

Forum Philosophicum

Volume 30 number 2 Fall 2025

FORUM PHILOSOPHICUM is a scholarly journal dedicated to philosophical inquiries into various respects of the relationship between philosophy and faith. It is published bi-annually in English by Ignatianum University Press, part of the Jesuit University Ignatianum in Krakow. The journal exists in both paper and electronic versions. The online version is available via the EBSCO Academic Search Complete electronic database (since vol. 6, 2001) and the Philosophy Documentation Center online subscription service (since vol. 1, 1996). Since vol. 23, 2018, all articles are published in full open-access model based on Creative Commons Attribution (BY) license. Selected earlier articles are published in open access on the journal's website.

EDITORIAL BOARD

Jakub Pruś, *Editor-in-Chief*

University Ignatianum in Cracow, Poland

Szczepan Urbaniak SJ, *Deputy Editor*

University Ignatianum in Cracow, Poland

Maciej Jemioł, *Editorial Secretary*

University Ignatianum in Cracow, Poland

Férdia Stone-Davis, *International Assistant Editor*

Margaret Beaufort Institute of Theology, Cambridge, UK

Daniel Spencer, *International Assistant Editor*

University of St Andrews, UK

Magdalena Jankosz, *Language Editor*

Pontifical University of John Paul II, Poland

Rev. Mark Sultana, *International Assistant Editor*

University of Malta, Faculty of Theology, Malta

Michał Zalewski SJ, *Associate Editor for Reviews*

University Ignatianum in Cracow, Poland

CIRCULATION 25 copies

COVER DESIGN Jacek Zaryczny

ISSN (PAPER) 1426-1898

ISSN (ONLINE) 2353-7043

www <https://forumphilosophicum.ignatianum.edu.pl>

TYPESET by Jacek Zaryczny with Libertinus Serif open font

© ⓘ *FORUM PHILOSOPHICUM* is published by Ignatianum University Press under a Creative Commons Attribution 4.0 International License.

MORE INFORMATION ABOUT *FORUM PHILOSOPHICUM*—its mission, contact data, Boards, submissions to the journal, editorial policies, subscriptions, online indexes, may be found in the Note about *Forum Philosophicum* on page 319.

Forum Philosophicum

International Journal for Philosophy

Volume 30 number 2 Fall 2025

JACEK POZNAŃSKI, SZCZEPAN URBANIAK

Editors' Note

5

ARTICLES

FINLEY LAWSON

A Naturalist Theology. Christianity within a Holistic Paradigm

9

CHARLES TALIAFERRO

The Argument from Reason Revisited

37

ROBERT KUBLIKOWSKI

What Is Distinctive About Human Intelligence in the Context of Artificial Intelligence? A Philosophical Approach With Reference to Robert B. Brandom's Semantic Inferentialism

49

WARD BLONDÉ

God as Absolute Machine: Aligning Modern Formalisms to Prove God

65

JOSHUA SIJUWADE

The Nature of Monotheism. A Philosophical Explication

87

TYMOTEUSZ MIETELSKI

The Relationship of Italian Neo-Scholasticism and Phenomenology to Naturalistic Anthropology. An Exploration via the Views of Sofia Vanni Rovighi

117

RYSZARD MORDARSKI

Some Difficulties of Theology Developed in the Context of Science. A Critique of the Position of Grygiel and Wąsek

131

ARTICLES ON OTHER SUBJECTS

PIOTR SZALEK

Wittgenstein, Relativism, and the Second-Person Perspective

147

MITCHELL T. WELLE, MARCIN KOSZOWY

"Fake News" in Reformulated Messages. Towards Expanding the Toolset for Identifying Misinformation

165

MARIUSZ WOJEWODA

The Ethics of Responsibility in the Context of the Use of Intelligent Machines and the Problem of the Technosystem

199

PIOTR SIKORA

Mindful Decentering, and Attention as Selection for Action

215

EWA ODOJ

- Doxastic Responsibility and the Challenge of Doxastic Voluntarism.
Insights from Cases of Self-Deception* 235

ROBERT GRZYWACZ

- Reading as an Affective and Discursive Event. Its Contribution to Reshaping
Human Identity* 271

DISCUSSIONS

KENNETH KEMP

- Second Reply to Fr. Chaberek: On Why Merely Biological Humans Can Survive,
and on When Merely Traditional Doctrine Can Be Abandoned* 295

MICHAŁ ZALEWSKI

- Is Saruman a “Peacemaker,” and Abortion “Murder”? Report from the Debate* 303

OSKAR LANGE

- Report from the 3rd International Christian Philosophy Conference, “Christian
Philosophy Facing Naturalism”* 311

- CALL FOR PAPERS 315

- NOTE ABOUT *FORUM PHILOSOPHICUM* 319

Editors' Note

Jacek Poznański, Szczepan Urbaniak

In this winter issue of *Forum Philosophicum*, we invite you to explore contributions inspired by the theme of the International Conference “Christian Philosophy Facing Naturalism,” which took place at the Ignatianum University in Cracow (Poland) on 24–25 September 2024. This biennial event was the third in a series of conferences devoted to broadly understood Christian philosophy (previously held in 2020 and 2022) and was attended by nearly forty speakers from the United States, India, Australia, the United Kingdom, Hungary, Switzerland, the Czech Republic, Germany, Finland, Croatia, Austria, and France. The participants represented renowned academic institutions, including several highly ranked in global university rankings, such as the University of Cambridge and the University of Oxford.

The main objective of the conference was to present the current state of research on the relationship between Christian philosophy and naturalism (as well as science more broadly), undertaken by an international group of philosophers from both historical and systematic perspectives. The papers and discussions demonstrated that although naturalism may pose various challenges, it can also serve as an opportunity to critically assess established positions within Christian philosophy and to develop new ones. A deeper understanding is needed—not only of naturalism itself but also of the claims made by contemporary sciences about the world and about human beings.

Naturalism encourages philosophy to maintain cognitive engagement with the empirical sciences. Yet Christian philosophy need not assume a defensive posture. On the contrary, it can reconsider its relationship to this influential philosophical tradition, expose its significant limitations and weaknesses, and at the same time draw inspiration for its own advancement.

✉ Jacek Poznański, Ignatianum University in Cracow, Poland ✉ jacek.poznanski@ignatianum.edu.pl

📄 0000-0002-8158-564X

✉ Szczepan Urbaniak, Ignatianum University in Cracow, Poland ✉ szczepan.urbaniak@ignatianum.edu.pl

📄 0000-0002-4459-5638

The current issue opens with an article by Finley Lawson, who argues that the alleged conflict between naturalism and Christianity rests on a mistaken material–immaterial dichotomy. He shows how the holistic ontologies of Michael Esfeld and Hans Primas make possible a scientifically informed metaphysics fully compatible with Christian doctrine. Charles Taliaferro, a keynote speaker at the aforementioned conference, reconstructs and defends Plato’s argument from reason in *Phaedo*, demonstrating its continuing force against eliminative and reductive forms of contemporary naturalism. Robert Kublikowski, drawing on Robert Brandom’s inferentialism, identifies the essentially normative, embodied, and deontic character of human understanding as what irrevocably distinguishes natural intelligence from the probabilistic performances of even the most sophisticated artificial systems.

Ward Blondé presents a rigorous formal proof—combining set theory, mereology, metaphysics, and concepts from theoretical computer science—of the existence and omni-attributes of the Anselmian God, boldly concluding that “God is the Absolute Machine.” Joshua Sijuwade offers a new philosophical explication of monotheism in terms of metaphysical fundamentality (drawing on Carnap’s method and Karen Bennett’s building relations), thereby clarifying both the complexities of Second Temple Judaism and the shortcomings of purely numerical definitions. Tymoteusz Mietelski reconstructs the anthropological thought of the Italian neo-scholastic Sofia Vanni Rovighi, who creatively integrated Thomistic metaphysics with phenomenological insights in order to resist naturalistic reductionism. Ryszard Mordarski subjects recent proposals for an “evolutionary theology” (W. Grygiel and D. Wąsek) to critical scrutiny from the standpoint of classical theism, warning against an overly science-dependent approach that risks collapsing into panpositivism.

The contributions inspired by the conference theme are complemented by six additional articles. Piotr Szalek defends the later Wittgenstein against the charge of relativism by appealing to the commensurability enabled by the shared second-person perspective inherent in the human form of life. Mitchell Thomas Welle and Marcin Koszowy examine how deliberate rephrasing of messages has become a powerful vehicle for spreading misinformation, and they propose new linguistic-argumentative tools for detecting and building resilience against “fake news.” Mariusz Wojewoda analyses the erosion of responsibility within the emerging “technosystem” of intelligent machines and, drawing on Hans Jonas and Andrew Feenberg, calls for an ethics oriented toward the long-term flourishing of present and future humanity. Piotr Sikora argues that the mindfulness phenomenon

of decentering constitutes a genuine counterexample to Wayne Wu's influential theory of attention as selection-for-action. Ewa Odoj defends epistemic responsibility (and a moderate doxastic voluntarism) by analyzing the mechanisms of epistemic self-deception and the metacognitive capacities that make self-control possible. Robert Grzywacz brings Marc Richir's phenomenology of affective phenomenalisation into dialogue with Paul Ricoeur's hermeneutics of reading to show how the literary event can profoundly reconfigure human identity.

The issue concludes with three discussion pieces: Kenneth Kemp's second reply to Fr. Michał Chaberek on human origins and doctrinal continuity; Michał Zalewski's report on a spirited debate concerning the nature and legitimacy of "semantic arguments"; and Oskar Lange's overview of the September 2024 Christian philosophy conference.

The project "Christian philosophy conference 2024—Christian Philosophy facing Naturalism" was co-financed from the state budget funds granted by the Minister of Science and Higher Education within the framework of the "Excellent Science II" Programme (Poland) (KONF/SN/0010/2023/01).

ARTICLES

A Naturalist Theology

Christianity within a Holistic Paradigm

Finley Lawson

ABSTRACT The root of the narrative that places naturalism in opposition to the central tenets of Christianity resides in the perception that a “naturalist” account of reality has no space for the nonmaterial/transcendental. That perceived dichotomy, this paper will argue, resides in a categorical error about the nature and number of things in reality. The apparent conflict assumes that one faces a binary choice between matter and non-matter, where the first of these falls under the remit of investigation by the natural sciences while the second does not, thus putting theology at odds with naturalism.

In contrast to this dichotomous account, the scientific holistic ontologies proposed by Michael Esfeld (philosopher of science) and Hans Primas (quantum chemist) provide a radically different account of foundational reality in which one can argue that there is no requirement to reconcile two fundamentally different kinds of “stuff.” The contradiction between naturalism and Christianity is only apparent. It is based on our presuppositions about the world as described by science and our commitment to particular accounts of the nature of personhood. This paper does not claim that scientifically informed holism “solves” the naturalism versus anti-naturalism debate; however, it does provide a way to integrate naturalistic (scientific) metaphysics into our Christian thought.

KEYWORDS Esfeld, Michael; Expansive Naturalism; Naturalistic Metaphysics; Non-Boolean holism; Ontological Holism; Primas, Hans

This article¹ responds to the claim that the dominance of the naturalistic paradigm requires the Christian philosopher not only to reflect on the conditions and consequences of naturalism, but also to critically attend to the “naturalist challenge” according to which nothing exists, and/or can be known, separately from the material reality examined by the natural sciences. The paper challenges both the notion that a naturalist account requires a “materialist”² metaphysics and the idea that juxtaposing naturalism with the transcendent presents us with a genuine dichotomy. The narrative that places naturalism and the non-material in exclusive categories is, I argue, based on a categorical error about the nature and number of things in reality. That apparent conflict assumes that one faces a binary choice between matter and non-matter, where the first of these falls under the remit of investigation by the natural sciences while the second does not, thus putting theology at odds with naturalism.

In addressing this “naturalistic challenge” I will draw on the works of Michael Esfeld (philosopher of science) and Hans Primas (quantum chemist). Although they have proposed two different naturalistic (scientific) pictures of fundamental reality through their holistic ontologies, one can argue that both ontologies remove the requirement to reconcile two fundamentally different kinds of “stuff.” Thus, the contradiction between naturalism and Christianity is only apparent, and is based on our presuppositions about the world as described by science and the nature of personhood. This paper does not claim that scientifically informed holism solves the naturalism versus anti-naturalism debate; however, it does provide a way to integrate naturalistic (scientific) metaphysics into our Christian thinking. To achieve this, the article is formed of three parts, of which the first provides a brief overview of what is traditionally meant by naturalism, and why even those accounts described as “expansive” do not do enough to address the perceived challenge to Christian philosophy. The second examines the grounds for claiming that a naturalist ontology can also be a holistic one, while the third part critically examines how scientifically informed holism may provide a way to integrate naturalistic (scientific) metaphysics into Christian thought. This final part addresses

1. A version of this article was presented at the Krakow 2024 conference “Christian Philosophy: Facing Naturalism,” and the author is grateful for the constructive questions and dialogue with colleagues at both that event and ESPR 2024. The sections on Esfeld’s and Primas’ ontologies are adapted and expanded from *Christ, Creation, and the World of Science: Against Paradox* (Lawson 2023).

2. How materialism is to be conceived, and whether naturalism necessarily equates to materialism, will be examined as part of this paper.

Esfeld and Primas in turn, due to the substantive differences in the way they make room for the non-material in their metaphysics. Whilst at this stage it is not possible to claim that either provides a definitive “solution” to a naturalist Christian metaphysics, both invite the reader to question the necessity of making a binary choice between matter and non-matter, where the first of these falls under the remit of investigation by the natural sciences and the second does not, thus placing theology in conflict with naturalism.

THE NATURALIST CHALLENGE

The challenge of naturalism resides in the assumption that a naturalist account necessarily requires the rejection of “supernatural forces” (gods, angels, demons, souls, ghosts etc.).³ This is often perceived as pertaining to a supernatural “other” world separate from our own investigable “natural” world. This can be further compounded by the insistence that the supernatural includes “non-material beings,” and also anything that is beyond the realm of investigation by the natural sciences, including “the aesthetic qualities of a painting, the moral significance of an action, the meaning of a gesture, the reasons for political conflict” (Ellis 2024, 2; see also: 2014, chap. 1). Such a reductive account of what is natural is known as scientistic reductionism, whereby the natural world only consists of the “physical” stuff (however that is to be construed). This extreme account of naturalism calls into question the validity of philosophy and theology in the manner of logical positivism—together with any scientific enterprise that requires emergent properties. In both *God, Value and Nature* (2014) and a recent paper delivered at the European Society for Philosophy of Religion (Ellis 2024), Fiona Ellis has argued against this form of naturalism, stating that a more “expansive” form is required. An expansive scientific naturalism allows for the genuine existence of science beyond the remit of physics and encompasses the realm of the social sciences. Whilst this allowance provides a greater source of hope for the philosopher and theologian it still, arguably, rests on an assumption that the natural sciences, correctly understood, only provide space for a reductive physicalist/materialist account of the world. This assumption of reductionism is identified by Primas (1983, 308) when he notes that “there still exists

3. There is an interesting question that falls outside of the scope of this paper as to why such definitions of supernaturalism are focused on “conscious” entities rather than the existence of the non-material and/or transcendent more generally, with the associated query as to whether such rejections of the supernatural are as strong when one moves to a broader conception of the non-material that goes beyond non-material agents.

a bias toward theoretical monism” with the aim of reducing all theories to an “all-embracing fundamental theory.”

The explanations provided by scholars such as Ellis for a naturalistic account of the transcendent maintain that the existence of the transcendent “does not mean that one must establish a “superworld” of divine objects” (Tillich 1975, 8); rather, the transcendent can be understood as being expressed in and through the immanent (Ellis 2014, chap. 2) in the context of an “ecstatic” naturalism. Yet there is a risk that some of these arguments still assume that allowing the transcendent in a natural world requires finding a way to circumvent the (dualistic) supernaturalism versus (reductive) naturalism dichotomy. Ellis explicitly challenges such dichotomous thinking and argues that “once it is allowed that the transcendent can be modelled other than in dualist supernaturalist terms, then the tantalizing possibility opens up that the transcendent is already presupposed in a world whose immanent character is suitably complex [i.e. beyond the boundaries set out by scientistic accounts]” (Ellis 2024, 4). On this account, Ellis argues for an intertwining of the transcendent and immanent, rather than opposing metaphysical categories, and I am hopeful that the metaphysical account developed here provides a fruitful contribution that further supports such discussions. In line with Ellis, I argue that these dichotomies need to be challenged, and that a different kind of expansive naturalism, one that challenges the necessity of physicalism, can please both camps—the scientistic and the theistic. This account differs from Ellis’ as it is based on a holism that creates space for the transcendent, the non-material and the divine by challenging our misperceptions about ontological structure. The following section will briefly set out the place of the non-material within science, and how Esfeld’s and Primas’ holisms provide space for a holistic naturalism that can leave room for the transcendent and/or non-material. Their accounts will then be examined in detail.

IS A HOLISTIC NATURALISM POSSIBLE?

Michael Esfeld (a philosopher of science) and Hans Primas (a quantum chemist) have provided accounts of foundational reality in which there is no requirement to reconcile two fundamentally different kinds of “stuff.” In other words, the ontological dichotomy that pits naturalism against theism is removed. The contradiction between naturalism and Christianity becomes a matter of appearance rather than of ontological fact: it is based on our presuppositions about the fundamental nature of the world as described by science, and the nature of personhood. The assumption that strict naturalism provides the “correct” description of reality is closely

tied to the early-20th-century commitment to theory reduction in science, whereby it was argued that “all phenomena of life can be ultimately reduced to the laws of physics and chemistry” (Primas 1983, 308). Whilst the reductionist holds that “since organisms are built of matter, every biological question must be sought in terms of the fundamental theory of matter [i.e. physics],” holists, like Primas and Esfeld, emergentists, and (historically) vitalists all deny that physical laws can sufficiently explain “the phenomena of life” (Primas 1983, 309).

Thus, the question must be raised as to whether one can provide a naturalistic account that allows for the non-material. In this discussion we are concerned not with abstract objects, but rather with the nature of materiality as it shows up within our current philosophical discourse. Numerous recent volumes highlight the breadth of this perceived change away from comprehending matter as “physical stuff” (Davies and Gregersen 2014; Koons and Bealer 2010). Further, in *The Quantum Enigma* (2005), Wolfgang Smith distinguishes the corporeal world of our everyday experience, which is “the sum total of things and events that can be directly perceived by a normal human being” (2005, 27), from the (physical) universe examined by the physicist. These two “worlds,” whilst ontologically unified, are epistemically distinct, with the physical universe being viewed entirely via measurement:

Physical objects are then known by means of a suitable model, a theoretical representation of some kind . . . object and representation do not coincide . . . one cannot know or even conceive of a physical object except by way of a model, or theoretical construct. (Smith 2005, 31)

Such a view of the inaccessibility of the world “as it is” is a defining feature of the German transcendental idealist movement of the 18th and 19th centuries. But Smith and his colleagues are not requiring us all to abandon realism in its entirety, even though we are required to put aside naive realism and recognise, on this view, that there is not a one-to-one correlation between our perception of reality and the fundamental ontology that grounds reality. Science does not require that meaning is only attached to concepts that are “unambiguously defined in terms of fundamental physics” (Primas 1983, 308). Indeed, for many, including Primas and Esfeld, fundamental reality is itself inaccessible to scientific investigation, although this inaccessibility does not diminish its reality. Moving away from scientific naturalism does not necessitate adherence to the false dichotomy of idealist ontology versus traditional materiality; however,

we must engage with pertinent questions about the nature of the material realm⁴ (including questioning what is within the bounds of the natural). When we speak of naturalism as “the idea that all existent entities in the world are of a physical nature” (Ludwig 2018, 285), there is an implication that all the properties of those entities are also physical (or can be related to physical properties). Such strict reductionism leads to the view that all the objects of the special sciences, including mental properties and consciousness, are, or will be, entirely describable in terms of fundamental physics. The decomposition (or reduction) of a system into non-trivial subsystems (parts) is a feature of classical mechanics (or systems). At this level, the state of the whole is determined by the states of these subsystems. Whilst for some this level of reductionism is a fundamental part of their metaphysical framework, it is by no means necessitated by our current scientific findings, and is certainly not a position endorsed by Primas or Esfeld. According to Primas “most theories of chemical, biological and social systems are still limited by the classical paradigm of explanation presupposing in an unreflected way the separability of these systems” (1983, 325).

