhD\_%282%29.pdf.

- Wittgenstein, L. (2009). *Philosophical Investigations*. Edited by P. M. S. Hacker and J. Schulte and translated by G. E. M. Anscombe, P. M. S. Hacker, and J. Schulte. Oxford: Blackwell.
- . (1947) Ts-229,448 Facsimile 1947. <http://www.wittgensteinsource.org/>
- . (2003). *Tractatus Logico-Philosophicus*. Translated by C.K. Ogden. New York: Barnes & Noble.
- . Tractatus Map. University of Iowa Tractatus Map. Accessed December 21, 2018. <http://tractatus.lib.uiowa.edu/>.
- Wittgenstein, L., and Y. Smythies. (2017) *Wittgenstein's Whewell's Court Lectures*. Edited by V. Munz. and B. Ritter, Malden and Oxford: Blackwell.



### 3. NORMATIVITY AND COLLECTIVE INTENTIONALITY



Vasiliki Xiromeriti

## COLLECTIVE DELIBERATION IN EPISTEMIC GROUPS: LESSONS FROM DELIBERATIVE DEMOCRACY

**Abstract:** The paper examines the role of collective deliberation in epistemic collaborations. While collective deliberation is extensively studied in political philosophy and democratic theory, its role in genuinely epistemic contexts remains underexplored. The paper argues that collective deliberation is vital for achieving understanding and justifiedness, particularly in situations where epistemic questions permit multiple solutions because of diverse background theories and methodologies. Building on Bratman's account of shared agency, the paper describes epistemic collaborations as joint actions, where deliberation plays a central role in shaping collective views. Collective deliberation plays an important role not only in structuring epistemic collaboration but also in justifying the collective results. The paper advances a dialectical or deliberative account of group epistemic justification, where the latter relies on a process of giving and asking for reasons. By integrating insights from deliberative democracy, normative standards for collective epistemic deliberation are proposed.

**Keywords:** collective deliberation, epistemic collaboration, deliberative democracy, group epistemic justification.

### 1. Introduction

It is widely accepted that knowledge is a social achievement. Scientific inquiry builds on collaborative interactions among experts who divide epistemic labor. Division of labor often combines with diversity and disagreement, which are thought to bear epistemic benefits to the scientific community (Zollman, 2010), and to be at the origin of scientific progress (Feyerabend, 1970, p. 203). Moreover, given the complexity of today's issues (climate change, for example), much research is based on

interdisciplinary collaborations, uniting scientists from different fields and educational backgrounds. Although it is clear that knowledge in these contexts is a collaborative achievement, the way in which individuals come to share and integrate knowledge is not sufficiently understood. How is it possible for individuals, coming from different disciplines and cultures, to successfully collaborate on epistemic questions? And, when they do, on what grounds are their collective views justified?

The paper aims at exploring the role of collective deliberation in epistemic collaboration and justification. During the past decades, collective deliberation has received much attention in normative philosophy – especially, in democratic theory. Deliberative democrats argue that collective deliberation – conceived of as an uncoerced discussion among equals – can lead to “better” collective outcomes (Habermas, 1993; Estlund, 2008; Landemore, 2017). The superiority of collectively deliberated decisions – to those resulting from voting, for example – owes to the fact that participants to deliberation are encouraged to assess alternatives by providing and comparing reasons for and against competitive claims. Decisions produced this way are thereby more likely to satisfy some epistemic requirements for political decision-making – i.e., they tend to conform to some notion of truth, namely the principles of justice or the common good.

Although deliberative democracy is often discussed as a theory of political justification, the normative role of deliberation in genuinely epistemic contexts remains underexplored. There is undoubtedly a good reason for this. One cannot really argue about facts of the matter. Unlike practical questions – with respect to which one is free to make up her mind – epistemic questions about how the world is are rationally settled in a compulsive way – by reference to agent- and context-independent criteria – namely evidence and proof. Evidence being objective, unlike personal commitments and preferences, disagreement should be considered as an indication of error or lack of information. Collective deliberation, involving weighing reasons for and against possibilities, is thus reserved to questions that call for a decision. On the contrary, collective inquiry seems to be a matter of integrating evidence from different sources, including higher-order evidence regarding each other’s expertise, and not deliberation.

However, it is still possible that there is more than one permissible explanations for a single phenomenon. The choice among them cannot be settled by solely appealing to evidence but remains dependent on background theories and methodological assumptions. There might

be different perspectives on what can count as evidence or on what one can reasonably infer from it (Longino, 2001). Indeed, the evidence available at some given time may be insufficient to determine what beliefs one should hold in response (the underdetermination thesis), or a body of evidence could support conflicting beliefs about the same proposition (the permissivism thesis). Diversity and disagreement of this sort are particularly salient in scientific collaborations – especially in interdisciplinary ones – and cannot generally be discounted to lack of evidence or intelligence.

If epistemic questions can have more than one permissible solution and cannot be tracked by appealing to some overriding principle (e.g., evidence), they need to be settled through argumentative interaction. This is why, according to Longino (2001, p.129), an adequate normative theory of knowledge should focus on the deliberative process through which knowledge is produced. However, a strictly procedural account of knowledge would be insufficient. One cannot possibly discount the epistemic requirements for truth, without inflicting violence on the concept of knowledge itself (Rescher, 2003, p.10).

This paper aims at clarifying the role of deliberation in epistemic reasoning in groups. By reinstating its role in scientific collaborations, it will build on the deliberative theory of democracy to propose an account of deliberative epistemic justification of collective views. The paper is structured as follows: It begins by addressing the role of collective deliberation in epistemic collaborations, by relying on Bratman's account of shared agency. Secondly, the paper focuses on the question of what group epistemic justification is. The normative account I defend can be described as dialectical or deliberative, since collective justification depends on a process of giving and asking for reasons – in other words, on argumentation. Finally, I will defend this epistemological account, by relying on the deliberative democratic theory. This way, I will respond to objections raised against dialectical accounts of epistemic justification, by showing that collective deliberation can respond to epistemic requirements.

## 2. Epistemic Collaborations

Collaboration is a salient phenomenon in science. Scientists often work together in research teams and publish their findings in multi-authored papers (Thagard, 2006). Moreover, scientific collaboration today seems to respond to a practical necessity (Rolin, 2015). Contemporary scientific

questions require research teams to integrate expertise from different specializations or disciplines, and to combine materials and resources (Andersen and Wagenknecht, 2013). Hence, groups play a central role in knowledge production. But what does it mean for agents to collaborate on epistemic questions? How are epistemic collaborations possible, especially among individuals with different expertise and epistemic perspectives?

Although epistemology generally takes group beliefs as a departure point and raises the question whether these beliefs can have the features of knowledge (truth and justification), I will begin by noticing that epistemic collaborations, such as research teams or authors publishing together, can be described as collaborative activities. This means that collaborations, including epistemic ones, are characterized by a shared intention. But what does it mean for a group of individuals to share an intention? There are many influential accounts for shared agency in the philosophical literature (Gilbert, 2000; Tuomela, 2007), but my explanation here will be broadly inspired by Bratman's analysis of shared intentional action (Bratman, 2014). For the purposes of the present enquiry, Bratman's approach presents two important advantages in comparison to competing accounts. First, Bratman's account makes the fewest possible normative assumptions on the ground level of explanation. For shared intentional action to be possible, it is not required that there is a strong institutional background or, more generally, a consensual evaluative or cognitive background among participants. Thus, it makes it possible to consider collaboration even in contexts where substantial disagreement prevails – in interdisciplinary groups, for example. Secondly, Bratman does not take shared intentions as given but addresses the process by which they come to be developed. Individuals may have a plan of acting together, but this plan needs to be “filled in” through reasoning on the part of the members of the group. However, unlike Bratman, I will be interested in epistemic, rather than genuinely practical, contexts.<sup>1</sup>

For Bratman, shared intentions should not be understood as attitudes in individual minds (i.e., “I intend to do my part in our shared action”), nor as attitudes belonging to a collective super-mind – for such a mind does not exist. Rather, a shared intention is a network of appropriately interconnected individual intentions. This way, Bratman proposes a non-summative, yet bottom-up account of shared agency. Following Gilbert's distinction (Gilbert, 1989), while summative views reduce the group's intentionality to an aggregation of the intentional states of its members,

---

1 Although I rely on Bratman's account of shared intentionality and I generally take collective views to be acceptances, rather than beliefs, I do not rely on Bratman's definition of acceptance (Bratman, 1992).

non-summativ accounts recognize a distinctive intentional standing to the group. At the same time, for Bratman, group intentionality remains responsive to the intentions and attitudes of the members of the group, the interplay of which is at the origin of the shared intention.

This is a plausible description of epistemic collaborations. Indeed, epistemic collaborations are characterized by a shared intention (e.g. a research project). However, this project is not a mere aggregation of the contributory intentions and actions of the members of the group. Although a research team has a shared plan, individual scientists contribute differently towards its realization, each having a distinct task – or contributory intention. After all, epistemic collaborations express views that are irreducible to those of their members. Collaborative research generally gives rise to conclusions that none of the members could ever individually reach or hold, but that emerge as a result of complex interactions among them (Palermos, 2022). A multi-authored research paper is not generated by simply adding up the authors' contributions. Also, even if there is a dissenting minority or the group view expresses a compromise that does not conform to anyone's opinion, the members can let it stand as their collective view on some epistemic question. The collective view is nevertheless responsive to the contributory intentions and actions. While some epistemological approaches (Kuhn, 1970), inspired by the sociology of science, explain individual contributions in terms of their relation to the cognitive resources of a scientific community, there are reasons to avoid trivially defining individual contributions this way. Indeed, in interdisciplinary groups, there is no single type of cognitive resource – contemporary research cuts across disciplinary boundaries. Moreover, it is the aim of science to constantly reflect and develop the cognitive resources and tools available within the community (Andersen, 2016).

The existence of an overall shared plan is a necessary condition for joint action, but not a sufficient one. For example, is not sufficient that we both plan to paint the room together for us to actually paint the room together (Bratman, 2000, p.122). In scientific collaborations, conditions for joint action as presented by Bratman do also hold. It is undoubtedly necessary that the group has a common project – e.g., to explain certain phenomena, to model a system, or to validate a hypothesis. Yet, a research team divides the labor among its members. Individual scientists or subgroups of scientists are generally supposed to contribute to the project according to their specialization and by accomplishing different tasks. Contributory intentions and actions are thus interdependent, in the sense that they are mutually efficient with a view to the realization of the project<sup>2</sup>.

2 In Bratman's vocabulary, one could say that contributory intentions "interlock".

Interdependence also involves beliefs about each other's commitment to the common project. Given the division of labor, collaborators need to trust each other that they will complete their task reliably for collaboration not to break down. Finally, the shared goal, as well as interdependency in efficiency and persistence, are matters of common awareness within the group (Bratman, 2014).

However, this complex intersubjective structure should not be taken for granted. Imagine that a research group is assigned a research project. The means by which this project will be carried out are not obvious. One cannot expect individual scientists to unilaterally work on some tasks according to their specialization and simply add up their findings to generate the collective result. Contributory intentions might conflict, and disagreements are likely to arise as to the relevant evidence, background theories and hypotheses, methods to be employed and specific goals to be achieved. Unless contributory plans and views are assigned an appropriate weight, their working together could hardly count as a genuinely collaborative activity. This is why Bratman introduces a further condition for shared intentional action – that of meshing subplans. In other words, participants to the joint activity need to be responsive to each other's reasons and try to make them mutually satisfiable and coherent.

This idea gives a central role to collective deliberation in building collaboration. Meshing subplans and achieving this kind of intersubjective rationality is a deliberative issue. Individuals need to provide and compare reasons for and against competitive claims in their effort to coordinate with a view to their shared intention. Research teams need to engage in reasoning of this sort. Division of labor requires scientists to appeal to each other, assess and organize partial contributions and testimony rather than acting by themselves. In case of epistemic divergence and disagreement, they need to combine their cognitive resources and make their contributions coherent for their collaboration to be efficient.

However, for shared reasoning to be possible, participants need to be able to evaluate and respond to each other's reasons. Thus, there should be some form of common ground, enabling mutual intelligibility and justifiability. Does this mean that collective deliberation is only possible among individuals of the same mind? It would be too strong to assume that, to be able to deliberate, individuals should have a background of consensual value judgments and beliefs. After all, many interdisciplinary collaborations succeed, while scientists have different cognitive resources, training, and conceptual devices.

Bratman suggests a weaker way of understanding this common ground. He claims that what is necessary for shared reasoning to be pos-



sible is not a background consensus on values and beliefs, but rather common policies about weights (Bratman, 2014). In other words, participants to deliberation should share an intention to treat certain types of considerations as mattering in their reasoning together, while excluding others. For instance, a scientific team would assign a proper weight to experimental findings and exclude considerations of religious nature, even if all members of the group are religious. When it comes to science, these might be shared epistemic standards for assessing contributions. These standards can be quite general in nature and defeasible, and there needs to be no agreement on their relative priority.

More generally, this common ground corresponds to a common background knowledge, which is procedural rather than propositional (Ryle, 1994; Thagard, 2006). This could involve policies of giving weight to certain types of considerations (epistemic reasons) or procedural rules (respect, consensus, etc.). Moreover, these shared commitments do not have a merely instrumental role in shared deliberation. They represent the group's point of view on an issue (Bratman, 2014, p. 141), according to which, in a context of common knowledge, different scientific assertions must be justified.

To summarize, epistemic collaborations can be described as intentional actions on the part of the group. Scientists often collaborate on common projects to combine epistemic resources and material means, to reach conclusions they could not unilaterally reach or could only reach with substantial costs. As such, contributory intentions, views, goals, and actions should be appropriately interrelated for the group to be performative. Following Bratman, this means that relevant attitudes and intentions should be appropriately responsive to intentions and actions of others. This is ensured through collective deliberation, which is required in several stages of the joint enterprise.

### 3. Group Epistemic Justification

Epistemic collaborations can be described as joint actions, characterized by a shared intention. They are indeed meant to address common epistemic questions, by adequately combining the cognitive resources and actions of their members. The main difficulty with this action-centered approach is that it leaves open the question of epistemic justification – that is, the question of whether the group is justified in holding a view that *por*, in other words, whether this view can have the features of knowledge (truth and justification). Indeed, within this approach, epistemic ra-

tionality can easily be understood as instrumental rationality, which does not exclude justification for pragmatic reasons. Shared deliberation has been described as aiming at coordination with a view to joint action, but agreement and efficiency are not by themselves epistemic values. Hence, one needs to understand how groups come to be epistemically justified in holding a view.

Collaborators in this case do not merely aim at acting together. Their goal is knowledge. Since they are meant to provide a response to epistemic questions – that is, questions of explanation and prediction – potential agreement on some view should be motivated by epistemic reasons. In other words, their collective views are subject to standards of epistemic rationality. The question is then how and when individuals working together can be said to *know*. On what grounds are the collaborative results epistemically justified? In the growing literature about knowledge in groups, one can distinguish two types of approach to group epistemic justification – *deflationary* and *inflationary* approaches (Lackey, 2016)<sup>3</sup>.

According to deflationary accounts, a group *G* is justified in holding a view that *p* as true if and only if all or most of its members are justified in holding a view that *p* as true<sup>4</sup>. Within this approach, one can also consider justification as a matter of degree, increasing with the number of individuals justifiedly holding the view in question (Goldman, 2014). In fact, deflationary approaches understand the justification of a group's view in terms of the justification of the individual members' views. They claim that justification of doxastic attitudes is the same for a group as it is for an individual. A group, as well as an individual, aims at the truth and the only means for accomplishing this goal is by considering all the relevant evidence. Since doxastic attitudes have a mind-to-world direction of fit, and since there are objective facts out there, when rational individuals are exposed to the same evidence, they will eventually converge on the truth. According to Meijers, if only epistemic reasons for believing are taken into consideration, it is impossible that there is a difference in content between individual beliefs and the beliefs of the group (Meijers, 2003). Therefore, deflationary approaches argue that, for a collective view to be justified, it needs to be responsive to individuals' justifiers – that is, individual mem-

---

3 Although these approaches are often denoted as summative and nonsummative – respectively, I avoid this notation in order not to confuse them with the notion of group employed. It is possible to hold a nonsummative deflationist view and a summative inflationist view of group justification.

4 Proponents of this account typically refer to beliefs. I reformulate “believing *p*” by “holding the view that *p* as true” to avoid the debate on whether groups can have genuine beliefs or mere acceptances.

bers' private evidence, including higher-order evidence about each other's expertise and reliability.

Deflationary approaches rightfully insist that the collective result, in order to be appropriately justified, should be responsive to the individuals' reasons. To the extent that epistemic collaborations are characterized by an epistemic intention – that is, they aim at the truth, ignoring or not assigning appropriate weight to relevant evidence is irrational. The problem is that this approach assumes that individuals' justifiers could be easily added up. This assumption underestimates diversity and disagreement within epistemic groups. Disagreement, here, is taken to be irrational. Rational individuals are supposed to be able to agree once all relevant evidence is taken into consideration. However, disagreement may well be justified by diversity in methodologies and background theories and assumptions (Dang, 2019). In research teams, scientists are likely to disagree on how to assess evidence or on what one can reasonably infer from the evidence collected. Collective epistemic justification is thus not identical to individual justification since it further requires an aggregation procedure and compromise.

Further limitations appear once one considers group justification in terms of aggregation. Imagine an experts' committee seeking to make a prediction about whether the planet's temperature will increase during the next decade. The proposition "The planet's temperature will increase during the next decade" ( $c$ ) is supported by a set of other propositions: "CO2 emissions will increase" ( $d$ ), "Increase in CO2 emissions causes rise in planet's temperature" ( $d \rightarrow c$ ). According to what has come to be known as the "discursive dilemma" (Pettit, 2001), majority voting on interconnected propositions may lead to inconsistent group judgments, even when individual judgments are fully consistent (List and Pettit, 2002). Consider the following example, where majority endorses  $d$  and  $d \rightarrow c$ , while rejecting  $c$ .

	$d$	$d \rightarrow c$	$C$
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
<b>Majority</b>	<b>True</b>	<b>True</b>	<b>False</b>

In other words, aggregation cannot secure a deductively closed and consistent set of collective attitudes. On the basis of justified individual

judgments, the group ends up justifiedly endorsing a contradiction. Thus, aggregation does not necessarily preserve rationality or justifiedness.

Such problems do not arise if one adopts an inflationary account of group justification. Inflationary accounts take the group as the epistemic subject and collective views are justified by means of reasons accepted on the collective level. In other words, the group justifiedly holds a view that  $p$  as true if and only if the group is itself justified in doing so, over and above the individual members' views. Thus, even when scientists within a research team do not actually agree on the conclusion, they have reasons to let it stand as the group's view. According to Schmitt, who defends such an account,

“A group  $G$  justifiedly believes that  $p$  if and only if  $G$  has a *good* reason to believe that  $p$ , and believes that  $p$  for this reason, where  $G$  has a reason  $r$  to believe that  $p$  if and only if all members of  $G$  would properly express openly a willingness to accept  $r$  jointly as the group's reason to believe that  $p$ .” (Schmitt, 1994, p. 265 – italics added)

This approach, also known as the joint acceptance account of group justification, argues that a group is justified in holding a view that  $p$  as true, if the members openly agree to let a reason  $r$  stand as the group's justifier. And, according to Schmitt, one should think of this reason as an epistemic one:  $p$  has been obtained through a reliable process, it is grounded in adequate evidence, and so on. It is important to notice that it is not necessary for individuals to actually express such an agreement. The endorsement of the relevant reasons can be entailed by their general commitments (Lackey, 2016, p. 347). In other words, joint acceptance does not need to be motivated by explicit acceptance of the relevant reasons through discussion and argumentation. It is sometimes enough to think of what individuals *would* accept. Agreement on the relevant reasons can thus be presupposed or assumed, given one's belonging to a group.

While group views can indeed be justified on their own right, independently from whether the individuals subjectively hold the views in question, this account faces some problems. The account appears to be both too strong and too weak for group justification. First, it makes group justification “too hard to come by” (Lackey, 2016, p. 350). Group justification here requires agreement on the reasons supporting a result. Yet, if groups have hard time agreeing on a conclusion, it is even harder to expect that they will agree on the reasons supporting it. And, sometimes, it is even possible to agree on a result even if they have different (epistemic) reasons for doing so. Hence, this view of group justification seems to neglect (or to suppress) fundamental diversity within epistemic groups (Dang, 2019).

Moreover, the requirement of joint acceptance can seem too weak for the group view to be justified, since there are no further explanations regarding the way joint acceptance has been obtained – it may be the result of manipulation or social pressure. Feminist philosophy of biology has revealed that broad scientific agreement has often dissimulated androcentric biases, resulting from the historically systematic exclusion of women from science. This has given rise to myths about female biology – regarding, for example, female orgasm (Lloyd, 2005). Agreement, or joint acceptance, is thus not constitutive of the correctness of the result.

Collaborative views indeed have a distinctive epistemic standing. They cannot be described as a mere aggregation of individual members' judgments but emerge as a result of complex interactions and reasoning among individuals who divide the labor, at multiple stages of the collaborative inquiry. Reasoning might lead to mutual adjustments and the discovery of novel solutions, which are not amenable to aggregative analyses. Therefore, group epistemic justification is distinct from and not reducible to individual justification. At the same time, group views need to be responsive to individuals' reasons. Not only it would be epistemically irrational not to assign the appropriate weight to individuals' justifiers or defeaters, but also diversity is generally thought to result in epistemic benefits to the group. It makes it possible to expose problematic background assumptions and sources of error, leading to epistemically better results (Longino, 2001). What matters is not whether the result is supported by generally accepted or presumably acceptable reasons, but whether the relevant justifiers have been adequately taken into consideration, compared, and evaluated. Hence, an appropriate account of group epistemic justification should also focus on the epistemic virtues of the procedures by which collaborative results are generated.

This suggests a procedural account of group knowledge. This idea is expressed in Popper's epistemology, where scientific knowledge advances toward the truth by improving tentative theories through a process of error reduction achieved by testing and intersubjective criticism (Popper, 1972). A theory is corroborated or can be provisionally considered as true if it has met the required burden of proof. The degree of corroboration of a theory is determined by factors such as the extent of critical discussion of the theory, its degree of testability, the rigorousness of the tests it has faced, and its resilience in enduring these tests (Popper, 1972, p.18). Knowledge in epistemic collaborations evolves in a similar manner. Participants to the collaboration pool all relevant evidence and submit their contributions to testing and argumentative assessment of evidence, hy-

potheses, and theories. A justified collective view is based on reasons that survive this critical process.

This weighs in favor of a dialectical or deliberative understanding of group epistemic justification. The group jointly accepts letting  $p$  stand as its view because the participants trust the procedure by which this result has been generated. In a dialectical or deliberative understanding of epistemic justification (e.g., Hakli, 2011; Beatty and Moore, 2010), justification is linked to the practice of giving and asking for reasons. Participants are encouraged to express their views, engage in open dialogue, and critically assess the evidence, assumptions, and reasoning behind each perspective. This critical discourse serves as a crucible for refining ideas, identifying weaknesses in arguments, and fostering a collective understanding that transcends personal convictions and biases. Through this process, the collective views that emerge are not only more robust, but are also reflective of a well-informed and scrutinized decision. Claims and views are justified to the extent that they survive a process of giving and asking for reasons. In other words, according to this account, a group  $G$  is justified in holding the view that  $p$  as true if and only if  $p$  has been successfully defended against reasonable challenges during collective inquiry and deliberation. What counts as a reasonable challenge or legitimate defense is determined by the epistemic principles characterizing the epistemic community (Hakli, 2011, p. 15).

It is, however, noteworthy that argumentation differs from demonstration. Indeed, according to this understanding, epistemic reasoning is defeasible. This means that the premises might provide good support to the conclusion, but they do not guarantee its truth. Collaborative results, as sets of propositions, that count as knowledge should not be considered as true, justified beliefs, but as corroborated statements which are subject to refutation once new evidence and better reasons become available. Hence, epistemic rationality is to be understood as “bounded procedural rationality” (Walton, 2016, p. 209). Given our limited cognitive abilities and the resources available in the specific epistemic context, scientific theories are rationally justified through interactive argumentation.

#### 4. Normative standards for epistemic deliberation

Collective deliberation seems to play a central role in structuring epistemic collaboration, especially when the group under consideration is significantly diverse. Although the dialectical understanding of epistemic justification captures the intuitive way of thinking of justification as pro-

viding and responding to reasons, the epistemic properties of collective deliberation have been extensively challenged. It has been argued that deliberation, which is subject to social pressure and confirmation biases, has the tendency to radicalize existing positions, instead of leading to more considered judgments (Sunstein, 2006). In the realm of science, this approach raises further worries because it seems to make collective views depend on negotiation, involving practical and political factors such as career interests and power relations (Pickering, 1986). Moreover, it seems to constitute a rather weak basis for epistemic justification, since a speaker can be quite persuasive when advancing false claims, or the audience can simply be gullible (Lackey, 2016, p. 349). It is thus necessary to defend the epistemic value of collective deliberation, by examining the conditions under which interpersonal argumentation might maximize the chances of getting to correct or true collective views. This issue has already been raised by deliberative democrats. So, it would be interesting to apply the ideas of deliberative democracy to thinking about collective deliberation and judgment in epistemic collaborations.

Deliberative democracy has been proposed as a theory of democratic justification.<sup>5</sup> One needs to notice that there is generally a conflict between democracy and truth – either we conceive it as factual truth, correctness, rightness, etc. On one hand, democracy, by definition, requires collective decisions to be responsive to citizens’ reasons. Moreover, admitting that citizens have an equal moral status requires that their preferences and views are treated in an equal way. Thus, democratic decisions should be produced through a fair procedure that takes into consideration individuals’ preferences and views – typically majority voting. However, without further qualification, voting seems like a garbage in – garbage out process: If individuals are irrational, immoral, or simply self-centered, the collective decision will probably reflect these properties. Democracy then risks violating some fundamental principles of justice – such as human rights. In this sense, democracy seems at odds with epistemic concerns about the quality of the decision. On the other hand, if one privileges this epistemic concern, she should be able to define some univocal procedure-independent principles that the collective decision should satisfy. If this were possible, collective decisions could easily be trusted with a group of moral experts. However, in a pluralistic society, it is likely that different individuals have different conceptions of the common good or of what justice requires. Democracy is thus a means to acknowledge diversity. In

---

5 “The notion of deliberative democracy is rooted in the intuitive ideal of a democratic association in which the justification of the terms and conditions of association proceeds through public argument and reasoning among equals.” (Cohen, 1989, p. 21)

response to this dilemma between inclusion and epistemic justification, deliberative democracy proposes itself as an alternative. Collective deliberation is presented as a meditation between reason and will (Habermas, 2010, p. 175). Collective decisions are responsive to individuals' preferences, but these preferences are no longer taken as primitives. They are informed and tested in a critical debate among equals (Elster, 2003) – eventually before voting.

Deliberative democracy makes democratic justification depend not simply on the fairness of the decision-making procedure, nor on some presumably consensual values or judgments. Justification is rather dependent on the reasoning supporting the collective results. In this sense, it suggests a means for justification of collective views that combines inclusion of the diverse points of view with collective justifiedness. Justification here is not simply a matter of coordination and agreement, but the results should be expected to have some epistemic qualities. In other words, argumentation should lead to “better” decisions. Assuming that there is more than one permissible solution, the aim of deliberation is to track the best response to the problem faced by the group. The question then is under what conditions these epistemic properties of deliberation are satisfied in epistemic groups. In the realm of epistemic collaboration, the integration of deliberative democracy offers a structured approach to decision-making that involves a careful balance between procedural (democratic) norms and substantial (epistemic) principles.

Deliberative democracy places significant emphasis on procedural norms, defining the rules and processes that govern deliberation. In the context of epistemic collaboration, these norms are vital for ensuring a fair, inclusive, and rational discourse among members. To begin with, *inclusion* is a democratic norm, which seems to have important epistemic value as well. Deliberative democracy underscores the importance of including diverse perspectives in decision-making processes. In a political context, this is certainly justified by the equal moral status afforded to each citizen. Yet, diversity, as it has already been explained, has an epistemic value as well. In epistemic collaborations, inclusion ensures a comprehensive exploration of ideas, minimizing blind spots, and revealing independent sources of error. Diversity acts as a safeguard against cognitive biases, promoting rigorous evaluation and critical thinking, while it stimulates creativity and innovative thinking.

Does this imply that contributory views and claims should be assigned an equal weight? According to Longino, the equality condition should be understood in a qualified or tempered way (2001, p. 131). Division of labor combines with different expertise across individuals and research



questions. In fact, what is required by democracy is that everyone is given an equal opportunity to influence the result. However, the actual influence she will have on it is mediated by epistemic considerations, such as her epistemic competence and expertise. This qualification of the equality condition follows, after all, directly from the argumentation practice. During collective deliberation, the mutual assessment of reasons and positions is supposed to discriminate between arguments. This is what Habermas calls “the unforced force of the better argument” (Habermas, 2000, p. 37). So, in case of disagreement with the collective view, members feel nevertheless that they had the opportunity to defend their points. Their position was simply not strong enough to convince others.