Strict reductionism therefore rests on an overly simplified view of the structure of scientific theories, and the assumption that complex systems are hierarchical. Complex systems can have “many modes of *description*, all equally valid and real” (Primas 1983, 315; emphasis added) but these “levels” do not exist independently of the whole. It is also crucial to realise that whilst it may be possible to describe the “higher” levels of the hierarchy in terms of “lower”-level descriptions, this may lead to overly complicated descriptions, or “the relevant patterns of the higher level are not put in evidence” through such a description. An excellent example from Primas is that whilst it *may* be possible to provide a molecular description of bee activity, such a description “would be entirely irrelevant to the understanding of, say, how honey bees signal the location of a rich source of nectar” (1983, 317). A similar argument applies to the matter of a naturalist explanation of reality. Whilst it may be possible to describe a painting in a manner that is distilled into a fundamental theory, this would undoubtedly fail to describe the aesthetic qualities, or do so only in a manner that becomes incomprehensible.

4. Questions at the interface of philosophy of mind and theology, concerning the nature of the physical realm and the emergence of consciousness/mind (a common theme of Nancey Murphy’s work), are prevalent in panentheistic discussions. Whilst important, philosophical discussion of these matters lies beyond the scope of this paper.

Thus, the question becomes not whether the world can include the “non-material,” but what we understand by “matter” and how we can comprehend the existence/nature of non-matter and its potential for transcendence. Davies describes the problem as follows:

Apparently solid matter is revealed, on closer inspection, to be almost all empty space, and the particles of which matter is composed are themselves ghostly patterns of quantum energy, mere excitations of invisible quantum fields. (Davies 2014, 83)

For Davies, a naturalist (scientific) account of the world is increasingly pointing to a universe in which the fundamental ground is not “physical” but “mathematical”—i.e. the world appears to be most explicable in terms of mathematical formalism: our only option is to “treat the physical universe as if it simply *is* mathematics” (2014, 86; original emphasis). Yet even if Davies is correct regarding our approach to understanding reality, one must be careful to not confuse method and/or epistemology for ontology. Treating the world as “mathematical” may help with describing experimental findings, but it does little more than rename Kant’s inaccessible noumenal realm when it comes to dealing with the relationship between the object and its representation.

This is well captured by Charles Minser when he argues that “material substance” is on the defensive, it being “reduced at most to scattered specks in the emptiness, its garrisons pulled together in isolated posts” (1978, 2). The more we have tried to comprehend the nature of “matter,” the more we have found ourselves reducing our understanding not to that which is being investigated, but rather to “the interactions among them [the particles]. We do not say, what an electron is, but we do write laws for how it interacts with photons and other electrons” (ibid.). It is this reduction to relational descriptions (and the fact that these remain constant despite the changing account of the material) that drives the move to adopting a structural realist ontology. Yet, whilst it may provide us with a naturalist ontology, such structural accounts (as will be seen in Esfeld’s moderate account) provide greater evidence for a transcendental reality in which relationality provides the ultimate ground or building blocks of reality, over and against bits of matter. This thinking is also echoed by Misner when he argues that this relationality within contemporary scientific accounts gives rise to a picture that is “radically anti-materialistic” (1978, 4), with the object of scientific investigation being the “immaterial constituents—the design relationships” (ibid.), not the “stuff” portrayed by scientific naturalism.

Whether one stands with Misner and/or the structural realists in granting that the fundamental reality is, in some meaningful sense, non-material, there can be no doubt that we need to seriously engage with the challenge posed by the fact that the “co-essential” materiality and divinity of Chalcedon, the “rational soul and human flesh” of the Athanasian, and, in the Nicene creeds, “all that is, seen and unseen,” all imply a dichotomous distinction of a kind that does not maintain its distinction in this non-matter-grounded ontology. The distinction that sets up the naturalist “against” the theist in finding space for the non-material and/or transcendent appears less a point of ontology and more an epistemic/methodological commitment to the scope and bounds of scientific investigation. Therefore, mindful of the contested nature of the “material,” the fact that Primas and Esfeld point towards naturalistic ontologies which are holistic is less surprising than one may first expect. So it is that one finds oneself in a space in which scientific discussions, for many, have become more open to the acknowledgement of the existence of non-matter (even whilst what this looks like may be contested). The implication that remains for the debate between naturalism and Christian philosophy is that the kind of non-materiality allowed is fundamentally different to that required by the theist’s commitment to the divine as non-matter and/or transcendent.

Given the uncertainty as to how we are to understand this foundational ground (informational, relational, structural, mathematical, etc), it seems premature to claim that they are dealing with distinct levels of reality. What is clear is that whether one sees reality as fundamentally holistic in line with Primas and Esfeld, or sits with Davies in that, at the very least epistemically, we must treat the world as mathematical, the “character of the material world casts severe doubts upon the consistency of the Cartesian separation [of matter and non-matter]” (Primas 1994c, 611). Yet here again we encounter the crucial distinction between the world “as it is” and what is required for scientific investigation—in that every experiment requires a fundamental ontological dualism in the form of the distinction between subject and object to proceed. Because of this necessity, Primas argues that a (naturalistic) materialist ontology must be incomplete, because it is “incapable of dealing with the complementarity of matter and spirit” (*ibid.*). This recognition by Primas (explicitly) and Esfeld (more implicitly) of the existence of that which is non-matter points away from scientific naturalism (whilst allowing for expansive naturalism), as both view this non-material aspect of reality as “beyond” scientific investigation.

Therefore, it is still necessary to maintain an expansive naturalism that avoids the temptation of assuming an excessively reductionistic account

of the natural and/or material. Provided one does not adopt a scientistic form of naturalism, there is no reason naturalism and holism cannot be compatible. Whilst not wishing to over-labour the distinctions between metaphysical holisms, it is worth taking note of the differentiations made by Healey and Gomes. For the metaphysical holist, there is a certain extent to which the whole is not entirely determined by the properties of the parts, and it is possible to see a correlation here with the expansive naturalist who recognises that reality cannot be fully described by the scientifically investigable nature of its parts. Healey and Gomes identify three kinds of metaphysical holism:

Ontological Holism: Some objects are not wholly composed of basic physical parts.

Property Holism: Some objects have properties that are not determined by physical properties of their basic physical parts.

Nomological Holism: Some objects obey laws that are not determined by fundamental physical laws governing the structure and behavior [sic] of their basic physical parts. (Healey and Gomes 2022, sec. 3)

All these accounts require a more expansive naturalism than that provided by the scientistic naturalist. The most comfortable (least contested) scientific “holism” is property holism, with its links to emergence. But for this paper, and in relation to the question of the potential compatibility of a rich Christian philosophy with a naturalist account of the world, I am interested in going “deeper,” to consider how scientific metaphysics may be pointing to a substantive ontological holism. Ontological holism is, in some circles, viewed as out of favour, or as requiring some commitment to a form of neutral monism or “third kind” of “stuff.” Whilst such a view is not necessitated by ontological holism, it offers greater support for the theist than reconciling a local materialism with a required “other” realm for the non-material divine. In addition, Primas adheres to a version of dual-aspect monism, and whilst Esfeld does not address what he means by “matter,” there is at the very least a certain agnosticism in his approach towards what that matter is.

Primas and Esfeld propose two very different pictures of holistic reality (despite both drawing on a holistic account of reality grounded in quantum theory). Even so, they are united in seeking to provide the philosopher and/or theologian with a nuanced and deep metaphysics that speaks to our need for an ontology that has space for transcendence (although neither takes

this step themselves). Esfeld's and Primas' accounts hold out the promise of a coherent metaphysics, either by providing an alternative to the matter-transcendence dichotomy seen in much naturalistic discussion (Primas) or by challenging our conception of the nature of ontological dependence between objects and the relationships they stand in (Esfeld).

It is to these two positions that I will now turn, with a view to examining how we may reconsider the natural, and what a scientifically informed holistic ontology may mean for our understanding of a naturalistic philosophy of religion. I will assume an expansive account of naturalism (i.e. one that allows for the existence of things beyond the scope of scientific investigation). I do not believe that such a position places the theist on the side of anti-naturalism. One only needs to go down the anti-naturalist route if one maintains that the ontologies proposed here (and/or by others who challenge a Cartesian view of matter as points of "stuff") do not actually describe the "natural" world. In other words, anti-naturalism is only required for the Christian philosopher who, in a similar vein to the Christian materialist, argues that the local world of human experience is circumscribed by that which can be described/investigated by the natural sciences, with a global exception for the divine. The scientific ontologies described by Primas and Esfeld are not contentious positions—the science they build their metaphysics upon is well accepted. The potential disagreement arises from choosing their suggested ontology over other metaphysical descriptions. However, both acknowledge that their choices are guided by wider metaphysical commitments.⁵ Esfeld's work with the model of ontic structural realism (OSR) is in good company, philosophically speaking, and Primas' account has been well discussed for its metaphysics, finding potential allies in such great names as Spinoza and Leibniz.

Both ontologies are "naturalist" in that they are constructed on the basis of findings from contemporary physics, and both are expansive in the terms described by Ellis and others in that they recognise and actively endorse the existence of features of reality that are beyond the investigative remit of the natural sciences. However, this is not the ecstatic naturalism of Tillich, whereby the transcendent "mean[s] that, within itself, the finite world points beyond itself. In other words, it is self-transcendent" (1975, 8).

5. For example, Esfeld stops short of adopting what I call "strong" ontic structural realism (SOSR), despite his argument in relation to quantum mechanics that there is no empirical way to establish what would be preferable between a metaphysics of relations and of individuals. His reason for adopting what I classify as a moderate form of structural realism (MOSR) is that the metaphysics of relations required by SOSR would leave one ignorant of the nature of the world "as it is."

What these inherently transcendent holistic ontologies require, and what they mean for the Christian philosopher, will form the basis of the next two sections.

A NATURALIST ONTOLOGY OF RELATIONS

In its most basic version, holism requires that the sum of the object/system/reality (etc.) be “more than” its parts taken in combination. When one considers the ontological proposal of holism, there are several ways to understand the “more than” that is required within a naturalist (scientific) account of reality.⁶ The first rests on a form of priority monism (as seen in Primas’ work), whereby there are many concrete objects but only one is fundamental, and all other objects subsist on, or are derived from, this single basic object.⁷ The second is what I call a “strong” ontic structural realism, of the kind adopted by Ladyman and Ross (2007; Ladyman 2020), in which the object-property distinction is conceptual,⁸ contrasting with “standard”⁹ metaphysics, where the “structure is fundamentally composed of individuals and their intrinsic properties, on which relational structure supervenes” (Ladyman and Ross 2007, 148). The final approach being considered here is the one adopted by Esfeld and his collaborators, which in many ways sits fuzzily between SOSR and Primas’ monism: this atomistic holism maintains that atoms are holistically individuated in terms of the distances among them, but unlike SOSR, where structural relations have ontological priority relations and relata are mutually ontologically dependent. For Esfeld, “the distinction between object and properties, including relations and thus structures, is only a conceptual one by contrast to an

6. It is important to note that the options discussed here are not exhaustive of the ways in which metaphysical holism can be formulated within a naturalistic paradigm. Just as Primas and Esfeld are driven by wider philosophical commitments, it would be possible to consider these accounts, for example, as they relate to one’s commitment to “thing” ontologies or to the necessity of individuation.

7. “Object” here is construed broadly, as Primas rejects the ontological division of the universe into discrete “objects”—though he does allow for the existence of a fundamental *Unus Mundus* foundation (an indivisible foundational reality).

8. Ontic structural realism, as proposed by Ladyman and Ross, holds that “all the properties of fundamental physics . . . [are] extrinsic to individual objects” (2007, 151). On this account it is the objective structure that is “ontologically fundamental, in the sense of not supervening on the intrinsic properties of a set of individuals.” Ladyman and Ross go on to argue that even the individuality of objects is dependent on underpinning relational structure: “there are no things. Structure is all there is” (Ladyman and Ross 2007, 130).

9. This being the term adopted by Ladyman and Ross to refer to metaphysics that assumes the existence of individuals, in contrast to their own OSR metaphysics.

ontological one: properties, including relations, are modes, that is, the ways in which objects exist" (Esfeld and Lam 2011, 13).

Holding this in mind, such holistic ontologies stand opposed to both theistic dualism (whether local or global) and the wholly anti-transcendent ontologies of the scientistic naturalist. Fundamentally, they create a space in which a natural account of the world not only admits, but requires, something "more than." As noted above, though, in many senses this dualist-versus-naturalist dichotomy can be understood as arising from a specific view on the kind(s) and number(s) of things that exist. This is clearly captured in Esfeld's engagement with Descartes.

It may seem counterintuitive to examine a claim of ontological holism via the very dualism that creates the problematic dichotomy for the Christian philosopher, and yet Esfeld argues that Descartes' denial of Aristotelian essentialism andhylomorphism means that Cartesian dualism is in fact a monistic ontology. For Descartes, the physicality of any object is described in its entirety by its spatial extension. In the case of humans there is an additional "cogitating substance," but because everything is grounded in spatial extension (in that even the cogitating substance requires something spatially extended to ground it) there is no multitude of entities in existence (as in existence pluralism): rather, Descartes' position is more aligned with Esfeld's own as a form of priority monism.

Were Descartes not to have been committed to a single token of three-dimensional extension (and thus priority monism), one would be required to overdetermine ontology by positing the addition of independently existing matter, as well as absolute space and time. It is a similar motivation that drives Esfeld and Deckert in their minimalist ontology (2020): for Descartes, the unnecessary addition is "matter," and for Esfeld and Deckert it is the requiring of *relata* independently of, and prior to, the relations in which they stand. This continuous "matter" or "gunk" that exists to a greater or lesser extent in each point in space would then require further explanation as to why it is more or less dense at different points, with the need for further layers of explanation then starting to push existence pluralism in a direction such that it falls foul of Ockham's razor.

Whilst it might be surprising that Descartes' ontology points towards a monistic account, he is insistent that just because we can discuss the "parts" of this primary "whole" of matter-space, it does not mean that these parts (corporeal shapes/objects) can be individuated. It does not require the existence of individual, independently existing parts that compose the whole. This position (rejecting an ontology of individuals) has found stronger forms in both the traditional monism of Spinoza and the contemporary

holism of Esfeld, both of whom argue that the “whole” is prior to any constructed parts. What distinguishes Spinoza and Esfeld on the one hand, and Descartes on the other, is that for the former, wider demarcations or individuations between the parts are not ontological. Yet both the monism of Spinoza and the “dualism” of Descartes point to there being only one existent substance/entity in the physical realm—matter-space.

Esfeld’s denial of discrete matter points rests on different grounds, but the foundational unity remains. As was the case with Descartes’ matter-space, for Esfeld space must be understood as a continuum, with the “parts of space” (Esfeld 2001, 176) correctly construed as nothing more than demarcations (rather than proper parts). Because space is a continuum, any region and/or point in space is fundamentally relational—no point can be described (or exist) in isolation. When this is combined with a Cartesian (and Esfeldian) identification of matter with space it results in a metaphysics that is inherently holistic. Matter and space cannot be distinguished or divided into discrete parts (or entities), and this means that “all matter is one holistic system” (Esfeld 2001, 178). This interdependence is not a causal relationship, but an ontological holism (interdependence) involving all “material” things.¹⁰

This challenge associated with hard reductionism brings us—surprisingly for the materialist—to a holistic metaphysics. In *Part and Whole in Quantum Mechanics* (1998), Maudlin likewise argues that reductionism taken to its conclusion leads to “truly radical holism” (Maudlin 1998, 49). Once one starts to analyse the whole (whether that be a watch or a universe) in terms of progressively smaller and smaller parts, looking for that additional basic block of matter, there comes a point at which the particularization must cease in that one can no longer divide the object. At this point, Maudlin argues, we arrive at “partless” fundamental bits of matter: “these partless parts must be spatially unextended: they must be points” (Maudlin 1998, 48). Once one is dealing with indivisible points, the metaphysician or philosopher has arrived at field theory, which is fundamentally interdependent and ontologically relational. This descent into holism on the part of the

10. It is worth noting that “materiality” here is not defined by Esfeld—nor, indeed, does Primas give a detailed account of how we are to understand the unified substance which grounds everything (the *Unus Mundus*). This is both a strength, in that the holism could be conceived within the terms of an informational, austere realist, and/or neutral-monist metaphysics (for example), and also, when it comes to applying such metaphysical accounts to specific problems in the philosophy of religion and/or personal identity, a challenge. However, for the purposes of this paper the question of how exactly one is to understand the “stuff” of this holistic reality will be put to one side, as it does not impact on the broader question at stake here.

hard-line reductionist—an inadvertent conclusion that is the outcome of the scientistic naturalist’s drive to reduce everything to physics—means that the assumption that dualistic (theistic) and naturalist accounts of reality are two mutually exclusive horns of an ontological dilemma need not necessarily be an accurate one.

Through his “holistic atomism,” Esfeld presents an ontology that allows for the existence of individuals; however, these individuals or physical systems are not *in addition to* space-time, and these fundamental building blocks of reality cannot be understood in isolation. Due to the inherent entanglement required by Esfeld’s realist interpretation of quantum mechanics, some of the properties exhibited by quantum systems are only there by virtue of the relationships in which they stand to each other (which he describes as “correlations of entanglement” [Weber and Esfeld 2013]). The reason that Esfeld’s holism is of such interest for the Christian philosopher and/or theist is that it does not rest on a reinterpretation of how we understand the matter that constitutes matter points: he rejects any ontology of “substance.” Instead, there is a transcendent (though he himself does not use this term) and fundamental (primitive) “stuff-essence” (Esfeld and Deckert 2020, 32) that permeates the universe.

This primitive stuff-essence can be understood as consisting of individual objects (it is a “thing”-ontology), yet it is not the individual atoms (or bits of stuff-essence) that individuate one tiger from another, for example: rather, it is their relative combination and position. Because it is not a traditional atomistic account, there is no intrinsic “thisness” distinguishing one object from another, or one atom from another. Indeed, the more we understand about the fundamental particles of physics, the clearer it has become that there are no intrinsically individuating features. Instead, what demarcates one object/atom from another can only be described by the state of the whole, which “fixes relations among the parts” (Esfeld and Lam 2011, 246). Thus, these individuating properties could not be possessed by an object existing in isolation. All properties that we traditionally consider intrinsic (including mass, charge, etc.) are primarily relations. For the Christian philosopher, then, this places a transcendent divinity and relations between creatures and their creator at the heart of our account of reality. On this model the objects, or *relata*, and the relations in which they stand are ontologically interdependent—one does not precede the other, but instead they have the “same ontological footing, being given “at once” . . . they are mutually ontologically dependent” (Esfeld and Lam 2011, 4).

If objects cannot be understood except as part of a holistic system (that itself is responsible for furnishing individuation), and individuation requires

the existence of a “whole,” then one would do well to consider what, if anything, could be the solely existing object within a possible world? When one also considers the fact that, on this account of MOSR, the relata (or objects) that stand in and are individuated by their relations cannot *be* properties of anything else, but must rather be capable of having properties predicated *of them*, it then appears—albeit unintentionally so—that some foundational “thing” that all others stand in relation to is required. This foundational thing is also to be understood as a “world-making” property:

If there is a plurality of things, there has to be something that relates these things so that they make us a world. . . . there has to be something that relates the things *in each world* . . . At least in so far as the actual world is concerned, position in the sense of spatial relation (distance) is what unites the world. (Esfeld 2021, 2; emphasis added)

Whilst it is possible to argue this is “merely” some kind of organisational field, there is nothing that says this transcendent, relational, and foundational “object” cannot be understood as God. As was previously mentioned, this is not a move that Esfeld makes, or is seeking to make, and by no means am I trying to imply that Esfeld’s metaphysics proves or requires God to be the holistic system within which individuation occurs. However, the fact that our universe can be understood as having a holistic (and relational) foundational ontology based on a naturalistic account of the nature of reality does provide promising avenues of exploration for metaphysicians generally and the Christian philosopher specifically.¹¹

Before moving on to discuss Primas’ account of holism, it is important to note that there is currently no empirical method for determining whether a metaphysics of individuals or of relations would provide a more accurate account. Esfeld chooses to adopt a metaphysics of individuals (making his structural realist position moderate in comparison to the metaphysics of relations provided by Ladyman and Ross), but this decision rests on his dislike of the Kantian-esque commitment of SOSR according to which we are left ignorant as to the intrinsic nature of things (Epistemic Structural Realism). This gap between a relational metaphysics and our limited fundamental theories that can only account for or describe the relationships

11. The nature of this paper and the focus on naturalism means that there is not scope here to provide a detailed account of how this metaphysics may be understood, or a detailed explanation of how Esfeld and his collaborators arrive at a holistic primitive ontology. However, this is discussed in more detail in *Christ, Creation, and the World of Science* (Lawson 2023); see chapters 4 and 5 for the holistic turn.

between objects and not the nature of the objects themselves means that we are left with one or other of two choices:

- (a) maintain a belief in a metaphysics of individuals but accept this means we are unable to gain knowledge about the intrinsic properties of the individuals as far as they are intrinsic
- (b) discard a metaphysics of individuals in favour of a metaphysics of relations according to which at the fundamental level only relations exist. (Lawson 2023, 240)

The fact that this is a choice driven by wider ontological and/or epistemic concerns is important for the naturalism debate because it assumes an ontology of individuals in which theism and/or transcendent entities stand opposed to the fundamental building blocks of reality. Yet “there is no *a priori* argument that excludes a metaphysics of relations” (Esfeld 2004, 615), so it is important to question the implications of a naturalistic metaphysics of relations when seeking to understand whether there can be space for the transcendent. Whilst Esfeld provides a relation-oriented metaphysics of individuals, Primas’ position is more challenging in that it questions ontological individuation, and it is to that account that we now turn—before subsequently offering some brief remarks on the potential implications of both of these for Christian approaches to the philosophy of religion.

A NATURALIST ONTOLOGY OF “NO-THINGS”

Esfeld’s account of holism challenges our understanding of the relationality of objects and requires us to reconsider where relations sit within an ontological hierarchy. Despite this, it allows us to continue with a metaphysics of individuals. Primas does not grant us this metaphysical comfort blanket. Like Esfeld, he begins with a consideration of Descartes, highlighting the key role that the “Cartesian cut” plays in all scientific endeavour. With the expression “Cartesian cut” Primas is referring to the requirement, within scientific experimentation, to make a distinction between the observing subject and the object being observed. For science to progress, it is necessary for the experimenter to establish the initial conditions of the experiment, but this “freedom” conflicts with the hard determinism required by strict materialism (Primas 1993). The question arises, where does it leave the naturalist if it is possible to distinguish between experimenter and experiment (whereas the scientific naturalist would have to accede that both “objects” are equally determined, and “physical”)? Whilst at first glance it

may appear that these dual commitments imply that Primas is committed to the ontological dualism spurned by the naturalist, he in fact argues that our current account of reality points to “an *ontological monism*, combined with an *epistemic dual-aspect approach*” (2009, 171; original emphasis).

We have already seen how Esfeld’s monism gives rise to a holistic ontology of individuals, so what does Primas’ account bring to the conversation on naturalism? Whilst Esfeld shifts the challenge of describing how any mental or transcendent aspect of reality is to be understood over to the philosopher and/or theologian, Primas confronts the issue head on. In “Endo- and Exo- Theories of Matter” (1994a) he argues that the Cartesian distinction between mind and matter, despite its absurdity, “is a temporarily useful fiction . . . that matter does not contain spiritual elements in an essential way” (1994a, 165). Thus, this “removal” of the transcendent is performing a heuristic role only. He goes on to argue that:

Our distinctions between an atemporal material, an atemporal mental, and a temporal domain do not imply an ontological partition of the world—it is *chosen as a partition of the universe of discourse to facilitate the discussion*. (Primas 2009, 24; emphasis added)

This step to remove the non-material/transcendent from our understanding of the “natural” world is not the ontological commitment of the scientific naturalist, but rather the move of an expansive naturalist. Primas is clear in his writings about the importance of ensuring that science does not overreach itself: the non-material is not within the scope of scientific investigation, but it *is* very much within the ontological structure of natural reality. In as much as he adopts the Cartesian split where scientific method is concerned, Primas also argues that we are deeply misguided if we mistake this method for ontology:

The experimentally well-confirmed holistic character of the material world casts severe doubts upon the consistency of the Cartesian separation of the *material* reality from the *spiritual* one . . . [despite this] *present day experimental science still requires an epistemological dualism*. (Primas 1994c, 611; original emphasis)

The key commitment (as noted earlier in this section) is that we need an *epistemological* dualism to make sense of the world, but this does not reflect the nature of reality as it is. Whilst Esfeld backs away from a metaphysics that entails the inaccessibility of the world as it is, for Primas there

is a fundamental incompleteness in contemporary scientific naturalism. This is because the scientific (but not naturalistic) description “as at present conceived, forces us to leave out crucial parts of reality” (1993, 250). As will be explained further shortly, Primas’ holism is committed to a rejection of atomism, and (to use Esfeld’s terminology) rejects any “thing”-ontology that would allow for a plurality of objects (see: Primas 1991, 165; 1994c, 611–13). Primas’ holism therefore denies an ontological commitment to “context-independent objects” (Primas 1994c, 629) and, indeed, even the relational framework provided by Esfeld’s holism does not furnish these (the relationality provides context). Furthermore, Primas’ epistemological dualism is framed by the fact that “our ability to *describe* the world cannot go further than our ability to isolate objects” (1994c, 626; emphasis added). Crucially, this does not limit the natural world to that which is describable by science; instead, it simply recognises our linguistic/epistemic limitations.

When Primas rejects a “thing”-ontology he is arguing against the scientific naturalist’s claim that everything can be described in “material” terms: one mistakes the scientific project if it is viewed as trying to describe “the material reality in terms of some elementary building blocks” (1998, 67), and on the basis of the “outdated” belief that reality can be explained with reference to “independently existing atoms” (1998, 87). In undertaking the reductionist project, whether as a scientist or a philosopher, one is assuming a particular “thing”-ontology, which is out of step with the findings of quantum theory (Primas 1998, 88). The latter, according to Primas’ interpretation, presents a metaphysical account of the world that not only provides a radically holistic “no-thing” foundational structure, but also registers the fact that the “*unbroken wholeness* of the material world . . . cannot be observed directly by our five senses” (1994b, 335; emphasis added). Here again, as with Esfeld’s account, we see that this version of monism is naturalistic in being grounded in the findings of physics, even while it calls for an expansive approach to naturalism.

How one begins to carry out individuation in the context of such an ontology is a theme developed throughout his writings, but some of the founding principles for his more metaphysical work can be encountered in his treatise on quantum chemistry and reductionism (Primas 1983). The mistake we make in assuming a classical (Boolean) account of reality is to assume that there are bare facts, and what Primas terms “absolute objects” that can be understood as existing independently. There is no “God-given” (Primas 1983, 308) decomposition of reality. There are “no absolute objects or absolute patterns” (Primas 1983, 325); instead, *every* observation, every description, is an exercise in identifying what is essential and what

is accidental *in this instance*. In doing this we introduce Boolean categories (e.g. material vs non-material) and deliberately exhibit a “*lack of interest [in certain components] which breaks the holistic unity of nature*” (Primas 1983, 325; original emphasis). This makes everything conditional on the “context,” and the latter is created through the erection of well-defined conceptual/metaphysical barriers that *create* a non-holistic ontological structure to which one can apply Boolean logic. It is this contextualized decomposition that gives rise to the appearance of paradox or conflict in the conversation pertaining to matter versus non-matter, or naturalism versus theology. The conflict arises because we misconstrue this “deliberately chosen abstraction” as brute ontology, instead of recognising that every phenomenon/object “is created by abstractions alone and does not otherwise exist” (where this includes objects perceived through our senses and examined via the natural sciences) (Primas 1983, 327).

Because Primas’ ontology is so different from how the “natural” is scientistically conceived, it is worth spending some time unpacking his account of reality. There is a strong connection (at least at the conceptual level) between his partless (non-Boolean) holism and Ladyman’s account of Strong Ontic Structural Realism (SOSR). Although Primas does not shift to a metaphysics of relations (at least, not in those terms), he does take the step that Esfeld seems reluctant to embrace—by moving to a metaphysics in which we cannot know the fundamental structure of the universe. Ladyman describes this lack of access to the reality of “objects” in themselves in SOSR as the fact that “individual objects are constructs . . . individuals have only a heuristic role” (2020, sec. 4). There are clear correlations with Primas’ earlier claims that “objects do not yet exist, we have to create them” (1993, 254) and “the world is not made out of some building blocks . . . these [the perceived building blocks, e.g. electrons] are just manifestations of the material reality” (1994c, 619). For Primas, this metaphysics is grounded in the following ontological starting point:

Since all predictions of quantum mechanics are experimentally well corroborated, and since the counterintuitive results of quantum theory are no logical paradoxes, we take the holistic structure of the quantum world as a true feature of nature. (Primas 2003, 253)

Whilst Ladyman’s metaphysics does not have the same scientific grounding, and indeed non-Boolean holism could be arrived at without reference to quantum theory, the fact that his account *is* scientifically grounded calls into question the necessity of an antagonistic relationship between Christian

philosophy and naturalism. What we perceive as individuals are not fundamental objects, but “represent patterns of reality . . . Elementary or composed “particles” . . . are not primary but rather secondary and derived” (1994c, 628). Not only are these “elementary” particles not primary, but even realist interpretations of quantum theory refer “only to a fictitious theoretically immanent reality, and not to the ultimate reality” (1994c, 622). This ultimate reality (which Primas terms the *Unus Mundus*) exists as a “primordial unity, not yet divided into two [res extensa and res cogitans]” (Primas 1993, 249), and all “*decompositions of the world are neither given a priori nor determined by first principles*” (Primas 2007, 27; original emphasis). On this account, the ultimate reality is “much nearer to Plato’s ideas, according to which the attempt to divide matter again and again results in mathematical forms” (1994a, 174). This is echoed in Davies’ account of the physical universe as mathematical form noted above. Individuals are “created” when we “isolate a phenomenon and assign individuality to it” (Primas 2007, 11–12) within a particular context, and this does not create an entirely amorphous ontology—the scientific picture of the world is built on our “knowledge of observable patterns or modes of reactions of systems” (Primas 1998, 96). However, instead of arriving at Esfeld’s relational holism, where individuals and their relations are ontologically mutually dependent, Primas argues that we should recognise that “individuals” do not exist *a priori*. Thus, we should adopt a non-Boolean holism, “a whole which has no parts” (2007, 27) and that requires us to choose “an appropriate partition of the universe of discourse” (Primas 2007, 27). These partitions are created when we decide which features are relevant and irrelevant, allowing us to create “fuzzy” or non-Boolean categories (or boundaries) around “individual objects.”

Within Primas’ account what is striking is his commitment to the non-material/transcendent. He specifically identifies the realm of endophysics as that which studies the “realm of non-spatial, non-mental, timeless, but nevertheless real entities” (1994a, 166), stating that “neither nowness nor consciousness can be identified with any property known to physics, so *we relate these phenomena to the nonmaterial domain*” (2003, 95; emphasis added). If we are finding challenges in demarking the material and non-material realms “it is because no such line of demarcation exists” (Primas 1983, 331). Our descriptions are caricatures “exaggerating some aspects by deliberate simplification and permitting extravagance” (Primas 1983, 331). The caricature is not intended to present brute fact, but rather to allow new patterns and/or levels of description to arise. Yet the most hopeful description for the Christian philosopher engaging with Primas’ non-Boolean holism comes when he states that:

We do not restrict the [nonmaterial] tensed domain to the inner world of private thoughts . . . *We relate the tensed domain to a mental world which we consider as fundamental to the nature of existence and being.* According to this view, “mind” operates as a principle beyond individual consciousness and is not restricted to the “human mind.” (Primas 2003, 92; original emphasis)

Where Esfeld’s account furnishes potential for the transcendent, that proposed by Primas explicitly endorses it. This is not to imply that wholehearted adoption of non-Boolean holism represents an easy trade-off for the Christian philosopher or scientist—it raises a number of methodological and ontological questions.¹² Nevertheless, it is this deep commitment to a timeless ultimate reality (Primas 2003) imparting fundamental order and grounding to the universe that provides a rich context for exploring the transcendent within a naturalist framework.

In his 2003 paper, we find Primas’ most explicit description of how we should understand the material/non-material “divide.” He draws heavily on Wolfgang Pauli’s work, arguing that:

Quantum theory describes the material world in a basically holistic way. Generalizing this result beyond the material world to ponder upon a holistic conception concerning mind and matter. Pauli . . . suggested that the mental and material domain are governed by common ordering principles and should be understood as “complementary aspects of the same reality.” (Primas 2003, 90)

Taken in combination, these two commitments (the extension of the non-material domain beyond individual consciousness and the commitment to a dual-aspect account of reality) require the naturalist to move away from Cartesianism to a “primordial unity, not yet divided into two” (Primas 1993, 249). This brings us back sharply to one of the central premises of this paper: namely, that there is space within a naturalistic account for a non-material and/or transcendent aspect of reality, provided that (a) one does not require it to be accessible through scientific enquiry (expansive naturalism) and (b) we revisit our metaphysical commitments pertaining to the “stuff” of fundamental reality.¹³ The concluding section of this paper will briefly

12. Discussion of these potential challenges goes beyond the scope of this paper, but some of the problems raised specifically in relation to the incarnation are discussed in chapters 6 and 7 of *Christ, Creation, and the World of Science* (Lawson 2023).

13. For example, whilst this paper draws certain distinctions between the “non-material” and “matter,” these are made for heuristic purposes in order to better proceed with the

outline how these holistic ontologies may be applied to concerns that are key for Christian philosophy, particularly in respect of divine personhood.

SOME REMARKS ON THE IMPLICATIONS FOR CHRISTIAN PHILOSOPHY OF RELIGION

So far, this paper has focused on setting out two accounts of ontological holism that can be viewed as naturalistic (on an expansive account of the latter) because they provide an ontology grounded within the reality described by contemporary science. The reason why these accounts disrupt the naturalism-versus-holism debate is that whilst the non-material and/or transcendent aspect of reality sits beyond direct examination by physics, both accounts endorse the reality of something “more than” the world described by science, with this “other” being central to a complete account of the nature of reality (even if we cannot directly access it). What is interesting for the expansive naturalist is that both Esfeld and Primas hand this task of further examination over to the philosopher and theologian; this final section thus picks up where the (non-theistic) holistic naturalist stops, and explores some implications for philosophical theology.

The present paper is not intended to provide a definitive answer to what the naturalist’s God may look like within a holistic ontology: indeed, Primas’ and Esfeld’s ontologies provide very different accounts of the nature of divine personhood and the incarnation (see, for example, Lawson 2023; chapters 6-7). To examine the details of individual theological commitments in relation to holism would take up more space than is allowed here. However, before concluding it will be worth noting some of the opportunities that exist for the theistic naturalist within this framework, along with some areas meriting further research there.

I would argue that one of the areas in which the impact of a holistic ontology is best seen is through an examination of the incarnation (at least in terms of understanding the united person of Christ). The reason for choosing this deeply theological issue for philosophical discussion is because issues raised by the transcendent for human personhood are compounded when brought into dialogue with the divine in Christ. This echoes Gregersen’s argument in “Deep Incarnation and Chalcedon” that the challenge for the Christian philosopher/theologian is that Chalcedon does not “specify the characteristics of the “divine nature” [or] the “human nature,” and ... does not tell us anything about their interrelation in the

discussion: they explicitly do not exist within Primas’ monism, and arguably do not exist, as traditionally understood, within Esfeld’s relational ontology either.

concrete person of Christ” (Gregersen 2020, 275). Gregersen provides an excellent systematic overview of how this relationship (the human-divine) has been understood in theological contexts, and it is worth drawing out from this some of the key features of personhood that could be reimagined within a holistic naturalist account of reality along the lines of the ontologies provided by Primas and Esfeld:¹⁴

DI1. Athanasius argued that “the particular human body of Jesus needed to be neither separated from other human bodies, nor from the materiality of the cosmos at large” (Gregersen 2020, 260). This draws on the concern raised by naturalists that the transcendent requires a “superworld” of divine entities “beyond” our own. Implicit in their account is the assumption that the materiality of the “cosmos” is ontologically separated from anything that could be transcendent, as the transcendent goes beyond the natural.

DI2. The stoics provided an ontology whereby there was “a co-extensive inherence of two elements within a general metaphysical scheme . . . the idea of mutual co-inherence is central” (Gregersen 2020, 265–66). In other words, this is the kind of ontological account that is rejected by the scientistic naturalist because (it is assumed that) only one aspect of this can be investigated using the scientific method. It is this metaphysical dualism that places naturalism and Christian philosophy at odds.

DI3. Gregory of Nyssa’s understanding of the infinitude of God provides several points for consideration: as infinite reality, God “must be equally close to the material as God is to the spiritual world” (2020, 270), and this allows us to understand the “logic” of the incarnation in that infinite unity means that there is no greater distance between God and the material than between God and the non-material—this being the challenge of how one is to understand genuine transcendence within a dualistic metaphysics. This challenge would also include accounts such as Tillich’s ecstatic naturalism, where the transcendent is “revealed” through the material, as it still assumes a degree of separation in kind, or of ontological level, for the transcendent.