Moreover, deliberative democracy emphasizes the importance of *rational discourse*, critical analysis, and the careful examination of possibilities to arrive at well-justified collective views. Argumentation is not simply meant to persuade or to win, although participants may be so motivated. In other terms, the normative role of argumentation does not rely on its ability to generate agreement, but on its truth-conducive properties (Betz, 2013). Indeed, what deliberative democrats focus on are the epistemic and sometimes social benefits of argumentation. During deliberation, participants are encouraged to formulate the reasons that support their claims, to defend their views and criticize those they disagree with. By doing so, and especially by revealing troubling implications or assumptions and blind spots in others’ positions, they reveal existing limitations and increase the information available within the group (Hafer and Landa, 2007). They also expect that it is those reasons that decide the fate of the propositions under discussion – and not, for instance, social power or coercion (Cohen, 1989). Agreement or coordination – if achieved – is thus rationally motivated. Others have argued that argumentation also has civilizing effects (Elster, 1995), which can be said to contribute to the epistemic quality of the results. In fact, although individuals may be self-centered, by being constrained to invoke reasons that are likely to persuade, they finish by adhering to these reasons. In this sense, argumentation promotes impartiality and objectivity.

These epistemic properties of argumentation are particularly salient in epistemic collaborations, where argumentation serves as a powerful tool for exploring, refining, and justifying collective results. However, one needs to consider an important objection. Argumentation, as one can easily realize by everyday practices, does not necessarily lead to these positive effects. People are not generally open-minded; rather, they tend to neglect dissenting arguments or to dismiss their sources as unreliable (Dutihl-Novaes, 2022). Under such circumstances, collective deliberation can be ex-

pected to have little effect on individuals' minds or the collective result, and its effects – if any – would be latent or slow (Mackie, 2006). This is why deliberative democracy draws attention to a central procedural requirement for collective deliberation: *mutual respect*. Also, participants need to be *open* to discussing their positions and revising their views when criticism succeeds. It is reasonable to suppose that these conditions are more easily satisfied in epistemic collaborations such as research teams, rather than in the political realm (Dutilh Novaes, 2022). Collaborators in science generally see each other as epistemic peers, though they might have different areas of expertise. Moreover, science being organized according to the principle of “organized skepticism” (Merton, 1973), researchers should welcome challenges.

While procedural norms provide the framework for deliberation, substantial principles focus on the underlying values and criteria that guide decision-making. More precisely, these are values and criteria according to which the advanced claims are defended or criticized, and decisions evaluated or justified. At least in its epistemic versions, deliberative democracy claims that procedure-independent criteria are needed so that deliberation can indeed constitute a theory of democratic justification. If the determination of rightness or truth is solely based on citizens' agreement – based on their preferences – it becomes challenging to substantiate key tenets of the deliberative ideal – such as the claim that collective deliberation improves the quality of collective decisions (Estlund 1997). One needs to appeal to some criteria, according to which the quality of decisions is to be assessed. For the same reason, epistemic deliberation needs to be constrained by epistemic criteria.

According to Longino, publicly recognized standards are indispensable for evaluating theories, hypotheses, and observational practices within a scientific community (2001, pp.130–131). They constitute the before-mentioned “common ground”, which being common knowledge among participants, structures shared reasoning and allows for mutual understanding and justifiability. In a deliberative context, it is against these standards that arguments are evaluated. Epistemic criteria, admitted by the epistemic community, determine what a reasonable challenge or a legitimate defense is in argumentation (Hakli, 2011, p. 15). They also allow us to distinguish what constitutes a valid contribution in the epistemic practice. The significance or relevance of the contribution is rooted in adherence to public standards, reflecting the community's cognitive aims and allowing for non-arbitrary evaluation in argumentation.

There is good reason to believe that not all considerations can count in epistemic deliberation. These standards distinguish the kind of reasons

one can reasonably advance in support of her claims or to challenge those of others. But what do these standards consist in? One solution is to derive those standards in an a priori way, by assuming that the goal of epistemic groups is truth and that the means by which one can reach it are univocal. In this sense, one could derive methodological criteria for science, or even the characteristics of a canonical method that could distinguish between science and pseudoscience. However, the history of science shows us that these abstract standards have little critical force in practice (e.g., Fagan, 2016). Research methods evolve according to the requirements of investigation in each epistemic context. Methods differ from one discipline to another and across experimental contexts.

There are no formal, univocal criteria for evaluating hypotheses and theories, especially in a collective epistemic practice. What groups dispose of is a set of “epistemological strategies” (Franklin, 2023), that is, a procedural background common knowledge that characterizes each epistemic group. Although social factors, such as power and career interests, can exert an important influence on scientific judgment, these considerations ultimately prove insufficient to provide justification in the long run. Justification of scientific results is ultimately based on reasons, generally accepted as epistemic by the scientists – even if this does not exclude disagreement as to the appropriate weight to assign to each of these reasons.

These strategies reflect general shared epistemological commitments. For example, in the case of physics, the standards constituting the common ground of argumentation appeal to the epistemology of experimentation – to considerations such as calibration, statistical studies of probable priors, elimination of plausible sources of error, independent confirmation and so on (Franklin, 2023). In interdisciplinary groups, these standards might be more abstract in nature, since participants do not belong to the same field, nor do they have the same educational background or methods of inference. However, there are some common epistemic values among researchers of different fields, and these are the basis for distinguishing points of agreement or even justified disagreement. These standards reflect the community’s cognitive aims and are thereby dynamic, subject to criticism and transformation based on shared values and goals. Their acceptance is an ongoing process, shaped through repeated acts of reflection and criticism, echoing the continuous evolution of knowledge (Longino, 2002).

A final point left to discuss is whether consensus is a promising normative goal for epistemic deliberation. Early theories of deliberative democracy have indeed considered consensus to be the ideal goal of deliberation (Habermas, 1993; Elster, 2013; Cohen, 1989). Rational consensus

seems to guarantee the two central claims of the deliberative ideal: autonomy (i.e., no one is constrained by decisions she has agreed to) and the common good (i.e., the common good or the truth is self-revealing through a reasoned debate). In science, consensus plays a central role as well, since it is considered as the touchstone of truth and attests of scientific authority. However, consensus has been criticized as a normative ideal. The demand for consensus is likely to create strong pressures towards unanimity for the wrong reasons – for example, social conformity – which contravenes the epistemic goals of deliberation (Sunstein, 2006). So, consensus is rather an indirect goal of deliberation, while participants directly aim for truth or better understanding (Landemore and Estlund, 2019, p. 124). What is required for epistemic justification of the collaborative views is not consensus, but joint deliberative acceptance (Beatty and Moore, 2010). In other words, what justifies the group in holding a view that  $p$  as true is that each has agreed on letting this view stand as the group's view, because of the process it has been generated by.

## 5. Conclusion

This paper has emphasized the central role of collective deliberation in epistemic collaborations. Collective deliberation appears to be crucial not only in organizing collaboration but also in justifying collective views. Relying on Bratman's account of shared intentionality, I showed how collective deliberation works in generating collective views, especially in a setting where diversity and disagreement are salient. The paper also proposed a deliberative account of group epistemic justification. Inspired by deliberative democracy, this approach asserts that the justification of collective views arises from the argumentative process by which they are generated. This approach was defended by demonstrating the epistemic properties of deliberation. The paper thus contributes to a deeper understanding of how collaborative interactions shape knowledge production in interdisciplinary settings and enriches discussions on the intricate processes involved in collaborative knowledge creation.

## References

- Andersen, H. (2016). Collaboration, Interdisciplinarity, and the Epistemology of Contemporary Science. *Studies in History and Philosophy of Science Part A* (56), 1–10.
- Andersen, H. and S. Wagenknecht (2013). Epistemic Dependence in Interdisciplinary Groups. *Synthese* 190(11), 1881–1898.

- Beatty, J. and A. Moore (2010). "Should We Aim for Consensus?". *Episteme* 7 (3), 198–214.
- Betz, G. (2013). *Debate Dynamics: How Controversy Improves Our Beliefs*. Dordrecht: Springer Netherlands.
- Bratman, M. (1992). Practical Reasoning and Acceptance in a Context. *Mind* 101(401), 1–16.
- Bratman, M. (2000). *Intention, Plans, and Practical Reason*. Stanford: CSLI.
- Bratman, M. (2014). *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press.
- Cohen, J. (1989). Deliberation and Democratic Legitimacy. In D. Matravers and J. Pike (eds.), *Debates in Contemporary Political Philosophy: An Anthology*. London and New York: Routledge, in Association with the Open University.
- Dang, H. (2019). Do collaborators in science need to agree? *Philosophy of Science* 86(5), 1029–1040.
- Dutilh Novaes, C. (2022). Argument and Argumentation. In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 ed.). Metaphysics Research Lab, Stanford University.
- Elster, J. (1995). The Strategic Uses of Argument. In K. Arrow, R. H. Mnookin, et al. (eds.), *Barriers to Conflict Resolution* (pp. 236 – 257). New York and London: W. W. Norton & Company.
- Elster, J. (2003). The Market and the Forum: Three Varieties of Political Theory. In D. Matravers and J. E. Pike (eds.), *Debates in Contemporary Political Philosophy: An Anthology*. London and New York: Routledge, in Association with the Open University.
- Estlund, D. (1997). Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority. In J. Bohman and W. Rehg (eds.), *Deliberative Democracy. Essays on Reason and Politics* (pp. 173–204). Cambridge (MA) and London: The MIT Press.
- Estlund, D. M. (2008). *Democratic Authority: A Philosophical Framework*. Princeton, New Jersey: Princeton University Press.
- Estlund, D. M. and H. Landmore (2018). The Epistemic Value of Democratic Deliberation. In J. Mansbridge, A. Bächtiger, J. Dryzek and M. Warren (eds.), *Oxford Handbook of Deliberative Democracy* (pp. 112 – 131). New York: Oxford University Press.
- Fagan, M. B. (2016). Stem Cells and Systems Models: Clashing Views of Explanation. *Synthese* 193 (3), 873 – 907.
- Feyerabend, P. K. (1970). *Against Method: Outline of an Anarchistic Theory of Knowledge*. Minneapolis: University of Minnesota Press.
- Franklin, A. and S. Perović (2023). Experiment in Physics. In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 ed.). Metaphysics Research Lab, Stanford University.
- Gilbert, M. (1989). *On Social Facts*. Princeton, New Jersey: Routledge.

- Gilbert, M. (2000). *Sociality and Responsibility: New Essays in Plural Subject Theory*. Lanham, Md: Rowman & Littlefield Publishers.
- Goldman, A. (2014). Social Process Reliabilism: Solving Justification Problems in Collective Epistemology. In J. Lackey (ed.), *Essays in Collective Epistemology*. Oxford: Oxford University Press.
- Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. William Rehg (trans.), Cambridge, MA: The MIT Press.
- Habermas, J. (2000). The Inclusion of the Other. *Studies in Political Theory*. Boston: The MIT Press.
- Habermas, J. (2010). La souveraineté populaire comme procédure. Un concept normatif de l'espace public. In C. Girard and A. Le Goff (eds.), *La démocratie délibérative. Anthologie de textes fondamentaux* (pp. 171–201). Paris: Hermann.
- Hakli, R. (2011). On Dialectical Justification of Group Beliefs. In H. B. Schmid, D. Sirtes, and M. Weber (eds.), *Collective Epistemology* (pp. 119–154). Frankfurt: Ontos.
- Kuhn, T. S. (2012). *The Structure of Scientific Revolutions*. Chicago: University of Chicago press.
- Lackey, J. (2016). What is justified group belief. *Philosophical Review* 125(3), 341–396.
- Landa, D. and C. Hafer (2007). Deliberation as Self-discovery and Institutions for Political Speech. *Journal of Theoretical Politics* 19(3), 329–360.
- Landemore, H. (2017). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, New Jersey: Princeton University Press.
- List, C. and P. Pettit (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy* 18(1), 89–110.
- Lloyd, E. A. (2005). *The Case of the Female Orgasm: Bias in the Science of Evolution*. Boston: Harvard University Press.
- Longino, H. E. (2001). *The Fate of Knowledge*. New Jersey: Princeton University Press.
- Mackie, G. (2006). Does democratic deliberation change minds? *Politics, Philosophy and Economics* 5(3), 279–303.
- Meijers, A. (2003). Why accept collective beliefs? *ProtoSociology* 18, 377–388.
- Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
- Palermos, S. O. (2022). Epistemic collaborations: Distributed cognition and virtue reliabilism. *Erkenntnis* 87(4), 1481–1500.
- Pettit, P. (2001). Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues* 11(1), 268–299.
- Pickering, A. (1986). Against correspondence: A constructivist view of experiment and the real. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1986*, 196–206.

- Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. New York: Oxford University Press.
- Rescher, N. 2003. *Epistemology: An introduction to the theory of knowledge*. Albany: State University of New York Press.
- Rolin, K. (2015). Values in science: The case of scientific collaboration. *Philosophy of Science* 82(2), 157–177.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Schmitt, F. F. (1994). The epistemic significance of disagreement. In F. F. Schmitt (ed.), *Socializing Epistemology. The Social Dimensions of Knowledge* (pp. 257–288). Lanham, MD: Rowman & Littlefield Publishers.
- Sunstein, C. (2006). *Infotopia: How Many Minds Produce Knowledge*. New York: Oxford University Press.
- Thagard, P. (2006). How to Collaborate: Procedural Knowledge in the Cooperative Development of Science. *The Southern Journal of Philosophy* 44 (S1), 177–196.
- Tuomela, R. (2007). *The Philosophy of Sociality: The Shared Point of View*. New York, US: OupUsa.
- Walton, D. (2016). *Argument Evaluation and Evidence*. Cham: Imprint: Springer.
- Zollman, K. (2010). The epistemic benefits of transient diversity. *Erkenntnis* 72(1), 17–35.





Ognjen Milivojević

## SEARLE AND THE CREATION OF SOCIAL NORMS

**Abstract:** John Searle's claim that institutional social reality, or what I call, normed social reality is established through a type of representation with the double direction of fit cannot be maintained in the light of Arto Laitinen's criticism to the effect that representations cannot have a double direction of fit. After presenting Laitinen's argument, I argue that the normed social reality is created by a collective statement of preferred normed collective action. I conceive normed collective action as a (necessarily) collective act of adding a normative aspect to some object, or an act that puts the actors in a normative relationship, stating which rights/ authorizations (etc.), i.e. obligations/duties (etc.) one agent has to another with regard to the object. In addition to not having a double direction of fit, the alternative speech act which I propose explains the fact that agents are willing to follow social norms and their expectation that social norms will be followed. At the end of the paper, I give a refined definition of normed social reality, from the resources of Searle's account, based on previous considerations and the identification of the perlocutionary force of such a speech act.

**Keywords:** John Searle, social reality, normed social reality, Arto Laitinen, speech acts, collective action, social norms.

### Introduction

In *The Making of Social World: The Structure of Human Civilization* (2010) John Searle claims that institutional social phenomena are established, reinforced, modified and abolished linguistically, i.e. through speech acts. Searle makes a distinction between institutional and non-institutional social reality: institutional assumes norms of behavior, while non-institutional social phenomena consist of spontaneous social events. According to him, the kind of speech act that creates social norms is the status function declaration. These speech acts, Searle claims, have a double direction of fit: through them, the reality is simultaneously represented

and changed. Throughout this paper I will use the term “normed social reality” to denote institutional social reality according to Searle, because it brings out the subject of consideration more closely – social phenomena constituted and governed by social norms.

Arto Laitinen argues against the notion of representation with the double direction of fit, finding it internally inconsistent. Since I agree with his argument, I am forced to look for an alternative speech act that creates social norms in general. My thesis is that the speech act that creates social norms is a collective assertion of the preferred normed collective action for each collaborator. Being a theoretical representation the speech act I propose doesn't have the double direction of fit, so it is immune to Laitinen's objection. The solution I propose supports and is supported by two conspicuous features of typical social activity: agents are motivated to comply with social norms, and they expect that others will comply with the same social norms as well. I will label them Motivation and Expectation respectively.

The paper is organized in the following manner. In Section 1 I give an account of Searle's general theory of social norms. In Section 2 a general account of representations will be given. Section 2 introduces some notions that are necessary for understanding Laitinen's criticism. Laitinen's criticism will be presented in Section 3. Finally, in Section 4, I give my argument for the claim that the speech act that creates social norms is a collective assertion of preferred normed collective action for each member of the collective. Also, in Section, 4 I will give a refined definition of normed society based (grounded) on Searle's foundation taking into account Laitinen's critic and the aforementioned two features of normed social reality that I identified and termed Motivation and Expectation.

I will not discuss the validity of Searle's theses about (1) the central importance of language in the construction of normed social reality, (2) Searle's understanding of the distinction between constitutive and regulative rules, and (3) ontological or epistemic priority of natural objects as Searle defines them. Although they can be questioned (e.g. Hindriks (2013) argues against (1) and (2), Morin (2013) questions (2), and Brandom's (2000) normative inferentialism can be seen as a challenge to (3)) I will assume their truth for the sake of argument.

## 2. Searle's general conception of social norms

One of Searle's key points in his analysis of social reality is that, originally, social norms are constitutive rules. As constitutive rules, social

norms take the form: an X counts as an Y in the context C. He claims that X and Y range over natural and social objects, and, loosely said, tools (more about these objects and their special relationship to variables later in this section). These rules, Searle claims, constitute, and even create the activity they regulate. He contrasts them with regulative rules that are introduced subsequently to regulate a certain activity:

“...the rule ‘Drive on the right-hand side of the road’ regulates driving in the United States, but driving can exist independently of this rule. Some rules, however, do not just regulate, but they also create the possibility of the very behavior they regulate. So the rules of chess, for example, do not just regulate pushing pieces around on a board, but acting in accordance with a sufficient number of the rules is a logically necessary condition for playing chess, because chess does not exist apart from the rules.” (Searle, 2010, pp. 9–10).

The conception of normed social reality is the central focus of Searle’s *The Making of Social World: The Structure of Human Civilization*. Searle presents it as part of a broader theory, the theory of social ontology, that deals with the social phenomena as such, both normed and not normed.

The basic distinction that Searle draws within his wider conception, and which is also transferred to the narrower theory, i.e. theory of normed social reality, is between natural and social phenomena. This distinction is made according to whether the phenomenon in question involves collective intentionality, which is the shared attitude of subjects toward the same object, such as ‘We believe it is raining.’ Examples of social phenomena or facts are animals hunting together, or, as Searle points out, two people pushing a car together, formal meetings, sports matches, etc. Unlike social phenomena, natural phenomena do not include collective intentionality. More precisely, natural objects, according to Searle, do not presuppose intentionality at all. They exist outside of consciousness. As Searle says: “there are many phenomena that are totally independent of the mind, such as mountains, molecules, and tectonic plates.” (Searle, 2010, p. 17)

Normed social facts are the ones that involve social norms, i.e. collectively accepted, expected and demanded patterns of behavior. The examples mentioned are non-normed social facts because, they in themselves, do not presuppose the existence of social norms. Unlike them, formal meetings, sports matches, etc. are normalized social phenomena; they imply the norms that regulate them.

The ability to posit and act according to constitutive rules presupposes the ability to conceive of a certain object as some other object; the ability to count a piece of wood as a bishop (in a game of chess) is an instance of this capacity. Searle, of course, acknowledges this by speculat-

ing that this ability of ours originates from an early ontogenic period and finds it explanatory for normed social reality:

“Small children can say to each other, ‘Okay, I’ll be Adam, you be Eve, and we’ll let this block be the apple.’ This, if one allows oneself to think about it, is a stunning intellectual feat. It was pointed out to me by Tomasello and Rakoczy and it seems reasonable to suppose that it is the ontogenetic origin of the human capacity to create institutional reality. If in fantasy we can count an X as a Y that it is not really, then with maturity it is not at all hard to see how we can count an X as a Y where the Y has a kind of existence, because it regulates and empowers our social life, even though the Y feature is not an intrinsic feature of nature.” (Searle, 2010, p. 121).

In the social norm schema, placeholder X is a natural object, causal agentive function<sup>1</sup> or status function, placeholder Y is a status function, and placeholder C is a particular environment. Both causal agentive functions and status functions, Searle says, are agentive functions – properties of objects that are “intentionality-relative” (Searle, 2010, p. 59). Namely, agentive functions are the properties of an object assigned by the subject when imagining the object as suitable for a certain purpose. Agentive functions of causal type are intentionality-relative properties that an object has solely in virtue of its intrinsic properties. Saws, cars and so on, are examples of this; a subject is required to count a collection of pieces of metal, plastic, rubber, etc., as a saw or a car, but the subject can use them for cutting or transportation based solely on the properties the objects already possess. In contrast, status functions are intentionality-relative properties that an object has solely by virtue of being accepted or collectively imagined to have it. About the characteristics of status functions, Searle says:

“The distinctive feature of human social reality, the way in which it differs from other forms of animal reality known to me, is that humans have the capacity to impose functions on objects and people where the objects and the people cannot perform the functions solely in virtue of their physical structure. The performance of the function requires that there be a collectively recognized status that the person or object has, and it is only in virtue of that status that the person or object can perform the function in question. Examples are pretty much everywhere: a piece of private property, the president of the United States, a twenty-dollar bill, and a professor in a university are all people or objects that are able to perform certain functions in virtue of the fact that they have a collectively recognized status that enables them

---

1 Searle doesn’t use the term “causal agentive function” in *The Making of Social World*, but in his earlier work in social ontology, *Construction of Social Reality* (1995). Still, the same concept is employed in the latter work, although not labeled by this term.

to perform those functions in a way they could not do without the collective recognition of the status.” (Searle, 2010, p. 7)

In order to fix the distinction between agentive functions of the causal type and status functions and explain in more depth the characteristics of the latter, I proceed with the following example, partly inspired by Searle. A natural border between two states is an agentive function of causal type (if some subject views it as a barrier), or simply a natural object (when taken in isolation from the interests of specific subjects), while the status function would be its political counterpart. In the case of a natural border, the transition between states is impossible due to the intrinsic properties of the border; in the case of a purely political border, the crossing is impossible due to the agreement of the relevant parties that it cannot be crossed.

The impossibility of crossing the border between two states, Searle would say, is not due to physical obstacles (here we assume that there are none), but to the normative force of this status function. Searle claims that status functions carry “deontology”, a special system of rights and obligations for those who agree to accept them. The status function of a political boundary that two groups of people add to a shallow, narrow watercourse entails an obligation on members of different groups not to cross the stream and the right to arrest and expel members of the opposing group if they do so.

It was said that agentive functions, i.e. both agentive functions of causal type and status functions, are added or assigned to natural objects. This is why Searle calls natural objects basic.<sup>2</sup> Also, it is said that the placeholder X is a natural object, agentive function of causal type, or a status function, according to Searle. The implications of these two theses are as follows.

First, status functions can be stacked on top of each other. Searle also argues that multiple mutually compatible status functions that are not superimposed can be assigned to natural objects, agentive functions of causal type or other status functions. An example of multiple non-stacked and stacked status functions is this: a citizen can have the status function of the president and a parent, which by definition are superimposed on the status function of a citizen, but not on top of each other: the president or

---

2 Searle also arranges natural objects into a hierarchy of basicity: “...basic facts are given by physics and chemistry, by evolutionary biology and the other natural sciences. We need to show how all the other parts of reality are dependent on, and in various ways derive from, the basic facts. For our purposes the two most fundamental sets of basic facts are the atomic theory of matter and the evolutionary theory of biology” (Searle 2010, p. 4)

a parent must be a citizen, but being the president does not necessarily mean being a parent and vice versa.

Second, although one status function may be assigned to another, making the first status function an X term and the second a Y term, in the social norm scheme, the initial X object in the chain of status functions must be a natural object or phenomenon. In our example, the status function of a citizen is assigned to a *homo sapiens*, a natural object.

Now we can propose the term *normed social reality* based on Searle's considerations: normed social reality is a network of assigned status functions connected in the two basic ways as described above, by stacking status functions on top of each other or by horizontally assigning multiple status functions to the same object, where the original nodes in the network are natural objects. Simply, a normed social reality is a network of social norms.

Searle takes that status functions are assigned through collective intentionality, more precisely through the collective linguistic action of changing reality by presenting it as changed. Searle calls these speech acts status function declarations (Searle, 2010, p. 13). For now, it is enough to state that Searle thinks that through them social norms, i.e. constitutive rules are established. In order to understand the nature of status function declarations and why they are flawed as Searle imagines them, we need to introduce a general account of representations, since status function declarations are supposedly speech acts and speech acts are representations.

## 2. General account of representations

Representations are the forms that the mind, that is, intentionality, takes in its focus on the world or parts of the world. Some artifacts, concrete objects created by the mind, are also considered representations when they serve as signs. The basic elements of representations are the mode, i.e. the way of referring, on the one hand, and the content i.e. what the mind refers to, on the other. (Examples of representations are beliefs, desires, assertions, etc. In these cases, the content is a proposition, the modes are *believing that X*, *wishing that X*, *asserting that X*, etc., where X is the content-proposition. When we talk about the mode of intention, the content is a particular act, not a proposition.)

By introducing the concept of representation, a fundamental difference is made between the mind or the representation and the external world, i.e. reality which is outside the subject's mind to which the subject's mind refers. The basic requirement imposed on the external world and its

representations is that they should or strive to fit in with each other, due to the fact that both are integral parts of Reality, which we believe is not disharmonious, but a unified whole.

The idea of the direction of fit can be traced back, at least, to G.E.M. Anscombe's *Intention* (1957), but the term itself was coined by John Austin (1953) a few years earlier. It is a property of a representation that tells what will happen or what should be done with the representation when there is no fit between it and the world. Anscombe distinguishes two directions of fit of a representation; one where the representation should change if it doesn't agree with (external) reality, and one where the world should change when there is a discrepancy between it and the representation. Her oft-quoted example of these two directions of fit is this:

“Let us consider a man going round a town with a shopping list in his hand. Now it is clear that the relation of this list to the things he actually buys is one and the same whether his wife gave him the list or it is his own list; and that there is a different relation when a list is made by a detective following him about. If he made the list itself, it was an expression of intention; if his wife gave it him, it has the role of an order. What then is the identical relation to what happens, in the order and the intention, which is not shared by the record? It is precisely this: if the list and the things that the man actually buys do not agree, and if this and this alone constitutes a mistake, then the mistake is not in the list but in the man's performance (if his wife were to say: ‘ Look, it says butter and you have bought margarine ‘, he would hardly reply: ‘ What a mistake I we must put that right ‘ and alter the word on the list to ‘ margarine ‘); whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record.” (Anscombe, 1957, p. 56)

Anscombe did not name the directions of fit of representations she observed. Over time, they were given names. The detective's record has, what is now officially called, the representation-to-the-world direction of fit. Here the representation needs or is expected to fit into the world. (If it does not match, the representation needs to change or tends to change. Types of this representation are beliefs, statements, predictions, prophecies, speedometers, etc. These representations are also called theoretical representations.) On the other hand, buyers shopping list has a world-to-representation direction of fit. The opposite is the case here; the world should or tends to fit with the representation. (If there is no alignment, the representation should be kept as is. Types are wishes, directives, commissives, etc. These representations are also called practical representations.)

Searle goes further than Anscombe in theorizing about the direction of fit by adding two more possibilities concerning this property of a rep-

representative. On his account a representation can have any of four mutually exclusive values regarding the direction of fit (Searle, 2010, p. 15, 97). In no particular order, these are:

- Representation-to-the-world direction of fit.
- World-to-representation direction of fit.
- No direction of fit. Here the representation has no direction of fit. (Examples of this are images of the imagination, expressions of gratitude, etc.)
- Double or opposite direction of fit. This is a property of status function declarations.

I will now give a concrete example of these 4 types of representative, according to Searle, using the same propositional content:

- a) Belief that “Pebbles are money” has the representative-to-the-world direction of fit. It is a description of a functional aspect of some society. It can be false or true. If it is false it means it doesn’t fit with the world and therefore should be rejected or modified.
- b) Desire that “Pebbles are money” has the world-to-representative direction of fit. If it is not satisfied, that is if it doesn’t fit the world it should remain as long as the relevant change occurs in the world.
- c) Entertaining of the thought “Pebbles are money” has no direction of fit. It’s nothing more than a passing idea or a mental exercise.
- d) The declaration that “Pebbles are money” has a double direction of fit. Through it, Searle claims, the reality is changed in the direction of pebbles being money by representing that pebbles are money.

### 3. Laitinen’s criticism of representations with the double direction of fit

Laitinen distinguishes three steps in his argument against the existence of representations with a double fitting direction:

1. The claim that only in the case of a discrepancy between the representation with the content A and the world does the direction of fit of the representation with the content A become apparent. Further, Laitinen distinguishes four possible states of the world when we speak of the relation between a representation with the



direction of fit and the world: the case where representation has content A and world is A, representation has content non-A and world is A, representation has content A and world is non-A, and representation is non-A and the world is non-A. (Laitinen, 2014, pp. 189–190)

2. The claim that the correct way to achieve the fit between a representation and the world depends on the kind of representation. (In the case of theoretical representations, fitting is done by modifying the representations. On the other hand, in the case of practical representations, the world should or tends to be modified when it does not fit.) (Laitinen, 2014, pp. 190–192)
3. The analysis of the failed representation with a double direction of fit. Since the representation has both directions of fit, both the representation and the world should or tend to change in the event of a mismatch. This, Laitinen notes, leads to a vicious circle. Namely, if the representation has content A and the world is not A, changing the representation to non-A and the world to A leaves the discrepancy. The attempt to eliminate this new discrepancy leads to the original one, and so on indefinitely. Simply, a representation with a double direction of fit is internally inconsistent. (Laitinen, 2014, pp. 192–194). As a side note, Smitz (2020) and Hindrinks (2015) are sympathetic to Laitinen's analysis.