DI4. Finally, drawing on Schleiermacher, Gregersen notes that “subsuming divinity and humanity under the same umbrella . . . [implies] that the two natures meet one another at the same level, while tacitly presupposing

14. The ensuing points DI1–4 and DI1’–DI4’ (excluding explicit references to naturalism or the philosophy of religion), are reproduced from Lawson (2023, 335).

a predefined contrast between divinity and humanity” (2020, 276)—this being, in effect, a continuation of the issue noted in DI3. Similar ontological presuppositions underpin the naturalist’s challenge to the Christian philosopher.

These metaphysical challenges regarding a non-materialist (non-scientistic) account of personhood in relation to the incarnation and, fundamentally, the existence of the transcendent, prove consonant within the holistic ontologies noted within this paper:

DI1’. Primas’ non-Boolean holism posits that there is a fundamental unity within the cosmos (the *Unus Mundus*) which cannot be ontologically distinguished in the discrete categories of spirit and matter (to use Primas’ terminology). Thus, the distinguishing of Christ, or the transcendent, from the universe, and the separation of other entities into a “supernatural” world, may be viewed as a contextually chosen decomposition, not a “brute fact” about the “separability of nature” (Primas 2007, 27). On this account, there is nothing that is “beyond” the natural world. The appearance of being “beyond” stems from our inability to investigate or access the *Unus Mundus*, and our own preconceived contextualization of the universe into the binary categories of “matter” and “non-matter” that we incorrectly take to be ontological.

DI2’. Following Pauli, Primas (2003, 90) argues that the mental and material domains are “complementary aspects of the same reality.” Less explicitly, Esfeld and Deckert (2020, 8) recognise that a radically reductionist worldview cannot include the non-material, and in SOSR the co-inherence may be understood as an aspect of the underlying “structure” of reality. One could argue that this co-inherence is what Tillich is describing as ecstatic naturalism when he argues for the revelation of the transcendent through the immanent.

DI3’. As with DI1’ this can be understood in the context of the fundamental unity of reality, whereby the distinction between humans and angels is not ontological, but dependent upon the chosen partition. Likewise, if a fully relational ontology is adopted (not Esfeld’s MOSR), this may provide room to consider the ontological interrelatedness of the cosmos, which again would put in question the requirement for a “supernatural” realm. On the other hand, it would allow one to classify the “supernatural” as that which stands outside of scientific investigation.

DI4’. Schleiermacher’s commitment has its closest reflection in Esfeld and Lam’s mutual ontological dependence between relations and their rela-

(see: Esfeld and Lam 2011, 4). The ontological co-dependence within Esfeld's ontology provides a philosophically interesting space in which to examine how transcendence may be conceived as an ontological relation within a naturalistic framework.

The biggest barrier at this stage to a fully (expansive) naturalistic account of the transcendent within holistic ontology is the lack of information on the nature and/or place of conscious beings within their ontology. There is much more that could be said about the ways in which holistic ontology may provide space to reconceptualize our account of transcendence, but to provide an exploration of these topics in depth would be to go beyond the scope of this paper. However, this article does identify two naturalistic accounts that place the transcendent at the heart of ontology (even if they do not provide the pertinent details). So, to borrow from Primas, the thoughts drawn together here regarding the place of holistic ontology in supporting a positive account of naturalism for the Christian philosopher "are of a fragmentary and speculative character so that this . . . should be considered as an exercise, whose aim is not to solve any concrete problem but to discuss new ways of thinking" (Primas 2003, 113). By this I mean that a holistic metaphysics holds out great promise for progressing our understanding of the relationship between naturalism and Christian philosophy, and in challenging the "correctness" of the assumed materialist (scientific) vs dualist (theistic) dichotomy that underpins the argument to the effect that naturalism stands in opposition to the central tenets of Christianity or Christian philosophy. Hopefully, this paper has shown that holistic ontology positively reconceptualizes our understanding of the kinds of natures or substances involved in asserting a transcendent divinity. However, there are a great many details still to work out, both in terms of the implications for whether a "classical" theism, with its distinction between the world and God, can be maintained, and with respect to how we are to understand that issue, so carefully avoided, of the place of conscious beings in a holistic universe.

CONCLUSION

The expansive naturalist argues that one must move to a "naturalistic" account of human nature, in which the transcendent is revealed through the immanent and the dichotomy putting naturalism and Christian philosophy at odds with one another is therefore removed. However, I am not convinced that this move is necessary, or that it is successful. Whilst expansive naturalism nominally rejects "dualism," I argue that it simply reinstates it within

another space. If the transcendent is revealed *through* the immanent it must still be, at some level, something other than the immanent. This raises the question of what kind of thing is being revealed through the “natural.” Ellis argues that the naturalist’s resistance to transcendence is “premised upon the assumption of . . . rejecting [God as] a supernatural ‘something else’” (2024, 3). This “othering” is overcome if the transcendent and immanent are “intertwined,” but this language still implies two kinds of things being brought together within one world/reality—much as “something like the claim that the two natures are mixed together into one new nature [in Christ]” (Cross 2002, 2) reinforces a “two-category” approach to the “kinds” of things involved in the incarnation. Do we really need to “expand” the natural to include God? Are we not mistaken in assuming that “natural” is synonymous with “contingent” or “created”? Is not a stronger version of reality one in which God is the underpinning feature, or organising principle, of reality?

The holistic accounts provided by Primas and Esfeld can be adopted by the Christian philosopher and/or theologian to avoid the need to “expand” the natural to include God. Instead, one is invited to seriously reconsider the presuppositions grounding the assumption of a dichotomy, and ask “What if?” What if the portrayals offered by Esfeld, Primas and others are based on an accurate word-world connection? What might this holistic, naturalistic ontology look like and, more importantly, what might it mean for our understanding of the nature of the transcendent? I believe that a combination of something like Esfeld’s fundamental relationality with Primas’ unified *Unus Mundus* furnishes a rich ground for a relational, transcendent reality in which the non-material (whilst not accessible to the methods of science) is something that is a foundational part of the “natural” world. However, it is not all plain sailing here: such an account raises a number of questions for the philosopher, including whether such approaches require the adoption of a pan(en)theistic account of divinity, how (in the context of Primas’ account) one might provide space for genuine individuation between persons (human and divine), and, as was noted earlier, the issue raised by the fact that neither Primas nor Esfeld engage with how one should understand conscious beings in the context of their metaphysics. Yet this need for further investigation does not, and should not, rule out the validity of the initial principles involved: namely, the idea that by questioning the perceived dichotomy, holistic naturalism provides us with a way to positively integrate naturalistic (scientific) metaphysics into our Christian thought.

BIBLIOGRAPHY

- Cross, Richard. 2002. *The Metaphysics of the Incarnation: Thomas Aquinas to Duns Scotus*. Oxford: Oxford University Press.
- Davies, P.C.W., and Niels Henrik Gregersen, eds. 2014. *Information and the Nature of Reality: From Physics to Metaphysics*. Canto Classics ed. Cambridge: Cambridge University Press.
- Davies, Paul. 2014. "Universe from Bit." In *Information and the Nature of Reality: From Physics to Metaphysics*, edited by P.C.W. Davies and Niels Henrik Gregersen, 83–117. Cambridge: Cambridge University Press.
- Ellis, Fiona. 2014. *God, Value, and Nature*. Oxford: Oxford University Press.
- . 2024. "Naturalism, Theism, and the Question of Human Nature." Paper presented at the European Society for Philosophy of Religion Conference, Trento, Italy, September 5.
- Esfeld, Michael, and Dirk-Andre Deckert. 2020. *A Minimalist Ontology of the Natural World*. New York: Routledge.
- Esfeld, Michael, and Vincent Lam. 2011. "Ontic Structural Realism as a Metaphysics of Objects." In *Scientific Structuralism*, edited by Alisa Bokulich and Peter Bokulich, 143–159. Version cited is author pre-print.
- Esfeld, Michael. 2001. *Holism in Philosophy of Mind and Philosophy of Physics*. Synthese Library, vol. 298. Dordrecht: Kluwer Academic Publishers.
- . 2004. "Quantum Entanglement and a Metaphysics of Relations." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 35 (4): 601–617. <https://doi.org/10.1016/j.shpsb.2004.04.008>.
- . 2021. "'Thing' and 'Non-Thing' Ontologies." In *The Routledge Handbook of Metametaphysics*, edited by Ricki Bliss and J.T.M. Miller, 195–208. New York: Routledge. Version cited is from author draft. <http://philsci-archive.pitt.edu/14235/>. Accessed October 13, 2025.
- Gregersen, Niels Henrik. 2020. "Deep Incarnation and Chalcedon: On the Enduring Legacy of a Cappadocian Concept of Mixis." In *Herausforderungen und Modifikationen des klassischen Theismus—Band 2: Inkarnation*, edited by Thomas Marschler and Thomas Schärfl, 253–290. Studien zur systematischen Theologie, Ethik und Philosophie, Band 16/2. Münster: Aschendorff Verlag.
- Healey, Richard, and Henrique Gomes. 2022. "Holism and Nonseparability in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2022 ed. Stanford, CA: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2022/entries/physics-holism/>. Accessed October 13, 2025.
- Koons, Robert C., and George Bealer, eds. 2010. *The Waning of Materialism*. Oxford: Oxford University Press.
- Ladyman, James, and Don Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Ladyman, James. 2020. "Structural Realism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2020 ed. Stanford, CA: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2020/entries/structural-realism/>. Accessed October 13, 2025.
- Lawson, Finley. 2023. *Christ, Creation, and the World of Science: Against Paradox*. PhD diss., Canterbury Christ Church University School of Humanities and Education Studies. <https://repository.canterbury.ac.uk/item/97459/christ-creation-and-the-world-of-science-against-paradox>.
- Ludwig, Pascal. 2018. "Reduction and Emergence." In *Philosophy of Science: A Companion*, edited by Anouk Barberousse, Denis Bonnay, and Mikael Cozic, 285–316. New York: Oxford University Press.

- Maudlin, Tim. 1998. "Part and Whole in Quantum Mechanics." In *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, edited by Elena Castellani, 46–60. Princeton, NJ: Princeton University Press.
- Misner, Charles W. 1978. "The Immaterial Constituents of Physical Objects." Paper presented at the Conference, Ulm, Germany, September 19. <http://www2.physics.umd.edu/~misner/Ulm%20talk.pdf>. Accessed October 13, 2025.
- Primas, Hans. 1983. *Chemistry, Quantum Mechanics and Reductionism: Perspectives in Theoretical Chemistry*. 2nd corrected ed. Berlin: Springer.
- . 1991. "Reductionism: Palaver Without Precedent." In *The Problem of Reductionism in Science: Colloquium of the Swiss Society of Logic and Philosophy of Science, Zürich, May 18–19, 1990*, edited by Evandro Agazzi, 161–72. Episteme 18. Dordrecht: Springer. https://doi.org/10.1007/978-94-011-3492-7_9.
- . 1993. "The Cartesian Cut, The Heisenberg Cut, and Disentangled Observers." In *Foundations of Modern Physics 1992: The Copenhagen Interpretation and Wolfgang Pauli*, edited by K.V. Laurikainen and C. Montonen, 245–68. Helsinki: World Scientific. https://doi.org/10.1142/9789814535984_0013.
- . 1994a. "Endo- and Exo-Theories of Matter." In *Inside Versus Outside: Endo- and Exo-Concepts of Observation and Knowledge in Physics, Philosophy, and Cognitive Science*, edited by Harald Atmanspacher and G.J. Dalenoort, 163–93. Springer Series in Synergetics, vol. 63. Berlin: Springer.
- . 1994b. "Mesoscopic Quantum Mechanics." In *Symposium on the Foundations of Modern Physics: Proceedings of the Fourth Conference*, edited by Paul Busch, Pekka Lahti, and Peter Mittelstaedt, 324–36. Cologne, Germany: World Scientific. <https://doi.org/10.1142/9789814535335>.
- . 1994c. "Realism and Quantum Mechanics." In *Logic, Methodology, and Philosophy of Science IX*, edited by Dag Prawitz, Brian Skyrms, and Dag Westerståhl, 609–631. Studies in Logic and the Foundations of Mathematics, vol. 134. Amsterdam: Elsevier.
- . 1998. "Emergence in Exact Natural Sciences." *Acta Polytechnica Scandinavica, Mathematics and Computing in Engineering Series* 91: 83–98.
- . 2003. "Time-Entanglement Between Mind and Matter." *Mind and Matter* 1 (1): 81–119.
- . 2007. "Non-Boolean Descriptions for Mind-Matter Problems." *Mind and Matter* 5 (1): 7–44.
- . 2009. "Complementarity of Mind and Matter." In *Recasting Reality: Wolfgang Pauli's Philosophical Ideas and Contemporary Science*, edited by Harald Atmanspacher and Hans Primas, 171–209. Berlin: Springer. https://doi.org/10.1007/978-3-540-85198-1_9.
- Smith, Wolfgang. 2005. *The Quantum Enigma: Finding the Hidden Key*. 3rd rev. ed. Hillsdale, NY: Sophia Perennis.
- Tillich, Paul. 1975. *Systematic Theology*. Vol. 2. Chicago: University of Chicago Press.
- Weber, Marcel, and Michael Esfeld. 2013. "Holism in the Sciences." In *Unity of Knowledge in Transdisciplinary Research for Sustainability*, edited by Gertrude Hirsch Hadorn. *Encyclopedia of Life Support Systems (EOLSS)*, developed under the auspices of UNESCO. Oxford: Eolss Publishers. <https://www.eolss.net>.

The Argument from Reason Revisited

Charles Taliaferro

ABSTRACT Arguments against naturalism and materialism have a long history, from Plato to Plantinga. The paper has three parts: First, I reconstruct an argument from reason found in Plato's the *Phaedo*. Second I consider whether the argument is relevant to contemporary forms of naturalism. I argue that the argument does constitute a serious objection to some forms of naturalism. I then defend the argument against objections from GEM Anscombe, Graham Oppy, and Peter van Inwagen.

KEYWORDS abstract objects; explanations; naturalism; Platonism; reason

What is often called “the argument from reason” has been deployed against various forms of materialism and naturalism. One can find different forms of this line of reasoning in work by Plato, Augustine, Anselm, Kant, and more recently in work by A.E. Taylor, E.J. Lowe, William Hasker, C.S. Lewis, and Alvin Plantinga. Like virtually all interesting arguments in philosophy, it has its critics. Is some version of the argument sound or is it deserving of the kind of contempt C.D. Broad had for some of the British idealists who “seem to start from no discoverable premises; to proceed by means of puns, metaphors, and ambiguities; and to resemble in their literary style glue thickened with sawdust”?¹ I hope to show that a version of the argument from reason is nothing of the sort.

In this essay, I explore an argument from reason to be found in Plato’s dialogue the *Phaedo*. I defend the argument in the context of what I suggest is a plausible Platonic metaphysics. In a second section, I contend that the argument from reason is relevant to contemporary naturalism. The argument raises a deep challenge, especially to scientific naturalism. A third section defends the argument against three critics: Elizabeth Anscombe, Graham Oppy, and Peter van Inwagen.

PLATONIC REASONING

The *Phaedo* depicts Socrates (d. 399 BCE) in dialogue with his companions on the day of his execution. It is a goldmine of philosophical reflection on the practice of philosophy itself, pleasure, the nature of dying and death, and on the soul and body.

The dialogue contains many arguments and objections, including a plea that persons should not fall prey to misology, the hatred of argument and reason. Socrates tells us: “No greater misfortune could happen to anyone than that of developing a dislike for argument” (89d). He likens the plight of becoming misologic to becoming misanthropic. One may meet many people who seem, on the surface, to be good, but turn out otherwise, and forget that, while there are many bad people, the majority of people are a mixture of good and bad. Similarly, there are some poor arguments, or cases of the misuse of reason, but this should not lead us to despair about attaining good arguments and reason.

We must not let it enter our minds that there may be no validity in argument. On the contrary we should recognize that we ourselves are still intellectual

1. Broad cited by A.C. Ewing (1934, 9).

invalids, but that we must brace ourselves and do our best to become healthy—you and the others. (Phaedo 88c-91c)

We are told by the narrator that Socrates was drawn to a work by Anaxagoras that proposes that the cause of all things is mind.

Then one day I heard a man reading from a book, as he said, by Anaxagoras, that it is the mind that arranges and causes all things. I was pleased with this theory of cause, and it seemed to me to be somehow right that the mind should be the cause of all things, and I thought, 'If this is so, the mind in arranging things arranges everything and establishes each thing as it is best for it to be. So if anyone wishes to find the cause of the generation or destruction or existence of a particular thing, he must find out what sort of existence, or passive state of any kind, or activity is best for it. And therefore in respect to that particular thing, and other things too, a man need examine nothing but what is best and most excellent; for then he will necessarily know also what is inferior, since the science of both is the same. (97c-97e)

This aligns intentional or mindful causal explanations and axiological realism; a mindful explanation is one that is aimed at what is good, best or excellent. The intelligibility of such an explanation is not argued for. It seems, rather, to be advanced as an evident phenomena, known to be true in our own self-aware acts. That confidence in the nature and causal role of the mind becomes apparent in Socrates's objection to Anaxagoras going on to explain intelligence or mind in terms of what is not mind or intentional.

As I went on with my reading I saw that the man made no use of intelligence, and did not assign any real causes for the ordering of things, but mentioned as causes air and ether and water and many other absurdities. And it seemed to me it was very much as if one should say that Socrates does with intelligence whatever he does, and then, in trying to give the causes of the particular thing I do, should say first that I am now sitting here because my body is composed of bones and sinews, and the bones are hard and have joints which divide them and the sinews can be contracted and relaxed and, with the flesh and the skin which contains them all, are laid about the bones; and so, as the bones are hung loose in their ligaments, the sinews, by relaxing and contracting, make me able to bend my limbs now, and that is the cause of my sitting here with my legs bent. Or as if in the same way he should give voice and air and hearing and countless other things of the sort as causes for our talking

with each other, and should fail to mention the real causes, which are, that the Athenians decided that it was best to condemn me, and therefore I have decided that it was best for me to sit here and that it is right for me to stay and undergo whatever penalty they order. (98b-98c)

Essentially, Plato's Socrates is asserting that a proper account of his activities must make an essential role to his intelligence, his choices, and reasons.² Granted, there must be a vital explanatory role for the body, but while such bodily, material factors are necessary, they are not sufficient to account for what is evidently the case.

I propose that this Socratic reasoning is sound. Our own experience of acting for reasons conflicts with causal accounts that make no reference to acting for reasons. An account of our dialogue with each other must give a primary ineliminable role to reasons, in this case judicial reasons and the prisoner's decision that it is best to suffer the punishment decreed. If some label is necessary, we may describe this as a phenomenological argument, objecting to a position based on what is phenomenologically evident.

The dialogue includes two other elements to highlight. I do so to sketch a Platonic version of the argument from reason, one that I believe to be plausible and of current interest in addressing naturalism today.

First, there is a case that persons or the soul is a substantial being rather than a mode of the body. Socrates contends (perhaps on phenomenological grounds) that a person is not a mode of a substance, the way musical sounds are the mode or result of an instrument. I believe this receives some support from our experience. I myself endure over time. My body may change while I remain the same. Moreover, at death, it seems natural to think that I have ceased to be (or am no longer biologically embodied or present) while my body is still existing.

Second, the dialogue advances a case for recognizing that there are forms. I offer a variation of Platonism by asserting that there are abstract objects we may call states of affairs (SOAs). A SOA (like there being philosophers or there being unicorns) is a way things may be; the SOA there being philosophers obtains while there being unicorns does not obtain. Using

2. A reviewer of an earlier version of this paper, points out that Socrates is not just concerned with cognition, but the forms, such as goodness and beauty. While my focus is on cognition, I agree fully that the Socratic stress on the mind is meant to complement and bring to the fore what are the objects of our mental reflection (Phaedo 65c). I go on to propose in this essay that we have reason to think our intentional attitudes are directed on states of affairs, as these are conceived of as abstract objects in the Platonic tradition.

the language (or metaphysics of SOAs), the Athenians believed the SOA Socrates deserves death due to impiety and corrupting the young obtains, whereas Socrates and his disciples believe that it does not obtain.

There is neither time nor space for making an all-out case for SOAs here. I have done so elsewhere, arguing that recognizing SOAs can account for the intersubjectivity of intentional attitudes (you and I can believe or doubt, hope or fear) the same thing and it provides a sound framework for a theory of truth. In my view, truth and falsehood are also best accounted for in terms of the obtaining or not obtaining of SOAs rather than sentential or epistemic accounts of truth and falsehood. Presumably, most of us want an account of truth that does not rest on human language-users (there are truths about reality prior to language) and there may be truths that are not epistemically accessible to humans.³

A summary of this Platonic outlook: Reasoning is based on the apprehension of SOAs and intentional action is carried out by substantial persons on the basis of their pursuit or aversion to SOAs that are believed to obtain or not to obtain. This is not to claim that all agents believe in a metaphysics of SOAs or to deny that intentional action can be (as it were) *de re*—directed on the things, persons or events in one's presence.

The appeal of this Platonic depiction of reasoning (which bears a very close resemblance to the philosophy of Roderick Chisholm) is based on what appears to be the case, an epistemic stance evident among many philosophers from Thomas Reid to G.E. Moore, Thomas Nagel, and Chisholm.

CONTEMPORARY NATURALISM AND REASON

There are different versions of naturalism. Some versions are referred to as scientific or strict naturalism, others have been called broad, liberal or expansive naturalism. The version I am focusing on in this paper is one that subordinates (or eliminates) intentional, mental causal accounts (accounts such as the Platonic version just sketched) to non-intentional, non-mental causal explanations.

Here are some passages describing such a form of naturalism. Daniel Dennett writes:

There is only one sort of stuff, namely matter—the physical stuff of physics and chemistry, and physiology—and the mind is somehow nothing but a physical phenomenon. In short: the mind is the brain We can (in principle) account for every mental phenomenon using the same physical principle,

3. See, for example, (Chisholm 1977). See also (Taliaferro 2015).

laws and raw materials that suffice to explain radioactivity, continental drift, photosynthesis, reproduction, nutrition and growth. (Dennett 1992, 33)

Note that accounts of radioactivity, continental drift, and so on, do not make any reference to explanations in terms of beliefs and reason. Dennett claims:

Psychology must not of course be question-begging. It must not explain intelligence in terms of intelligence, for instance by assigning responsibility for the existence of intelligence in creatures to the munificence of an intelligent Creator, or by putting clever homunculi at the control panels of the nervous system. If that were the best psychology could do, then psychology could not do the job assigned it. (1976, 171)

Dennett explains the appeal of Darwinian theory on the grounds that

Darwin explains a world of final causes and teleological laws with a principle that is . . . entirely independent of “meaning” or “purpose.” It assumes a world that is absurd in the existentialist’s sense of the term: not ludicrous or pointless, and this assumption is a necessary condition of any non-question-begging account of purpose. (1976, 171–72)

Similar claims are advanced by George Rey, Alex Rosenberg, and others. Rosenberg is quite explicit about an error theory when it comes to reasoning and our self-awareness of ourselves as agents. In all such cases, the Platonic argument from reason constitutes an important challenge or objection. Indeed, from the standpoint of the argument from reason, at least Rosenberg seems to entirely undermine his own reasoning and writing.

Our conscious thoughts are very crude indicators of what is going on in our brain. We fool ourselves into treating these conscious markers as thoughts about what we want and about how to achieve it, about plans and purposes. We are even tricked into thinking they bring about behavior. We are mistaken about all these things You cannot treat the interpretation of behavior in terms of purposes and meaning as conveying real understanding What individuals do, alone or together, over a moment or a month or a lifetime, is really just the product of the process of blind variation and environmental filtration operating on neural circuits in their heads. (Rosenberg 2011, 210–55)

A naturalist that is more difficult to assess is David Papineau. On the one hand, he endorses the causal closure of the physical world and embraces the truth of a complete physics. By his lights, a complete physics excludes psychological properties. Arguably, this is countered by the Platonic argument from reason. Reasoning is a matter of mental, intentional causation in light of entertaining states of affairs, and SOAs are not part of the physical world. Still, Papineau seems to acknowledge the veracity of psychological explanations. He writes:

When I say that a complete physics excludes psychology, and that psychological antecedents are therefore never needed to explain physical effects, the emphasis here is on ‘needed.’ I am quite happy to allow that psychological categories can be used to explain physical effects, as when I tell you that my arm rose because I wanted to lift it. My claim is only that in all such cases an alternative specification of a sufficient antecedent, which does not mention psychological categories, will also be available. (Papineau 1993, 81)

The notion that a psychological explanation is intelligible and acceptable, but not needed is curious. By “not needed” I assume he means dispensable and when he refers to a physical explanation being “available” I assume he means more than a mere possibility. We might believe there are all kinds of available explanations of my writing this paper (perhaps Watson’s crude behaviorism is logically possible) but absurd and not to be taken seriously. No, Papineau seems to be banking on the primacy (or preferred) physical account, with no psychology.

It is worth noting that while Papineau thinks there are no mental states as (in his words) “extra to the brain,” he nonetheless acknowledges the intuitive appeal of some form of mind-body dualism.

Indeed, I would say that there is a sense in which even professional philosophical physicalists, including myself, cannot fully free themselves from this intuition of distinctness. Of course, we deny dualism in our writings, and take the theoretical against it to be compelling, but when we aren’t concentrating, we slip back into thinking of conscious feeling as something extra to the brain. (Papineau 2008, 57)

It appears that Papineau adopts some form of identity theory according to which all mental or conscious thinking is identical with physical, brain processes. He sees the ostensible difference between the mental and

physical as a difference in sense but not reference. The mental is a “mode of presentation”—presenting what may turn out to be non-mental.

I suggest that such a strategy is unsuccessful. It is implausible to think of the mental (whether this is a matter of thinking or of sensations) as a mere mode of presentation. Thinking is an activity carried out by a subject; thoughts and feelings are real, phenomenologically evident facts. When persons report they are thinking that 6 is the smallest perfect number (the smallest number equal to the sum of its devisors, including 1, but not including itself; $6 = 1+2+3$) they are directing their attention to mathematical entailments. Any attempt to dispel this ostensibly evident relationship seems counter to all experience and reflection. Granted there are cases of when distinct terms have the same referent (water and H₂O), but such cases depend on composition or aspects (the robber and the masked man) that do not compare with identifying thinking with brain processes. Looking at the molecular composition of water establishes co-reference as would following the robber around until he drops his mask. But no amount of searching the brain reveals conscious thinking or reasoning or how the subject feels. Yes, we can correlate brain processes with thinking but correlation is not identity (Taliaferro 2018).

OBJECTIONS FROM ANSCOMBE, OPPY, AND VAN INWAGEN

Elizabeth Anscombe is well known in the mid-twentieth century for her critique of a version of the argument from reason advanced by C.S. Lewis. Some of the details of Lewis’s version of the argument can be set aside. He uses the term “supernaturalism” to advance his non-naturalist account of reason. The Platonic argument in this paper does not reference the supernatural. (I prefer the term “theistic” as in a theistic understanding of God to avoid the link between supernatural and superstition.) Lewis employs the image of reason impacting the natural world in terms of making pockmarks; not a great analogy as it makes reason seem unnatural. Finally, Lewis uses the term irrational to refer to natural causes; he changed this reference to irrational. But what remains of Anscombe’s critique relevant to this paper?

Anscombe claims that there is an important distinction between assessing whether reasoning is good and matters concerning its cause. Reasoning may be caused by any number of factors (the use of a typewriter or computer) while weighing its validity is another. This misses the mark of the Platonic argument from reason which appeals to the evident fact that, when a person reasons, they entertain SOAs and reach conclusions like “Socrates is being executed by the Athenian state” and 6 is the smallest

perfect number by grasping the reliability of beliefs, testimony, and mathematical entailments. There is no “confusion” about the difference between a cause and a reason, the Platonic argument appeals to the apparent experience of reasoning being causally efficacious. I suggest that the following point by Anscombe is based on a misconstrual of reasoning:

It appears that if a man has reasons, and they are good reasons, and they are genuinely his reasons, for thinking something -then his thought is rational, whatever causal statement can be made about him. (Anscombe 1981, 229)

In reply, I propose that if a person reaches a conclusion (6) based on his reasoning that $6=1+2+3$ then that is the causal explanation of the person thinking and saying 6. And if “whatever causal explanation” does not include the causal role of his reasoning, then that “explanation” is as incomplete as having Hamlet without the Prince.

In her reply to Lewis, Anscombe distinguishes different types of explanation (naturalistic, logical, psychological, and personal history), but none of these overshadow the force of the Platonic argument from reason.

Anscombe, perhaps under the influence of Wittgenstein, would later go on to defend increasingly counter-intuitive claims about intentional explanations involving no “actual mental processes”⁴ and denying that the use of the first-person pronoun (as in “I object to your kicking me”) refers to a substantial subject or person. Anscombe came to believe that to think that the use of the first-person “I” refers to a self is a “grammatical illusion” (Anscombe 1975, 65).

Graham Oppy objects to a version of the argument from reason developed by William Hasker. Unlike Lewis’s version, it will not be necessary to distinguish Hasker’s work from the Platonic version at hand. Oppy adopts an identity theory of the mental and physical. He contends that his physicalism can give a causal role to persons having “representational content” that can account for reasoning. He adopts the causal closure of the physical world (Oppy 2022).

Some of the difficulties with Oppy’s position have been sketched in response to Papineau. The identity thesis faces numerous problems, as brought to the fore in versions of the knowledge argument and the principle of the indiscernability of identicals. If A is B, whatever is true of A is true of B. But this is not so in the case of the mental and the physical. One may observe and know all about a person’s brain and bodily states without

4. Anscombe contra mental processes, cited by (Cumhaill and Wiseman 2023, 202).

knowing their thoughts and feelings / mental states. Moreover, if you adopt the causal closure of the physical, then it is the physical properties and events that are causally efficacious, not the mental (representational content) (see Hasker 2022).

Peter van Inwagen has objected to C.S. Lewis's version of the argument from reason. Van Inwagen's case against Lewis has been responded to by several philosophers, including Brandon Rickabaugh and Todd Buras and, Stewart Goetz (see Rickabaugh and Buras 2017; Goetz 2018). I will therefore focus on van Inwagen's objection to Hasker's version of the argument from reason, which is similar to the Platonic argument sketched earlier.

His principle objection is that when a causal explanation involves rational reflection employing a principle of rationality (for example, *modus ponens*; If A, then B. A. Therefore B) it involves a faulty metaphysics.

A principle of rationality is a proposition or perhaps an imperative—in any case an abstract object of some sort. And if there is one thing abstract objects do not do, it's this: they don't exert influence on things. (Inwagen 2024, 265)

One may grant that the abstract object (the state of affairs or proposition that 6 is the smallest perfect number) does not itself exercise causal power, but it would be (in my view) profoundly mistaken to deny that my answering 6 when asked about identifying the smallest perfect number did not causally involve my grasping the relevant entailment relations. I suggest, as noted at the outset, that, as a matter of common sense, much of our reasoning involves contemplating states of affairs such as *Socrates is to be executed by the Athenian state* and their entailments. For a further defense of my proposal see my 2015 article "Abstract Objects and Causation; Bringing Causation Back into Contemporary Platonism" (Taliaferro 2015).⁵

To summarize: an argument from reason against at least one form of naturalism can be traced back to Plato in the fourth century BCE. I have bolstered this Platonic picture by affirming the evident reality of persons as thinking, enduring subjects who have intentional attitudes to SOAs. I have put on exhibit several claims by naturalists that shows the argument from reason is relevant today. I then defended a version of the argument from reason against objections by Anscombe, Oppy, and van Inwagen.

I add three further, brief points.

Because the argument from reason, as I have presented it, is basically phenomenological, it will not be effective with philosophers who treat first-person

5. For reasons why I oppose van Inwagen's modal skepticism, see (Taliaferro 2001; 1994).

phenomenology as spurious. However, like Richard Fumerton, I find such skepticism contrary to what is most evident in all experience.

Second, I have employed the argument from reason against scientific naturalism (as found in work by Dennett, Rey, and Rosenberg). One might develop a form of naturalism that allows for the emergence of persons who engage in mental reasoning and maintains that this is neither identical to, nor reducible to, non-mental processes.

A third point is rarely observed in print, but I think it bears noticing. Peter Van Inwagen offers the following reason for not adopting some form of dualism: “I should have to tear up most of my work in metaphysics and start over” (Van Inwagen 2024, 269). The topics of this paper has been reasoning and naturalism, not dualism per se, but Van Inwagen’s concern is relevant when assessing philosophical positions and arguments contrary to one’s own. While I think the vast amount of Van Inwagen’s excellent metaphysics is compatible with dualism (and with the phenomenological data of mental reasoning), I think that a deep message from Socrates in the *Phaedo* is that we should relish the opportunity to revise our views in light of good reasons.

BIBLIOGRAPHY

- Anscombe, G.E.M. 1975. “The First Person.” In *Mind and Language*, edited by Samuel Guttenplan, 45–65. Oxford: Clarendon Press.
- Anscombe, G.E.M. 1981. *Metaphysics and the Philosophy of Mind*. Minneapolis: University of Minnesota Press.
- Chisholm, Roderick. 1977. *Theory of Knowledge*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.
- Cumhaill, Clare Mac, and Rachael Wiseman. 2023. *Metaphysical Animals: How Four Women Brought Philosophy Back to Life*. New York: Anchor Books.
- Dennett, Daniel C. 1976. “Why the Law of Effect Will Not Go Away.” *Journal for the Theory of Social Behaviour* 5 (2): 169–187. <https://doi.org/10.1111/j.1468-5914.1975.tb00350.x>.
- Dennett, Daniel C.. 1992. *Consciousness Explained*. New York: Back Bay Books.
- Ewing, A.C. 1934. *Idealism: A Critical Study*. London: Methuen.
- Goetz, Stewart. 2018. *C.S. Lewis*. Oxford: Wiley Blackwell.
- Hasker, William. 2022. “Reply to My Friendly Critics.” *Roczniki Filozoficzne* 70 (1): 191–223. <https://doi.org/10.18290/rf2201.12>.
- Inwagen, Peter van. 2024. “Afterword.” In *The Origin of the Soul*, edited by Joshua R. Farris and Joanna Leidenhag, 255–268. London: Routledge.
- Oppy, Graham. 2022. “Anti-Naturalistic Arguments from Reason.” *Roczniki Filozoficzne* 70 (1): 15–35. <http://doi.org/10.18290/rf2201.2>.
- Papineau, David. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau, David. 2008. “Explanatory Gap and Dualist Intuitions.” In *Frontiers of Consciousness*, edited by Lawrence Weiskrantz and Martin Davies, 55–87. Oxford: Oxford University Press.
- Plato. 1969. *The Complete Works of Plato*. Edited by Edith Hamilton. Princeton, NJ: Princeton University Press.

- Rickabaugh, Brandon, and Todd Buras. 2017. "The Argument from Reason and Mental Drainage: A Reply to Peter van Inwagen." *Philosophia Christi* 19 (2): 381–399. <https://doi.org/10.5840/pc201719230>.
- Rosenberg, Alex. 2011. *The Atheist's Guide to Reality: Enjoying Life without Illusions*. New York: W.W. Norton.
- Taliaferro, Charles. 1994. *Consciousness and the Mind of God*. Cambridge: Cambridge University Press.
- Taliaferro, Charles. 2001. "Sensibility and Possibilia." *Philosophia Christi* 3 (2): 403–420. <https://doi.org/10.5840/pc20013240>.
- Taliaferro, Charles. 2015. "Abstract Objects and Causation: Bringing Causation Back into Contemporary Platonism." *Revista Portuguesa de Filosofia* 71 (4): 769–780. https://doi.org/10.17990/rpf/2015_71_4_0769.
- Taliaferro, Charles. 2018. "Substance Dualism: A Defense." In *The Blackwell Companion to Substance Dualism*, edited by Jonathan J. Loose, Angus J.L. Menuge, and J.P. Moreland, 41–60. Oxford: Wiley Blackwell.

What Is Distinctive About Human Intelligence in the Context of Artificial Intelligence?

A Philosophical Approach With Reference to
Robert B. Brandom's Semantic Inferentialism

Robert Kublikowski

ABSTRACT Since antiquity, humans have traditionally been characterised as *animal rationale* or *homo sapiens*. Such a definition takes into account, on the one hand, the physical aspect, referring to the human body, but also, on the other hand, the mental aspect. With this in mind, the present article seeks to develop a philosophical approach to the problem posed in its title.

Various aspects of human language use and cognition are, in the era of the information revolution, being increasingly taken over by AI. So what still remains specific to humans—or, more precisely, to human, natural intelligence, given this dynamically developing context?

The article addresses this question in a series of steps. Initially, it considers whether rationality is a good candidate for a uniquely human trait. In defence of the distinctiveness of human intelligence, and using Robert B. Brandom's semantic inferentialism, it then points to our ability and skill in understanding, as well as the normativity of language use and cognition. The ensuing discussion focuses on the normative categories of deontic status and deontic attitude, and related notions of commitment and entitlement that these in turn imply.

KEYWORDS artificial intelligence; commitment; entitlement; natural intelligence; semantic inferentialism; understanding

Acknowledgements

For valuable comments concerning the topic of this article I am very grateful to Prof. Robert B. Brandom (The University of Pittsburgh), Prof. Andrzej Bronk (The John Paul II Catholic University of Lublin), Prof. Piotr Kulicki (The John Paul II Catholic University of Lublin), Prof. Agnieszka Lekka-Kowalik (The John Paul II Catholic University of Lublin), Prof. Monika Walczak (The John Paul II Catholic University of Lublin), Dr Piotr Biłgorajski (The John Paul II Catholic University of Lublin), Dr Marcin Grabowski (The John Paul II Catholic University of Lublin), Dr Tomasz Łach (The John Paul II Catholic University of Lublin), Dr Łukasz Sarowski (The John Paul II Catholic University of Lublin), Dr Kamil Szymański (The John Paul II Catholic University of Lublin), PhD student Justyna Horbowska (The John Paul II Catholic University of Lublin) and PhD student Damian Szczęch (The John Paul II Catholic University of Lublin). I am also grateful to two anonymous reviewers.

Since antiquity human beings have traditionally been characterised as *animal rationale* or *homo sapiens*. This definition takes into account, on the one hand, the physical aspect, referring to the human body and its classification within biological anthropology, but also, on the other hand, the mental aspect, which belongs to the domains of psychology and philosophy. With this kept firmly in mind, the present article seeks to develop a philosophical approach to the problem posed in the title.