Apart from giving very general possible directions, Laitinen's argument is a negative critique of Searle's account of speech acts that create social norms. In the next section I give my positive contribution to the problem.

#### 4. Speech act that creates social norms as a (collective) assertion of preferred normed collective action

Laitinen's critique limits what characteristics the speech act that creates social norms can have. It cannot have the property of opposite direction of fit. In order to discover other features of the speech act that creates social norms, I suggest that we start from our experience of normative life and ask how it is established by speech acts?

Two related phenomena concerning typical social activity stand out in particular: subjects are motivated to act in accordance with social norms, and subjects predict or expect that others will follow the same social rules as they do.

Therefore, the speech act that creates social norms must result, at least, in the motivation of the subjects to comply with the norms, in the common knowledge that the social norms will be respected, and cannot have the double direction of fit.

Before introducing the speech act which I believe has these implications, I will introduce what I will call the *law of collective action*. The law says that individuals associate with each other, they perform collective action, if and only if everyone makes a net gain in their well-being from association. (The individual gain I am talking about includes altruistic desires). Collective action, therefore, when actual, is necessarily something that individuals prefer to non-cooperation.

A few definitions are needed:

- Def. 1) A collective action is defined as an action that cannot be performed individually, in other words, one that requires collective intentionality.
- Def. 2) The preferred collective action is the most beneficial for all collaborators among the alternatives. We assume it always exists. If only one collective action is possible, it is trivially the one preferred.
- Def. 3) Normed collective action is a collective action according to the aforementioned social norms, counting some X as Y in the environment C. A preferred normed collective action exists.

Having explained relevant definitions and the limitations of a speech act that creates social norms I aim to consider a speech act that satisfies the definitions given. The speech act that creates social norms is a collective statement of preferred normed collective action. Its form is: we claim that each of us derives the maximum possible welfare by counting X as Y in an environment C and we believe that this is common knowledge.

Being a theoretical representation, the above representation may not fit into the world, in which case it is false and should be discarded or modified. This representation also takes into account the motivation of the relevant subjects to treat X as Y in the context C – it is believed to be most beneficial for the subjects involved. I label this feature of normed social reality Motivation. In addition, this speech act explains why the parties expect the norm(s) to be obeyed by everyone involved – there is common knowledge about the maximum benefit of relevant cooperation for all collaborators. This second feature I label Expectation.

Austin draws our attention to the fact that when speaking, three acts are performed (Austin, 1962, pp. 108–109). A “locution” is an utterance

with a fixed meaning in the context of a linguistic community; “illocution” or “illocutionary force” of a speech act is the intention of the speaker when uttering the locution; “perlocution” or “perlocutionary force” of a speech act is the effect of the speech act on the speaker. When the sentence “It’s cold in the room” is spoken, with the intention that the listener closes the window, and in a situation where the listener responds to the request, the locution is the sentence “It’s cold in the room” which is an assertion with its grammatical meaning, a request and an implicit sentence “Close the window” is an illocution, and the fulfillment of the request by the listener, i.e. the implicit sentence “I will close/I close the window” is a perlocution.

The perlocutionary force of the aforementioned speech act that creates social norms is the collective intention to count X as Y under the assumption that the desire-belief model of intention is correct, i.e. the account of intention according to which available actions are chosen based on their expected utility, the strength of the desire multiplied by the subjective assessment of the probability of fulfilling it with the available means. An example of the account is intending to go to the grocery store, which is available means, to enjoy an ice cream, which is desired, since the expected utility of going to the grocery store to get an ice cream is greatest among the alternatives, for example staying at home to read a book. The collective intention to count X as Y does not necessarily have to be verbalized during the establishment of institutions, although such a collective speech act certainly exists implicitly.

From everything said so far, including the truth of Searle’s claim about the ontological or epistemic priority of natural objects, it follows that a normed social reality is a structure made up of the intentions that certain objects, primarily natural ones, carry specific status functions in order for each interested party to derive the maximum possible benefit. The system of intentions is based on or caused by a shared and commonly known belief system about which status functions to assign to certain, primarily natural objects, in order to achieve a maximally beneficial outcome for each member.<sup>3</sup>

---

3 My attention has been drawn to the problem of reconciling the explanation I offer here with the Prisoner’s Dilemma scenario. In the Prisoner’s Dilemma, there is a shared understanding among criminals of the common benefit of maintaining a social norm of non-cooperation with the police. However, when they pursue their individual interest, criminals override that social norm. The problem can be solved by distinguishing between the social norm of non-cooperation with the police before arrest and the absence of that norm, i.e. establishing a social norm of cooperation with the police by criminals when they are arrested. In both cases there are certain social norms that actors establish considering their individual interests. Also, iterated

## Conclusion

My main aim in the paper was to fill in the gap in, what I call, Searle's theory of normed social reality left by Laitinen's refutation of status function declarations which Searle considers as speech acts that create social norms. I showed that this gap should be filled by claiming that the speech act that creates social norms is a collective statement of preferred normed social action, based on three facts: the proposed speech act, being a statement, has one direction of fit, it perfectly explains the motivation of agents to follow social norms (Motivation) and why agents who follow social norms expect others to do the same (Expectation). My secondary objective was to give a definition of a normed social reality, on Searle's grounds, informed by the previous considerations and by positing that the perlocutionary force of a such speech act is a collective intention to count X as Y (in context C). My ultimate conclusion is that a normed social reality is a system of intentions to hold certain Xs, originally natural objects, as Ys in contexts C, a system based on a shared belief system about which status functions to assign to certain, originally natural objects, in order to achieve a maximally beneficial outcome for each agent.

## References

- Anscombe, G.E.M. (1957). *Intention*. Cambridge, Mass.: Harvard University Press.
- Austin, J. L. (1953). How to Talk-Some Simple Ways. *Proceedings of the Aristotelian Society*, 53, 227–46.
- Austin, J. L. (1962). *How To Do Things With Words*, Oxford: The Clarendon Press.
- Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.
- Collins, Rory W. (2022). The Prisoner's Dilemma Paradox: Rationality, Morality, and Reciprocity. *Think*, 21(61), 45–55.
- Hindriks, F. (2013). Restructuring Searle's Making the Social World. *Philosophy of the Social Sciences*, 43(3), 373–389.
- Hindriks, F. (2015). Deconstructing Searle's Making the Social World. *Philosophy of the Social Sciences*, 45(3), 363–369.
- Laitinen, A. (2014). Against representations with two directions of fit. *Phenomenology and the Cognitive Sciences*, 13(1), 179–199.

---

prisoners' dilemma can make cooperation among criminals likely, as the payoffs of available actions change. The iterated prisoner's dilemma reflects the reality of certain criminal circles, hardened criminals who are unwilling to tarnish their reputation by cooperating with the police or who have a reasonable suspicion that cooperating with the police will result in drastic sanctions from the rest of the underworld.

- Morin, O. (2013). Three Ways of Misunderstanding the Power of Rules. In M. Schmitz, B. Kobow and H. B. Schmid (eds.), *The Background of Social Reality* (pp. 185–199). Dordrecht: Springer.
- Schmitz, M. (2020). Of layers and lawyers. In M. Garcia, R. Mellin and R. Tuomela (eds.), *Social Ontology, Normativity and Philosophy of Law* (pp. 221–240). Berlin: De Gruyter.
- Searle, J. (1995). *Construction of Social Reality*. New York: The Free Press.
- Searle, J. (2010). *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.
- Sinhababu, N. (2012). The Desire-Belief Account of Intention Explains Everything. *Noûs* 47 (4), 680–696.



## 4. NORMATIVITY AND LOGIC





Aleksandra Vučković

## NORMATIVITY AND TRUTH IN NATURALIZED EPISTEMOLOGY

**Abstract:** In his famous article “Epistemology Naturalized” (1969), Quine established a novel take on the position of epistemology in philosophical and scientific discourse. According to his views, epistemological questions are a subset of psychological questions, and psychology in itself is a branch of natural science. Thus, epistemology, as understood in the Quinean sense, threatens the very idea of its normative aspects, as natural science is empiristic and, as a result, relies on purely descriptive claims. Hence, the following question arises: Does the naturalized account of epistemology entail the rejection of epistemic norms? In this research, we explore the three possible answers to this question and argue there is a sense of normativity in Quine’s naturalized epistemology, but only insofar as we are willing to accept his imperfect notion of the truth.

**Keywords:** naturalized epistemology, normativity, truth, Quine.

### 1. Traditional vs. Naturalized Epistemology

Ever since Ancient Greece, up to Descartes’ distinction between *res cogitans* and *res extensa*, or – perhaps most importantly Kantian distinctions between analytic and synthetic propositions and *apriori* and *a posteriori* knowledge – philosophers have approached the concept of human knowledge in a rather peculiar way. While some were strong proponents of a scientific approach and others had more of a “soft spot” for metaphysics, one thing remained the same. Throughout its various incarnations across multiple centuries and philosophical schools, traditional epistemology has maintained its central position as a discipline independent of any other scientific enterprise. As a result, it always relied on metaphysical grounds.

Even Carnap, famous for his strong physicalist inclinations, had fallen into the trap of a never-ending metaphysical loop. While trying to separate philosophical questions from the rest of the inquiries concerned with human knowledge of scientific facts (Carnap, 1950, pp. 22–23), he unintentionally “admitted” that philosophy belonged to a domain other than physics and the rest of natural science. Quine did not appreciate Carnap’s reductionism, nor did he believe that we have sufficient epistemic reasons to justify the separate class of philosophical questions (Quine, [1960] 2013, p. 217); thus, he proclaimed the entirety of Carnap’s project redundant and based on metaphysical grounds (Quine, 1968, p. 203).

Apart from abandoning Carnap’s reductionism, Quine has, in his most cited paper to this day – “Two Dogmas of Empiricism”, rejected the established distinction between analytic and synthetic propositions by claiming that such a duality could not be justified through reliance on any other more fundamental conceptual category (Quine, [1951] 2000). First published in 1951, “Two Dogmas” not only challenged the distinction that was considered a holy grail of Western philosophy but paved the way for the wholly novel outlook on epistemological questions and their position in the holistic web of knowledge<sup>1</sup>. Nearly two decades later, in 1969, Quine published the essay titled “Epistemology Naturalized” which took an even more radical approach to the old philosophical question of knowledge.

According to this new approach, as Quine argues, epistemological questions ought to be understood as part of psychology, which, furthermore, should be understood as a branch of natural science (Quine, 1969, p. 83). Such a classification of epistemology as a mere chapter of a chapter in a major “book” of knowledge has roots in Quine’s life-long endorsement of naturalism – a philosophical stance according to which, if there is such a thing as the truth, then we should look for it within (natural) science.

Although naturalism in itself has various forms – from a comprehensive ontological theory that postulates only those entities that are recognized in natural science, to a mere methodological stance that proclaims we should follow the methods of natural science (Forrest, 2023; Spiegel, 2023) and although the debate between proponents of the two types of naturalistic doctrine is still ongoing – this research hypothesizes that Quine endorses its former (and stronger) kind. Quine’s acceptance of ontological naturalism<sup>2</sup>

---

1 The thesis on *holism*, according to which *all* our beliefs and knowledge propositions are consistently interconnected, was also first introduced in “Two Dogmas of Empiricism”.

2 In one of my previous works, I have provided a more elaborate explanation of *methodological* reasons behind Quine’s loyalty to naturalism (Vučković, 2016).

arises from his firm rejection of the first philosophy and, as such, serves as an ontological basis for his defense of physicalism. This brings us to the second hypothesis of this research according to which, Quine's ontology is, in its essence, physicalistic and minimalistic in the sense that he does not posit more entities than are necessary for theoretical consistency.

Let us now return to the distinction between traditional and naturalized epistemology. We can notice that these two approaches differ not only in their ontological background (that is, philosophy as a unique kind of knowledge, as opposed to philosophy based on natural science) but also in their understanding of normativity. While traditional forms of epistemology face no trouble establishing epistemic norms as they are generally accepted as *apriori* knowledge that requires no further justification, the specification of these norms seems more of a challenge for Quine's naturalized epistemology. The reason for this difficulty is, of course, Quine's rejection of the concept of *apriori* knowledge, which means that his epistemology either has to deny normativity or base its existence somewhere other than *apriorism*.

Quine was reluctant to accept the absence of normativity, possibly out of fear that it would leave his epistemology hollow. The lack of epistemic normativity could, after all, lead to the lack of reliable epistemic methodology, which would result in his epistemology turning into quite a messy study field. In "Reply to Morton White", he writes:

"[N]ormative epistemology is a branch of engineering. It is the technology of truth-seeking [...] There is no question here of ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction."

(Quine, 1986, pp. 664–665)

These excerpts provide two valuable insights into his understanding of the relation between naturalized epistemology and the concept of normativity. First, there is a branch of philosophy he refers to as *normative epistemology*, and, second, it is concerned with leading us to the truth. So, the question is: *Do we agree that there is normativity in naturalized epistemology? And if we do, where do we seek the source of this normativity?*

## 2. The Zero or Low Normativity Cluster

The *zero or low normativity cluster*, as we shall call it in this research, includes the solutions that claim that Quine's naturalized epistemology either:

- a) entails no normativity, and, thus, it is questionable whether we should classify it as epistemology at all (zero normativity), or
- b) uses the term “normativity” very loosely, so it does not cover all the usual meanings and uses of the notion (low normativity).

What both of these types of criticism have in common is that they claim that Quine’s understanding of normativity is vastly different from the traditional understanding of the notion. As a result, his epistemic norms –or so they claim – fail to serve their primary purpose and, as such, jeopardize the entire project of naturalized epistemology.

One of the most famous proponents of the *zero* normativity solution is Jaegwon Kim, who claims that naturalized epistemology does not deserve its recognition as an epistemological theory, as it excludes the very concept of knowledge, which is the central focus of *any* epistemology. Following the traditional definition of knowledge as “justified true belief”, Kim argues that the very essence of the concept of “justification” lies in its normativity. Or more precisely, the standard request for justification is essentially nothing short of normative (Kim, 1988, pp. 382–383). The idea behind Kim’s reasoning – the one which will become even more prominent in Kornblith’s criticism of Quine – is that naturalized epistemology cannot offer normativity because it is as descriptive as any other branch of natural science.

Hilary Kornblith addresses the very same concern regarding the naturalized epistemology’s claim on normativity – if epistemology is a part of psychology and psychology is nothing more than a chapter of natural science, epistemology is, then, descriptive by its definition. However, analyzing the aforementioned Quine’s response to Morton White, Kornblith recognizes that the abandonment of normativity was never the intention behind the naturalization of epistemology. Nevertheless, we should not take Quine’s alleged acceptance of normativity at face value, Kornblith claims, as he has to justify the source of that normativity. Furthermore, Quine needs to explain why we should believe that the pursuit of truth is the only purpose of epistemology (Kornblith, 1993, p. 358).

Kornblith’s resolution falls into the *low* normativity cluster insofar as he accepts that the quest for truth as an ultimate goal of naturalized epistemology does warrant this doctrine at least some type of claim on normativity. He writes:

“Since many people do clearly care about the truth of their beliefs, Quinean epistemic norms, construed as imperatives contingent upon valuing truth, will carry normative force for a great many people.”

(Kornblith, 1993, p. 365)

The implication behind the use of the words “many” and “a great many” is more than intentional in this case; Quine’s naturalized epistemology fails what may be described as a *universality test*. While the naturalized approach to the knowledge issue may fit the expectations that some or even *many* people have of epistemology, there is nothing inherently non-disputable about this project. If we circle back to Kim’s criticism, then it would seem that naturalized epistemology has nothing to offer to those who are more interested in the *justified* than the *true* part of “justified true belief.” Thus, Quine has offered little support for the claim that naturalized epistemology is superior or at least more convincing than the traditional.

However, these solutions tend to overlook the full meaning behind Quine’s ontological naturalism. The ontological claim in his naturalism is that there is *no* truth besides the one found in science. The reason for that is the absence of the first philosophy or any other metaphysical doctrine. Metaphysical doctrines, in their essence, have very little to do with knowledge and everything to do with our belief in them. Beliefs also fail the universality test since they vary among people and across cultures. As a result, naturalism, in the Quinean sense, is not only the source of the truth but also the source of justification. That is, anything that requires justification can only be justified through the scientific method.

### 3. The Medium Normativity Claim

The *medium* normativity claim entails that there is normativity in naturalized epistemology, although not in such a strong sense as is the case in traditional epistemology. According to this cluster of solutions, there *are* norms in naturalized epistemology, but instead of being unique to Quinean theory, they are borrowed from natural science. Epistemic norms are, thus, the same as scientific norms, and natural science is the source of them all.

One of the proponents of the medium normativity claim is Paul Roth, who argues that Kornblith’s solution to Quine’s issue with normativity is not satisfactory, as he fails to provide the answer to Quine’s original question of what is so special about epistemology that makes it deserving of different treatment than any other scientific theory. As Roth argues, Kornblith failed to prove there is a “special method” to epistemology, and thus, he cannot explain why it should not be naturalized like the rest of the science (Roth, 1999, p. 90).

Inspired by Maffie's (1990) and Kitcher's (1992) take on naturalism, Roth focuses on the methodological aspect of this doctrine and claims there is a methodological continuum between epistemology and natural science (Roth, 1999, p. 91). This continuity in methods warrants the same kind of normativity claims in epistemology and scientific theories, as they are all rooted in the same approach. Methodological naturalism is a weaker claim than its ontological version that Quine accepts, and Roth acknowledges, as he speaks of Quine as a "radical naturalist" (Roth, 1999, p. 94). If one adheres to the radical version of naturalism, then there is no harm in saying he also follows its less extreme consequences; in this case, it is naturalistic methodology.

However, Roth interprets Quine's project of naturalized epistemology as less revolutionary than the other philosophers make it out to be. He argues that the essay "Epistemology Naturalized" is not to provide a unique grounding of epistemic norms but only to remind us that we need to work from within. If there is justification for epistemic normativity, then we should seek it intra-theoretically, that is, within the domain of empirical knowledge (Roth, 1999, pp. 95–96). Since empirical knowledge is a matter of natural science, we should seek normativity too in natural science.

Another proponent of the "middle ground" solution is WyboHoukes, although he could also fall into the *zero normativity cluster*, depending on which of his two interpretations we choose to follow. Nevertheless, both of his interpretations are focused on Quine's comparison of epistemology to "engineering" and "technology of truth-seeking", as we have seen in the excerpts from his "Reply to Morton White".

Houkes argues that such comparisons are either to be taken literally or metaphorically. If we interpret the word *engineering* in its literal sense, then we also need to acknowledge that the main purpose of engineering is to design artifacts. But what *is* the artifact of epistemology? Artifacts are, after all, physical objects that we create, and we do so with certain intentions in mind (Houkes, 2002, p. 261). We build houses and bridges with a clear purpose on how we will use them, so this intentionality is the source of normativity when it comes to artifacts.

A literal interpretation of the word *engineering* will not suffice in explaining the roots of normativity in naturalized epistemology for two reasons. First, it is hard to see what knowledge, truth, or whatever else may be the topic of naturalized epistemology has in common with physical objects. Second, if we are to understand knowledge as an artifact, it would require intentionality, and Quine's view on that matter is pretty straightforward – natural science is devoid of any intentionality whatsoever (Houkes, 2002, pp. 262–263).

It means that if we are to interpret the words of *engineering* and *truth-seeking technology* in any favorable way, then, Houkes argues, we need to understand them as metaphors. These metaphors serve a purpose to signal the continuity between natural science and epistemology and remind us that epistemic and scientific norms are the same. Truth-seeking technology is a metaphor for scientific methodology, and the source of normativity in naturalized epistemology is reliance on these methods (Houkes, 2002, p. 259).

Therefore, one of the solutions to Quine's problem of normativity is to accept natural science as the primary and *only* source of epistemic norms. It is certainly a step forward from zero and low normativity claims that do not seem to amount to what Quine was trying to explain in his discussion with Morton White. However, one can still wonder if there is more to his idea of epistemic normativity. Perhaps the question we need to ask is this: Can we agree with the Quinean notion of continuity between natural science and epistemology and *still* believe some *purely epistemological questions* exist and require its unique type of approach and source of normativity?

#### 4. The High Normativity Claim

The high normativity claim – as we shall call it in this research – entails not only that there *are* norms in naturalized epistemology but also that they are, at least in some aspect, different from the rest of the scientific norms. For this solution to work, we need to prove that the continuity between natural science and epistemology does not obligate us to the claim that epistemic norms are identical to scientific ones. The best course of action would be to try to show that there are at least some purely epistemological questions that can be answered without relying on scientific methodology.

Such a solution would work in this case, as we could claim that in the *ontological* sense, natural epistemology is still grounded in science insofar as its ultimate purpose is to discover the truth, and the only reliable source of truth is science. Nevertheless, this ontological commitment does not entail that we should follow the methodology of natural science, as we could argue that the answers to epistemological questions require their own methods. What we will find at the end of the road will be the truth in the naturalist sense, but we could argue that it is not the same road for a scientist and a philosopher.

One of the interpretations of normativity in naturalized epistemology that aligns with our idea of a high normativity claim is offered by Richard Foley in his essay “Quine and Naturalized Philosophy” from 1994. Foley argues that Quine, in his work “The Web of Belief” which he had co-written with Joseph Ullian (1970), not only does not claim that epistemology is the same as psychology but quite clearly makes the distinction between the two. While psychologists focus on thought processes, epistemologists are concerned with evidence, which shows that these two are entirely *different* disciplines. Foley also believes that the claims in “Epistemology Naturalized” are exaggerated on purpose to underline the importance of understanding that epistemic knowledge is inseparable from knowledge of science (Foley, 1994, pp. 248–249). However, it does not mean that we need to follow the same methodology in psychology and epistemology, nor even that they explore the same part of the experience. Psychology is, thus, natural science in the full sense of it, while epistemology is concerned with evidence that supports claims from natural science.

Foley also argues that his view is supported in Quine’s later works, such as *Pursuit of Truth* (1990), where Quine claims that we can tackle the knowledge questions even if we are not familiar with the physicalist ontology of nerve endings (Foley, 1994, p. 250). Finally, Foley offers a new *holistic* explanation of epistemic norms, according to which they cannot be known *a priori* – as Quine himself claimed – nor are they the same as scientific ones – as we have seen that the topics of psychology and epistemology are rather different – but are an essential part of the holistic web of knowledge and, as such, can be revised when necessary (Foley, 1994, pp. 256–257).

What are we to make from this solution? In Section 1, we claimed that Quine’s ontology was physicalist, as there was no first philosophy that could serve as the basis for ontological commitment. We also claimed that Quine’s ontology is minimalistic in the sense that it does not presuppose more entities than necessary. It means that if physicalist ontology offers just enough entities to provide sufficiently good predictions about the world around us, then we should refrain from positing any other type of entities. Quine’s allegiance to physicalism is inseparable from his claim on naturalism, and we do not need to know physicalist ontology to be naturalists and interested in epistemological questions.

In an *ontological* sense, epistemology is based on natural science no less than any other field of research. However, in the *methodological* sense, epistemologists are free to establish any kind of norms, as long as they contribute to the pursuit of truth, which is the final goal of this project. The purely epistemological questions exist, and they are concerned with



the evidence for natural science. There is no *apriorism* in naturalized epistemology, though, insofar as it is ontologically grounded in natural science. Nevertheless, epistemic norms are not equal to scientific norms, as they are directly determined by the questions that epistemology explores.

## 5. The Feasibility of Truth Pursuit

The high normativity claim is – as we have seen – inseparable from the idea that the pursuit of truth is the main quest of the philosophical and scientific enterprise. However, this idea is, perhaps, of even greater importance to Quine’s naturalism in general – whether he is talking about normativity or any other philosophical project; to the point that even his last big piece of writing – first published in 1990 – was appropriately titled *Pursuit of Truth*. Therefore, it comes as no surprise that the notion of normativity, in the context of Quine’s naturalized epistemology, cannot be evaluated independently of his take on truth. But what if his concept of truth is not feasible within the scope of naturalized epistemology? Perhaps, since the beginning of this debate, we have been asking the wrong question of whether naturalized epistemology can withstand the concept of normativity. What if we should turn to the notion of truth instead and assess whether it can retain its usual meaning in this new naturalized environment?

Putnam’s 1982 essay “Why Reason Can’t Be Naturalized” seems like a good starting point for this assessment<sup>3</sup> for several reasons. First, Putnam – not unlike Quinean high normativity claim – believes that normativity is an essential aspect of human thought. To eliminate normativity, according to him, would be to attempt “mental suicide” (Putnam, 1982, p. 20). Second – and perhaps this is Putnam’s most obvious similarity to Quine – the concept of *normativity* is inseparable from the notion of *truth* (Putnam, 1982, p. 21). Finally, Putnam was the one who challenged the feasibility of Quine’s take on truth within his naturalism. Nevertheless, we shall argue that Quinean and Putnam’s understanding of the truth – at least when it comes to the practical use of this concept – have more in common than it seems at first glance.

As the title of Putnam’s essay suggests, the central notion behind his analysis is the *reason*, which he uses as a starting point for the evaluation of many different philosophical standpoints – from the very modest project of evolutionary epistemology (Putnam, 1982, p. 4) to more complex theories like Quinean positivism (Putnam, 1982, p. 15) and finally natu-

---

3 I would like to thank the anonymous reviewer for this recommendation.

ralized epistemology (Putnam, 1982, p. 19). And it is precisely our ability or virtue of reason which makes the concept of human thinking inseparable from the notion of truth and the pursuit for it (Putnam, 1982, p. 21). The truth is, furthermore, essential to any discussion regarding normativity, hence, normativity claim of any level is inseparable from the tendency to claim truthfulness. So, what is the issue with the Quinean concept of the truth that it does not align with Putnam's thoughts on the matter?

According to Putnam, the main problem behind Quine's loyalty to the pursuit of truth lies in the disbalance between how he interprets the notion of truth as opposed to the other concepts of a similar ontological background. Quine famously rejects concepts of justification, rational acceptability, warranted assertability, etc., on the grounds of them being based on metaphysical realism, and yet he proceeds to use the notion of the truth in the very same metaphysically realist sense (Putnam, 1982, p. 20). The only concept of truth Quine can coherently adhere to within the context of his naturalism needs to be defined in Tarskian sense, that is to be understood as a syncategorematic concept that allows for "semantic ascent" (Putnam, 1982, pp. 19–20). In more colloquial terms, the truly naturalist notion of truth would entail nothing more than a semantic device that allows switching from one linguistic level to another.

Putnam, therefore, believes that the Quinean notion of the truth goes beyond what Tarski had in mind for that concept, and seems rightfully so. While Quine's epistemology and even ontology (or the lack of it) focus on the truth, he is skeptical regarding the idea of a language fully grounded on syncategorematic terms. Such a language – according to Quine's earlier paper "A Logistical Approach to the Ontological Problem" – would lack reference (Quine, 1966, pp. 66–67). He is even more wary when it comes to the conception of "semantic ascent" – a strategy he famously criticized Carnap and logical positivism for to the degree that he dedicated the entire chapter of *Word and Object* to disparage the idea that we can semantically ascent from scientific to philosophical questions (Quine, [1960] 2013, pp. 249–254). But if Quine's notion of truth should not be understood as a semantic and syncategorematic device *and* if Putnam is right to claim that its metaphysical and realist version does not suit a naturalist environment, where does it leave us on the Quinean interpretation of this concept?

One route would entail not taking Quine's words at face value but in a more metaphorical sense – starting from his rejection of the traditional epistemological tools to even his use of the word *truth*. The latter, in particular, fits his later writings on anomalous monism, a type of token physicalism he borrows from Donald Davidson and which allows him to mix physicalist ontology with non-physicalist vocabulary when it is neces-

sary (Quine, 1992, pp. 71–73). This tokenism hints at his original reasoning behind physicalist ontology that proliferates in the absence of the first philosophy while simultaneously permitting the use of the realist notion of the truth. Furthermore, even if we accept that the concept of the truth is an exception to the rule of non-metaphysical vocabulary, it should be noted that it is not *the only* exception Quine makes. While he famously excluded most of the universals from the domain of existing things, he still needed to allow the existence of the sets to preserve a meaningful talk of the properties.

Naturalist epistemologists can, however, take a simpler route, as proposed by Putnam's liberal and anti-reductionist take. According to him, normativity – despite being an essential aspect of human thought – cannot be grounded in physics, biology, or any other particular field of scientific research. The quest for epistemic normativity can be understood as a general long-term goal that, due to the various challenges, we cannot achieve for the time being (Putnam, 1982, p. 21). This limitation, however, does not mean that we need to subscribe to a reductionist approach nor that we are unable to tackle the philosophical questions until the whole system is complete. It only means that, for now, we have to work with what we have and make the most of it.

Nevertheless, if we adhere to the aforementioned (and more relaxed) interpretation of Quine's physicalism, we can see that it does not presuppose more than Putnam's view, at least when it comes to its practical implementation. In that case, a naturalist epistemologist does not need to believe in the ultimate truth of any particular scientific theory as long as she understands that the truth, *in general*, needs to be concerned with what we can know based *only* on evidence. This interpretation allows us to understand Quine in a non-reductionist sense – physicalism is a token theory that serves as a source of an ontological commitment in the absence of the first philosophy.

## 6. Conclusion

The question of normativity in naturalized epistemology is inseparable from one's takes on ontology, methodology, and the general understanding of the role of naturalism. While the high normativity claim does seem to fit within the context of Quine's entire ontology, it may not be as efficient without the background idea of the truth as the main quest in science. And even Quine's notion of the truth is imperfect in its own sense. We can either understand it as a metaphysical and realist addition

to otherwise non-metaphysical ontology, or we can subscribe to a weaker type of (token) physicalism. Finally, we can take Putnam's route and maintain that epistemic normativity is a goal we should strive to achieve and keep the high normativity claim.

Most naturalists – that is, anyone who does not have as developed an ontological system as Quine nor as liberal take on the idea of naturalism as Putnam – would probably fit in the medium normativity cluster, as they would not be willing to give up on normativity in its entirety but would possibly still refrain from ontological statements. Zero and low normativity solutions do not seem to fare as well, given the criticism they have received over time from the proponents of both medium and high normativity claims. However, just like many other interpretations of Quine's philosophy, they serve their purpose as a thought experiment or, more precisely, a reminder of why the normativity question needs to be taken seriously.

## References

- Carnap, R. (1950). Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie*, Vol. 4, No. 11 (Janvier 1950), 20–40.
- Foley, R. (1994). Quine and naturalized epistemology. *Midwest studies in philosophy*, 19(1), 243–260.
- Forrest, P. (2023). Methodological Naturalism Undercuts Ontological Naturalism. *American Philosophical Quarterly*, 60(1), 99–110.
- Houkes, W. (2002). Normativity in Quine's Naturalism: The technology of truth-seeking? *Journal for General Philosophy of Science*, 33, 251–267.
- Kim, J. (1988). What is "Naturalized Epistemology?". *Philosophical perspectives*, 2, 381–405.
- Kitcher, P. (1992). The Naturalists Return. *The Philosophical Review*, 101(1), 53–114.
- Kornblith, H. (1993). Epistemic Normativity. *Synthese*, 94, 357–376.
- Maffie, J. (1990). Recent Work on Naturalized Epistemology. *American Philosophical Quarterly*, 27(4), 281–293.
- Putnam, H. (1982). Why Reason Can't be Naturalized. *Synthese*, 3–23.
- Quine, W.V.O. (1966). A Logistical Approach to the Ontological Problem, reprinted in *The Ways of Paradox and Other Essays*. 64–69.
- Quine, W.V.O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays* (pp. 69–90). New York Chichester, West Sussex: Columbia University Press.
- Quine, W.V.O. (1968). Ontological Relativity. *Journal of Philosophy*, 65(7), 185–212.
- Quine, W.V.O., (1990). *Pursuit of Truth*. Harvard University Press.

- Quine, W.V.O. (1986). Reply to Morton White. In L. E. Hahn and P. A. Schlipp (eds.), *The Philosophy of W. V. Quine* (pp. 663–665). Chicago and La Salle, Illinois: Open Court Publishing.
- Quine, W.V.O. and Ullian, J.S. (1970). *The Web of Belief*. New York: Random House.
- Quine, W.V.O. ([1951] 2000). Two Dogmas of Empiricism. In *Perspectives in the Philosophy of Language: A Concise Anthology* (pp.189–210). Peterborough (Canada): Broadview.
- Quine, W.V.O. ([1960] 2013). *Word and Object*. Cambridge MA: MIT Press.
- Roth, P. (1999). The Epistemology of “Epistemology Naturalized”. *Dialectica*, 53(2), 87–110.
- Spiegel, T.J. (2023). Why Naturalism cannot (Merely) be an Attitude. *Topoi*, 42(3), 745–752.
- Vučković, A. (2016). Quine’s Dilemma Between Ecumenical and Sectarian Response. *Theoria*, Beograd, 59(3), 64–80.



Miroslava Trajkovski

## NORMATIVITY, VALIDITY AND SEMIOTIC IMPLICATION

**Abstract:** Validity of an inference depends on the implication involved, as Hartry Field (Field, 2009, p. 342, p. 349) stresses “our views about implication constrain our views about *how we ought to reason*, or (perhaps better) about *the proper interrelations among our beliefs*.” Hence, the relation of implication “has a broadly normative component.”

A non-standard implication is introduced, I call it *semiotic implication*. It will be argued that semiotic implication is an important interpretative tool and some formal characteristics will be discussed. Since the latter differentiate it from standard implication the sign “ $\div$ ” for it is introduced. I define semiotic implication ( $x$ ) ( $P(x)\div Q(x)$ ) as: P is an index of Q, for any P and Q, and Q is an icon of P, for Q different from P.

Semiotic implication can be linked to semiotic validity (cf. Trajkovski, 2024) which differs from deductive validity. In this context, it can be examined whether and in what way semiotic implication calls into question the normativity of classical logic. Especially, the enthymematic potential of semiotic implication supports abductive validity.

Hence, we can distinguish three deviations from the norm in connection with semiotic implication: deductive validity is defined through an index; the enthymeme might be taken as a category of natural reasoning implying abductive validity defined through an icon. However, it will be argued that the deviations are not necessary, with certain limitations, the notion of semiotic implication can be adapted to the norm.

**Keywords:** normativity, enthymeme, validity, semiotic implication.

### Introduction

Semiotic implication was introduced in “The Origin of Semiotic Validity – Peirce and Aristotle on Reasoning by Signs” (Trajkovski, forthcoming 2024) where I discussed the notion of semiotic validity. There I

argued that a universal premise “All M is P”, i.e. “For all x, if x is M, then x is P” in deduction, is to be read: M is an index for P, while, in abduction, it is to be read: P is an icon for M. Hence, in general the conditional states a semiotic relationship between M and P which says that M is an index of P, and P is an icon of M.

Semiotic validity has to do with the semiotic connection between the premises and the conclusion entailed; this is the theme of Section 1. In Section 2, I point out the flaw in Peirce’s logical argument for the claim that the set of premises of induction as an index signifies the conclusion. Section 3 analyzes logical reasons that might support Peirce’s later but metaphysically grounded view that the set of premises of induction as a symbol signifies the conclusion, and points out a mistake in this argument as well. Section 4 gives the solution, and semiotic implication is introduced. Section 5 discusses normative and formal aspects of semiotic implication. Finally, the conclusion summarises the results of previous sections and stresses why semiotic implication is of substantial importance.

## 1. The premises signify their conclusions

The idea that the premises signify their conclusion is natural, for premises point to the conclusion, they refer to it. In a Brentanian framework, it could be said that the premises intend the conclusion, in Husserl’s that they constitute it, in Peirce’s the premises represent or signify the conclusion – premises are the sign of the conclusion.

More generally, the conclusion is the object to which the premises refer. Observing the conclusion as an object represented by the premises, it is clear that different types of argument will differ according to the way in which their premises denote their object: the conclusion.

With regard to the relationship between the sign and the object, Peirce distinguishes three ways in which the sign indicates the object: through similarity with the object, or by being in a real relationship with the object, or by convention. They are, respectively: icon, index and symbol. Since Peirce distinguishes three basic types of reasoning – deduction, abduction and induction– his question is how their premises are related to their respective conclusions. When he asks this question, Peirce does not look for an answer in anything like modern predicate logic. First, in “On a New List of Categories” (1868), he tried to find the solution in the syllogistic framework, later; in *Minute Logic* (1902) he sought it in phaneroscopy – the study of three categories of reality: firstness, secondness and thirdness. Briefly, these two approaches of Peirce can best be presented through quotations.



## 1) The 1868 view:

“In an argument, the premisses form a representation of the conclusion, because they indicate the interpretant of the argument, or representation representing it to represent its object. The premisses may afford a likeness, index, or symbol of the conclusion. In deductive argument, the conclusion is represented by the premisses as by a general sign under which it is contained. In hypotheses, something like the conclusion is proved, that is, the premisses form a likeness of the conclusion. Take, for example, the following argument:

M is, for instance, P', P'', P''', and P'''';  
 S is P', P'', P''', and P''''':  
 [Ergo,] S is M.

Here the first premiss amounts to this, that “P', P'', P''', and P''''” is a likeness of M, and thus the premisses are or represent a likeness of the conclusion. That it is different with induction another example will show.

S', S'', S''', and S'''' are taken as samples of the collection M;  
 S', S'', S''', and S'''' are P:  
 [Ergo,] All M is P.

Hence the first premiss amounts to saying that “S', S'', S''', and S''''” is an index of M. Hence the premisses are an index of the conclusion.” (Peirce, 1868, pp. 296–297)

## 2) The 1902 view:

“Argument is of three Kinds: Deduction, Induction and Abduction (usually called adopting a hypothesis.)

An Obsistent Argument, or Deduction, is an argument representing facts in the Premiss such that when we come to represent them in a diagram we find ourselves compelled to represent the fact stated in the Conclusion; so that the Conclusion is drawn to recognize that quite independently of whether it be recognized or not the facts stated in the premisses are such as could not be if the fact stated in the conclusion were not there; that is to say the Conclusion is drawn in acknowledgement that the facts stated in the Premiss constitute an Index of the fact which it thus compelled to acknowledge. Deduction is Obsistent in respect to being the only kind of argument which is compulsive...

Originary Argument or Abduction is an argument which presents facts in its Premiss which present a similarity to the fact stated in the Conclusion but which could perfectly well be true without the latter being so much more without its being recognised; so that we are not led to assert the Conclusion positively but are only inclined towards admitting it

as representing a fact of which the facts of a Premiss constitute an Icon... (An Abduction is Originary in respect to being the only kind of argument which start a new idea)

A Transuasive Argument or Induction is an Argument which sets out from a hypothesis resulting from a previous Abduction and from virtual prediction drawn by Deduction of the results of possible experiments and having performed the experiments concludes that the hypothesis is true in the measure in which those predictions are verified this conclusion however being held subject to probable modification to suit future experiments. Since the significance of the facts stated in the premisses depends upon their predictive character which they could not have had if the conclusion had not been hypothetically entertained; they satisfy the definition of a Symbol of the fact stated in the conclusion. This argument is Transuasive also in respect to its alone affording us a reasonable assurance of an ampliation of our positive knowledge.” (Peirce, 1902, pp. 136–139)

A few pages earlier Peirce explains the terms Originality, Obsistence, Transuasion:

“Obsistence (suggesting obviate, object, obstinate, obstacle, insistence, resistance, etc) is that wherein secondness differs from firstness; or is that element which taken in connection with Originality makes one thing such as another compels it to be.

Transuasion (suggesting translation, transaction, transfusion, transcendental, etc.) is mediation or the modification of firstness and secondness by thirdness taken apart from the secondness and firstness; or is being in creating Obsistence.

Although Originality is the most primitive, simple and original of the categories it is not the most obvious and familiar.” (ibid, p. 128)

There is no difference between these two solutions regarding abduction. But according to 1), the premises of deduction represent the symbol of the conclusion, and the premises of induction are the index of the conclusion, while according to 2), it is the other way around. More precisely:

In 1) we see that (despite using the statistical terms of sample and collection) the arguments are presented syllogistically and it is claimed that the sign of the middle term is the sign of the conclusion. (The middle term is the one that appears in both premises but not in the conclusion. It should be emphasized that the reference to middle terms does not require syllogistic logic, as can be seen from the definition.) A meta-reason for the 1868 account could be the fact that deduction does not depend on a context, nor does symbol, while index and induction do. These could be the reasons why some Peircean scholars still do not question the view that the middle term of the induction is an index. (cf. Paavola, 2004, 2011)

In 2) we see that Peirce relates the kinds of argument to the categories of firstness, secondness, thirdness. Unlike the complexity of Peirce's terminology, the idea presented in 2) is very simple. The category of the first is potentiality, some quality, so it corresponds to an icon; the second is actuality, something individualized, it corresponds to an index; the third is generality that seeks convention – it corresponds to a symbol. Yellow, for example, as a quality *can* represent the Sun, my hand pointing to the Sun *is* representing the Sun, and the word "Sun" *is to* represent the Sun. Indeed, the conclusion of abduction is a possibility, of deduction the conclusion is that which is the case, and the conclusion of induction is a derived, the projected generalization.

## 2. The flaw in Peirce's argument

Although he starts from the natural idea that the semiotic relationship of the premises as a whole with the conclusion has its counterpart in the semiotic relationship within the premises, Peirce seeks this relationship in the relationship of syllogistic terms, thus taking a step backwards. Moreover, despite Peirce's competence, the argument that he makes in 1868 is inherently logically flawed. With regard to his argument, two things should be emphasized:

- (i) Peirce looked for the key to the semiotic relationship within the premises in the logic of terms, not in propositional/predicate logic.
- (ii) The semiotic relationship that Peirce identifies between the terms in the syllogism is not set consistently.

It is clear from (i) that the semiotic implication that I am introducing is not present in Peirce's analysis, from (ii) it follows that even the translation into modern logic would not have given rise to it.

Let us see this more closely.

In the given abductive reasoning: *M is, for instance, P', P'', P'''*, and *P''''*; *S is P', P'', P'''*, and *P''''*, Ergo: *S is M*; the term "*P', P'', P'''*, and *P''''*" is obviously the middle term, and Peirce says it "is a likeness of *M*", which means, in his later terminology, that it is an icon of *M*.

In the given inductive reasoning: *S', S'', S'''*, and *S''''* are taken as samples of the collection *M*; *S', S'', S'''*, and *S''''* are *P*, Ergo: *All M is P*; "*S', S'', S'''*, and *S''''*" is the middle term. For this term Peirce says it "is an index of *M*".

Note that Peirce sets up the generalization: from *S', S'', S'''*, and *S''''* are taken as samples from the collection *M*, Peirce concludes that "*S', S'', S'''*, and *S''''*" is an index of *M*.

But note that the 1868 account relies on another generalization as well: the sign of the middle term corresponds to the semiotic relationship in which the premises stand to the conclusion.

Now, if the premises as an index point to the conclusion (just as “now” points to the moment of an utterance, or “I” to the person who utters it) it does not look like a relationship between the premises of the induction and its conclusion, which might not follow. More precisely, as Peirce correctly observes in 1902, such is the relationship between the premises of a deduction and its conclusion.

To conclude, it seems clear that Peirce, because his 1868 solution does not satisfy the conditions he set, had to abandon it. These conditions are:

- I) The middle term is the argument sign;
- II) The premises of the argument signify its conclusion in the manner of the argument sign.

A correct solution should meet both requirements.

### 3. One solution

Applied to each of the three signs, the requirements I) and II) give:

- 1) If the middle term is an index, the premises as an index point to the conclusion.
- 2) If the middle term is an icon, the premises as an icon point to the conclusion.
- 3) If the middle term is a symbol, the premises as a symbol point to the conclusion.

Specifically: if it is claimed that the relationship between the premises of the induction and its conclusion is the relationship of the symbol to its object, as claimed by Peirce in 1902, then the middle term of the induction must be a symbol as well. In view of one specific assumption present in the 1868 account it appears that this would mean that the sample of a population, instead of being an index, is a symbol of the population. Previously I defended this position in *Creative Enthymeme* (Anđelković, 2007, pp. 25–33) but also in the paper “Reasoning by Signs: Peirce and Aristotle” presented at Charles S. Peirce International Centennial Congress, Lowell, 2014. The solution that I offered in this presentation was accepted and cited in Lorenzo Magnani’s “The eco-cognitive model of abduction II – Irrelevance and implausibility exculpated”, *Journal of Applied Logic* 15 (2016) and “Naturalizing the logic of abduction”, *Logic Journal of the*

IGPL 24 (4) (2016). In coming to this solution I used Aristotle's examples for arguments based on signs as a guide.

Aristotle's text is a good guide for two reasons: a) Aristotle's approach is syllogistic like Peirce's and b) because Peirce himself, in order to defend the introduction of abduction, pointed out the connection between deduction, abduction and induction and the first, second and third syllogistic figures respectively. It should be emphasized though, that Peirce does not draw the connection between his and Aristotle's teaching on signs.

a) In *Prior Analytics* (II 27), Aristotle says:

“An enthymeme is a syllogism from probabilities or signs; and a sign can be taken in three ways—in just as many ways as there are of taking the middle term in the several figures: either as in the first figure or as in the second or as in the third. E.g., the proof that a woman is pregnant because she has milk is by the first figure; for the middle term is ‘having milk’...The proof that the wise are good because Pittacus was good is by the third figure. ...The proof that a woman is pregnant because she is fallow is intended to be by the middle figure; for since fallowness is a characteristic of women in pregnancy, and is associated with this particular woman, they suppose that she is proved to be pregnant.” (Aristotle, 1962, p. 525)

b) As we saw above, Peirce presented induction and abduction syllogistically in the form of the second and third figures. Aristotle's syllogism *Barbara* Peirce takes as “the primitive type of inference” corresponding to deduction and consisting of three statements expressing a rule, a case and a result. (cf. Peirce, 1931–35, 2.710)<sup>1</sup>

Aristotle's examples looked as if they could provide a direct solution is suggested already at the level of language. For “milk” in the first figure, Aristotle uses the term *tekmérion*, which is translated into Latin as “index” (cf. Weidemann, 1989, p. 343, Fidora, 2014, p. 12). As the middle term of the third figure, Aristotle takes Pittacus, which can be seen as a sample of the population of sages, but this cannot be understood as a random sample. Pittacus can be seen as a paradigmatic case of the wise man, or simply as a symbol of the wise man.

Later, after looking more broadly at Peirce's dilemma, I saw a flaw in this explanation. Concretely, dealing with the question of semiotic validity in general, I noticed that the material implication in predicate logic can be interpreted semiotically. I motivated it in the presentation “The Origin of Semiotic Validity” I gave at the Symposium *Validity throughout History*, UCLA, 2019. In the paper “The Origin of Semiotic Validity – Peirce and

---

1 More about this in Trajkovski 2024.

Aristotle on Reasoning by Signs” which followed (Trajkovski, forthcoming 2024) I formally introduced the semiotic implication. There I argued that a universal premise “All M is P”, i.e. “For all x, if x is M, then x is P” in deduction is read: M is an index for P. In abduction, it reads: P is an icon for M. Since in general the conditional can be read as stating a semiotic relationship between M and P, I called it semiotic implication.

#### 4. The solution

The translation of Aristotle’s examples into predicate notation highlights the fact that in induction the middle term is an individual constant, not a predicate one as in the case of deduction and abduction. Hence its middle term is a name whose referent is the name bearer, which means that it functions as a symbol. But then, “Pittacus” is a symbol of Pittacus, not the symbol of wise men. So in Trajkovski (2024) I conclude that the names serve as middle terms in induction; therefore, induction is based on symbols.

Aristotle’s examples presented in predicate logic are:

Proof I:    M(s)  
              (x)(M(x)⊃P(x))  
              [Ergo]  
              P(s)

where “P” stands for “is pregnant”, “M” for “is having milk”, and “s” for “this woman”.

Proof II:    P(s)  
              (x)(M(x)⊃P(x))  
              [Ergo]  
              M(s)

where “P” stands for “is sallow”, “M” for “is pregnant”, and “s” for “this woman.”

Proof III:   s is a sample of M  
              P(s)  
              [Ergo]  
              (x)(M(x)⊃P(x))

where “P” stands for “is good”, “M” for “is wise”, and “s” for “Pittacus”.

If in proof I, the major premise  $(x)(M(x) \supset P(x))$  we read “M is an index for P”, and in proof II, the major premise  $(x)(M(x) \supset P(x))$  we read “P is an icon for M”, the two-way semiotic reading of the implication  $(x)(Y(x) \supset Z(x))$  is almost imposed.

Let us note that the semiotic relationship establishing the sign of the middle term in I and II is set in their major premises. To see this we need to introduce some relevant notions:

“The term which is contained in both premises is the middle term; the predicate of the conclusion is the major term; and the subject of the conclusion is the minor term. The premise which contains the major term is the major premise, and the premise containing the minor term is the minor premise. The order in which the premises are stated does not, therefore, determine which is the major premise. In the syllogism *All mystery tales are a danger to health, for all mystery tales cause mental agitation, and whatever is a cause of mental agitation is a danger to health* the conclusion is stated first, and the major premise last. It is usual, however, to state the major premise first.” (Cohen & Nagel, 1955, pp. 77–78)

If we now look at proof III and the premise “Pittacus is wise”, we see that it has been omitted like the premises “Having milk means being pregnant” and “Sallowiness is a characteristic of women in pregnancy” in the quote from Aristotle. This is understandable because Aristotle gives these examples in his definition of enthymemes (as inferences based on signs) and enthymemes omit what can be implied.” If only one premiss is stated, we get only a sign; but if the other premiss is assumed as well, we get a syllogism.” In the footnote attached, it says: “strictly an enthymeme”, and in the margin there is a comment: “a sign may regarded as a syllogism with one premise suppressed.” (Aristotle, 1962, p. 525)

Let us return to the question of what is the semiotic role of the middle term in induction. If the basic semiotic relation should be sought in the omitted premise, then in the case of proof III it should be sought in “Pittacus is wise”. Then Pittacus can be seen either as a symbol or as an index of the population of wise men. The former would be in line with the solution I give in *Creative Enthymeme*. The latter would be in line with Peirce’s 1868 view that the semiotic relationship of the sample to the population is indexical.

We should go back and look once more at which premise gives the ground for the sign of an argument. Peirce gives examples of abduction and induction and insists that the semiotic relationship is found in the first premise. He says:

in abduction “the first premiss amounts to this, that “P’, P”, P””, and P””” is a likeness of M”, and

in induction “the first premiss amounts to saying that “S’, S”, S””, and S””” is an index of M”.

It seems that Peirce conversationally implied that the key to the semiotic relation is given in the major premise of an argument. In the example of abduction, this is the case, because the premise “M is, for instance, P’, P”, P””, and P””” contains the major term M, that is, the predicate of the conclusion, so it is the major premise. In the case of induction, the premise “S’, S”, S””, and S””” are taken as samples of the collection M” contains a minor term M, namely, the subject of the conclusion, hence it is the minor premise.

It is obvious that in the examples given by Aristotle, premises that are known have been omitted, and this as an epistemic criterion need not be logically uniform. And really, in proofs I and II the major premises are omitted while in proof III the minor premise is. If we set the condition that the criterion for the premises in which the sign of the argument is given should be logically uniform, then Peirce’s 1868 argumentation is clearly inconsistent.

Let us set a consistent criterion:

If the middle term in the major premise is index, icon, or symbol, then the relationship of the premises to the conclusion is index, icon, or symbol, respectively.

Now we see that in Proof I the middle term is M, and the major premise is  $(x)(M(x) \supset P(x))$ , so the semiotic role of M in that premise determines the sign of the argument. Having milk is a sign that a woman is pregnant. It is a sure indicator, that is, an index.

We can see that in Proof I the middle term is P, and the major premise is  $(x)(M(x) \supset P(x))$ , so the semiotic role of P in that premise determines the sign of the argument. Being pale can be an indicator that a woman is pregnant, but it is not a sure indicator because someone can be pale for other reasons, a pale woman looks like a pregnant woman, paleness is an icon here.

Finally, we see that in Proof III the middle term is s, and the major premise is P(s), therefore the semiotic role of s in that premise determines the sign of the argument. The role of s here is to denote Pittacus, it functions as a symbol.

To conclude: the names serve as middle terms in induction; therefore, induction is based on symbols.



## 5. Some normative and formal aspects of semiotic implication

This section is about formal importance of semiotic implication, but also stresses how it deviates from the norm. The symbolic notation for semiotic implication is introduced, as well as its definition and some basic principles. Finally, I emphasize that abductive validity does not follow from the principles of semiotic implication, which only allows the expansion of the framework in such a way that non-standard validity can be grounded on rational reasons.

The formal importance of the semiotic implication is reflected in the fact that it enables writing in the object language that, for example, milk is an index of pregnancy, or that a human is an index of an animal, while an animal is an icon for a human being. Qualities are semiotically related to other qualities, for example, smoke is a sign of fire.

Thus, the index and the icon as semiotic factors are introduced into the symbolic language. Properties in relation to each other do not function as symbols, but as indices or icons, for properties are indicators of other properties not by convention.

Deviations from the norm:

1. The argumentation that I have presented allows deduction to be defined through index as an argument in which the sign of the argument is an index. Thus, instead of deductive validity, one can speak of indexical validity, which would be a deviation from the norm. About how far-reaching this deviation is, is an issue that requires special consideration.
2. The enthymematicity of intuitionistic implication motivated me to argue that the semiotic implication is enthymematic in a complementary way. Intuitionistic implication, as a major premise, is sufficient according to Anderson and Belnap (1961, p. 719). In Trajkovski (2024) I argued that semiotic implication as a major premise is not necessary.<sup>2</sup> In short, the point I am making is that if we take the sentence “For any  $x$ , if  $x$  is  $M$ ,  $x$  is  $A$ ” (“ $M$ ” stands for “a man”, and “ $A$ ” stands for “an animal”) as a missing premise in the argument:  $t$  is  $M$ ;  $s$  is  $A$ . I take as a starting point a prag-

---

2 These ideas were presented in M. Trajkovski, “Semiotic implication” (1st International Congress on Logic, Epistemology & Methodology, 2021, Philosophy Department of the University of Costa Rica) and “Semiotic, material and logical validity” (3rd Context, Cognition and Communication Conference Varieties of Meaning and Content, 2022, Faculty of Philosophy, University of Warsaw).

matic understanding of attributing to a cognitive subject that s/he has the concept  $X$ . If this is the case then the subject must connect this concept with at least one other concept which either implies  $X$  or is implied by  $X$ . Hence, the subject who claims “ $M(t)$ ” must at least, for some  $Y$  and  $Z$ , have an understanding of either “ $(x)(Y(x) \supset M(x))$ ” or “ $(x)(M(x) \supset Z(x))$ ”.

Therefore, an abridged deduction  $M(t)$ ; [*Ergo*]  $A(t)$ , or an abridged abduction  $M(t)$ ; [*Ergo*]  $B(t)$  can be assumed. The point is, as is argued, that claiming “ $M(t)$ ” assumes that “ $M$ ” as a sign is, by the subject, connected with at least one more sign “ $X$ ” in a sense that “ $M$ ” is either an index or an icon for it. As a conclusion it follows that semiotic implication as a major premise is not necessary if the subject making the inference believes and understands the minor premise.

3. The principle of identity  $(x)(P(x) \supset P(x))$  does not hold.

This conditional would be semiotically read as:

A)  $P$  is an index of  $P$

and

B)  $P$  is an icon of  $P$ .

While A) seems natural, B) does not: I wouldn't say that I am my own icon.

This creates the need for a special sign for semiotic implication, so I introduce the following sign:  $\div$ .

### Definition and basic principles

$(x)(P(x) \div Q(x))$  reads:  $P$  is an index of  $Q$ , for any  $P$  and  $Q$ , and  $Q$  is an icon of  $P$ , for  $Q$  different from  $P$ .

This leads to the basic principle:

$(P(x) \div Q(x)) \leftrightarrow ((P(x) \supset Q(x)) \wedge P \neq Q)$ , for all  $x$ .

Other principles are:

Contraposition:  $(x)((P(x) \div Q(x)) \leftrightarrow (\neg Q(x) \div \neg P(x)))$ , it says that if an  $x$  does not have property  $Q$  which is the icon of property  $P$ , then it is not  $P$ .

Transitivity:  $(x)((P(x) \div Q(x)) \wedge (Q(x) \div R(x))) \supset (P(x) \div R(x))$ , it says that if  $Q$  is an icon of  $P$  and index of  $R$ , then  $P$  is an index of  $R$ .

Finally, the deviations are not necessary, for semiotic implication does not entail the validity of abduction since the conjunction  $P(x) \div Q(x)$  and

$Q(x)$  does not entail  $P(x)$ . Namely, if an  $x$  has property  $Q$  which is an icon of property  $P$ , it does not mean that it has property  $P$ . This is in accord with the general definition of an icon as a sign – it can be present in an absence of its referent.

## Conclusion

In the text, I discussed some formal features of semiotic implication; first of all there is non-reflexivity due to which the semiotic relationship of the predicates in the implication cannot be expressed by a standard implication. Hence, I introduce a special symbol “ $\div$ ” for the semiotic implication which I define as  $P$  is an index of  $Q$ , for any  $P$  and  $Q$ , and  $Q$  is an icon of  $P$ , for  $Q$  different from  $P$ , formally:  $(x)(P(x)\div Q(x))$ . This implication is transitive and contraposes. I accept Peirce’s idea that different arguments correspond to different signs and agree with Aristotle and Peirce that the sign should be sought in the middle term, I point out the key error in Peirce’s argumentation and show that the sign should be sought in the major premise. I also underline that the syllogistic approach that Peirce opted for directed my attention to Aristotle’s understanding of syllogisms based on signs that I referred to in my analyses, but that it made it impossible for Peirce himself to see the semiotic implication. As Aristotle talked about signs in the context of enthymemes and Anderson and Belnap about intuitionistic implication as being enthymematic, I asked the question about the relationship between semiotic implication and enthymeme. I present the thesis that the semiotic implication is enthymematic in a way complementary to the intuitionistic one: when the major premise in the syllogism is an intuitionistic implication, the minor one is not necessary, while when the semiotic implication is the major premise, then the minor premise is not necessary. I also highlighted several aspects in the system based on semiotic implication that deviate from the norm. I emphasized, however, that with certain limitations, the notion of semiotic implication can be adapted to the norm.