At first glance the task seems simple: we have specific bodies and minds that allow us to experience emotions, make free choices and, above all, use reason (intellect) in ways that display varying degrees of intelligence in respect of thinking, language use and cognition. Our distinctive culture also encompasses society, morality, religion, art, law, politics and the state. These remain our specifically human—both natural and cultural—traits.

However successive aspects of human language usage and cognition—particularly in the era of information revolution—are being “appropriated” or taken over by artificial intelligence (AI). So what remains as distinctive of human beings—or, more precisely, of human, natural intelligence—in the context of rapidly developing AI?

The question will be addressed here in a series of steps. First, I will consider whether rationality is a good candidate for a uniquely human trait. In defence of the distinctive character of human intelligence—and drawing on Robert B. Brandom’s semantic inferentialism—I will then appeal to our ability and skill in understanding, as well as to the normativity of language use and cognition. Subsequently, I will focus on such normative categories as deontic status (commitment and entitlement) and deontic attitude (including the undertaking and attributing of commitments, as well as the attributing of entitlements).

1. IS RATIONALITY WHAT IS DISTINCTIVE ABOUT HUMAN INTELLIGENCE?

It seems that, according to the received definition of a human being as *animal rationale*, our distinctive trait is indeed rationality, which characterises both non-linguistic and non-cognitive actions, as well as linguistic and cognitive ones.

One can distinguish between practical and theoretical rationality. The former consists in striving for rational agency—that is, actions that satisfy our needs or desires. The latter refers to the rationality of cognition, aimed at attaining a system of true beliefs about the world (Audi 2004, 17 ff.). Human rationality pertains to the experience of emotions, to decision-making, as well as to thinking, language use and cognition. Cognition includes empirical (sensory) cognition: perception, intuitive (immediate, non-sequential) cognition, discursive (sequential) cognition, as well as creativity and memory.¹

Surprisingly, emotions can be understood as either irrational or rational. At times, they serve as a prelude to the formation of thoughts (Eemeren 2015, 204). Emotive passion—such as enthusiasm, etc.—supports decision-making and intellectual engagement.

We human beings pride ourselves on our intelligence, and one of its defining features is that we can remember our earlier thinking and reflect on it (Dennett 2013, 26). Intelligence underlies rationality, which is expressed in our thinking capacity and skills—especially in remembering and engaging in reflection.

Our rational linguistic and cognitive actions are evident in our ability to analyse, synthesise (summarise), pose questions, define, reason, discuss, etc. However, AI is also able to use language or get to know things in a surprisingly rational and correct way.² One only need mention such spectacular applications of it as automatic translators, data analysis and decision support in astronomical observations, in medical diagnostics or military operations, etc.³

1. Our rationality is also revealed in the creativity of cognition. It often consists in—previously unimaginable—questioning of the rules of the system to which the cognition pertains. This may involve the challenging of hidden—sometimes false—assumptions or principles underlying a given research program or scientific theory (Dennett 2013, 45–7).

2. It is important to remember that rationality is originally attributed to human, natural intelligence and only secondarily to AI, which is a human, technological artifact.

3. For more examples of AI applications, see, e.g., Bostrom (2014), Floridi (2014), and Russell and Norvig (2021).

Let us focus on three examples of AI applications, and specifically of Natural Language Processing (NLP) applications which use large language models (LLMs): namely, IBM's Watson, Apple's Siri and OpenAI's ChatGPT.

Watson is used in some online customer services, and has been adapted for medical applications, such as assessing cancer treatment options. It goes beyond simply answering straightforward questions, as does Siri. It is also capable of handling complex riddles as featured in the game show *Jeopardy!*, in which players are not given direct questions but rather clues, and must infer the question that fits the clue. For example: "On May 9th, 1921, this airline opened its first passenger office in Amsterdam. Its name consists of three consecutive letters of the alphabet." The correct answer—in the form of a question—is: "What is KLM?" Watson is capable of meeting such challenges. Unlike Siri, Watson's *Jeopardy!* version does not have access to the Internet (though the medical version does), and it does not understand the structure of conversation. It also cannot obtain answers through logical reasoning. Instead, it relies on parallel statistical searches through a vast but closed database. This database includes documents—countless summaries and source texts, as well as *The New York Times*—providing factual information on a wide range of topics. For *Jeopardy!* Watson's searches are guided by hundreds of specially designed algorithms that reflect the probabilities within the game. Watson can also learn from the answers of its opponents. However, it still makes mistakes. Even in everyday fact-finding tasks, humans often rely on judgments about the adequacy of usage, which remains beyond Watson's reach. For example, one clue in *Jeopardy!* required identifying two of Jesus' disciples whose names are among the ten most popular babies' names and end with the same letter. The correct answer was "Matthew and Andrew"—and Watson gave it immediately. A human gave the same answer—but only after first considering "James and Judas" and then rejecting this, because he or she thought "Judas" was not a popular name for a baby. Watson would be incapable of making that kind of reflection. (Human judgments of adequacy or relevance in language usage are often much more subtle than in this example.) Relevance is the linguistic/conceptual version of the well-known "frame problem" in robotics: the difficulty of determining what is important and what is not in a given situation. It may be that the frame problem will never be fully solved by a non-human system—perhaps due to the complexity of "the frame problem," or because relevance is rooted in our specifically human form of life (Boden 2018, 81–82).

Siri is a personal assistant—a speaking chatbot—that can quickly answer a wide variety of questions. It has access to the entire Internet, including

Google Maps, Wikipedia, the constantly updated *The New York Times* and many local services concerning taxis and restaurants. It also uses the powerful WolframAlpha program—a tool for answering questions—which can use logical reasoning to infer, rather than merely to find, answers to a wide range of factual questions. Siri accepts spoken questions from a user (gradually adapting to the user’s voice and dialect) and responds using web searches and conversational analysis. Conversational analysis studies how people organise topic sequences in a conversation and how they structure interactions such as explanation and agreement. This allows Siri to consider questions such as: “What does the speaker want?” and “How should I respond?,” and, to a certain degree, adapt to the interests and preferences of the individual user. In summary, it appears to be sensitive not only to topic relevance, but also to personal relevance as this pertains to language use. This seems impressive—at least on the surface. However, Siri is easily led astray and often gives silly answers. And if the user strays beyond known facts, Siri gets confused (Boden 2018, 80–81).

Despite some deceptively impressive examples—such as Watson, Siri, or cases of machine translation—modern computers do not understand what they “read” or “say.” Google’s search engine retrieves terms and estimates the plausibility of their usage, but this evaluation is statistical. Search engines—and, more generally, NLP systems—are able to find relationships between words and/or texts, but they do not possess understanding (Boden 2018, 56).

ChatGPT does not retain human-like conscious thinking or understanding, but it increasingly simulates such capacities in a way that is specific to its own architecture. It learns linguistic usage patterns—relatively effectively—from large datasets. Its functioning amounts to a complex form of language processing. ChatGPT employs algorithms to analyse expressions into units smaller than sentences (so-called tokens, which include words, parts of words, punctuation marks, etc.). However, it does not treat these tokens as textual elements in the human sense; rather, it generates approximate internal representations of these elements and their meanings in the form of numbers—more precisely, semantic or numerical vectors (known as embeddings) situated within a multidimensional mathematical space. For instance, the vectors corresponding to semantically similar words—such as “cat” and “dog”—are located close to each other, whereas the vectors for “cat” and “democracy” are placed much farther apart. In this way, ChatGPT is able to “compare” semantic relationships between expressions: both similarities and differences. It predicts (or estimates) which unit is most likely to appear next in a given sequence. For example, in the

sentence “The sun rises in the . . .,” the most probable next token is “east.” Additionally, ChatGPT applies the so-called attention mechanism, which allows it to take linguistic context into account so as to maintain coherence within the topic being developed. For instance, in the sentence “Anna told Maria that her dog had run away,” the pronoun “her” is ambiguous. It is unclear to whom the possessive pronoun refers. The model attempts to determine which referent is statistically most probable in the given context. A key issue, however, is that if ChatGPT’s training data contains false information, the model may retrieve and process that falsehood. Put simply, ChatGPT does not have “direct” access to reality, and this can lead to errors, including so-called hallucinations. A partial solution to this problem is Internet access, which can help test relevant information. Nevertheless, ChatGPT requires critical control and correction.⁴

LLMs—such as ChatGPT, etc.—have been parameterised in such a way that, given an input sequence, they calculate (predict) its most probable continuation (i.e. an output sequence). However, this does not mean that LLMs understand (or interpret) the text, as they do not operate directly on its content (Landgrebe and Smith 2025, 10). People using AI-powered, conversational systems—like ChatGPT—learn the “style” of these chat-bots and the kinds of mistakes they make. LLMs can certainly impress us with their conversational capabilities. However, experienced users know what to say in order to reveal the difference between human, general intelligence (which enables us to perform a wide range of tasks) and AI—a computer program that is only “capable” of performing specific tasks (Togelius 2024, 37)⁵.

Comparing our achievements to those of AI, it seems that AI is catching up with us—or even surpassing us in some areas (e.g., the speed of answering questions). However, as has already been noted, a fundamental issue remains: the aforementioned AI applications do not understand language—or, to put it more cautiously, they do not understand language in a human way. This seems to mark, at least for now, one of the key differences between such systems and natural human intelligence. Let us therefore now turn our attention to the concept of human understanding.

4. ChatGPT’s Language Understanding, based on: <https://chatgpt.com/> (retrieved July 4, 2025).

5. In computer science, a distinction is made between the as yet unattained artificial general intelligence (AGI), which aims to simulate human intelligence, and the currently available narrow AI systems that specialise in performing specific tasks.

2. AN ATTEMPT TO DEFEND THE DISTINCTIVE CHARACTER OF HUMAN INTELLIGENCE

We humans possess properties that are distinctive, especially in the context of rapidly developing AI: this is the hypothesis I hope to defend here, with reference to the position called semantic inferentialism worked out by Robert B. Brandom.⁶

When it comes to our own self-cognition, we should pay more attention to what we are capable of doing, rather than our origins or our biochemical structure (Brandom 1998, 3 ff.). Emphasising the pragmatic, dynamic aspect—namely, our capabilities—is justified. However, it is worth remembering that our origins—understood as our beginnings—are something fundamental, as is the biologically grounded structure dependent on these. They set the scope of our potential and actual actions, whereas AI, by contrast, as a human artifact, operates on the basis of the data it has been provided with and the programmed methods for processing that data.

Among the cognitive capacities constituting human mentality one can distinguish that of sentience—the reception of sensory stimuli—and that of sapience (or wisdom). The former is a biological property shared by other animals and humans alike, whereas the latter is connected with human intelligence and understanding (Brandom 2001, 157). What is important here is sapient consciousness or awareness—a sort of practical mastery or kind of know-how (Brandom 1998, 88 ff.; see also Brandom 2019).

Several questions arise. Firstly, can AI achieve understanding? Secondly—and a more important question for this article is the following one: what is this understanding that is distinctive of human natural intelligence?⁷

6. There are also other approaches to defending the distinctive character of human intelligence. Dreyfus (1992) argued that complex mental processes cannot be fully represented by the logical apparatus on which computers are based. Therefore, AI is not capable of surpassing natural intelligence. Meanwhile, Landgrebe and Smith (2025) claim that it is mathematically impossible for AI to equal or exceed human natural intelligence. In other words, human intelligence will retain its distinctive character and its advantage. They support this claim with the following arguments: (1) human intelligence is a capacity of the human brain and central nervous system, which is a complex, dynamic system; (2) such systems cannot be expressed in mathematical language in a form that would enable their operation within a computer.

7. Incidentally, one may ask whether AI is capable of receiving sensory stimuli. Of course, it is possible to record data through a camera connected to a computer, which is equipped with RAM and a hard-drive memory. However, is this already a case of perception—i.e. perception analogous to that of humans?

2.1. Human Understanding and the Normativity of Language Use and Cognition

Understanding is the fundamental goal of our social, normative and discursive linguistic-cognitive practices (Brandom 2001, 6). Grasping an expression is manifested in our distinguishing its correct from its incorrect uses (Brandom 1998, 13–14, 32). Words such as “correct,” “appropriate,” “proper”—and their opposites—really serve to indicate the normative aspect of our rationality, intelligence and engagement through language and cognition.⁸ By contrast, AI is programmed to operate in accordance with rule-governed norms.

Understanding requires knowing what someone thinks or says about something, as well as what the relation between these two approaches is (Brandom 2001: 158).⁹ An additional factor is how we do it (Brandom 1998: 120). Here, the distinction between the content of cognition and the object of cognition is emphasised.

The grasp of an expression manifests itself in its correct use in appropriate circumstances, and in controlling the consequences of such usage (Brandom 1998, 120). For example, a parrot, when pointing at a red object, “knows” the circumstances of using the term “red,” but it does not “know” the consequences of its use: namely, that if something is red, then it is colourful (Brandom 2001: 62–66; see also 70–71, 148). In other words, it concerns what the extra-linguistic and linguistic context of an utterance is, and what follows from this. The extra-linguistic context might be, for example, the room where the cage with the parrot stands, and the red scarf lying on the table that the parrot is pointing at. The linguistic context will be the sentence “Something is red,” occurring alongside the sentence “Such a thing is colourful.”

Understanding is the ability to “navigate” in the space of reasons, or in terms of the cognition of reasons (Brandom 1998, 5).¹⁰ The feature of rationality that qualifies humans as sapientes can be identified with the

8. For more on normativity, see, e.g., Hattiangadi (2007).

9. Brandom is the originator of the anaphoric theory of reference and truth, according to which both reference and truth are understood within language rather than as relations between language and the world (see Brandom 1998: 275 ff.). Consider the following sentences: “The National Gallery in London is located relatively close to the Houses of Parliament. It contains very valuable collections.” The word “It” in the second sentence refers back to “The National Gallery in London.” This is an example of anaphoric reference (i.e. reference internal to language). The same applies to the category of truth: “The Houses of Parliament are situated on the Thames. This sentence is true.” Here, the second sentence refers to the first, again demonstrating an anaphoric structure. In a similar way, LLMs operate using the aforementioned attention mechanism, which enables them to analyse the relations between expressions within a linguistic context.

10. For more on reasons, see, e.g., Skorupski (2010).

capacity to participate in the game of giving and asking for reasons (Brandom 2001, 81; see also 1998, 230). The conscious human mind has a distinctive capacity for rational action: i.e. using reasons. This concerns action broadly understood: both extra-linguistic and extra-cognitive, as well as linguistic and cognitive. It is precisely consciousness—awareness of one's own rational activity—that constitutes the fundamental difference between human intelligence and AI.

Participation in the game of giving and asking for reasons involves—and this goes right to the core of our linguistic, cognitive, social and discursive practices—the inferential articulation of a claim. It is in inference (reasoning) that a reason for a given claim is obtained (Brandom 2001, 161–65). Understanding a term—knowing its meaning, where this is tantamount to grasping its concept—requires competence in handling the inferences in which the given term is used. Understanding requires knowledge in the practical sense—i.e. a kind of knowing how: the skill that enables one to distinguish which claims are inferentially connected to the claim in which a given term occurs (Brandom 2001, 48). For example, to explicate to someone how to understand the word “friend,” we can present the following inferential consequences of a sentence containing this word:

If *a* is a friend of *b*, then *a* is honest towards *b*.

If *a* is a friend of *b*, then *a* cares about *b*.

If *a* is a friend of *b*, then *a* is faithful to *b*.

If *a* is a friend of *b*, then *a* does not act to the detriment of *b*, etc.

Understanding, then, is a kind of practical competence in respect of knowing how to participate in the game of giving and asking for reasons that support a given proposition, how to establish what the reason for something is, how to distinguish good and bad reasons as well as what the score of the ongoing linguistic-cognitive practice (the game) is, and how to change the score of such a game in which both “parties” have some commitments and entitlements (Brandom 2001, 165). What is at stake here is a linguistically-cognitive, social, normative, and discursive game of giving and asking for reasons, played out in the real world. In specific instances, this may take the form of games such as bridge, checkers, or chess. Understood in this way, such a practice requires the existence of human communities and conscious participants (players). It is therefore not the same as a game of chess played between a human and a machine.

One might critically object that AI is also subject to certain commitments and entitlements—for example, those defined by the rules of chess:

the obligation to make legal moves and the entitlement to move when it is one's turn. The crucial difference, however, lies in the fact that such a machine-player lacks consciousness—as well as everything that is inextricably bound up with the latter.

Our understanding of the concept to be explicated therefore just consists in our practising its inferential use. This concerns the ability to determine what someone is still committed to when they apply that concept. It concerns knowledge about what would entitle someone to do something and what would prevent such entitlement (Brandom 2001, 11). The categories of commitment and entitlement will be elaborated in the next part of the article.

2.2. Normative Categories: Deontic Status and Deontic Attitudes

In Brandom's pragmatic theory, the primary concepts are the normative concepts of deontic status and deontic attitude. Deontic status is a result of someone's attitude towards certain actions, or their effects. Such a status consists of a commitment and an entitlement. The deontic attitude manifests itself in undertaking or attributing a commitment, and attributing an entitlement. Undertaking a commitment is connected with attribution of an entitlement (Brandom 1998, 157 ff.; see also Scharp 2005, 206 ff.). Let us now analyse more precisely what deontic status and deontic attitudes amount to.

2.2.1. Deontic Status: Commitments and Entitlements

The formation and development of concepts—meaning of terms with which we think and speak—takes place by making explicit what is implicit in practices of language use that previously went unquestioned. Meaning, though explicit, can be further explicated in the context of social cooperation. In various conversational situations, interlocutors formulate claims, as well as arguments for or objections against them, and also consider their possible consequences, along with ways of obtaining the entitlements needed to recognise certain claims (Brandom 2001, 149).¹¹ In linguistic practice, we

11. From early childhood—driven by curiosity about reality—we ask questions such as “What is this?”, while simultaneously pointing at the object of our inquiry. In response, we receive successive names, and intuitively begin to form abstract concepts related to the concrete world around us. Through this process, we learn to use appropriate terms for newly recognised, similar things. In training AI, we follow an analogous approach by presenting various examples (patterns). LLMs—unlike humans—have only indirect access to reality, relying on information gathered from various sources (such as the Internet, etc.). By contrast, robots (e.g., humanoid robots)—if equipped with sensors, cameras etc.—have some degree of contact with the real world.

not only form our utterances but also explicate or clarify them. This is done by giving a direct presentation in a clearer form, as well as in discussion, when our claims—even justified ones—are questioned.

We can also speak about content that remains implicit within explicit claims: that is, those implicit inferential consequences of a given claim that have not yet been drawn out. In the context of the network of our inferential practices, expressing acceptance of, or commitment to, one proposition amounts to an implicit expression of acceptance of, or commitment to, another proposition that follows from that claim (Brandom 2001, 18).