## References

- Anđelković, M. (2007). *Kreativni entimem*. Beograd: Institut za filozofiju, Filozofski fakultet.
- Anderson, A. & Belnap, N.D. (1961). Enthymemes. *The Journal of Philosophy*, 58(23), 713–723.

- Aristotle. (1962). *The Categories, On Interpretation, Prior Analytics* (Translated by H. P. Cook– *The Categories, On Interpretation*, H. Tredennick– *Prior Analytics*). Cambridge, MA: Harvard University Press.
- Cohen, M.R. & Nagel, E. (1955). *An Introduction to Logic and Scientific Method*. London: Routledge & Kegan Paul Ltd.
- Fidora, A. (2014). Signs vs. Causes? An Epistemological Approach to Prognosis in the Latin Middle Ages. *Tópicos, Revista de Filosofía* 47, 9–23.
- Field, H. (2009). Pluralism in Logic. *The Review of Symbolic Logic*, 2(2), 342–359.
- Paavola, S. (2004). Abduction through Grammar, Critic, and Methodeutic. *Transactions of the Charles S. Peirce Society*, 40(2), 245–270.
- Paavola, S. (2011). Diagrams, Iconicity, and Abductive Discovery. *Semiotica*, 186(1/4), 297–314.
- Peirce, C.S. (1868). On a New List of Categories. *Proceedings of the American Academy of Arts and Sciences*, 7, 287–298.
- Peirce, C.S. (1931–35). *The Collected Papers of Charles Sanders Peirce*, vols. 1–6 (eds C. Hartshorne & P. Weiss) Cambridge, MA: Harvard University Press.
- Peirce, C.S. (1997). *Pragmatism as a Principle and Method of Right Thinking: The 1903 Harvard Lectures on Pragmatism*, (ed. P. A. Turrisi). Albany: State University of New York Press.
- Peirce, C.S. (1902). *Minute Logic*. unpublished manuscript, <https://fromthepage.com/jeffdown1/c-s-peirce-manuscripts/ms-425-1902-minute-logic-chapter-i>.
- Ross, W. D. (ed.) 1951. *Aristotle's Prior and Posterior Analytics– A Revised Text with Introduction and Commentary*. Oxford: Clarendon Press.
- Trajkovski, M. (forthcoming 2024). The Origin of Semiotic Validity – Peirce and Aristotle on Reasoning by Signs. In G. Ciola and M. Crimi (eds), *Validity throughout History*. Philosophia Verlag.
- Trajkovski, M. (forthcoming 2024). Semiotic Implication as an Enthymematic Implication – Semiotic, Material and Logical Validity. *Theoria* Volume 67, Beograd.
- Weidemann, H. (1989). Aristotle on Inferences from Signs. *Phronesis*, 34(3), 343–351.



BALKAN  
ANALYTIC  
FORUM

**BAF+**: Normativity  
of Art



## 5. ART WORKS AND COGNITION





Ted Kinnaman

## NORMATIVITY IN ART IN KANT'S AESTHETICS

**Abstract:** Like all books, Kant's *Critique of Judgment* is of its time. On the topic of art in particular, Kant assumes that great art is beautiful art, and that artists attempt in some sense or other to represent nature. It is reasonable to ask whether Kant's theory of beauty can illuminate art in our time, when these assumptions have mostly been set aside. I will argue that it can, given a proper understanding of the theory. In particular, Kant's account of normativity in aesthetic judgment in general is the central thread also in his account of fine art. The crucial point is that beauty, for Kant, is more intellectual than sensuous, so consequently whether a particular object is beautiful is independent of its sensible qualities. I begin by explaining Kant's account of aesthetic normativity in general. Then I turn to what he says about art, which he treats as a special case within the broader theory of taste. Crucially, Kant thinks that beauty is the same thing in art as in nature, and thus so is the basis of normativity. I conclude by showing how Kant's theory can, perhaps surprisingly, shed light on our appraisal of recent works of art.

**Keywords:** Kant, normativity, aesthetics, beauty, art, genius.

Like all books, Kant's *Critique of Judgment* is of its time. On the topic of art in particular, Kant assumes that great art is beautiful art, and that artists attempt in some sense or other to represent nature. It is reasonable to ask whether Kant's theory of beauty can illuminate art in our time, when these assumptions have mostly been set aside. I will argue that it can, given a proper understanding of the theory. In particular, Kant's account of normativity in aesthetic judgment in general is the central thread also in his account of fine art. The crucial point is that beauty, for Kant, is more intellectual than sensuous, so consequently whether a particular object is beautiful is independent of its sensible qualities.

I will begin by explaining Kant's account of aesthetic normativity in general. Then I will turn to what he says about art, which he treats as a

special case within the broader theory of taste. Crucially, Kant thinks that beauty is the same thing in art as in nature, and thus so is the basis of normativity. I will conclude by showing how Kant's theory can, perhaps surprisingly, shed light on our appraisal of recent works of art.

## I.

Beauty, for Kant, is found primarily in nature, and only secondarily in works of art. The central problem in Kant's theory is reconciling the subjective and objective dimensions of taste: We say that "there is no disputing about taste," but also expect others to agree with our judgments of taste. This is a problem of normativity. Kant wants to show that this expectation is justified, which is why the issue of normativity is so central to his aesthetic theory. His solution is to borrow, in effect, the normativity of taste from the critical account of cognition. In cognitive judgments, e.g. "That bird is a gray catbird," as in judgments of taste, e.g. "That catbird is beautiful," we expect others to agree with our judgment. "Nothing," Kant tells us, "can be universally communicated except cognition, and representation so far as it belongs to cognition." (5:217) From this he infers that the "determining ground" of the judgment of taste "can be none other than the state of mind that is encountered in the relation of the powers of representation to each other insofar as they relate a given representation to *cognition in general*." (5:217) But what is "cognition in general [*überhaupt*]?" Kant cannot mean that taste is cognition, full stop, because then judgments of taste could be supported by appeal to rules, for example, "That sound is the characteristic song of the gray catbird." At the same time, though, it must be cognition in some very significant sense, for otherwise it could not demand the assent of others, which is to say it would carry no normative force.

My suggestion is that we take "cognition in general" to refer to the effort to systematize empirical cognition. The problem of the normativity of taste in the *Critique of Judgment* has its roots in the center of Kant's critical project. In the Appendix to the Transcendental Dialectic in the *Critique of Pure Reason*, Kant tells us that just as the understanding unifies appearances under concepts, reason seeks to unify empirical concepts into a system:

[W]hat reason quite uniquely prescribes and seeks to bring about concerning (the understanding) is the **systematic** in cognition, i.e. its interconnection based on one principle.

What Kant means here by a system is his version the unification of science. Kant's argument, in the earlier *Transcendental Analytic*, has the consequence that every object of experience can be subsumed under some concept or other. A tiny few of these concepts are the well-known pure categories of the understanding, the objective validity of which is a condition of the possibility of experience. This validity is established by the *Transcendental Deduction of the Categories*. More commonly, though, we subsume objects of experience under empirical concepts, such as 'table,' 'stone,' and 'catbird.' Reason's systematizing task is to find unity among these contingent concepts, for example by grouping 'catbird' under 'birds,' 'vertebrates,' and so on, and distinguishing within the grouping between gray catbirds, northern catbirds, etc. The goal, to which we can approach only "asymptotically" (A663/B691), is a hierarchical structure of concepts organized around an idea of pure reason, which defines as "a necessary concept of reason... to which no congruent object can be given in the senses." (A327/B383)

But why do empirical concepts require unification in a system? Kant calls systematic unity the "touchstone [*Probierstein*] of truth" for empirical cognition, which I take to mean that systematicity is Kant's construal of the demand for coherence in empirical cognition. The systematization of empirical cognition is thus an essential part of the task of the systematization in general. And this task is the work of the *Critique of Pure Reason* itself: investigating the possibility of cognition through pure reason. Systematic unity is thus crucial for the larger, well-known project of setting metaphysics on the "secure course of a science," (Bvii) for, he tells us near the end of the work, "systematic unity is that which first makes ordinary cognition into science." (A832/B860)

With what right, if any, do we suppose that our empirical concepts can in fact be unified in a system? That this unification is possible is a contingent proposition, and thus cannot be known through mere transcendental reflection, like the categories. But the possibility of systematic unification is, as I have explained, necessary for the critical project of putting metaphysics on the secure course of a science. Kant takes up this problem in the 1790 *Critique of Judgment*. In the introduction to the third *Critique* Kant tells the reader that the task of the book is to provide a transcendental ground for the possibility of systematizing empirical cognition:

[T]he power of judgment must... assume it as an *a priori* principle for its own use that what is contingent for human insight in the particular (empirical) laws of nature nevertheless contains a lawful unity, not fathomable [*nichtzue-rgründende*] by us but still thinkable, in the combination of its manifold into one experience possible in itself. (5:183–4)

I call this assumption the principle of the purposiveness of nature. Furthermore, of the two major divisions of the work, the “Critique of Aesthetic Judgment” is the one that belongs to that task “essentially,” since “this alone contains a principle that the power of judgment lays at the basis of its reflection on nature entirely *a priori*.” (5:193) So we should expect to find something in Kant’s aesthetic theory that supports the goal of system-building, and I propose that it is found in §9, which Kant calls the “key to the critique of taste.”

Kant’s question in §9 whether, in a judgment of taste, the pleasure precedes the judgment, or the reverse. In the former case, the pleasure would be merely agreeable rather than beautiful, and have only “private validity, since it would immediately depend on the representation through which the object is given. (5:217; emphasis in original) Therefore, Kant concludes, the judging must precede the pleasure. But what must this pleasure be such that it gives rise to a pleasure we can impute to everyone? “Nothing,” Kant says, can be universally communicated except cognition, and representation in so far as it belongs to cognition,” and so the judgment of taste must have a relation to “cognition in general [*Erkenntnis überhaupt*].” (5:217) Unfortunately, Kant is rather vague about what this claim means, so interpreters of Kant have faced a dilemma. If “cognition in general” is just cognition, then judgments of taste ought to be supportable by appeal to determinate rules, which Kant consistently denies. It also threatens the unappealing consequence that all objects must be beautiful, since, according to Kant’s argument in the Transcendental Deduction of the Categories in the first *Critique*, every object is subject to the pure principles of cognition, the categories. On the other hand, to the extent that we take “cognition in general” to cognition *strictu sensu*, Kant is open to the objection that the appeal to cognition does not entail the normativity that he needs from it.

On my reading, this dilemma is resolved by taking “cognition in general” to refer to the goal of systematizing empirical cognition. This allows us to avoid the first horn of the dilemma, because the principle of the purposiveness of nature is not a condition of the possibility of experience. Rather, the imperative to systematize empirical cognition is a regulative principle that guides our investigation of nature.<sup>1</sup> We also avoid the second horn of the dilemma, because systematization is of central importance to Kant’s overall account of cognition. So although Kant does not explicitly say that this is what he means by “cognition in general,” this reading fits the stated needs of his theory quite precisely.

---

1 This is the point of the first section of the Appendix to the Transcendental Dialectic in the *Critique of Pure Reason*.

I want to emphasize two aspects of Kant's view as I understand it. First, Kant's statement in §9 that in the judgment of taste the judging precedes the characteristically aesthetic pleasure has the consequence that the beautiful object does not need to be in itself particularly pleasant. Aesthetic appreciation is a mode of cognition, and the pleasure results from the (quasi-) cognition. I think Kant means to draw our attention to this when he tells us that the judging has the pleasure as its "consequence [*Folge*]." (5:217) Kant surely thought of beauty, including beautiful art, as starting with pleasant things, but this is not required by his theory. This will be important later on for applying Kant's theory to trends in modern art.

The second point I want to emphasize is that, as I read Kant, beauty is a property of objects, albeit an indeterminate one. This point in turn has two aspects: First, judgments of taste *refer* to things in the world, but (second) the only evidence we can give for them is the feeling of pleasure we felt upon judging. To the first point, Kant says quite plainly in the first section of the *Analytic of the Sublime* that "we express ourselves on the whole incorrectly if we call some **object of nature** sublime, although we can quite correctly call very many of them beautiful." (5:245) His task here is to explain the place of the sublime in his "Critique of Aesthetic Judgment," so although this is just one spot in the text, it is a crucial one where Kant was surely writing very carefully. In addition, we should consider the many places in the text where Kant refers, seemingly without irony, to things as beautiful. If he thought it were an error to call objects beautiful, we would expect him to say so, and he does not. To the second point: taste is subjective in the sense that it is based on a feeling of pleasure (or displeasure), which Kant says is "the subjective aspect of a representation **which cannot become an element of cognition at all.**" (5:189; emphasis in original) Here is how he presents his view in resolving the "Antinomy of Taste"—again, a crucial spot where Kant is writing carefully:

The judgment of taste doubtlessly contains an enlarged relation of the representation of the object (and at the same time of the subject), on which we base an extension of this kind of judgment as necessary for everyone, which must thus be based on some sort of concept, but a concept that **cannot** be determined by intuition, and which thus also **leads to no proof** for the judgment of taste. (5:339 – 40; emphasis in original)

He goes on to tie taste to the rather murky notion of the "supersensible substrate of appearances," a "pure rational concept" which, he says, "grounds the object (and also the subject) as an object of sense, consequently as an appearance." (5:340)

In sum, Kant holds the view that (some) objects are beautiful; that we have a right to expect, in a strongly normative sense, that others agree

with our judgments of taste; and that this normativity is grounded in the role of taste as a variety of cognition.

## II.

After developing his aesthetic theory in relation to natural beauty, Kant begins his discussion of art in §43 of the “Critique of Aesthetic Judgment.” There he faces a problem left over from §16, where he distinguished between “free” and “adherent” beauty. Free beauty—beauty in the truest sense, for Kant—presupposes no concept of what the object is supposed to be, whereas adherent beauty does presuppose a concept, so that the concept provides a standard against which the object can be judged. Kant gives the example of a church, the very concept of which places constraints on how a church should look. This is in keeping with a central idea in Kant’s aesthetic theory, already noted, that judgments of taste cannot be supported by appeal to rules. These rules would presumably have the form ‘Every object that is  $p$  is beautiful;  $A$  is  $p$ ; therefore,  $A$  is beautiful.’ But for Kant, artistic production, like all intentional production, requires that the artist start from a concept of what she wants to produce, so it might seem as though it should be impossible for a work of art to be freely beautiful, and that any beauty there is in works of art must be different from beauty in nature.

Kant does not accept this consequence, and in fact affirms that beauty is, in a very strong sense, the same thing in works of art as in natural objects. He allows for the possibility of judging a work of art while ignoring, or being ignorant of, the artist’s intention for the work. The key is that, to be beautiful, art must seem to be nature: Referring back to his earlier account of beauty in nature, Kant says, “[n]ature was beautiful, if at the same time it looked like art; and yet art can only be called beautiful if we are aware that it is art and yet it looks to us like nature.” (5:306) Note that Kant characterizes his own view by saying that “nature resembles art.” This is a reference to the thesis of the *Critique of Judgment*, which is the claim that “particular empirical laws... must be considered in terms of the sort of unity they would have if an understanding (even if not ours) had... given them for the sake of our faculty of cognition, in order to make possible a system of experience.” (5:180) Nature is beautiful in so far as it seems to be made for us, as rational beings. Works of fine art actually are made for us, and it is central to Kant’s theory that, at least in so far as it is successful, art speaks to us as rational beings, just as beauty in nature does.

The univocity of beauty is also reflected in Kant’s account of artistic genius. Adopting a key concept of the *Sturm und Drang* movement as well as, later on, Romanticism, Kant says that “beautiful art is art of genius,”

and he defines genius as the talent “*through which* nature gives the rule to art.” (5:307; emphasis in original) Genius requires “spirit,” which is the source of the originality of its works (“genius is entirely opposed to the *spirit of imitation*” [5:308]), but also taste, which

is the disciple (or corrective) of genius, clipping its wings and making it well behaved or polished; but at the same time it gives genius guidance as to where and how far it should extend itself if it is to remain purposive. (5:319)

Taste, that is, ensures that works of genius are beautiful as well as original. Kant's description of the nature of genius depends on the assumption that the taste guiding the genius is the same taste outlined earlier in the “Critique of Aesthetic Judgment,” and that the beauty she produces is the same beauty.

Thus beauty is the same in nature as in art, and so is the normative basis for each, namely the connection to cognition. Kant distinguishes beautiful from art that is merely “pleasant (*angenehm*)” by saying that while in the case of the latter “pleasure accompanies the representations as mere sensation,” in the former it accompanies them as “*kinds of cognition*.” (5:305; emphasis in original) Note that Kant refers to beauty here not merely as related to cognition but as a type of cognition (*Erkenntnisart*). Though Kant does not say so explicitly, it is clear that the motivation for contrasting ‘pleasant’ with ‘beautiful’ by means of their connections, respectively, to sensation and cognition is to emphasize that in all contexts cognition is the basis for imputing judgments about beauty to everyone.

Kant explains the genius' spirit as “the faculty for the presentation of aesthetic ideas.” These in turn are defined as “representations of the imagination that occasion much thinking though without it being possible for any determinate thought, i.e. *concept*, to be adequate to it.” (5:314) The aesthetic idea is the “pendant” of the idea of reason, which he defines, consistently with the *Critique of Pure Reason*, as a representation of reason to which no intuition can be adequate. Kant's thought here is that a beautiful artwork offers a presentation to the senses that leads us to think beyond that presentation. But not just any thought will do; Kant emphasizes that beautiful art “brings the faculty of intellectual ideas [i.e. reason] into motion,” helping us to “feel our freedom from the law of association.” (5:314) Here too, then, the mark of good art is its appeal to the intellect.

To summarize the view of art I am attributing to Kant: Good art is beautiful art, and beauty is the same thing in works of art as in nature. Beauty in both contexts is the suitability of an object for integration into a system of empirical cognition. Objects are beautiful, but the evidence we give for our judgments of taste is the subjective feeling of the productive

interplay of our cognitive faculties that Kant calls the “harmony of the faculties.” The role of the beautiful object in Kant’s account is important because it is what makes beauty *normative*: Kant assumes throughout his critical writings that our goal in cognition is to get the world right, that is, to represent it accurately. The aspect of cognition most relevant for the *Critique of Judgment* is the systematization of empirical concepts. We cannot prove that such a system is possible, but we can feel that we are making progress toward it. This is what Kant thinks we feel when we have the experience of beauty.

I will conclude this section with a bit of a digression on the topic of normativity of art. We can ask about the basis for normative judgments about particular works of art, but philosophers have also debated the value of art as such. The original impetus for this discussion is Book X of the *Republic*, where Plato excludes artists and poets from the *kallipolis* because fine art is not a craft. Art appeals to our senses with beautiful falsehoods, undermining the rule of reason in the soul and leading us away from the truth. By connecting art to cognition, Kant’s theory suggests a response to Plato. As I read him, Kant connects art to the pursuit of truth without reducing art to a tool of science or morality. Contrary to criticisms from, e.g. Hans-Georg Gadamer, Kant does justice to the subjectivity of taste without reducing it to ‘mere feeling.’

### III.

Equipped now with a proper understanding of Kant’s account of normativity in art, we are in a position to take a fresh look at how Kant’s aesthetic theory can be used to look at some important works of 20th-century art. To put it very broadly, many of the greatest works of modern art are great for reasons other than their propensity to be pleasant to experience. But on the present reading of Kant’s theory of taste, the pleasure characteristic of positive aesthetic, i.e. beauty, attaches to our reflection on the object rather than to the object or artwork itself. In the final section of this paper, I will show how Kant’s theory of art, suitably understood, can shed light on particular works of art exemplify a few of the trends in art in the 20th century. It might seem as though Kant, with his emphasis on art as the production of beauty, his ranking of the various art in terms of their “aesthetic value,” (5:326) and his unfortunate fondness for the poetry of Frederick the Great, would have little to offer to discussion of art in our time. I hope to show that he does.

Let me make clear what I am doing here. I have selected a few works of 20th-century art, some of them, I hope, familiar to most of us. Since my topic is normativity in art, I have chosen works that are generally regarded



as good, that is, successful. Of course, one might quarrel with my judgments, but I think that at least any theory of art that has the consequence that all of them are unsuccessful is not a plausible theory. I will rely to some extent on certain interpretations of the works I discuss, but I do not pretend that these interpretations are original with me, nor that they are unassailable. What matters is that the interpretations are plausible, and Kant's aesthetic theory can help us make sense of them.

Consider, for a first example, perhaps the most significant work of visual art in the 20th century, Marcel Duchamp's *Fountain*. In 1918, Duchamp exhibited *Fountain* in the Independents Art Show in New York. It consisted of an ordinary porcelain urinal, turned on its head and signed, whimsically, 'R. Mutt 1917.' One could be enraptured by *Fountain's* smooth white surfaces, but in fact *Fountain* owes its exalted status not to its appearance but rather to the statement Duchamp makes by means of it: Art is whatever the artist says it is, or whatever an artist does. This is a claim about the nature of art, and as such it might appear to be too abstract to be the basis for a claim of beauty, for Kant or indeed for anyone. This appearance is misleading, however. For Kant, a beautiful thing is one that gives pleasure by stimulating the intellectual faculties. To give one's attention to the sensuous surface of the work would be to miss the genuine significance of it. (Tomkins, 1996, Chap. 12)

The 20th century, in fact, saw a number of attempts, in a variety of media, to erase the artist from the artwork. The American composer John Cage, for example, composed *Music for Change* by having visitors to his New York apartment throw the I Ching and thereby determine notes in the piece. Cage's thought, apparently, was that the work was composed by chance, rather than by himself or his guests. Similarly, Surrealist artists played a game called Exquisite Corpse, in which multiple people would draw on a surface without seeing what the others had drawn. The result was supposed to be a work produced by the collective unconscious of the group, rather than a collaboration between individuals.<sup>2</sup> Now, Kant certainly assumes that artworks are produced by rational agents; indeed, he considers this to be definitive of works of art as opposed to natural objects. (5:303) Interestingly, though, Kant might accept the 'death of the artist,' because rational agency plays no role in explaining a work's success, that is, its beauty. This follows from two aspects of Kant's aesthetic theory already mentioned. First, "beautiful art is an art to the extent that it seems at the same time to be nature." (5:306) In other words, beautiful art does not seem like it is intentionally made. Second, "beautiful art is art

---

2 Breton, André (7 October 1948). Breton Remembers. Archived from the original on 27 January 2008. Retrieved 28 May 2024.

of genius,” and genius, as Kant defines it, is the mental faculty “*through which* nature gives the rule to art. (5:307; emphasis in original) Genius is a nonrational faculty, so to the extent that an artwork issues from genius (and thus, to the extent that it is beautiful), the rational agency, and thus the personhood of the artist drops away. What is crucial for evaluation of the work is its propensity to stimulate the free play of the understanding and the imagination.

Finally, consider what Kant might say about works of art that are non-representative. Jasper Johns’ *Flag*, for example, looks like an American flag, but in fact it is actually a flag. The artist seems to be making a bit of a joke about the fact that any surface with the appropriate arrangement of color planes counts as a flag. Abstract art, too, does not represent anything, that is, it does not carry the mind to anything other than itself. Jackson Pollock’s drip paintings, for examples, are just records of the artist’s activity in creating them—hence the term ‘action painting.’ Both “Flag” and the drip paintings are examples of artworks for which there is no gap between the artwork (in both of these cases, painting) and the thing represented. “Flag” is actually a flag. Both it and the drip painting are, as Hume would say, “original existences,” not representations. Neither has what Descartes calls “objective reality.” (Danto, 1981, Chap. 1) But this does not matter for Kant’s theory. Both works are causally attributable to specific people. Even if they weren’t, remember that for Kant beauty is the same thing in nature and in art, and beauty is the propensity to stimulate the free play of the cognitive faculties.

## References

- Breton, A. (1948). Breton Remembers. Archived from the original on 27 January 2008. Retrieved 28 May 2024.
- Danto, A. (1981). *The Transfiguration of the Commonplace*. Cambridge, MA, USA: Harvard University Press.
- Gadamer, H.-G. (1989). *Truth and Method*, 2nd rev. edition (1st English ed. 1975, trans. by W. Glen-Doepel, ed. by John Cumming and Garret Barden), revised translation by J. Weinsheimer and D.G. Marshall. New York: Crossroad.
- Kant, I. (1997). *Critique of Pure Reason*, trans. and ed. by Paul Guyer and Allen W. Wood. Cambridge, UK: Cambridge University Press.
- Kant, I. (2000). *Critique of the Power of Judgment*, trans. and ed. by Paul Guyer and Eric Matthews. Cambridge, UK: Cambridge University Press.
- Kinnaman, T. (2024). Kant on Aesthetic Normativity. In E. Valdez (ed.), *Rethinking Kant* Volume 7. Newcastle: Cambridge Scholars Press.

- Kinnaman, T. (2018). Kant on the Cognitive Significance of Genius. In V. L. Wai-  
bel, M. Ruffing and D. Wagner (eds.), *Natur und Freiheit: Akten des XII.  
Internationalen Kant-Kongresses* (pp. 3021 – 3028). Berlin: De Gruyter.
- Plato (1992). *Republic*. Indianapolis: Hackett Publishing.
- Tomkins, C. (1996). *Duchamp*. New York: Henry Holt and Company, Inc.



Isidora Novaković

## PHILOSOPHICAL VALUE OF LITERATURE: MACHIAVELLI AND SHAKESPEARE

**Abstract:** The aim of this paper is to show the relationship between philosophy and literature by interpreting relevant plays by Shakespeare. It will be shown that philosophy and literature can talk about the same significant truths, e.g. about human nature, motivation and ambition. The debate between aesthetic cognitivists and aesthetic non-cognitivists that concerns the cognitive value of art, the question of whether art can truly teach us something is significant. While asking about the normativity of art, we should also examine whether its value depends on the social context. This will be the content of the first part of the paper which will deal with the more universal philosophical questions, most notably with the epistemological question of whether we learn through reading literature. And although literary approach to the same topics differs from the philosophical, literature can give us more vivid particular examples that can help us understand philosophical ideas and theories better. Thus, Shakespeare's work can guide us towards a greater understanding of philosophical concepts. For example, we can look into Shakespeare's tragedies for his understanding of virtue and compare it to the philosophical exploration of the same topic. The second part of the paper will explore specific ideas shared by Shakespeare and Machiavelli. Dealing with the relationship between the two authors, it will be inevitable to ask whether Shakespeare's texts can offer us a basis for normative claims about Machiavelli's ideas. Does Shakespeare provide us with a set of norms and rules for the (right) interpretation of Machiavelli's texts?

**Keywords:** philosophy, literature, Shakespeare's plays, Machiavelli, politics, fortune, *virtù*, irony, conspiracies.

### 1. Truth in philosophy and literature

We often hear comments on artworks. A play can be evaluated as profound, a science fiction novel celebrated for the depiction of a faraway world that is so similar to ours, and which we wouldn't have known with-

out that book. These propositions about the works of art have led to the appearance and development of the debate between aesthetic cognitivists and aesthetic non-cognitivists. Baumberger highlights the fact that aesthetic cognitivists accept and take seriously propositions about artworks (Baumberger, 2014, p. 1). Their position is best understood as a conjunction of an epistemic claim that “[a]rtworks have cognitive functions” (Baumberger, 2014, p. 1) and an aesthetic claim that the “[c]ognitive functions of artworks partly determine their artistic value” (Baumberger, 2014, p. 1). On the other hand, Baumberger highlights that the aesthetic non-cognitivists deny and reject one or both of these claims. Thus, there are different versions of this standpoint (Baumberger, 2014, p. 1).

According to Baumberger, the claims that form the basis of aesthetic cognitivism neither state that all artworks have cognitive functions (e.g. romantic comedies), nor that the cognitive functions of artworks necessarily increase their value. And indeed, like he believes is the case with biographies, we will be capable of learning something about history (Baumberger, 2014, p. 1) if we read or watch a play, e.g. *Richard II* and *Richard III*, but that does not (necessarily) make them better plays than *Hamlet* and *Macbeth*. There are valuable artworks that serve purposes that are not cognitive. Examples for that would be propaganda films that have political functions, but do not really convey truth and have no cognitive functions. Apart from cognitive functions, Baumberger believes that an artwork can also have “practical, decorative, political, and economic” (Baumberger, 2014, p. 1) functions, as it can have, apart from cognitive (profoundness, shallowness etc.), other values (Baumberger, 2014, p. 1), e.g. moral, political etc.

Lamarque emphasises that no one denies that the truth has *some* kind of relationship with literature. According to him, fiction reports truth and we can get to truth through fiction (e.g. we can learn how it was to live at some time in some part of the world). However, it seems to him that it is not always the task of literature to report the truth, literature doesn't have to have a didactic purpose. “To the extent that truth is better than falsehood and learning better than ignorance then conveying truth is valuable and works that convey truth have value in that regard.” (Lamarque, 2010, p. 367) He stands behind the idea that all of this is acceptable and indisputable, but the problems emerge once we ask ourselves whether truth contributes to the *literary* value of a work, which leads to even more problems. This then becomes a dispute about the value of literature, because the value of truth isn't problematic, but we are concerned with whether there is some specific kind of truth for literature (Lamarque, 2010, p. 367).