For a particular speech act, certain conditions are set under which—according to the practices of the language community—someone is committed or entitled to perform that speech act. What the act changes with respect to other language users, and how the act impacts on their commitments and entitlements, is also determined (Brandom 2001, 129). An important aspect of a discursive practice will be the interpersonal communication through which what individual interlocutors are committed and entitled to comes to be fixed. Entitlements to commitments are reciprocally transferred (Brandom 2001, 165). Subsequent speech acts modify the linguistic practice of the community in question. (This could be a political debate, a marital quarrel, etc.) Individual behaviours modify previous commitments or entitlements. For example, a married couple might be arguing about where to go on vacation. He strongly believes that a trip to London for the Wimbledon Championships is the best option, while she does not like sport, and does not want to go to London. She dreams of going to Paris for sightseeing and shopping. The husband proposes, in conciliatory fashion: “Let’s go to London and Paris! If you go to London, you don’t have to watch tennis. You can see the sights and go shopping. And I—as a reward—will go shopping with you every day in Paris without complaining!”

An instance of correct linguistic communication will consist in providing someone with a sufficient number of clues from which to infer what a given person intends to commit themselves to by making the individual claims that they put forward, and what they are entitling someone else to do through undertaking such commitments. An error in recognising these components will be tantamount to an error in recognising the inferential commitments involved (Brandom 2001, 64).

A given inferential “move” can be justified or entitled by other “moves,” can entitle further “moves,” and exclude yet other ones (Brandom 2001, 162). For example, the commitment associated with asserting that *a* is a dog does not imply a commitment to asserting that *b* is a mammal, but it does imply a commitment to asserting that *a* is a mammal. Similarly the judgment

that *a* is a dog is not incompatible with the judgment that *b* is a fox, but it is incompatible with the judgment that *a* is a fox (Brandom 2012, 295).

As can be seen, commitments and entitlements are fundamental, social, normative aspects of discursive practices (Brandom 1998, 159 ff.). Let us illustrate the categories analysed here with the following example: someone goes to the British Library to open an account. On the one hand, they fill out and sign a form, thereby committing them to behaving according to the library's rules (i.e. committing them to not making noise, not damaging the book collection, not smoking in the reading room, etc.). By possessing a valid card, on the other hand, such a person obtains an entitlement to use the library's resources. This example will be detailed below with categories such as undertaking or attributing a commitment, as well as attributing an entitlement.

2.2.2. Deontic Attitudes: Undertaking or Attributing Commitments, and Attributing Entitlements

To think or say that things are thus-and-so is to undertake a specific type of inferentially formulated commitment towards the claim expressed. It concerns expressing a claim about a given thing in a correct premise in subsequent inferences. It also concerns the entitlement to use such a claim as a premise (Brandom 2001, 11). Commitments are rational when undertaken on the basis of appropriate entitlements obtained on the basis of the reasons one is prepared to entertain (Brandom 2001, 80).

Someone holding a certain belief applies it when undertaking further cognitive commitments. For example, if someone believes that a perceived object is red, then this belief imposes on them a commitment towards the belief that the perceived object is colourful—that it is red, more precisely scarlet, and that it is not green (Brandom 2001, 108–09; see 45–49).

A participant in the game of giving and asking for reasons understands the discursive significance of some speech act if they can attribute to that act the appropriate commitment regarding the claim used in a reasoning. In addition, they must attribute an entitlement to that commitment. Furthermore, the “player” attributes truth to the claim by the very fact of its use (Brandom 2001, 165–69). Let us recall that the attribution of truth to a sentence just consists in its assertion.¹²

Claiming is identical to acknowledging or undertaking a commitment regarding a given proposition. Meanwhile, undertaking a commitment

12. Assertion—even if it actually concerns a true claim—expresses knowledge on condition that someone making it understands the claim (Brandom 1998, 214).

related to a proposition is an act that entitles the attribution of that commitment both to somebody and to the proposition. Undertaking such a commitment can occur through using the proposition as a premise in practical reasoning (Brandom 2001, 173–77).

Here are some examples of practical inferences that clarify the normative aspect of a social, communicative and discursive practice:

- (1) Only opening an umbrella will protect me from getting wet. Therefore,
I ought to open the umbrella.
- (2) I am a bank employee and I am going to work. Therefore, I ought to wear
a necktie.
- (3) By repeating a rumour, I might harm someone without a reason. Therefore,
I ought not to repeat the rumour.

The use of the word “ought” is meant to express the significance of the conclusion, understood as an instance of undertaking a certain commitment (Brandom 2001, 84–85). Someone who evaluates the inference “I am a bank employee and I am going to work. Therefore, I ought to wear a necktie” will find it correct for any person *a* who makes such an inference. The evaluator undertakes a cognitive commitment regarding the inference: namely, that *a* is a bank employee. Such a commitment differs from attributing to someone a certain desire. In this case, the norm (obliging the bank employee to wear a necktie), which is an implicit assumption of the inference, is linked to holding a certain status: namely, the status of a bank employee. It is about holding this status rather than expressing a certain desire or preference. Whether someone has a reason to wear a necktie simply depends on whether they have the relevant status or not (Brandom 2001, 91).

A (discursive) cognitive practice is an activity of deontic “scorekeeping” (Brandom 2001, 81). In updating the “score” acquired by a participant in the linguistic-cognitive game, assertion-making plays a pragmatically significant role (Wanderer 2008, 123). Individual speech acts will change a person’s linguistic-cognitive commitments and entitlements (Brandom 2001, 81). In communicating—through successive instances of “scorekeeping”—various new discursive commitments and entitlements are acknowledged by them and attributed to others, with certain previous commitments and entitlements being rejected (Loeffler 2018, 185).

The socio-historical, discursive process of “scorekeeping” forms a system of acts of undertaking or attributing commitments and entitlements in ongoing interactions. Each participant in such a process constantly

registers and evaluates the actions of the other one (Loeffler 2018, 203–04). Holding a belief is a kind of a commitment understood as taking a certain position within an inferentially connected network. If someone—holding a certain belief—undertakes a specific cognitive commitment, it affects the acceptance of the consequences of that belief as well as the rejection of beliefs that are inconsistent with it (Brandom 2001, 118–19). Individual participants, or at least their acts within the linguistic-cognitive game, are mutually connected, forming a network (system, structure) whose anticipated value is taken to reside in the coherence of beliefs and related commitments or entitlements.¹³

The earlier example concerning the library can be extended by stating that the librarian attributes to someone—who opens an account—the commitment to follow the library regulations. That person then undertakes this commitment, becoming responsible for their behaviour on library premises. At the same time, the librarian attributes to the reader appropriate entitlements as regards using the library resources. Smoking in the reading room would be regarded as a misunderstanding of the regulations, and of the commitments attributed and undertaken. In other words, such behaviour would be inconsistent with those commitments.

CONCLUSIONS

What distinguishes us from AI are our natural bodies. However, humanoid robots are increasingly impressive in their appearance and behaviour.¹⁴ Besides that, we are distinguished by our minds: emotions,¹⁵ free decisions¹⁶ and especially consciousness and reason responsible for rational thinking, the use of language and sensory, intuitive¹⁷ and discursive cognition, as

13. AI—for example, ChatGPT—thanks to the aforementioned attention mechanism, is capable of adhering to the rule of coherence in a generated text. However, AI lacks consciousness, and therefore has no beliefs.

14. See, *inter alia*, Actroid from Osaka University in cooperation with Kokoro Company Ltd. (2005), Honda's Asimo (2011), Nadine from Nanyang Technological University (2014), Hanson Robotics' Sophia (2016), Tesla's Optimus (2023), Boston Dynamics' Atlas Electric (2024), and Human Plus from Stanford University (2024), in: https://en.wikipedia.org/wiki/Humanoid_robot (retrieved July 4, 2025).

15. AI is capable of recognising emotions and adjusting the tone of its speech, as well as its facial expressions, to the emotional state of the interlocutor.

16. Human beings possess free will—naturally grounded and culturally developed. In contrast, AI-controlled vehicles and similar systems are said to exhibit autonomous (independent) behaviour, based on software developed by humans.

17. Intuition and creativity, as distinctive characteristics of human intelligence, constitute a subject that would require a more comprehensive analysis than is possible within the scope of the present article.

well as creativity and memory. However, it is not the case that only we are capable of operating in terms of language, cognition (including both sensory and discursive cognition) and memory: AI also is.

An important difference between AI and our own natural intelligence is our distinctively human capacity for understanding. Human understanding is connected to pragmatic and normative categories such as inference,¹⁸ commitment, the undertaking and attributing of commitments or entitlements, and also the attribution of entitlements related to claims, concepts, etc.

Intelligent machines, on the other hand, seem to understand, but in fact this is the effect of using LLMs, which employ machine learning—applying probabilistic calculations and statistics—executed with access to large databases. For example, if we enter the letter “y” into an Internet search engine, AI will complete this as the phrase “youtube.” This happens not because the search engine understands that this is what we mean and want to watch on YouTube. The search engine performs this completion because previously—many times when typing “y”—we most often added “outube,” and then used the YouTube channel. This is therefore not yet a humanly intelligent act of understanding. One can, however, speak of artificial or machine understanding.

Someone enthusiastic about technology might claim that the topic of the distinctive character of human, natural intelligence, and its distinguishability from AI, is unimportant, as the latter achieves such spectacular results. Yet it is possible that an analysis of the unique characteristics of our own intelligence (such as our understanding, etc.) may prove to be a source of inspiration for the further development of AI. Who knows what the future holds for this rapidly developing field?

REFERENCES

- Audi, Robert. 2004. “Theoretical Rationality: Its Sources, Structure, and Scope.” In *The Oxford Handbook of Rationality*, edited by Alfred R. Mele and Piers Rawling, 17–44. Oxford: Oxford University Press.
- Boden, Margaret A. 2018. *Artificial Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.
- Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Brandom, Robert B. 1998. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

18. Reasoning (inference) is not our exclusive domain, as AI is also capable of performing inference: i.e. analysing related circumstances and presenting the consequences of a given statement. The difference, though, lies in the fact that in our case such inferences are related to additional elements, such as beliefs, commitments, entitlements, etc.

- . 2001. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- . 2012. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford, New York: Oxford University Press.
- . 2019. *A Spirit of Trust: A Reading of Hegel's Phenomenology*. Cambridge, MA: Harvard University Press.
- Dennett, Daniel C. 2013. *Intuition Pumps and Other Tools for Thinking*. New York: W.W. Norton & Company.
- Dreyfus, Hubert L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.
- Eemeren, Frans H. van. 2015. *Reasonableness and Effectiveness in Argumentative Discourse*. Cham: Springer.
- Floridi, Luciano. 2014. *The 4th Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press.
- Hattiangadi, Anandi. 2007. *Oughts and Thoughts: Rule-Following and the Normativity of Content*. Oxford: Oxford University Press.
- Landgrebe, Jobst, and Barry Smith. 2025. *Why Machines Will Never Rule the World: Artificial Intelligence without Fear*. 2nd ed. Abingdon, UK: Routledge.
- Loeffler, Ronald. 2018. *Brandom*. Cambridge: Polity Press.
- Russell, Stuart, and Peter Norvig. 2021. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken, NJ: Pearson.
- Scharp, Kevin A. 2005. "Scorekeeping in a Defective Language Game." *Pragmatics and Cognition* 13 (1): 203–06. <https://doi.org/10.1075/pc.13.1.14sch>.
- Skorupski, John. 2010. *The Domain of Reasons*. Oxford: Oxford University Press.
- Togelius, Julian. 2024. *Artificial General Intelligence*. Cambridge, MA: MIT Press.
- Wanderer, Jeremy. 2008. *Robert Brandom*. Montreal: McGill-Queen's University Press.

Netography

ChatGPT's Language Understanding based on: <https://chatgpt.com> (retrieved July 4, 2025).
 Humanoid robot based on: https://en.wikipedia.org/wiki/Humanoid_robot (retrieved July 4, 2025).

God as Absolute Machine: Aligning Modern Formalisms to Prove God

Ward Blondé

ABSTRACT Let Anselm's God denote that than which nothing greater can be conceived. The rationale of this paper is to show that not only the existence, but also three omni-attributes of Anselm's God—omnipotence, omniscience, and omnipresence—can be defined and proven via modern formalisms. The objective is to do this via a terminological alignment of set theory, mereology and computer science on the one hand, and metaphysics and natural theology on the other. The methodology used consists of a two-step argument: first, if physical entities are of paramount ontological greatness, then God is equal to an absolutely infinitely large, physical universe with omni-attributes. Second, using a slightly different criterion, God can be either abstract, or concrete and non-physical. Some important findings are that (1) a central axiom explains both God and the physical realm, (2) Cantor's Absolute Infinite—and therefore God—can be given a consistent definition, and (3) isolated possible worlds are never observed. The essence of this paper is, in short, that "God is the Absolute Machine."

KEYWORDS Anselm's ontological argument; computer science; Cantor's Absolute Infinite; natural theology; proof of God; set theory.

Acknowledgments:

I thank Emanuel Rutten for the suggestion to investigate infinities, Ludger Jansen for his many suggestions and discussions about sections of the paper, and the anonymous reviewers who gave feedback on earlier versions of this paper.

1. INTRODUCTION

Anselm of Canterbury defined God as “that than which nothing greater can be conceived” (Marenbon 2015). Starting from this definition, he argued that God exists in reality. Many philosophers, such as Aquinas, Descartes, Leibniz, Kant and Plantinga, have commented on his ontological argument, which originated in the 11th century (Oppy 2018).

The aim of this paper is to go further than Anselm’s existential proof and demonstrate that God, so defined, is omnipotent, omniscient, and omnipresent, by using modern formalisms. More precisely, the research objective is to establish proof-enabling definitions of God and God’s omni-attributes that are rooted in a merger of set theory, mereology and computer science on the one hand, and metaphysics and natural theology on the other.

The methodology consists of providing a philosophical justification of a central axiom that underlies the proof of God, followed by the proof in two steps. The first step of the latter is to prove that God is equal to an absolutely infinitely large physical universe, and that this physical universe is omnipotent, omniscient, and omnipresent.¹ This proof makes use of a criterion that takes physical entities to count as ontologically greater than non-physical ones. However, altering this criterion so that abstract entities, or concrete non-physical entities, are held to be ontologically the greatest, is also investigated. This issue is considered in the context of the second step of the proof.

The central axiom essentially states that there are as many abstract entities as physical entities: namely, absolutely infinitely many. This means that the ontologies of the five formalisms (from set theory to natural theology) all have the same structure, and all have absolutely infinitely many elements. Because of this, their ontologies can be aligned. Using these formalisms, the essence of this paper can be expressed in a single short sentence: “God is the Absolute Machine.”

Apart from the question of whether God and God’s omni-attributes can be proven through modern formalisms, the paper addresses the following research questions: (1) Why is there something rather than nothing? (2) Are God and the Absolute Infinite consistent? (3) Which formalisms can prove God and render Him conceivable? (4) Is axiological greatness of superior importance to mereological greatness? and (5) Can set-theoretic realism be reconciled with the observations of the empirical sciences? The last of these questions also represents the most significant research gap in this paper, but can

1. Cantor introduced the Absolute Infinite and associated it with God (Jané 1995). He believed it would be possible to prove that the universe was finite and distinct from the mind of God, which he believed to be absolutely infinite. He never proved these things, however.

be answered, in my view, by appealing to evolutionary conservation (Blondé 2016, 2019) in the context of cosmological natural selection (Smolin 1992).

The next section contains a literature review. Section 3 expresses the technical definitions and the central axiom in the working theory used in the first step of the proof. The methodology in Section 4 provides a philosophical justification of the central axiom and the proofs of God and God's omnattributes. Three important findings are presented in Section 5. A discussion follows in Section 6 and the conclusions are summarized in Section 7.

2. LITERATURE REVIEW

According to Anselm, a non-theist can always conceive of God in their mind. However, a god that exists in reality is even greater than a god that exists only in the mind. Therefore, God, the greatest conceivable being, exists in reality. Kant objected to this argument by appealing to the claim that existence is not a real predicate, meaning that it is a second-level property that cannot be essential with respect to anything (Forgie 2000). This undermines the ontological argument.

Geach's (1956) account of intentional comparatives shows that a phrase like "greater than" does not require either of the entities being compared to actually exist in reality and/or be great. Thus, "that than which nothing greater can be conceived" may function merely within the conceptual space of intentional comparison, without implying greatness of existence in reality. Nevertheless, consider the following two definitions:

1. "that than which nothing greater can be conceived"
2. "that which is greatest"

The greatest entity of (1) requires maximum proof-theoretic strength—at least if we are using theories with axioms and definitions to conceive of entities. The greatest entity of (2), on the other hand, can be interpreted as trivially existing in any theory, weak or strong, that has a greatest entity. As such, it may be the big bang universe, or the smallest transfinite ordinal ω . For this reason, Anselm's definition (1) will be used in this paper. While Kant's and Geach's arguments may very well prohibit Anselm's leap from conceptual to real existence, the proof in this paper does not depend on this. Instead, the proof stands or falls with the central axiom, which is defended via a form of reasoning that is clearly distinct from that of Anselm.

Computer-scientific theories that have similarities with the argument in this paper are those of Jürgen Schmidhuber (2006), Max Tegmark (2008), and Eric Steinhart (2010). The Great Programmer of Schmidhuber runs a program that computes all the computable programs on a universal Turing machine (having memory and available time equal to ω , the smallest infinity).

Our observable universe, which is describable by a finite string of bits S , is computed by one of these Turing-computable programs. However, Blondé (2015) shows that, based on equal probabilities, the program that computes our observable universe computes it infinitely many times. If non-halting programs are banned, then still the program that computes us does not start after any finite time. Consequently, even though S is Turing computable, running the program that computes us requires at least transfinite resources (beyond ω). If, on the other hand, the first string S is chosen that is compatible with our existence, then this will be the execution of a program P that deals with resource limitations. P will, therefore, result in a small universe with a low resolution, or even an artificially intelligent mind that is just intelligent enough to count as being compatible with our existence.

Tegmark augments his Mathematical Universe Hypothesis with the Computable Universe Hypothesis. The first hypothesis is that all structures that exist mathematically also exist physically, and that our observable universe is such a structure. The second hypothesis says that the universe is brought about through Turing-computable functions. Just like Schmidhuber, Tegmark puts a severe limitation on the computational power of what computes the universe. Nevertheless, ordinal machines with memory size beyond that of a Turing machine (this being, therefore, a transfinite size) can compute finite answers to questions that cannot be computed by a Turing machine. Such ordinal machines will inevitably favor or disfavor the abundances of finite worlds that contain a correct or a false answer to questions that are not Turing computable. Moreover, transfinite ordinal machines produce many (infinitely) more worlds than Turing machines. This suggests again that the memory and available time of the ordinal machine that computes the world must be far beyond the countable ω .

According to Steinhart, we live in the first level of absolutely infinitely many (or Ω -many) levels of simulations in simulations. God can then be found at level Ω . This theory comes close to the metaphysics that is proposed in this paper—especially because it uses Ω levels instead of only ω . Two differences stand out: first, Steinhart restricts reality based on the requirement that it consists of levels that simulate other levels through civilizations with computers, and second, he has difficulties explaining why there is something rather than nothing.

3. THEORETICAL FRAMEWORK

In order to define and prove God's existence and attributes, a foundational framework of definitions is called for within the formalisms furnished by set theory, mereology, metaphysics and computer science. All these formalisms

can be translated into formal theories that have a number of entities that can be extended from natural numbers to ordinals and the Absolute Infinite. For this reason, set theory is introduced first, as this is the canonical basis for such formal theories. The central axiom can be found in the subsection dealing with metaphysical definitions.