Philosophical texts, such as Hume's *Treatise on Human Nature*, but also historical texts such as Thomas Macaulay's *History of England* fall

under literature, as Lamarque highlights, despite the fact that they aim at truth. If we try to restrict ourselves to “the relevant class of *fictional* works” (Lamarque, 2010, p. 368), we lose lyric poetry from sight, even if we do not try to make a distinction between fictional and non-fictional works and realize how hard of a task that is (Lamarque, 2010, p. 386). Because of that, Lamarque suggests that we focus on the question of *literary* value and to judge the artworks *from the literary perspective*. That’s when we come to an understanding that what has a *literary* value in the case of Hume’s work, is not its philosophical aspect, but the way in which the *Treatise* is written (Lamarque, 2010, p. 368). Another reason for the connection between the truth and literature is a long tradition that Johnson illustrates by saying: “Poetry is the art of uniting pleasure with truth.” (Lamarque, 2010, p. 368, quoted from Samuel Johnson, *Milton (Lives of the Poets)*) Still, Lamarque indicates to us that the truth in poetry can be thought of here as a kind of seriousness of poetry: the seriousness of the reflection on the work, thoughts on the new possibilities and the development of imagination. But we have to be careful, because there is no proof of the connection of this seriousness with truth (Lamarque, 2010, p. 368).

Still, Lamarque emphasizes that there are those who oppose the idea that fiction, unlike science and history, can communicate any *truth*, because it stands opposite to literature in some key respects. Literary value rests on “inventiveness, imagination, clever plots, or engaging characters.” (Lamarque, 2010, p. 369) However, “works of fiction are usually set in the real world, often referring to real places, events, or famous people” (Lamarque, 2010, p. 369), they get their ideas from the real world and so they can teach us geography or history. By using means similar to thought experiments in philosophy, literature can teach us truth through parables, “or moral tales told to children” (Lamarque, 2010, p. 369).

Yet, Lamarque points out that there are radical skeptics who completely deny that there is truth or they do not find it particularly important, so the truth would be a system “of metaphors, metonyms, and anthropomorphisms.” (Lamarque, 2010, p. 369, quoted from Friedrich Nietzsche, “On Truth and Falsity in Their Extramoral Sense”). Statements like these, obviously describe the abovementioned concepts extremely pejoratively compared to the previously mentioned understandings of the topic.

Interesting for us are the stances of those who expressed their thoughts through the perspectives of both literature and philosophy. Because of that, Lamarque quotes Iris Murdoch, philosopher and writer, who found truth, perhaps even for her own fiction, in the great works of literature and understood it as clarity:

[W]hat we learn from contemplating the characters of Shakespeare or Tolstoy . . . is something about the real quality of human nature, when it is envisaged, in the artist's just and compassionate vision, with a clarity which does not belong to the self-centred rush of ordinary life. . . . [T]he greatest art . . . shows us the world . . . with a clarity which startles and delights us simply because we are not used to looking at the real world at all. (Lamarque, 2010, p. 371, quoted from Iris Murdoch, *The Sovereignty of Good* (London: Routledge, 1970), p. 65)

Lamarque points us to another way in which we can interpret the truth in literature – if it corresponds to something, so if it's faithful to life, or, perhaps, human nature. That is achieved through believable and “recognizable characters and situations, [it] must avoid implausibility in plot structure . . . and must conform to norms of action and motivation.” (Lamarque, 2010, p. 372) Moreover, in literature, we can take everything that is expressed beautifully and clearly as true (Lamarque, 2010, p. 372).

Sometimes truth will not be conveyed explicitly, but, as Lamarque emphasizes, the reader's effort to fill the gaps so as to recognize the truth in the context will be necessary. The reader will be expected to recognize certain generalizations at the level of topic. For example, a literary work can contain general statements about the human nature that would lead the reader to the conclusion that (Lamarque, 2010, pp. 374–5), e.g. men are corrupt. Therefore, we can conclude that there is some kind of truth in literature, but it can be interpreted, as Lamarque states, as honesty, clarity, authenticity, probability of the structure of the plot, et al. As such, it can contribute to the literary value, but it is not the same as that necessary truth that would satisfy us in philosophy or history (Lamarque, 2010, pp. 372–3). It is some kind of “universal” truth that needs no proof, something that is not a true *factual* proposition (Lamarque, 2010, p. 376).

Another approach to thinking about the relationship between philosophy and literature comes from Dorothy Walsh and can be clarified with examples from Shakespeare's plays. In her opinion, in addition to *knowledge that* and *knowledge how*, we can also speak of *knowledge what it is like*. In that way we could understand what it means to lose a child or change our faith and that would be the knowledge characteristic of literature (Wilson, 1983, pp. 491–2, according to Dorothy Walsh, *Literature and Knowledge* (Middletown, Ct: Wesleyan University Press, 1969), 96.). The examples from Shakespeare's works could then be useful to us precisely so we would learn, *know what something is like* from Lear's loss of Cordelia or Jessica's conversion from Judaism to Christianity. Walsh believes that this knowledge doesn't lack anything, it doesn't need proof, so it would be absolutely unnecessary to look for it and its lack wouldn't be problematic



at all (Wilson, 1983, p. 492). If we understand knowledge that we find in literature in this way, Shakespeare's works would give us an almost inexhaustible source of knowledge. However, Wilson emphasizes that we have to take into account that even this understanding is not complete and that there are many challenges for it to face. It is not hard to *know what it is like* on a superficial level, let us use examples from *King Lear*, to divide the property between ungrateful daughters and consequently lose a home, money and mind, but Wilson appeals to us that we can truly know this in the 'strong sense' only if we make big changes to our thoughts and conduct in the light of such circumstances. That is to say that, if we were to know this in the strong sense, we would have to modify our understanding of homelessness, poverty and mental stability, and with them our thoughts and conduct because only then can we broaden and apply our knowledge on new cases. We would, thus, have to think of ways to find housing, gain money and seek professional help, and act accordingly: to get a job, rent a place and visit a therapist. Only then does this knowledge become a tool for reacting to situations, problems and questions (Wilson, 1983, pp. 492–5).

Baumberger believes that the knowledge gained from literature doesn't need to be interpreted in the same way we usually interpret knowledge, namely, as a propositional knowledge composed of justified (or reliable) true claims. It is absolutely enough to say that reading literature leads us to understanding and that already suggests that we have made some cognitive progress (Baumberger, 2014, p. 2). For cognitive progress it is enough to ask questions or pose problems for which we don't need to give conclusive answers or solutions, but to give some clarifications – this is especially important for philosophy (Baumberger, 2014, p. 5).

And, since artworks can teach us something (Baumberger, 2014, p. 9), e.g. how England looked during the rule of Richard II, they particularly help the development of our cognitive faculties when they deepen our understanding of things. Baumberger, thus, emphasizes that Don Quixote refers to a hopeless, although noble, man who acts absurdly, and Don Juan refers to a great lover. We will sometimes meet such people in real life and we will be able to give them names of these heroes as predicates (Baumberger, 2014, p. 10). So a noble scholar who constantly delays his goals might remind us of Hamlet, maybe we will even begin to use that name to refer to him.

Baumberger highlights the fact that literature has more means that can lead us to a better understanding of something. It can use specific accentuation, deliberate subtraction and addition, deforming and alienation of different characteristics of objects so as to grip our attention (Baum-

berger, 2014, p. 10). E.g. we can find specific accentuation and deforming in the descriptions of Richard III that are used to focus our attention to his foulness (*Richard III*, I. i. 1–31; IV. iv. 166–173). Addition may be found in *Macbeth*, if the witches were there to grip the king's attention. And Hamlet finds himself an alien in the world from which quite a lot has been subtracted, his father most notably. So, “[b]y carefully selecting and describing fictional incidents, actions and characters, [literary works] provide perspectives on real people and their relationships and interactions.” (Baumberger, 2014, p. 11)

Art, and with it literature, could serve as a thought experiment (Baumberger, 2014, p. 16). Its relationship to philosophy is clear to that extent because, as Baumberger says, philosophy and science can “contain and use thought experiments” (Baumberger 2014: 16). And insofar as thought experiments in these disciplines can lead us to cognitive progress, Baumberger doesn't see why that wouldn't be the case with literature. Literary works will prompt us to explore different possibilities, see the consequences of assumptions, illustrate and support ideas and hypotheses (Baumberger, 2014, p. 15). That can lead us to ask more questions, change our opinions on a topic, give us a new perspective, show us what it means to find ourselves in a situation etc. (Baumberger, 2014, p. 16) Baumberger believes that literature can help the development of different cognitive faculties by encouraging our imagination, but we will also be able to reflect about that of which an artwork speaks, we will learn from examples that we read, and maybe we will ponder some problems that we usually wouldn't, e.g. some moral dilemmas. All of this can lead to an adoption of new knowledge, and the enhancement of our memory (Baumberger, 2014, p. 17) as we later on think back on the work or reproduce it.

Of course, Baumberger states that the aesthetic non-cognitivists do not agree with this understanding of literature, and they will claim that the truths and beliefs acquired through reading literature are trivial. For instance, *Crime and Punishment* would not teach us that murder is bad, we ought to know that beforehand so as to recognize and accurately interpret these ethical ideas (Baumberger, 2014, p. 18). Hence, according to aesthetic non-cognitivists, a parallel would be to take *The Merchant of Venice* and claim that we came to an insight that greed is bad through it. On the other hand, Baumberger highlights that some non-cognitivists even claim that we cannot base our beliefs on literary works, because they could never be justified, especially not by simply referring to the work. We can change Baumberger's example and say, according to some aesthetic non-cognitivists, to read *Richard III* isn't sufficient to permit us to say that

we know how England of his time looked. For that we need history books (Baumberger, 2014, p. 18).

But by using this example we can illuminate Baumberger's position, and say that *Richard III* undoubtedly is some kind of biography. Shakespeare's historical play, or tragedy as it was previously classified (Greenblatt, 2016, p. 296), refers to the real political figures and is set in a real place, so it can provide us with a basis for beliefs about them, and with that it will offer us some points of view that the historical texts do not offer us (Baumberger, 2014, p. 19). *Richard III* will, thus be a thought experiment that will use historical figures, places and events in its setup. Furthermore, the accuracy of the certain historical information found in the tragedy can provide us with a reason for the better evaluation of this play (Baumberger 2014: 20). "Hence, we can learn about aspects of the world through imagination, and ordinary or literary thought experiments can aid these imaginings." (Baumberger, 2014, p. 20)

Novitz believes that there can ultimately be no knowledge without imagination, and imagination is precisely that which lies in the foundations of fiction (Carroll, 1990, p. 167). Fiction, thus, can help us form hypotheses.

Furthermore, if the fanciful imagination operates this way with respect to knowledge acquisition across the board, the fact that some of our hypotheses are concocted in fictions should serve as no impediment epistemically so long as those hypotheses turn out to be successful – successful, that is, with respect to illuminating the world. (Carroll, 1990, p. 167)

In Novitz's understanding, propositional knowledge is not the only form of knowledge (Carroll, 1990, p. 168), and

literature may impart beliefs about values, practical skills – knowledge of *how* to do *x* (strategic skills) or new ways to think about *x* (conceptual skills) – and empathetic skills (the ability to experience what it feels like to be caught up in certain situations), as well as deepening, and perhaps complicating, our understanding of our own values by exploring them in relation to challenging situations. In all these different ways, we can learn from fiction[.] (Carroll, 1990, p. 168)

As opposed to Novitz, Carroll sees the knowledge acquired in this way as a sort of induction (Carroll, 1990, p. 169). It does seem to me that the works of fiction can inspire us to form hypotheses, and if Carroll is correct in his view of this knowledge as a product of induction, we could rely on it just as much as we can rely on astronomy or, in general, sciences that rest on induction. And since we usually don't doubt science, maybe it is time to put a bit more trust in literature.

## 1.1. Universal/particular

According to an ancient idea, literature presents “the universal through the particular.” (Shiner, 2010, p. 25) Most notably, Aristotle understands this relation as follows:

[A] historian and a poet . . . differ in that one speaks of what truly happened, and the other of what could have happened.

That is why poetry is a more philosophical and a more serious thing than historiography, because poetry shows more that which is general, and historiography that which is particular. General is when we say that the person with these or those traits has to speak or act in this or that way by probability or by necessity; and poetry pays attention to that, when it gives persons names. (Aristotel, 2015, pp. 71–2)

Shiner differentiates three understandings of the relationship between the universal and the particular. Actual particulars could, thus, be characters that are not fictional, but are real historical figures or might have been real: Julius Caesar, Anthony and Cleopatra; Hamlet and Macbeth. In dealing with this problem like Shiner, we come to an understanding that, since Prospero from *The Tempest* never existed, he is not a particular, despite the fact that such owls and frogs showed in *Macbeth* do exist (Shiner, 2010, p. 25). Abstraction from these particulars leads us to “any ambitious man or woman, anyone who is torn between passion and duty,” (Shiner, 2010, p. 25) or just any owl. Yet, there is still one more mode of presentation, so to say, of the relationship between universal and particular through which the “ontological function” is reflected. Through it, we can find, not only characteristics of a certain type of person or a certain danger to man, but persons in general and general dangers to them. This is “ontological” precisely because it deals with how things are in the ultimate principle, human nature in general (Shiner, 2010, p. 25). Once we allow such an understanding, we can say that even science fiction can intelligibly provide us with a viewpoint of our world by indicating something else entirely, thus illuminating parts of our own world (Shiner, 2010, p. 28).

The general claims an artwork makes are typically made implicitly by the work’s treatment of particulars. They are displayed in the fine-grained descriptions or representation of particular characters and events. Hence, the cognitive contribution of artworks does usually not primarily consist [of] conveying general beliefs considered in abstraction from the particulars of the narrative and its characters. It rather lies in the detailed descriptions or representations of particular (and often imaginary) cases that suggest the general beliefs. (Baumberger, 2014, p. 20)

Hunt explains that approaching a work of fiction from a philosophical standpoint offers us a plethora of examples, particulars to which we can apply our ideas or beliefs (Hunt, 2006, p. 400). He showcases his ideas on an episode of *The Twilight Zone*, but they can be slightly modified so as to serve the purpose of our examination of literature. So, when we come in contact with a play, we encounter particulars that may have an impact on our cognition.

Since these particulars may differ widely from those in which the viewer [or reader] first acquired these beliefs and values, this process might produce surprising results. These results can affect beliefs they hold when they are no longer viewing the motion picture [, reading or watching a play] and actively contemplating this fictional world, because that world might well be logically relevant to what their beliefs ought to be. (Hunt, 2006, p. 400)

What happens here is that particulars might influence our real life beliefs. In that way king Lear's particular circumstances are able to modify our beliefs in regards to the division of property or make us acquire new beliefs about it. And not only that, but "the narrative itself, or part of it, is the example that drives the argument." (Hunt, 2006, p. 400) That would mean that we could even formulate an argument against the division of property as long as we are alive based on the particular case of king Lear.

What the narrative contributes to the viewer's [or reader's] cognition is not so much the abstract and universal as the concrete and particular. To the extent that it works on the viewer's [or reader's] mind in an argument-like way, it works as an argument-by-example. (Hunt, 2006, p. 401)

Because of that, we are allowed to explain further or base our philosophical arguments on the particular examples from literature.

## 2. Machiavelli and Shakespeare

In this part of the paper I attempt to show that philosophical understanding can be improved by taking literary works into account. By using excerpts from selected Shakespeare's plays I endeavor to clarify some of Machiavelli's more universal ideas.

Similar topics, affinities and interests that can be found in the works of both authors can undoubtedly, at least partially, be explained by their seemingly similar lives. They highlighted both past and present events and compared them, and they both studied history so as to understand current political events better. "Shakespeare's plays were always decisively . . . and are, in the world and of the world . . . [H]e also wrote scripts that were

intensely alert to the social and political realities of their times.” (Greenblatt, 2016, p. 12) Shakespeare and Machiavelli knew very well both the highest and the lowest social classes. And, besides writing about the rulers and for the rulers (and wore (literally or figuratively) their clothes (see Makijaveli, 2018, p. 8)), they wrote about the citizens and directly spoke to the ordinary (or more ordinary) people, one of them at the theatre, and the other primarily in *Discourses on Livy*. Because of these similarities, not only Machiavelli’s works, but also Shakespeare’s plays can be read as real-political (Čavoški, 2019, p. 374).

Despite the fact that Shakespeare referred directly, but very rarely, to other thinkers, it is obvious that he was very well read and informed. Given that he mentioned the names Socrates and Aristotle at least once in his works, (Rowe, 2010, p. 174, p. 178) and a play on the name Machiavelli – Machevil (Grady, 2000, p. 124, according to William Shakespeare, *The Riverside Shakespeare* (Boston: Houghton Mifflin, 1974), 3.2.193), we can be sure that he knew of them and some of their ideas. Obviously this use of Machiavelli’s name wasn’t kindhearted and it might have led to further misconceptions about the theory relayed in *The Prince*. Chavoshki says as follows: “Shakespeare knew well who Machiavelli is and what kind of diabolical political teaching he expounded.” (Čavoški, 2019, p. 375) In contrast to Chavoshki, I would say that Shakespeare did not know who Machiavelli truly was, because were Shakespeare actually aware of Machiavelli’s words, I doubt that he would ever use Machiavelli’s name like that and yet write in such a similar vein.

## 2.1. Fortune

One of the key notions of Machiavelli’s political philosophy is fortune and it is one of the “two important independent notions that stand in the beginning of every political action.” (Uzelac, in Makijaveli, 2020, pp. 164–5) Fortune is an external force, independent of man; fortune is a necessity that is impossible to control and escape (Uzelac, in Makijaveli, 2020, p. 165). Apart from fortune as a notion of political philosophy, Machiavelli speaks of personified Fortune, which is the remnant of olden times:

[F]ortune being changeful and mankind steadfast in their ways, so long as the two are in agreement men are successful, but unsuccessful when they fall out. For my part I consider that it is better to be adventurous than cautious, because Fortune is a woman . . . and it is seen that she allows herself to be mastered by adventurous rather than by those who go to work more coldly. She is, therefore, always, womanlike, a lover of young men, because they are less cautious, more violent, and with more audacity command her. (Machiavelli, 2014, pp. 121–2)

We find a very similar conception of Fortune in Shakespeare's plays. On one hand, Fortune determines greatly the destiny of political figures and is capable of changing the course of history. Thus, she is understood as an unpredictable force. On the other hand, she can also be a "personified superhuman force with the matching will, [a force] that *chooses* on a whim, that is, caprice, what and when she will grant or deny someone, make happy or harm, give good or inflict evil." (Čavoški, 2019, p. 68)

All unavoided is the doom of destiny. (*Richard III*, IV. iv. 218)

Hamlet testifies to the idea that there is no escape from fortune when he starts following the ghost, despite Horatio's plea (Čavoški, 2019, p. 72):

My fate cries out[.] (*Hamlet*, I. iv. 83)

And despite the fact that he is so rational, Hamlet, who is on a mission to "bring the world to the law and justice", he sees fortune as inescapable. Of course, Hamlet was very hesitant by nature (Čavoški, 2019, p. 56) which is easily noticeable in his delay of bringing the world to the law and justice, and so it seems that he is ruled, maybe contradictory, by both reason and destiny. "Because, when he couldn't live according to his ideal of rational stoicism, he let himself be destiny's hostage just in the way he predicted her:" (Čavoški, 2019, p. 56)

[B]lest are those  
Whose blood and judgement are so well commingled,  
That they are not a pipe for fortune's finger  
To sound what stop she please. (*Hamlet*, III. ii. 73–76)

Machiavelli tells us that "everything changes according to the circumstances" (Makijaveli, 2018, p. 93) and Shakespeare testifies to that. Fortune constantly turns the wheel,

so she raises some, and others she throws down, and in the order that only she knows and determines. . . . In other words, those who are engaged in politics and aspire to high positions ought to closely follow the wheel of fortune, so as to find out where their allies and adversaries are. When the wheel goes downhill, those who fall and perish should be abandoned immediately. If, on the other hand, it rises uphill, one should join the lucky ones so as to quickly prosper. (Čavoški, 2019, p. 76)

## 2.2. *Virtù*

Still, there is a way to fight Fortune. The other key notion that lies in the foundations of every political action is *virtù*. This notion is usually translated as *virtue* (Uzelac, in Makijaveli, 2020, pp. 164–5), which is not

adequate for it differs from the ancient, and ethical in general, notion of virtue. *Virtù* is Machiavelli's technical term that signifies "a part of man's character, his boldness to react or not to react to the elements that fortune brings into events . . . i.e. the ability to adapt to fortune and her game of chance, even to impose our law on fortune." (Uzelac, in Makijaveli, 2020, p. 165) And while fortune is a necessity, *virtù* is a condition of our free will, that is, it provides space for our decisions despite and against fortune. "[V]irtù is an answer to her, and it depends on our *virtù* whether we will be able to respond effectively or if the response will ultimately be a prelude to our defeat." (Uzelac, in Makijaveli, 2020, p. 165) Still, Machiavelli speaks of virtue in the same sense as we do so today, as of positive characteristics of men (Machiavelli, 2014, p. 56), but he also speaks of *virtù*, roughly, as of courage and morale (Machiavelli, 2014, p. 60).

Shakespeare also speaks of virtue as of "fundamental moral and political value" (Čavoški, 2019, p. 164). It is exactly because both of these authors use the term 'virtue' to describe positive characteristics of a person, and use it in political contexts and some actions evaluate highly, that we can say that both of them speak of *virtù*. "[*Virtù*] [does] not manifest and prove itself in everyday, but only in unusual circumstances, when [a man] is put to great trials." (Čavoški, 2019, p. 164) When it comes to Shakespeare's plays, like in the work of Machiavelli, we can find the use of the notion of virtue that was characteristic for the Romans, in order to signify mainly courage (Čavoški, 2019, p. 164; *Coriolanus*, II. ii. 85–86). Because of that, *virtù* "stands out and proves itself [best] in the war that Othello exalts and glorifies." (Čavoški, 2019, p. 165) Such an understanding of *virtù* is illustrated, for example, by Hamlet's words:

Rightly to be great  
Is not to stir without great argument,  
But greatly to find quarrel in a straw  
When honour's at the stake. (*Hamlet*, IV. iv. 53–56)

*Virtù* is especially important for the new rulers. And since Macbeth proved himself as full of *virtù* in the battles, he later gained the ability to establish himself as the usurper.

[B]ut all's too weak;  
For brave Macbeth, –well he deserves that  
name, –  
Disdaining fortune, with his brandisht steel,  
Which smoked with bloody execution,  
Like valour's minion,



Carved out his passage till he faced the slave;  
Which ne'er shook hands, nor bade farewell to him,  
Till he unseam'd him from the nave to th'chops,  
And he fixt his head upon our battlements. (*Macbeth*, I. ii. 15–23)

Yet, Machiavelli believes that this leonine characteristic, this courageousness and the ability to scare away the enemies and wolves (gentry), isn't sufficient. He believes that a ruler ought to think ahead, to anticipate traps and act cunningly and twofacedly like a fox (Machiavelli, 2014, pp. 83–4). And there is no bigger fox in Shakespeare's plays than Richard III who was always and everywhere a fraud (*Richard III*, I. i. 90–95, I. i. 147–150; I. iii. 325–340). On the other hand, Macbeth was no fox, which was best seen in act II, scene iii when no one believes him (although they do not openly tell him so) that he wasn't involved in the murder of king Duncan.

As a political realist, Machiavelli realized that virtuous rulers sometimes fail, while the vicious persevere. Sometimes evil is necessary (Machiavelli, 2014, p. 73), and Shakespeare comes to the same conclusion and has Iago tell Cassio: “[Y]ou are but now cast in his mood, a punishment more in policy than in malice” (*Othello*, II. iii. pp. 269–270). But we ought not to be seduced by the notion of *virtù* and think that Machiavelli knew of no more virtues than of that bravery/morale/resolve/readiness for battle. Still, it is more important to act like a fox and pretend that one has them, than to actually possess them. Next to *virtù*, there are still some crucial virtues (Machiavelli, 2014, pp. 85–6). “[A] prince ought to . . . appear to him who sees and hears him altogether *merciful, faithful, humane, upright, and religious*. There is nothing more necessary to appear to have than this last quality” (Machiavelli, 2014, p. 85, our cursive). No one demonstrates that better than Richard III, who is presented as disciplined in war, wise in peace, bountiful, virtuous, fair and humble (*Richard III*, III. vii. 16–17), when he says:

But then I sigh; and, with a piece of Scripture,  
Tell them that God bids us do good for evil:  
And thus I clothe my naked villainy  
With old odd ends stoln out of holy writ;  
And seem a saint, when most I play the devil. (*Richard III*, I. iii. 335–339)

### 2.3. Conspiracies

A hasty ruler, the one who is not careful enough, can become hated, and general hatred of the people is the cause number 1 of conspiracies (Makijaveli, 2020, p. 10). And “[i]f fortune leads the conspiracy, *virtù* is an

answer to her, and it depends on our *virtù* whether we will be able to respond effectively or if the response will ultimately be a prelude to our defeat.” (Uzelac, in Makijaveli, 2020, p. 165) A ruler puts himself at risk with the onslaughts “on a man’s life, his property, or his honour.” (Makijaveli, 2020, p. 11) For instance, Richard II stole the land of Bolingbroke’s father which ultimately led to his demise. If a ruler is to attack someone, Machiavelli states, he then ought to extinguish the lineage of the former ruler (Machiavelli, 2014, p. 10). Not doing that was the mistake of Macbeth’s and Claudius’. Furthermore, Claudius offends Hamlet by marrying his mother, his father’s widow, and by calling himself Hamlet’s father (Šekspir, 2016, pp. 15–9). So Claudius’ demise was only natural.

“Conspiracies are figured out either by uncovering or by guessing.” (Makijaveli, 2020, p. 27) That’s why no one believed Macbeth that he had no involvement in the murder of Duncan, despite the lack of evidence for or against him – that conspiracy was figured out by guessing. “[I]n order to ensure that the ruler is not under threat, people should either be treated kindly or [he has to] get rid of them” (Makijaveli, 2020, p. 27). To ensure that their fathers will not be avenged, Richard III orders the murder of his nephews (*Richard III*, IV. iii. 24–44) and decapitates his subjects left and right when he doesn’t like what they’re saying, which should serve as a reminder for anyone who would try to conspire against him (*Richard III*, III. v. 46–48).

DUKE OF GLOSTER: Then be your eyes the witness of their evil:

Look how I am bewitched; behold mine arm

Is, like a blasted sapling, wither’d up:

And this is Edward’s wife, that monstrous witch,

Consorted with that harlot-strumpet Shore,

That by their witchcraft thus have marked me.

LORD HASTINGS: If they have done this thing, my gracious lord, –

DUKE OF GLOSTER: If! thou protector of this damned strumpet,

Talk’st thou to me of ‘ifs’? Thou art a traitor: –

Off with his head! (*Richard III*, III. iv. 68–77)

## 2.4. Irony

When we call someone ironic or sarcastic in our everyday life, we usually mean that he says, more or less humorously, and expressed by means like intonation or gesticulation (e.g. by rolling his eyes), something opposite from what he actually thinks. Yet another kind of irony is also present in the works of Machiavelli and Shakespeare. Namely, Machiavel-

li's irony usually boils down to the question of his actual political stance. Was he a monarchist or a republican? He undoubtedly gives great advice for both sides. But it is this overemphasis that can guide us toward the answer to the question. Something similar can be found in Shakespeare's plays as well, but there will also be some differences.

Machiavelli wants, first and foremost, the common good. And even when the ruler or tyrant himself is virtuous, this common good is unobtainable in a monarchy. Because of that, Skinner points out that the only way to freedom lies in the republic, a community of free people (Skinner, 1990, p. 141). Viroli states that, according to Machiavelli, someone's status and class should in no way keep him from participating in politics. On the contrary, politics should be something available to the best of the citizens, and not only to the high ranking, gentry and nobles. All those who have proven themselves and gained their reputation in public office are worthy of political life, and not those who gained it simply by being wealthy and born in a certain family. One can, thus, prove himself by offering wise counsel and by doing acts that benefit the citizens (Viroli, 1990, pp. 155–6).

But Shakespeare didn't think highly of the common people. *Coriolanus* shows us people terrified before the battle, but vulturous after it (Čavoški, 2019, pp. 297–8), incapable “of doing anything good without the right lead” (Čavoški, 2019, p. 298). The people are imprudent, irresponsible and disunited because of their great individual differences. All of that makes them unskilled in politics, fickle and unreliable (Čavoški, 2019, pp. 298–300). Coriolanus is especially harsh toward plebs:

I muse my mother  
 Does not approve me further, who was wont  
 To call them woollen vassals, things created  
 To buy and sell with groats; to show bare heads  
 In congregations, to yawn, be still, and wonder,  
 When one but of my ordinance stood up  
 To speak of peace or war. (*Coriolanus*, III. ii. 7–14; Čavoški, 2019, p. 301,  
 quoted from  
 Šekspir, *Sabrana dela* (Beograd: Službeni list SRJ i Dosije),  
*Koriolan*, III čin, 2. Scena, 8–14.)

But even in his dislike of the common people, Shakespeare still manages to be humorous. As Chavoshki highlights, Shakespeare shows the people as morally corrupt and lacking guilt for their own mistakes, such as the murder of Cinna the poet, who didn't conspire against Caesar (Čavoški, 2019, pp. 302–3).

CINNA: Truly, my name is Cinna.

FIRST CITIZEN: Tear him to pieces; he's a conspirator.

CINNA: I am Cinna the poet, I am Cinna the poet.

FOURTH CITIZEN: Tear him for his bad verses, tear him for his bad verses.  
(*Julius Caesar*, III. iii. 27–30)

All of this leads us to the conclusion that, while Machiavelli was a republican, Shakespeare was most probably a monarchist.

### 3. Conclusion

Although they differ in method, philosophy and literature share some common topics, so we can conclude that their goal – understanding – is the same. Literature is a thought experiment, and thus has a cognitive function. Truth finds its place in literature, at least in the sense that fiction takes some elements from the real world, so it can teach us, say, history and geography. And if we were to see imagination as the basis for the formulation of hypotheses, we could conclude that literature can inspire us to create arguments based on induction, much like sciences do. We can also learn the truth from metaphors, which are not only stylistic devices, but are also characteristic of philosophy. One of the tasks of literature is illumination of our universal principles and concepts through particular examples of situations and characters.

Philosophical interest in art was born around the same time as philosophy itself. Aristotle's definition of tragedy dates from the ancient era, but tragedy moved farther away from it later on. This move from Aristotle's suggestions for a good tragedy can be found in Shakespeare's plays. Despite that, we will find a plethora of philosophical ideas in his works. A special attention was given to the comparison of Shakespeare's ideas with the ideas from the political philosophy of Machiavelli.

Hence, the chosen plays also have a political value. That was primarily shown in the comparative analysis of them and Machiavelli's ideas from *The Prince* and *Discourses on Livy*. These ideas were sorted by topic so as to follow the notions and problems of fortune, *virtù*, conspiracies, and irony. Thus, both authors speak of virtue as positive characteristics of a man, but also, in a more technical sense, they speak of *virtù*, i. e. conquering and militant bravery and morale. Both of them, however, believe that fortune also has a giant impact on one's rule. Fortune is an unpredictable force capable of turning the political situation over. The influence of fortune and *virtù* together is best seen in the examples of conspiracies, which both

Machiavelli and Shakespeare especially scrutinize as the greatest threat to the ruler. And aside from all of that, both writers are very humorous and sharp, although not at all radical, which is indicated by their use of irony. Irony in their works can also be interpreted in a strictly political sense when we try to discover their own personal stances. The interpretation of the nature of men lies in the centre of their political stances and here they finally diverge. All of these topics provide us with a framework in which we can interpret the words of one more easily through the words of the other, primarily Machiavelli's through Shakespeare's examples.

## References

- Aristotel (2015). *O pesničkoj umetnosti*. Beograd: Dereta.
- Baumberger, C. (2013). Art and Understanding. In Defence of Aesthetic Cognitivism. In M. Greenlee, R. Hammwöhner, B. Köber, C. Wagner and C. Wolff (eds.), *Bildersehen. Perspektiven der Bildwissenschaft* (Regensburg: Schnell + Steiner), link:  
[https://envphil.ethz.ch/content/dam/ethz/special-interest/usys/ied/environmental-philosophy-group-dam/documents/Art\\_and\\_Understanding\\_Homepage.pdf](https://envphil.ethz.ch/content/dam/ethz/special-interest/usys/ied/environmental-philosophy-group-dam/documents/Art_and_Understanding_Homepage.pdf) (accessed 14 August 2023).
- Carroll, N. (1990). Review. *The Journal of Aesthetics and Art Criticism* 48 (2), 167–169.
- Čavoški, K. (2019). *Politička filozofija u Šekspirovim dramama*. Beograd: Catena Mundi.
- Grady, H. (2000). Shakespeare's Links to Machiavelli and Montaigne: Constructing Intellectual Modernity in Early Modern Europe. *Comparative Literature* 52 (2), 119–142.
- Greenblatt, S. (2016). *Will in the World*. New York & London: W. W. Norton & Company.
- Hunt, L. H. (2006). Motion Picture as a Philosophical Resource. In N. Carroll and J. Choi (eds.), *Philosophy of Film and Motion Pictures: An Anthology* (pp. 397–405). Malden, Oxford & Victoria: Blackwell Publishing Ltd.
- Lamarque, P. (2010). Literature and Truth. In G. L. Hagberg and W. Jost (eds.), *A Companion to Philosophy of Literature* (pp. 367–384). Chichester: Wiley-Blackwell.
- Machiavelli, N. (2014). *The Prince and Other Writings*. San Diego: World Cloud Classics.
- Makijaveli, N. (2020). *O zaverama*. Beograd: Službeni glasnik.
- (2018). *Vladalac*. Beograd: Akia Mali Princ.
- Parel, A. J. (1992). *The Machiavellian Cosmos*. New Haven and London: Yale University Press.

- Rowe, M. W. (2010). Iago's Elenchus: Shakespeare, *Othello*, and the Platonic Inheritance. In G. L. Hagberg and W. Jost (eds.), *A Companion to Philosophy of Literature* (pp. 174–192). Chichester: Wiley-Blackwell.
- Shakespeare, W. (2007). *The Complete Works of William Shakespeare*. Hertfordshire: Wordsworth Editions Limited.
- Shiner, R. A. (2010). Philosophy and Literature: Friends of the Earth?. In G. L. Hagberg and W. Jost (eds.), *A Companion to Philosophy of Literature* (pp. 22–37). Chichester: Wiley-Blackwell.
- Skinner, Q. (1990). Machiavelli's *Discorsi* and the pre-humanist origins of republican ideas. In G. Bock, Q. Skinner and V. Maurizio (eds.), *Machiavelli and Republicanism* (pp. 121–141). Cambridge: Cambridge University Press.
- Šekspir, V. (2016). *Velike tragedije*. Beograd: Vulkan.
- Viroli, M. (1990). Machiavelli and the republican idea of politics. In Gisella Bock, Q. Skinner and V. Maurizio (eds.), *Machiavelli and Republicanism* (pp. 143–171). Cambridge: Cambridge University Press.
- Wilson, C. (1983). Literature and Knowledge. *Philosophy* 58 (226), 489–496.

## 6. ART WORKS AND COLLECTIVE INTENTIONALITY





Milan Popadić

## CAN A MONUMENT BE BAD? NORMATIVITY AND COMMEMORATIVE VALUES IN PUBLIC SPACE

**Abstract:** A monument is usually understood as an entity (sculpture, building, landmark...) erected (or recognized) as a sign of memory of a person or event. This is applicable regardless of the type of monument we are considering (private, public, cultural); what differs is the type of memory (which, in this sense, again can be private, public, or cultural). From that basic division, all other divisions of monuments into different types are derived (for example, by form, by historical period, by social function...). That is why when we talk about monuments, it is always about memory of someone or something. Hence, the basic value attached to monuments is commemorative value. Commemorative value can be understood as the content (memory) that is kept in our minds via the monument. Thus, it is possible to say: if it has a commemorative value, then it is a monument. Or in normative terms, a monument is something that ought to have commemorative value. That seems clear and understandable. However, there are many examples, some very recent, of monuments being destroyed or removed because of their commemorative value. In other words, they were considered unacceptable as public monuments precisely because they met their monument “norms”, namely to commemorate someone or something. In our time, monuments are most often destroyed or removed because they allegedly represented symbols of racism, colonialism, or hegemony. Their commemorative values do not match the current social values of the public. But does this mean that the monuments are morally problematic? Does that make them “bad”? Thus, in this paper we discuss the relationship between different aspects of the value of monuments, their transformations over time and ways to recognize their essential commemorative function in public space.

**Key words:** monuments, commemorative values, public, removal and destruction of monuments.

## Introduction: The rise and fall of monuments

Nineteenth-century artist Horatio Greenough (1802–1852) is occasionally recognized as the first American sculptor (Wright, 1963). During his lifetime, he was best known for two sculptures commissioned by United States government. The first one, *The Rescue* (1837–50), depicts a pioneer family rescued from vicious native American (“indian”) attackers, meant to “commemorate the dangers and difficulty of peopling our continent, and which shall also serve as a memorial of the Indian race” (Boime, 2004, p. 527). The second one was seated figure of George Washington (1840), an American Founding Father, ichnographically modeled in the monumental form of Olympian Zeus.



Fig. 1. Horatio Greenough's *The Rescue* (1837–50) and *George Washington* (1840)

Both sculptures primarily occupied significant spots in Washington D. C.: the first one was located at the entrance to the Capitol building, the second one in its Rotunda. Representative motifs displayed in prominent places made them national monuments, but also made Horatio Greenough famous during his lifetime. But the life of a monument is measured by different standards than the life of a man. In fact, the controversy surrounding the George Washington sculpture began immediately after its erection. The half-naked body of the first American president was not universally approved. It also turned out that the light in the Rotunda was not good for this type of sculpture, so it was moved to the lawn in front of

the Capitol. Nevertheless, dissatisfaction with the appearance of the sculpture continued, so in the beginning of the twentieth century the sculpture was moved to the Smithsonian Institution Building. In the second half of the same century it finally settled in a quiet second floor in The National Museum of American History in Washington D.C. But *The Rescue* had a much more dramatic fate. In 1939, there was discussion in the U.S. House of Representatives on the “joint resolution to remove a monument now standing at the right of the east entrance to the National Capitol, representing the American Indian”. One of the speakers even recommended that *The Rescue* should be “... ground into dust, and scattered to the four winds, that no more remembrance may be perpetuated of our barbaric past, and that it may not be a constant reminder to our American Indian citizens ...” (Fryd, 1987, p. 17). It did not happen, but the sculpture was finally removed in 1958. That was not the end: in 1976, while moving it to a new Smithsonian storage a crane accidentally dropped *The Rescue*, reducing it to several fragments (Fryd, 1987, p. 17). Thus, two (once very respectable) monuments designed by Greenough, the first American sculptor, and commissioned by a patron of the highest status (the United States government), today are out of the focus of the public, placed in the shadows of the museum display or in the deep darkness of storage units. Were they that bad?

## What is a monument?

The illustration we started with is not unknown or rare in the history and culture of monuments. There are countless example of removal or demolition, whether it is a Christian break with pagan heritage, the iconoclasm of the French Revolution, or the reckoning of Eastern Europe with bronze or stone giants from the era of the Soviet Union (Nelson, R. S. & Olin, 2004). In the domain of the contemporary global culture of monuments, the first quarter of the twenty-first century was marked by the specific phenomenon of removing landmarks that after decades, sometimes even centuries, are now recognized as undesirable (Shahvisi, 2021, pp. 453–468). They became “bad”, often due to their racist or colonialist background and thus invite discussions about their *ethics* (Demetriou & Wingo, 2018, pp. 341–355). Sometimes, as in the case of Greenough’s aforementioned sculpture of George Washington, there are also *aesthetic* disagreements (Lehtinen, 2019, pp. 30–38). But are these ethical and aesthetic norms, as important and socially relevant as they may be, actually what defines the vital idea of a monument and its public connotation? Thus, maybe we should start with the first and basic question: what is a monument?

A monument is usually understood as a sculpture, building or landmark *erected* in memory of a person or event. The more considerate ones would agree, but would also add “erected or *preserved* in memory of...”. In the first case, for example, we are talking about monuments like The Lincoln Memorial in Washington, D.C., while in the second case we are talking about the remains of the Berlin Wall (which were not erected as a monument, but were preserved to be one). The first are most often called *intentional*, the second *non-intentional monuments*. However, the most one careful would perhaps replace “erected or *preserved*” with a word that unites them, but also essentially complements them. That word is *recognition* (cf. Riegl, 1903/1982; Young, 2003, pp. 234–247). Because, we can ask ourselves, is it significant (in the context of culture of monuments) that something was erected or preserved if it is not *recognized* as a sign of memory of a person or an event? So, to begin with, here are two key terms: *memory* and *recognition*. We can notice that in contrast to the physical constitution of the monument (most often it is built with the idea that it should last “sub specie aeternitatis”), memory and recognition give the impression of “fragility”. This dualism is the basis of the dynamics of the culture of monuments.

The result of the blend of *recognition* and *memory*, that special capacity that makes a monument a monument, we usually call *commemorativeness*. If we were to play with words, we could translate the term commemorativeness (from Latin, of course) literally as “call to remembrance”. The values derived from that capacity are *commemorative values* (Riegl, 1903/1982; Harrer, 2017, p. 31). In a few quick words, commemorativeness would be the recognized as the content (memory) that is kept in the minds via the physical reality of a monument, while commemorative values represent the *current aspect* of the commemorative property. To summarize: a monument is a physical structure that has the property of commemoration; commemorativeness is a permanent property of a monument; commemorative values during the lifetime of the monument are variable. For example, a memorial dedicated to an army general can at one moment be a monument to a liberator, at another a monument to a conqueror.

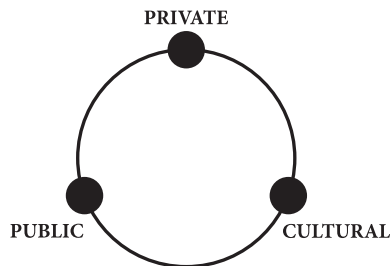


Fig. 2. Three aspects of commemorativeness

Generally speaking, through the blend of recognition and memory, commemorative values can be defined in three ways: private, public and cultural (cf. Dewey, 2005, p. 24; Assmann, 2011, pp. 15–27). It seems that every monument always has these three aspects, and what differs is their mutual relationship (Popadić, 2023, pp. 257–258). For example, the tombstone we erect for a deceased family member is most certainly a private monument, or rather a dominantly private monument. Because, most often (except on very rare exceptional occasions), tombstones are erected in cemeteries, and cemeteries are public spaces and a certain community (family, relatives, friends), therefore part of the public, usually participates in the commemoration. The shape of a tombstone and the material from which it is built are also an expression of the cultural patterns of a certain community. Sometimes they are standardized, “pre-designed”, so that all monuments in the cemetery are uniform. A *public monument* was initially intended directly for the public and was erected in a public space. It can be initiated by a private person, but it is generally raised in the name of a community (local, national, state...) for the purpose of public commemoration. However, in addition to these dominant characteristics, it becomes a part of people’s everyday life or part of their private lives. For example, if we pass a public monument on our way from home to work and back every day, it can hardly be considered a commemorative ritual related to the public function of the monument, and it will undoubtedly create a certain private commemorative content. It will become “our” place and “our” monument, no matter how much it is in public space. Like private monuments, public monuments are always an expression of cultural patterns. Finally, public monuments are often understood immediately after their construction as cultural monuments, although they become such only through the administrative process of establishing cultural values. In other words, a *cultural monument* is a “title” determined by officially authorized acts, which a monument acquires after a formally prescribed procedure. Let us note that here the adjective *cultural* means the confirmation of the institutional-administrative apparatus of a certain community and puts monuments in the corpus of formally identified cultural (or natural) heritage (the decision on the recognition of cultural monuments is made by a representative body depending on the importance of the monument, e. g. an institute for the protection of monuments, a parliament, a government...). Cultural monuments are therefore always public monuments, and in the same way as “ordinary” public monuments, they also have their own private aspect. Perhaps one clarification should be made here: not all monuments are “cultural monuments” in the formal sense, but all monuments are an expression of culture. That is why, whenever we talk about monuments and their commemorativeness, we always talk about the intertwining of those three aspects: private, public and cultural.

## The Good and the Bad: A short comparison of two Belgrade monuments

If we would like to summarize the previously stated in normative terms, we could say *a monument is something that ought to have commemorative value*. That seems clear and understandable. However, there are many examples, some very recent, of monuments being destroyed or removed because of their commemorative value. In other words, they were considered unacceptable as public monuments precisely because they met a “monument norm”, namely to “call to remembrance” someone or something. In our time, monuments are most often destroyed or removed because they are recognized as symbols of racism, colonialism, or hegemony. Their commemorative values do not match the current public values. But does this mean that the monuments are morally problematic? Does that make them “bad”? Are we confusing commemorative value with celebration or glorification? Do we blame monuments for human faults? After all, monuments cannot be “racist” or “colonialist”; people can. Monuments do what they are supposed to do – they commemorate. Paradoxically, aside from the affirmative commemorative rituals, it is by removing monuments that their true (“monumental”) nature is confirmed. This means that they were removed purposely because they were “good” monuments, that is, they performed their commemorative function. If it they did not, it would be only removal of a pile of stones, metal, or concrete, and that would have no public purpose or importance other than clearing the ground. However, blocking that “call to remembrance” it is often justified by ethical or aesthetic reasons.

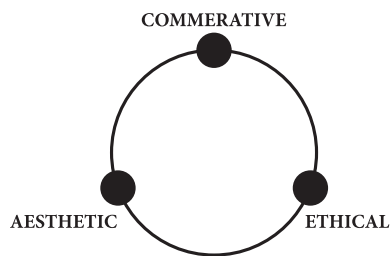


Fig. 3. Three aspects of the value of monuments

We will use two examples in an attempt to explain this combination of commemorative, aesthetic and ethical values. In both cases, we are talking about monuments in Belgrade, the capital of the Republic of Serbia, but with very different current statuses. These two monuments are: *Po-*

*bednik* / ‘The Victor’, casted in 1913 and erected in 1928, and the *monument to Stefan Nemanja* erected in 2021 (basic references about these two monuments: Vučetić-Mladenović, 1999, pp. 110–123; Makuljević, 2022, pp. 212–237). The first one is likely to be the most recognizable monument in the city and often serves as an unofficial symbol of Belgrade. The second, almost a century younger, is a frequent subject of public polemics about the current culture of monuments in Serbia. In other words, today the first (despite its not so glorious beginnings) is often seen as an example of a “good”, the second of a “bad” monument. Well, let us see what this is all about.



Fig. 4. Two Belgrade monuments: ‘The Victor’ and *Monument to Stefan Nemanja*

For almost a century ‘The Victor’ has stood on the plateau of the Upper Town of the Belgrade Fortress. It was designed by the sculptor Ivan Meštrović: a standing bronze male figure in the nude with symbols of peace and war (a falcon in the left hand and a lowered sword in the right), commemorating Serbia’s victory over the Ottoman and Austro-Hungarian empires during the Balkan Wars and the First World War. Architect Petar Bajalović, who designed the pedestal (a transposed version of the Doric column) is also accountable for monument’s adequate spatial arrangement. The total height of the monument is 14 meters, but the impression of significantly larger dimensions was created with the help of its position at the high terrace of the Belgrade fortress (above the confluence of the Sava River and the Danube River) and its visibility. The stylized but understandable expression of the figure, clear commemorative content, spatial accessibility and visibility from afar (which results in an iconic silhouette), make this monument recognizable and effective.

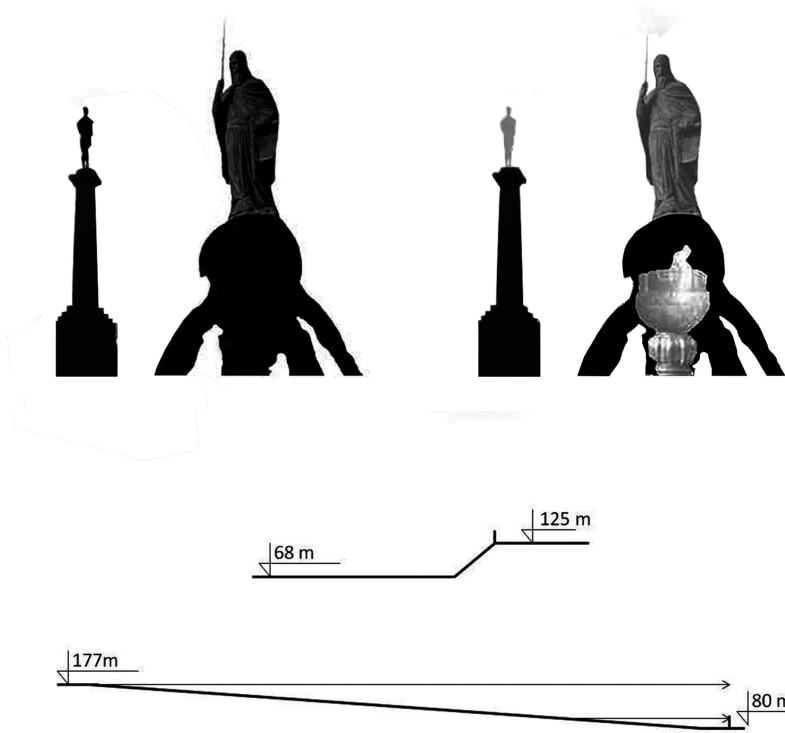


Fig. 5. *'The Victor'* and *Monument to Stefan Nemanja*: comparative analysis of visual and spatial characteristics

The monument to Stefan Nemanja is located on Sava Square, one of the central squares of Belgrade in the immediate vicinity of the Belgrade Waterfront (a project that completely changed this area of the city in a style of brutal and arrogant investor urbanism). The monument is the work of the Russian sculptor Alexander Rukavishnikov and is dedicated to the medieval ruler of Serbia, Stefan Nemanja (c. 1113–1199), the founder of the Nemanjić dynasty (who ruled between 1166 and 1371). The monument has a very complex pedestal: on the scepter of Saint Sava (the first Serbian archbishop and the youngest son of Stefan Nemanja), there is a broken Byzantine helmet; on the inside of the helmet are scenes from the life of Stefan Nemanja. On the pedestal is a figural representation of Stefan Nemanja, who holds a raised sword in his right hand and in the left the Hilandar Charter (the founding charter of the Hilandar monastery, important for the establishment of the Serbian Orthodox Church and the main endowment of Stefan Nemanja and Saint Sava). The total height of the monument is slightly more than 23 meters. The monument was placed at the bottom of the River Sava



slope in a very specific urban environment. In consequence, the monument, which has the height of an eight-story building, remains very difficult to see and it is almost invisible except from the immediate surroundings. Only the side views of the monument can be seen to some extent without obstruction. But we see a slightly bent figure in contrast to the authoritative attitude visible from the close-up foreground. However, from close-up, a spectator is puzzled by the pedestal that visually and narratively competes with the figure instead of highlighting it. Thus, an observer, either from close-up or from afar, gets, to say the least, confused, if not disorientated about the commemorativeness of this landmark.

Compared to the monument to Stefan Nemanja, ‘The Victor’ seems as a textbook example how a monument should look like and how should be placed. Although ten meters smaller (for the height of a two-story building) ‘The Victor’ seems more monumental, while the memorial to Stefan Nemanja, despite its “monumental” dimensions, is suffocated by the dissonance of its own symbolism, by eternal jams of the city traffic and by the conflict of the miscellaneous layers of the urban matrix. But are we overlooking something?

## How to live with monuments?

We have already mentioned that ‘The Victor’ is almost a century older. Before it was erected at the Upper Town of the Belgrade Fortress, the monument had a whole prehistory that caused public controversies. But, leaving that aside, even when it was placed in its current position, it was not without criticism. In 1933, Miodrag Grbić, a prominent inter-war archaeologist and curator of the National Museum in Belgrade, wrote: “[‘The Victor’] is mercilessly swallowed by the Upper Town and it admirably affects only the observer from the side, without connection to the city. This is not about the quality of the monument, but about the quality of the place where it is located. It must retreat from the corner of the city to the place where it will dominate” (Grbić, 1933, p. 286). What is this about? How is it possible that one of the important advantages of the positioning of this monument was not recognized at the time? The answer is more than simple. In the early thirties of the twentieth century, the dynamics of the city of Belgrade were completely different. The left bank of the Sava River was entirely undeveloped. Meanwhile, just a few decades later, new prospects (urbanistic, political and cultural) opened up, primarily with the expansion of New Belgrade (a new administrative center, but also with an immense residential neighborhoods) on the left bank of the

River Sava. Now ‘The Victor’ was seen not only by visitors of the Belgrade Fortress, but by tens of thousands of citizens who every day migrated from the left to the right bank of the River Sava and vice versa. In 1958, ‘The Victor’ even appeared on the first title card of the newly founded TV Belgrade, the first television broadcaster in the country, thus soon to become a trademark seen daily by millions of viewers. It was no longer “in the corner of the city”, but in the center of the urban and media landscape. The aesthetic qualities of the monument, as well as the universally understandable theme of victory, unequivocally contributed to that recognition.

Can the monument to Stefan Nemanja hope for the same fate? It is always difficult to predict and it is also beyond the competence of this paper. What could be said is that in this case *challenges* (a convenient word to avoid the word “problem”) are much higher: the monument to Stefan Nemanja is literally in a *depression* (in terms of urban morphology), and in addition, its iconography requires concentration that goes beyond the average observer (it is necessary to perceive three visual elements – a helmet, a scepter and a human figure – in three different proportional scales at the same time). But even with all those flaws, it is not without commemorative value. It seems that this depressed monument testifies so clearly and directly about our confusion on values of the past in the present times, about the replacement of the idea of historical meaning by mere material grandeur, about the substitution of eloquence by the accumulation of content, about the swap of cultural and national development by political authoritarianism... It may not be the best memorial to Stefan Nemanja, but it seems an appropriate monument for Serbia in the first quarter of the twenty-first century. But, above all, it is there, in the city center and, by all odds, we will have to learn to live with it.

If we have to find a way to live with monuments, how can we reconcile their commemorativeness with the values we consider acceptable today? Here are some examples. In Tübingen, in Germany, there is monument dedicated to Friedrich Silcher (1789–1860), a German composer whose works were infused with local folklore, and who reached the zenith of his creativity precisely in Tübingen, where he died. Celebrating folk musical motifs and composing works inspired by patriotism, he gained the status of a recognizable national bard in the spirit of the romantic culture of the first half of the nineteenth century. Almost a century later, his patriotism and creative expression were recognized by the ruling Nazi ideology as a predecessor of their own principles. Designed in 1939, Silcher’s monument was finally erected in 1941. It is a bulky, almost six-meter stone block, from which emerges the composer’s figure engrossed in work, and a family, a woman, a man and a child – illustrations of motifs from Silcher’s works – but in the recognizable National Socialist iconography.

In the period after the Second World War and the military collapse of the National Socialist ideology, despite the law that ordered the removal of symbols that bore the recognizable characteristics of that period, the monument to Silcher was not removed. Was it the cultivation of the Nazi spirit or Silcher's work? Controversies continued and continue. Thus, at the beginning of 2020, the art collective *Neue Dringlichkeit* and the residents of Tübingen had an action in which they installed an interpretive panel with the aim of changing the dedication of the monument. The new one read: "a monument against the appropriation of art by racist and nationalist forces." In other words, they decided to snatch it from the jaws of destructive ideology and appropriate it (Silcher Monument, [S.d.]).

Here is another example. When in April 2015 Ukraine adopted a package of laws requiring, among other things, the removal of communist monuments, the statue of Lenin in Odessa, located in an old factory yard on the outskirts of the city, avoided such a fate. Local artist Oleksandar Milov "encased" it in a new titanium costume, creating what is said to be the world's first monument to Darth Vader, the fictional hero from the "Star Wars" movies. The posture of the existing feature was fully utilized and Lenin's long coat became Darth Vader's swirling cloak, a lightsaber fit into a clenched fist, and helmet covered the statute's face. The statue of Lenin has no doubt been removed from view, but it is also preserved under the costume of the anti-hero of global popular culture (Macdonald, 2015).



Fig. 6. Monument to Friedrich Silcher (Tübingen, Germany), Monument to Darth Vader (Odessa, Ukraine) and Monument to Matija Gubec (Krško, Slovenia)

In 1973, during the socialist period of former Yugoslavia, in Krško, Slovenia, a monument was erected to Matija Gubec, a well-known leader of the Croatian–Slovene Peasant Revolt of 1573. In the modern era, Gubec was celebrated as a national hero, but in the socialist period he would also be a symbol of all the oppressed classes through history that finally,

after the communist revolution and with the help of the inevitable logic of dialectical materialism, freed themselves. Nevertheless, the end of the socialist era and the new age of global capitalism brought oblivion to former heroes, as well as to oppressed classes. But citizens of Krško did not give up their monument. The monument was recognized as “theirs” by the supporters of the local football club. Often, before a match, the figure (significantly larger than life size) is dressed in the jersey of local team or the Slovenian national team, depending on who is playing. The monument and the square become the gathering place, a symbolic rallying point from where fans go to the stadium. The monument again became a leader, but this time a leader of football supporters (Posavski obzornik, 2010).

## Conclusion: A circle of commemorativeness

If we delve a little deeper into the previous examples, we can notice that they are intertwined with the aspects of the appearance of monuments that we talked about in the first part of the paper: private, public, cultural, aesthetic, ethical and of course commemorative. All these aspects are part of one normative chain, which in the context of culture of monuments we can call *a circle of commemorativeness*.

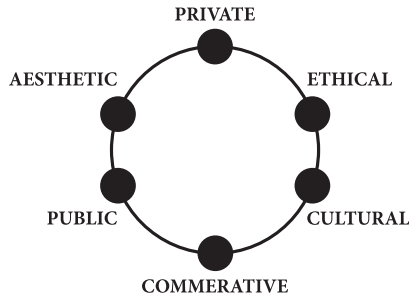


Fig. 7. A circle of commemorativeness

There is no doubt that monuments are created to manipulate memory. Although the word manipulation means “to treat or operate with or as if with the hands” (for example, I’m manipulating the computer as I write this), much more often we use the word in a connotation that assumes certain “ulterior motives”, like abuse, deception and similar shenanigans. It is the same with monuments. By erecting a monument, someone use their position of power to impose certain representations of the past on the public. But, as we said at the beginning, commemorativeness is conditioned by recognition. *Recognition is the moment of connecting the monu-*

*ment with our view of the world.* And our view of the world may or may not coincide with the intention of the one who erected the monument. This is what sets the circle of commemorativeness in motion. It is the wheel that has upturned the fate of the Horatio Greenough's monuments (which we met at the very beginning) and many others.

When we talk about the merging of the mentioned aspects of commemorativeness (private, public, cultural, aesthetic, ethical), it means that we accept monuments in all their complexity, as an expression of high (representative) culture, but also as a phenomenon present in everyday life (one that is unburdened with ideas of "eternity" and takes care of immediate needs). The "call to remembrance" sent to us by the monuments cannot be authoritatively normed (even though it seems so at the moment of erection) and is necessarily received in the context of current public norms (Petovar, 2022, pp. 58–59). Although it looks like a symbol of the power of the one who raises the monument, the monument is actually an expression of the public's power to deal with wanted and unwanted memories. As an experienced connoisseur of monuments, the Croatian art historian, Ivo Babić says: "In the very concept of a monument, regardless of its debatable and diffuse scope, the concept of duration, warning, will and speech that will not be silenced is embedded; the concept of survival, personal, group, human in general. A monument opposes entropy, defies time. Its essence is transcendent, if not transcendental. By questioning the essence of a monument, we also question the essence of a human being" (Babić, 1988, p. 707). In other words, ("essentially") whenever we talk about monuments, we are actually talking about human beings.

Thus, understanding the circle of commemorativeness can help us to learn to live with monuments we "disagree" with, not because they glorify unwanted ideas, but because they remind us of our own weaknesses and delusions. By fulfilling their function and thus reminding us of the dark side of human nature, monuments (which we imprudently accuse of praising unacceptable values) actually allow us a better understanding of the human condition. Hence, in the end, we have to ask ourselves: do "bad" monuments make "better" people? If so, how can these monuments be "bad"?

## References

- Assmann, J. (2011). Communicative and Cultural Memory. In P. Meusburger, M. Heffernan, E. Wunder (eds.), *Cultural Memories. The Geographical Point of View* (pp. 15–27). Dordrecht: Springer.
- Babić I. (1988). Uvod: Za jedno antropološko shvaćanje spomenika [Introduction: For an anthropological understanding of monuments]. *Pogledi*, 18 (3–4), 703–708.

- Boime, A. (2004), *A Social History of Modern Art, Volume 2: Art in an Age of Counterrevolution, 1815–1848*. Chicago: University of Chicago Press.
- Demetriou, D. & Wingo, A. (2018). The ethics of racist monuments. In D. Boonin (ed.), *The Palgrave Handbook of Philosophy and Public Policy* (pp. 341–355). London: Palgrave Macmillan.
- Dewey, J. (2005). *Art as experience*. New York: Perigee Books.
- Fryd, V. G. (1987). Two Sculptures for the Capitol: Horatio Greenough's 'Rescue' and Luigi Persico's 'Discovery of America'. *American Art Journal*, 19 (2), 16–39.
- Grbić, M. (1933). Srpska akropola [Serbian acropolis]. *Beogradske opštinske novine*, 4, 284–286.
- Harrer, A. (2017). *The Legacy of Alois Riegl: Material Authenticity of the Monument in the Digital Age*. *Built Heritage*, 1, 29–40.
- Lehtinen, S. (2019). New Public Monuments: Urban Art and Everyday Aesthetic Experience. *Open Philosophy*, 2(1), 30–38.
- Macdonald, f. (2015). *The man who turned Lenin into Darth Vader*, BBC, 23rd October 2015, <https://www.bbc.com/culture/article/20151023-the-man-who-turned-lenin-into-darth-vader>.
- Makuljević, N. (2022). *Memorija i manipulacija: spomenička politika u Srbiji 1989–2021* [Memory and manipulation: monument politics in Serbia 1989–2021]. Beograd: Biblioteka XX vek, 2022.
- Nelson, R. S. & Olin. M. (eds). (2004). *Monuments and Memory, Made and Unmade*. Chicago: University of Chicago Press.
- Petovar, K. (2022). Javno dobro i gradski prostor (Public good and urban space). In D. Milovanović Rodić, Lj. Slavković, M. Maruna (eds.), *U potrazi za javnim interesom: dometi urbanizma* (pp. 41–61). Beograd: Arhitektonski fakultet.
- Popadić, M. (2023). Šta čini dobars pomenik? Teorijska polazišta za heritološko tumačenje komemorativnih vrednosti u prostoru grada [What Makes a Good Monument? Theoretical Starting Points to the Heritological Interpretation of Commemorative Values in the City Space]. *Zbornik radova Filozofskog fakulteta u Prištini*, 53(1), 253–273.
- Posavski obzornik, (2010). *Tudi Matija Gubecnavija za Slovenijo...*[ Matija Gubec also supports Slovenia], 30.11.2010. <https://www.posavskiobzornik.si/panorama/matija-gubec-spet-v-nogometnem-dresu>
- Riegl, A. (1903). *Moderne Denkmalkultus: Sein Wesen und seine Entstehung*. Wien: K.K. Zentral-Kommision für Kunst- und historischeDenkmale.
- Riegl, A. (1982). The Modern Cult of Monuments: Its Character and Its Origin. *Oppositions*, 25, 21–51.
- Shahvisi, A. (2021). Colonial Monuments as Slurring Speech Acts. *Journal of Philosophy of Education*, 55/3, 453–468.
- Silcher Monument, Tübingen, Germany. [S.d.]. <https://www.atlasobscura.com/places/silcher-monument>

- Vučetić-Mladenović, R. (1999). Pobeđeni Pobjednik. Polemike uoči postavljanja Meštrovićevog spomenika [Defeated “Victor” (Polemic on the eve of the erection of Mestrovic’s monument)]. *Godišnjak za društvenu istoriju*, VI (2), 110–123.
- Wright, N. (1963). *Horatio Greenough: the first American sculptor*. Philadelphia: University of Pennsylvania Press.
- Young, J. E. (2003). Memory/Monument. In R. S. Nelson and R. Shiff (eds.), *Critical Terms for Art History* (pp. 234–347). Chicago and London: The University of Chicago Press.

## Table of figures

- Fig. 1. Horatio Greenough’s *The Rescue* (1837–50) and *George Washington* (1840)
- Fig. 2. Three aspects of commemorativeness
- Fig. 3. Three aspects of the value of monuments
- Fig. 4. Two Belgrade monuments: ‘*The Victor*’ and *Monument to Stefan Nemanja*
- Fig. 5. ‘*The Victor*’ and *Monument to Stefan Nemanja*: comparative analysis of visual and spatial characteristics
- Fig. 6. *Monument to Friedrich Silcher* (Tübingen, Germany), *Monument to Darth Vader* (Odessa, Ukraine) and *Monument to Matija Gubec* (Krško, Slovenia)
- Fig. 7. A circle of commemorativeness





Ivan Popov

## WHEN IS ART INTERACTIVE?

**Abstract:** The contemporary Bulgarian art scene is dominated by the parlance of artistic interactivity, whose main goal is seen in overcoming and replacing traditional modes of dealing with art, which consist mainly in its passive (visual) contemplation. The paper combines an analysis of the term “interactive art” with a critical reconstruction of the opposition between the eye and the body, which seems to rest on an incorrect understanding of the historical tradition of Western art as well as on a dubious transfer of a dispute from philosophical metaphysics into the domain of art theory. The conclusion is that in the case of traditional art, in the sense in which contemporary Bulgarian authors understand the latter, there has never been a principled impossibility for the audience to transcend the passive role of the observer. Moreover, interactivity has to be defined much narrower in order to make sense as an autonomous artistic category. The paper ends with a reflection on how history of ideas and philosophy can complement each other when conceptual issues of the kind exemplified by the speaking of “interactive art” are at stake.

**Keywords:** interactive art, art for contemplation, Western European artistic tradition.

In this paper, I would like to take a position that is situated between the disciplines of the history of ideas and philosophy. I hereby pursue the following goals:

1. To present certain trends that have crystallized in recent years both in Bulgarian art and in the debates that are taking place in Bulgaria regarding the place, role and functions of contemporary art. To summarize, one main interest of the artists and the audience is oriented toward a greater role of the interactivity between audience and artwork.

2. To critically reconsider some basic theoretical premises related to the term “interactive art”, which in recent years has become established in Bulgaria and is being instrumentalized among other things for the purpose of conceptualizing contemporary aesthetic and artistic trends.

3. To try to draw some generalizations about the way we should think about the interaction between artistic practices and the accompanying theoretical reflection; the two are not the same and should not be identified accordingly.

In the contemporary debates on the question of the direction in which Bulgarian art should develop, so that it exceeds the limitations of the national tradition, a major role is played by the opposition between art intended for visual perception and, accordingly, for contemplation, and art where the author and/or the viewer's body plays a major role, being drawn into various actions, thus erasing the boundary between a work of art and its audience. An example can be given with the theoretical work of the contemporary Bulgarian artist Venelin Shurelov (Венелин Шурелов), who discusses the need for art to focus on problems and issues that are relevant to the problems of contemporary society (Shurelov, 2020, p. 8). Shurelov of course does not speak as a philosopher, but as an artist conceptualizing his own artistic activity. Nevertheless his example is significant, given that in the quoted text the so-called interactive art is contrasted with the traditional one, in which the artwork functions as an object and is even fetishized (*ibid.*). I believe that the quoted article expresses a way of thinking that has informally dominated the contemporary Bulgarian debate on the visual arts in the recent years.

In this case we are talking about the processes that take place in a particular artistic tradition, but I think that careful analysis is able to reveal something about the proper ways of conceptualizing emerging art forms *per se*. So I will critically trace the genesis of said opposition and try to answer the question of how adequate it is and how much the term "interactive art" itself, understood in the given specific way, is able to become an artistic category, with the help of which we can give thought to various phenomena and trends from the contemporary world of art.

From the point of view of the history of ideas, it is obvious to me that the contrast between looking (and thus the eye, i.e. visual perception) and the body, in its capacity as a direct participant in the processes of the world, sounds very close to a common conception of the history of Western philosophy that contrasts Martin Heidegger with the intellectual tradition that preceded him. To put it in the most simplistic terms, the thesis states that traditional Western, Cartesian metaphysics postulates the existence of two substances – matter and soul – and therefore believes in the objective and ahistorical knowability of the world by the sciences. In contrast, Heidegger's critique of metaphysics overcomes the subject-object divide, declaring untenable the claim that the self constitutes a kind of

tabula rasa, which discovers the external world only in its encounter with it. Subject and world are intrinsically linked; any attempt to take an absolute viewpoint “from nowhere” is doomed to failure.

The details of Heidegger’s critique of Cartesian dualism are well known – a useful reconstruction of it is to be found on the Stanford Encyclopedia of Philosophy website (Wheeler, 2020), where, among other things, a special article dedicated to the specifics of Heidegger’s aesthetics is available (Thomson, 2019). At this point, I am unable to go into detail about the German philosopher’s conception of art and his claim that artworks reflect on the cultural and axiological “picture” of a given social community at a particular moment in history, which is at least arguable as to its ability to capture some basic intuitions of the interested artworld public. In any case, Heidegger makes use of the opposition between the work of art, which we perceive as an entity separate from other social and cultural practices, converting it thus into an object of aesthetic perception, and true art (ibid., see Section 2.2. “Heidegger’s Critique of the Aesthetic Approach”). I do not claim that every contemporary Bulgarian author and critic does refer explicitly to these specific theories, but I am of the opinion that attempts to break with a – real or imaginary – tradition of alienating the viewer from the artwork should be thought of in the light of the various intellectual fashions that have dominated our artistic and academic community since 1990. A major part in the course of this process has been played by the reception of Martin Heidegger’s philosophical project of overturning the subject-object dualism.

As a next step I will go straight to some critical remarks that are relevant when attempting to transfer said philosophical dispute to the realm of art. First, here we are dealing with a discussion that, as is usually the case in philosophy, is far from reaching its final resolution (which perhaps will never happen anyway). At the risk of further oversimplifying the exposition, I will say that in today’s philosophy of mind the so-called dualism remains one of the possible options with respect to the mind-body-problem and it has its supporters. The German philosopher Holm Tetens, for example, discussing the peculiarities of philosophical argumentation in his book of the same name, ranks linguistic dualism among several other options for the treatment of the philosophical problem in question – eliminative materialism, functionalism, etc. (Tetens, 2004, p. 274). The conclusion is that unlike the development of the other sciences, where the new encompasses but also elaborates on what has been reached so far for the understanding of a given slice of reality, in the case of philosophical metaphysics it cannot be said that a given account can undo what another aims to achieve. For better or worse, philosophical knowledge progresses in a completely differ-

ent way than scientific knowledge: over time, different positions and theories become more and more precise, able to more fully “map” the conceptual interdependencies that make up our thinking. However, the fact that theoretical reflection always remains underdetermined by empirical data means that it is difficult to arrive at a unified and definitive philosophical vision of the world. Like advances in chess, knowledge of the different types of positions in the game is increasingly refined, with no universal algorithm devised that leads to victory every time. This issue is very convincingly dealt with in Pigliucci (2016), where the specific differences between natural science and philosophy are elaborated in detail.

To repeat, for me the important question is what justifies the transfer of philosophical problems into the sphere of art? If we assume that in the place of “art for contemplation” should come “interactive art”, then the analogy with philosophical debate is fundamentally inadequate, since there, as we have seen, progress does not consist in replacing one system unconditionally with another.

Second, it is fundamentally questionable to what extent looking, i.e. the visual perception of fine art, not only has its origins in, but in some way reflects the claims of Cartesian metaphysics. As is very well known, the viewer’s gaze within the art world is never innocent, and this is probably even true of those examples of “first” art that have played such an important role in a number of theoretical debates in the second half of the twentieth century, such as the one about the definition of the term “art”. We do not need to resort to elaborate theories to convince ourselves of the truth of the claim that, as a cultural practice, visual arts emerged long before Descartes and his metaphysical convictions. Even if we assume that the French philosopher was fatally mistaken in his aesthetic views (though I am not aware that he has ever spoken on this point), this would tell us something only about the quality of his theory and is far from proving that he influenced thinking about art in Europe over the last four centuries. In principle, it does not sound very convincing to assume that Descartes has almost single-handedly somehow succeeded in steering the development of art in a direction that should be overcome today. Neither a theoretical analysis of the ways in which we learn and use the concept of art nor empirical data are compatible with generalizations of this kind.

A further major problem in this case is the fact that the visual perception of art is only thought of as a cultural phenomenon, not a biological one. In contemporary philosophy, there are numerous attempts to apply the knowledge and discoveries of the natural sciences, especially cognitive science, to the arts. The processing of visual information by the eye and the brain, respectively, is being studied, and empirically detectable reac-

tions of the human organism in its interaction with art are being sought. These problems have been presented recently in a concise and accessible way in the section on “art and science” in Noel Carroll’s and Jonathan Gilmore’s volume on the philosophy of fine art and sculpture (see quotation below), where light is shed on various debates that deal with the relationship between the visual arts and the findings of neuroscience. Attempts to reduce art, and therefore culture, to processes occurring at a biological level in the human brain can certainly be seriously criticized – something similar is done, for example, by David Davies who, when discussing the relationship between empirical data and philosophical reflection, claims, that “[...] the principal philosophical task is to arrive at a codification of the practices thereby evidenced” (Davies, 2018, p. 74). I agree that the goals of philosophical analysis are fundamentally different from those of empirical science, but this is not the important point here. I would like to emphasize that the reasons for conceptualizing visual perception exclusively in terms of its real – or imagined – susceptibility to ideological constructs are far from clear. Such a perspective is highly simplistic and one-sided; the natural sciences offer a very different interpretation of the phenomenon, with which, regardless of any stipulations and remarks, modern theory should comply.

Adopting a more general perspective, it is necessary to ask whether, in the history of art, one can speak at all of the abolition of the old by the new. Is it not more adequate to assume that as a result of extremely complex processes of a cultural, aesthetic and intellectual nature, the historical development here is expressed in the displacement/marginalization of some artistic and aesthetic norms and expectations (“paradigms”) by others? As in philosophy, so in art, there are developments and accumulations: it is obvious that we are not in a position to “rewind the tape” at will, resurrecting in all its peculiarities, for example, the age of Mannerism. Ultimately, however, nowadays happenings and paintings exist as co-equal artistic phenomena, rather than being situated in a diachronic, much less teleologically organized, line of historical progress. Different artistic forms are no doubt capable of realizing a variety of purposes, but experience shows that expanding the repertoire of possibilities for relating to other spheres of life does not put an end to the tradition that has already been formed and continues to exist.

From the point of view not only of the direct participants in the processes in the modern world of art, but also of the interested (i.e. professional) public, it is necessary and also valuable to have both art for viewing and art in which the viewer – in whatever form – can participate. For quite a few representatives of contemporary Bulgarian art, whether they are artists,

critics or theorists, the latter is more important than the former (see Shurelov's article quoted above). In agreement with what I have just said, however, I am of the opinion that the real progress lies not in moving entirely into the paradigm of interactive art, but in enriching the existing aesthetic landscape with it. According to Jerrold Levinson's article (Levinson, 2017) it is exactly at this point that the positions of the artist and the "aesthete" differ in a fundamental way. Levinson's analysis reminds us of the important fact that the various inhabitants of the art world pursue their own interests, which do not always overlap (*ibid.*, p. 483) – an interested observer may welcome a diversification of the artistic forms in a given cultural context, ignoring the practicing artist's perspective, which usually tends to remain exclusive and shaped by a "tunnel vision" (*ibid.*, p. 482).

The sociological observations made so far should be supplemented by a conceptual analysis, to which I intend to devote myself in the remaining part of this text. As is well known from the work of Dominic McIver Lopes, the very term "interactive art" is quite problematic. Taken too broadly, interactivity loses its sharpness and ceases to do philosophical work (Lopes, 2001, p. 67), because ultimately every engagement with a work of art is in some way interactive – the reading and interpretation of a novel, for example, in which the reader's body obviously plays a very minimal role, requires a cognitive effort, the activation of the imagination and so forth, so here we are already justified to talk about a manifestation of activity on the part of the audience. Lopes therefore proposes a much narrower definition of the term (Lopes, 2010, p. 37), generally defining as interactive the audience-induced change in the structure of the work's so-called vehicle (be it a narrative, a visual image, a musical structure, etc.). Further distinctions follow, e.g. between "strong" and "weak" interactivity, the differences between which are explained by using analogies drawn not least from the field of computer (role-playing) games. A similar account of interactive art is offered by Shelby Moser in the volume by Carroll/Gilmore already mentioned. In contrast to Lopes, she uses the terms "digital" and "interactive" synonymously (Moser, 2023, p. 52), directing her efforts again towards the formulation of criteria with the help of which artistic interactivity can become an operative category which could be implemented in our commerce with art. Without going into details, I would like to stress that both authors speak of interactivity in terms of the modification of the medium of the artwork, but not the effect this process has on the physical state of the viewer (i.e. the interactor). Provoking a bodily reaction on the side of the audience can certainly be an element of the overall conception of the particular work, but obviously does play a minor role in the philosophical analysis of the notion of artistic interactivity.

For my purposes, it is important to reiterate that I do not deny the existence of art in which the audience is somehow involved and invited to participate directly, nor that I doubt the value and meaning of such art. On the contrary, it may very well turn out that an artwork, which the viewer does not just look at (that is, contemplate it, as René Descartes probably would have done...), but does something, plays a very important role in making sense of a number of cultural, political and social processes taking place in the contemporary world. But the important question from a philosophical point of view is whether interactive art understood in this specific sense can be conceptualized as a separate category.

Let us return to the notion of the “body”, which obviously plays a major role in the postmodern thinking, being opposed to the non-participating eye of the adherent of traditional metaphysics. In the analysis of the term “interactive”, as it is used in contemporary Bulgarian discourse, it remains fundamentally unclear whether the body – be it the body of the author or that of the spectator – is thought of as the subject of the respective work or as a factor in the appreciation of the latter. We would hardly find anyone who would dispute the claim that there are masses of varieties of art forms and practices, both pre and post Descartes, in which the spectator becomes a participant on a corporeal level. Dance is just one example, but the same is true in cases where the artwork is part of a ritual – e.g. the worship of icons in an Orthodox church, processions with icons, etc. The active role of the audience is a necessary element of *some* artistic traditions, but not of art in general. Bulgarian contemporary art, for example, discovered the body in various manifestations since the early 1990s, in which a major role has been played precisely by the questioning of the norms and conventions of the socialist period, in which art was understood as the traditional forms of painting, sculpture, etc. Of course, such trends and their significance cannot be explored without taking into account the historical and cultural context of their emergence and development, and here one cannot help but acknowledge that the socialization of the body for aesthetic purposes is an important moment in the course of the synchronization of Bulgarian art with the cosmopolitan tendencies (see Zankov 2019). But such observations are of a historical/empirical nature and do not help us much in our search for the conditions for the adequate use of the term “interactive art”.

Either way, if we shift the historical gaze from these particular examples towards those artistic periods and phenomena that do not belong to the context of “high” art, we find that there has never been a monolithic tradition requiring the viewer to passively contemplate art. Moreover, it is obviously an incorrect historical generalization that only modern art

forms have thought to make the spectator a participant. We arrive at the problem that has preoccupied more than a few philosophers of art from the second half of the last century onwards, which consists in the observation that neither conceptually nor in purely empirical terms should the world of art be thought of as obeying universal rules that apply once and for all and everywhere.

Finally I will present one more remark, which may sound trivial, but seems to me quite important in this case. If we proceed from the assumption that there are interactive tendencies in contemporary art, taking Heidegger's critique of Western metaphysics, then one of the conditions for defining the phenomenon as an autonomous artistic and aesthetic category is postulating in the audience an explicit awareness of the philosophical debate in question and of its particularities. It seems to me, however, that in actual practice one is unlikely to find many spectators/participants who attend a performance or happening with the thought that they are having a metaphysical dispute with the Cartesian tradition, and that through their actions they are presenting first-person counterarguments against this specific dualist ontology. At best, these issues are the domain of a very small circle of authors and viewers, and it is generally debatable whether the latter have the necessary professional knowledge to be able to go beyond the parameters of the specific work and address a range of objections and criticisms of the kind I present in the current paper.

In conclusion, it can be said that the reconstruction of the ideological genesis of the talk about interactive art, taking place in the contemporary Bulgarian intellectual and cultural context, is a different task compared to the attempts to define it as an appreciative kind (again in the terminology of Dominic Lopes, see Lopes, 2010, pp. 17–18). The encounter of peculiar difficulties in the course of this endeavor, of course, in no way means that the interest of contemporary Bulgarian artists in the body cannot and should not become an object of research interest. However, the latter should maintain awareness of its own historical and sociological character. Following the distinction made by the British philosopher Derek Matravers between “contextual” and “acontextual” questions that can be asked regarding art understood as a social practice (Matravers, 2014, pp. 4–5), the research in this case should focus on the particularities and dynamics of the specific context, without pretense of arriving at statements about art in general.

Thus, in the end, we come to the fundamental question, whether and to what extent the introduction of new artistic and/or aesthetic terms only reflects, or also influences the processes actually taking place in the world of art. There is an obvious difference between the identification of certain artistic phenomena, which, as we have seen, is often served by insuffi-



ciently precise formulations, and the systematic theorizing on the latter. I hope that with my paper I have been able to shed more light on the way in which, at least in my view, the history of ideas and philosophical analysis should complement each other when the here and now of the art world is under scrutiny.

## References

- Davies, D. (2018). 'This is Your Brain on Art': What Can Philosophy of Art Learn from Neuroscience?. In G. Currie, M. Kieran, A. Meskin, and J. Robson (eds.), *Aesthetics & the Sciences of Mind* (pp. 57–75). Oxford: Oxford University Press.
- Levinson, J. (2017). Artist and Aesthete: A Dual Portrait. *The Journal of Aesthetics and Art Criticism*, 75:4, 479–487.
- Lopes, D. (2001). The Ontology of Interactive Art. *The Journal of Aesthetic Education*, 35:4, 65–81.
- Lopes, D. (2010). *A Philosophy of Computer Art*. New York: Routledge.
- Matravers, D. (2014). *Introducing Philosophy of Art in Eight Case Studies*. New York: Routledge.
- Moser, Sh. (2023). On Regarding Digital Art. In N. Carroll and J. Gilmore (eds.) *The Routledge Companion to the Philosophies of Painting and Sculpture* (pp. 49–59). New York: Taylor & Francis.
- Pigliucci, M. (2016). *The Nature of Philosophy. How Philosophy Makes Progress and Why it Matters*. Stoa Nova Publications.
- Shurelov, V. (2020). Deytafitsirana interaktsia (Dataficated Interaction). *Izkustvo i kritika*, I, 7–14.
- Tetens, H. (2004). *Philosophisches Argumentieren. Eine Einführung*. München: C. H. Beck.
- Thomson, I. (2019). Heidegger's Aesthetics. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), URL = <https://plato.stanford.edu/archives/fall2019/entries/heidegger-aesthetics/> (Accessed 15. 12. 2023).
- Wheeler, M. (2020). Martin Heidegger. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), URL = <https://plato.stanford.edu/archives/fall2020/entries/heidegger/> (Accessed 15.12.2023).
- Zankov, V. (2019). Tyaloto na hudozhnika v izkustvoto na prehoda. URL: <https://openartfiles.bg/bg/topics/3084-the-body-of-the-artist-in-the-art-of-the-transition-period> (Accessed 15. 12. 2023).
- Занков, В. (2019). Тялото на художника в изкуството на прехода. URL: <https://openartfiles.bg/bg/topics/3084-the-body-of-the-artist-in-the-art-of-the-transition-period> (Accessed 15. 12. 2023).
- Шурелов, В. (2020). Дейтафицирана интеракция (Dataficated Interaction). *Изкуство и критика*, I, 7–14.

CIP – Каталогизација у публикацији –  
Народна библиотека Србије, Београд

1(082)

7.01(082)

BALKAN Analytic Forum (1 ; 2023 ; Beograd)

Normativity & normativity of art : Conference proceedings /  
Balkan Analytic Forum, 19–29. X 2023. Belgrade, Serbia ; edited by  
Miroslava Trajkovski, Emily C. McWilliams. – Belgrade : University,  
Faculty of Philosophy : Center for Contemporary Philosophy –  
Balkan Analytic Forum, 2024 (Belgrade : Službeni glasnik). – 255 str.  
: ilustr. ; 24 cm


Tiraž 100. – Napomene i bibliografske reference uz apstrakte. –  
Bibliografija uz svaki rad

ISBN 978-86-6427-286-5

а) Филозофија — Зборници

б) Филозофија уметности — Зборници

COBISS.SR-ID 148359177



The Balkan Analytic Forum aims to bring together experts in analytical philosophy from the Balkans to exchange their ideas, but it is also open to approaches that establish connections between analytical philosophy and other philosophical traditions, as well as to interested experts from other parts of the world. The activity of the forum is to establish, through conferences and accompanying publications, a platform for discussion where scientists from the Balkans, and all those interested in analytical philosophy, can meet regularly and present the texts they are currently working on and their new publications. The mission is to carry out basic, applied and developmental research in the domain of analytical philosophy; to publish the results of this scientific research and professional work; to include young researchers and doctoral students at the Faculty in the implementation in this through participation in the programs implemented by the Center for contemporary philosophy – Balkan Analytic Forum of the University of Belgrade, Faculty of Philosophy (<https://www.f.bg.ac.rs/instituti/baf>); to participate in the organization of gatherings, symposiums, professional meetings and workshops for the purpose of training researchers in the field of analytical philosophy; as well as to cooperate with other institutions in the country and abroad, especially with countries from the Balkans.

Miroslava Trajkovski



The publication “Normativity and Normativity of Art” originated from the inaugural Balkan Analytic Forum Conference, organized by the University of Belgrade with a focus on analytical perspectives in philosophy. This comprehensive work delves deeply into the concept of normativity, encompassing epistemological, logical, and aesthetic dimensions. The contributions within this publication present a remarkable demonstration of erudition, precise reasoning, and adept composition, providing enlightening perspectives on the multifaceted nature of normativity. I highly endorse this scholarly work for its exceptional quality and the valuable contributions of esteemed scholars, predominantly from the Balkan region, yet not exclusively so.

Irina Deretić, *University of Belgrade, Serbia*

This volume is an important contribution to analytic philosophy. It features articles on a wide range of issues that can be broadly construed as pertaining to the topic of normativity. That’s a key strength of the book. A variety of positions are represented, and no single view within any single sub-specialization of normativity is given priority. Scholars working on any of the following issues could benefit from consulting this volume: norms of philosophical explanations, norms of belief for individuals as well as for collectives or groups, norms of inference, aesthetic values, and the normative value of works of art.

Amber Riaz, *Lahore University of Management Sciences, Pakistan*

The key value of this volume (as well as the Balkan Analytic Forum Conference, from which the volume originated) is that it epitomizes what is distinctive of all philosophical disciplines: a normative aspect. The volume gives an excellent overview of how the notion of normativity features in philosophical explanation, belief/judgment (including ethical and aesthetical judgments), logical reasoning, meaning, and intentionality.

Any reader of the volume will gain a good roadmap of the current issues in the discourse on normativity across the philosophical spectrum: metaphilosophy, epistemology, ethics, aesthetics, philosophy of social science/intentionality, and logic.

Ivana Simić, *University of Florida, USA*