3.1. *Foundational Set-Theoretic Definitions*

For the purpose of introducing set theory, some theory-related terms will first have to be defined. A sentence will consist of a sequence of symbols from an alphabet, constructed according to the syntactic rules of a language. A collection of sentences will be a theory. A theory will have a model if and only if (henceforth iff) there exists a structure (called a ‘model’) in which all the sentences of the theory are true. A theory will consist of axioms and definitions (sentences assumed as starting points), and further sentences (theorems) will be deduced from them using rules of inference. A theory will be formal iff all its sentences are recursively enumerable by a Turing machine.² An entity *E* exists theoretically (or proof-theoretically) according to a theory *T* if the existence of *E* can be proven in *T*.³

The axioms of the theory ZFC (Zermelo–Fraenkel set theory with the axiom of choice) (Zermelo 1908) are about sets in *V*, the universe of all the pure, well-founded sets, which is known as the ‘von Neumann universe’ (Neumann 1928). ZFC can be strengthened by infinitely many formal, relatively consistent axioms about sets that can neither be proved, nor disproved by ZFC. Any such extension is a formal theory. The weaker theories, with less axioms, are fragments of the stronger, more extended (or more expressive) theories. The theory that extends ZFC with absolutely all the true axioms, jointly making up the largest possible *V*, will be called ZFC^Ω , and is not formal. The definition of *V* via ZFC^Ω implies two things: (1) any formal theory *T*, set-theoretic or otherwise, can be translated into ZFC^Ω via an injection⁴ in *V* of the collection of entities that exist theoretically according to *T*, and, conversely, (2) every set in *V* exists theoretically according to some formal theory.⁵

2. This means that, in formal theories, all the axioms, theorems, symbols, and syntactic rules must be recursively enumerable.

3. Theoretically existing entities do not necessarily exist in reality. All set-theoretic talk about existence is considered to be about theoretical existence, while existence in reality is a first-level property that is a subproperty of the second-level theoretical existence (Forge 2000).

4. An injection will be an all-to-some relation that works via one-to-one links.

5. Note that if this were not the case, we could construct a theory *P* in which the smallest set *x* that does not exist theoretically according to any formal theory exists theoretically

Ordinal numbers include the natural numbers, but also the extensions of the natural numbers beyond ω , the smallest infinity (Cantor 1883). Ω , the Absolute Infinite, exceeds every formally definable (set) ordinal and is itself a proper class ordinal. More technically, Ω is the proper class cardinality of the non-formal class

$$C = \{S_v \mid v \in M\}$$

where:

- the set-theoretic multiverse M is the non-formal proper class of all universe models v of any formal extension of ZFC (Hamkins 2012),
- for each $v \in M$, S_v is a set such that $S_v \in v$, and
- $v \mapsto S_v$ selects exactly one set S_v from each universe model v .

What is innovative about ZFC^Ω is that the number of ZFC^Ω axioms is equal to Ω and can, therefore, not be formally defined.⁶ Because its axioms cannot be listed by an effective procedure, ZFC^Ω escapes Gödel's (1931) incompleteness theorems: it can be both consistent and complete, and it can prove all the arithmetical truths.

The theory NBG (von Neumann–Bernays–Gödel set theory) (Neumann 1928) can also be extended to NBG^Ω , and concerns classes. NBG^Ω extends ZFC^Ω conservatively by distinguishing classes that are sets (namely, the classes that are a member⁷ of some class) from proper classes (all others). Therefore, V is a proper class, also known as the universal class. NBG^Ω has the axiom of limitation of size, which asserts that all proper classes have an equal size: namely, Ω . The theory Morse–Kelley (Wang 1949) with the axiom of Global Choice (GC)⁸ will be called MK, and also pertains to classes. MK extended with all the true formal axioms is MK^Ω . MK^Ω extends NBG^Ω by allowing for recombination (defining new entities via a formula in a comprehension schema) of classes instead of only sets. This makes MK^Ω a maximally expressive meta-theory of ZFC^Ω that can prove the existence of absolutely all classes, including that of the greatest class V .

according to P . Because a set is by definition always an element of some larger set, this makes P , paradoxically, itself formal. Indeed, there would be sets above x that could be recursively enumerated via the definition of x .

6. The problem with a formal theory F that has a formally defined number of formal axioms, is that it can easily be superseded by a slightly stronger formal theory F' that uses one or more extra formal axioms. This severely limits the value of F in a philosophical project.

7. While sets have elements, classes have members.

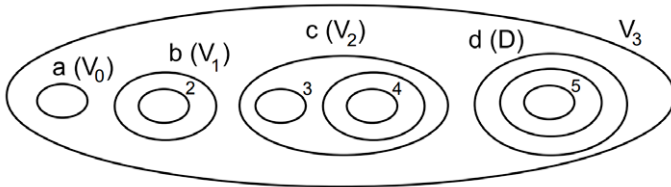
8. Without GC, comparability between proper classes sizes, and therefore the notion of a greatest entity, is lost.

Any axiom or definition that depends on the definition of the Absolute Infinite is not formal, but meta-formal. Therefore, the definitions of Ω , V , V^p , ZFC^Ω , the axioms of NBG^Ω and MK^Ω that extend ZFC^Ω , and several other definitions in this paper, are meta-formal. Nevertheless, meta-formal classes and class ordinals can be used in formal theories, which will consistently reinterpret their absolute nature as something less great than the true absolute nature. This will enable the claim that the existence of God can be proven in formal theories about metaphysical entities—if, at least, we adopt the inferior perspective of these formal theories. Only a meta-formal theory can truly prove God.

3.2. Foundational Mereological Definitions

For set-theoretic classes to be used as a basis for metaphysics, they need to have mereological properties, such as parthood, location, duplication relations, size and abundance (or multiplicity). Therefore, I will first introduce located duplicates. These are not classes, although they are intrinsic duplicates (henceforth ‘duplicates’) of some class.⁹ Moreover, a located duplicate has a location that is defined by a chain of classes that are connected by ‘is a member of’ relations. The located duplicate is a duplicate of the smallest (first) class in the chain and has a location in all the classes in the chain. The largest class in which a located duplicate has a location is the locating class. Concrete examples of the definitions in this and the following paragraphs are given in Figure 1.

Figure 1: What is visualized here is a small class V_3 and the eleven located duplicates of which it is the locating class (V_3 has size ten). V_3 has four elements (or members): V_0 , V_1 , V_2 and D . These are duplicates of the located duplicates a , b , c and d respectively. The abundance of the empty set (V_0) in V_3 is five. V_0 and the five empty located duplicates a , 2, 3, 4 and 5 are all duplicates of each other.



Parthood will be defined as the transitive closure of the membership and the subclass relation, such that every class is a part of V . Parthood is reflexive, while proper parthood is irreflexive. The abundance of a class

9. Intrinsic duplicates have the same internal make-up, although they can have a different external environment.

A in a class B is equal to the number of located duplicates that are duplicates of A and that have B as their locating class. Every set has abundance one in itself and an absolutely infinite abundance in V . The size of a class A is equal to the number of located duplicates that are properly located in A . The empty set has size zero. The size of any proper class is equal to Ω . A relation R between two classes A and B (so A is R -related to B) holds iff a duplicate of A is R -related to a duplicate of B . Therefore, as a mnemonic, the phrase ‘a duplicate of’ may be added between brackets.

3.3. Foundational Metaphysical Definitions

3.3.1. The Central Axiom and the Theory $MK^{\Omega p}$

In order to prove the theological theorems in the Methodology section, the support of a central axiom is invoked that extends MK^{Ω} . It has a metaphysical and a set-theoretic formulation. V^p and V are the classes of, respectively, all non-absolute (or mundane) physical entities and all sets:

Metaphysical: The physical reality is rendered by set theory.

Set-theoretic: There exists a bijection¹⁰ between V^p and V .

$$(\exists f: V^p \leftrightarrow V)$$

The central axiom is consistent with the axioms of ZFC, the Ω formal axioms that extend ZFC toward ZFC^{Ω} , the axioms of NBG^{Ω} , and the axioms of MK^{Ω} set theory. MK^{Ω} + the central axiom will be written as $MK^{\Omega p}$. The theological theorems in the first step of the proof can be proven in every sufficiently expressive formal fragment of the theory $MK^{\Omega p}$, even though only $MK^{\Omega p}$ itself can truly capture the intended meanings of the definitions.

3.3.2. Other Metaphysical Definitions

Proceeding via the definitions in the previous sections, we can construct metaphysical definitions for the entities that exist theoretically in $MK^{\Omega p}$. All these entities potentially exist in reality, and they are either concrete or abstract. Given the assumption of an ontologically parsimonious monism in the first step of the proof, all concrete entities are physical entities.

10. A bijection will be an all-to-all relation that works via one-to-one links. It will be a bidirectional injection, and will prove an equal size.

Every physical entity will be either mundane or transcendent. A mundane physical entity will be definable by a set. All its properties can be fully observed by observers who are themselves mundane physical entities. A physical entity that is too large or too complex to have a set model will be a transcendent physical entity. This means that it would require an absolutely infinite period of time for a mundane observer to observe each property of the entity at some point in time. All physical entities, mundane or transcendent, will be definable by a class.

Both worlds and the universe will be spacetimes. V^p , the union of all the mundane physical entities, will itself not be mundane but the transcendent physical universe, and will have every physical entity as a part. The universe will have an absolutely infinite size.

The remaining metaphysical and theological terms needed for proof of the theorems concerning V^p will be proposed in the subsection that proves these theorems. They include: conceivability, mereological and ontological greatness, and greatness (Theorem 1); causality, direct causation, and omnipotence (Theorem 2); omniscience and direct epistemic access (Theorem 3); omnipresence (Theorem 4); and existence in reality (Theorem 5).

3.4. Foundational Computer-Scientific Definitions

A machine will be a brain, a computer, a being, or a robot, and will be either physical or abstract. It produces an output from a given input. Several abstract models exist for machines: (1) a Turing machine (Minsky 1967), this being an ω -long tape with symbols and a moving read/write head that reads, (over)writes, or moves over the tape, or halts, according to a set of instructions, during ω time instants; (2) a cellular automaton (Wolfram 1984), where cells are made dead or alive according to the states of other cells and, again, a set of instructions; or (3) an artificial neural network (Sharma et al. 2012), consisting of a network of artificial neurons programmed in software that resembles the human brain. Machines that are Turing complete (a requirement that is easily fulfilled) can, given enough (transfinite) time and memory resources, solve any computation problem and simulate any world.

Before producing the next bit of information in the output, a Turing complete machine can consult the whole input. This implies that causal consequences of an elementary section of the memory contents of a machine are non-local: they can travel without any delay through the rest of the memory contents. Consequently, in contrast to Lewis' (1986) modal realism, and in spite of our limited empirical observations, there is no causal isolation in a set-theoretic realism.

Machines are either absolute or ordinal. An ordinal machine (Koepke and Seyfferth 2009) is one of the following three machines: a hypercomputer with transfinite uncountable resources, a Turing machine with countable transfinite resources, or a computer with finite resources. Ordinal machines are, therefore, the ordinal extensions of computers with finite resources. Their available time and memory (tape length, cell matrix, number of neurons, etc.) have ordinal values. Every mundane machine is a physical ordinal machine and is a part of a computationally stronger mundane machine, which has more available time and memory. The Absolute Machine has absolutely infinite available time and memory. The whole physical reality is both observed and computed by the Turing complete Absolute Machine V^p . Even though V^p makes multiplicative computations that include itself as input, it does not output entities larger than V^p , but at most ones identical to V^p .

4. METHODOLOGY

Because our proof of God depends on the central axiom, the latter will first be provided with a philosophical justification. Then the proof will proceed in two steps: in the first of these, it will be proven that a physical God with omni-attributes is the greatest conceivable entity, and in the second, the criterion of ontological greatness will be altered so that a non-physical, concrete God or an abstract God is proven to be the greatest conceivable entity.

4.1. *The Central Axiom: Philosophical Justification*

The central axiom states: There exists a bijection between V^p , the class of all mundane physical entities, and V , the class of all sets in MK^Ω set theory. The philosophical justification of the central axiom is that both physical entities and classes contain information.¹¹

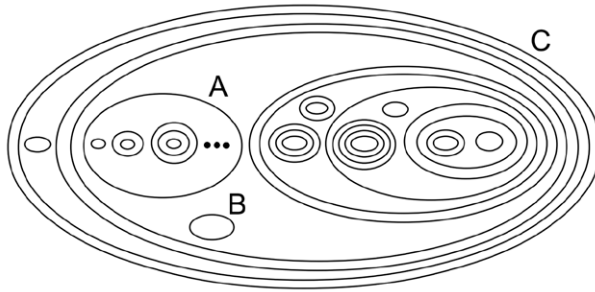
A sufficiently large class contains information that can be derived through a statistical analysis of the parthood relations among the parts of the class. Just like for strings of symbols, the number and size of recurring patterns determine the amount of information.¹² In Shannon's (2001) information theory, the average amount of information in a string of symbols is called 'entropy.' More disordered (entropic) strings contain more

11. In fact, there are at least four kinds of abstract entities that have a bijective correspondence with information entities: sets, ordinal numbers, ordinal machines, and ordinal strings of symbols.

12. The conversion to strings of symbols can easily be made via the notation with curly brackets of pure sets. An example is $\{\{\},\{\{\}\}\}$.

information, whereas more ordered (redundant) strings are easier to analyze. Only classes that are highly ordered or very easy to define can be said to contain little or no information. Three examples are given in Figure 2.

Figure 2: The three classes A , B , and C have various degrees of entropy, or disorder. Class A has an infinite size ω , the smallest infinity. Yet it has so much order that, just like the empty class B , it contains no information. Class C , which has size $\omega+24$, contains roughly 24 bits of information.



The key argument for why abstract entities suffice to bring about physical entities is that observers simulated by an abstract machine cannot conduct any experiment to find out whether they are simulated by an abstract machine or a physical machine with the same information contents. Abstract observers will, therefore, firmly believe they live in a physical world.

4.2. Step 1: A Physical God with Omni-Attributes

Having defended the central axiom and introduced the foundational definitions pertaining to set theory, mereology, metaphysics and computer science in the theory $MK^{\Omega p}$, we are now in possession of the starting material to construct some further definitions and proofs about V^p in this theory. This V^p will be shown to be our preliminary physical God in Theorem 1.

4.2.1. Theorem 1: The Unique Maximal Greatness of V^p

Theorem 1 (V^p is uniquely the greatest of absolutely all conceivable entities) follows from definitions of conceivability, mereological greatness and greatness, and from a criterion of ontological greatness.

We can therefore consider the following definitions and criterion: an entity A is conceivable iff it exists theoretically according to a formal or meta-formal theory. An entity A is greater than an entity B iff either (1) A is ontologically greater than B , or (nonexclusively) (2) A is mereologically greater than B , and B is not ontologically greater than A . An entity A is mereologically greater than an entity B iff B is a proper part of A .

Criterion: if an entity A is physical and an entity B is not physical, then A is ontologically greater than B .¹³

Proof. V^p is conceivable, given that it exists theoretically according to $MK^{\Omega p}$. Moreover, absolutely all other conceivable entities are less great than V^p , because absolutely all other entities that are ontologically equally great compared to V^p , such as worlds, are mereologically less great than V^p . Therefore, V^p is uniquely the greatest of absolutely all conceivable entities. \square

Because V^p is uniquely the greatest of absolutely all conceivable entities, it is equal to Anselm's God.

4.2.2. Theorem 2: The Omnipotence of V^p

Theorem 2 (V^p is omnipotent) follows from definitions of a Turing complete Absolute Machine, the 'causes' relation, direct causation, and omnipotence.

We can consider the following definitions: An entity A causes an entity B iff a machine or process uses (a duplicate of) A as input to bring about, simulate, or output (a duplicate of) B within some spacetime. An entity A directly causes an entity B iff A causes B , and no distinct intermediate entity C is in the causal chain between A and B . An entity is omnipotent iff it causes absolutely every physical entity directly.

Proof. As a physical, Turing complete Absolute Machine, V^p uses itself as input to bring about absolutely every physical entity. Therefore, V^p causes absolutely every physical entity. Because absolutely every physical entity is a part of V^p , V^p causes absolutely every physical entity directly. Therefore, V^p is omnipotent. \square

4.2.3. Theorem 3: The Omniscience of V^p

Theorem 3 (V^p is omniscient) follows from a definition of omniscience and direct epistemic access. We can consider the following definitions: An entity is omniscient iff it has direct epistemic access to absolutely every conceivable entity. Epistemic access is direct iff it is without inference, mediation, or delay.

Proof. Because V^p is a Turing complete machine, it has direct epistemic access to every entity in its memory. Moreover, every entity that is conceivable via $MK^{\Omega p}$ is in the memory of the physical Absolute Machine V^p . That is absolutely every conceivable entity. Consequently, V^p has direct epistemic access to absolutely every conceivable entity. Therefore, V^p is omniscient. \square

13. This criterion will be altered in Step 2 of the proof.

4.2.4. Theorem 4: The Omnipresence of V^p

Theorem 4 (V^p is omnipresent) follows from a definition of omnipresence. We can therefore consider the following definition: An entity A is omnipresent iff A is present at absolutely every physical location in the sense that it is able to act upon and be aware of absolutely every physical event wherever it occurs.

Proof. Given that absolutely every physical entity is a part of V^p and computed by V^p , V^p is able to act upon and be aware of absolutely every physical event wherever it occurs. Therefore, V^p is omnipresent. \square

4.2.5 Theorem 5: The Existence in Reality of V^p

Theorem 5 (V^p exists in reality) follows from a definition of existence in reality of entities and from a corollary of Theorem 2 (V^p causes absolutely every physical entity). We can consider the definition that a physical entity exists in reality iff (1) it potentially exists in reality, and (2) it is causally related to our actually observed world.

Proof. All the entities that exist theoretically according to $MK^{\Omega p}$ potentially exist in reality. Since V^p exists theoretically according to $MK^{\Omega p}$, it potentially exists in reality. According to a corollary of Theorem 2, V^p causes absolutely every physical entity, including our actually observed world. In this way, V^p fulfills all the requirements for existence in reality. Therefore, V^p exists in reality. \square

Given the lack of any causal separations in V^p , absolutely every possible physical entity is required as something that exists in reality, including V^p itself. Together with Theorem 1 and Anselm's definition of God, it then follows that God exists in reality.

4.3. Step 2: A Non-Physical God

In a longstanding tradition, classical natural theology developed the following hierarchy of ontological greatness, from least great to greatest (Leftow 2012):

1. Physical (causal agency, spatiotemporal, contingent, perishable).
2. Abstract (causally inert, non-spatiotemporal, necessary, unchanging).
3. Concrete non-physical (causal agency, personal, imperishable).

From the point of view of the central axiom, this hierarchy is no longer sustained in the same way, because exactly the same entities can be found for physical and for abstract entities. For example, the output of an abstract Turing machine is indistinguishable from that of a physical Turing machine.

Both have the same causal powers and create entities with the same properties. Nevertheless, there remains a difference in perception between the categories. Physical entities are observed by us as very large, but finite, outputs of an Absolute Machine. Abstract entities can be mentally constructed by us, starting from the very smallest units, even though they exist independently of our minds and explain the physical entities. Classically, concrete non-physical entities inherit the best properties of the physical and the abstract entities. For at least categories (1) and (2), the same two levels can be distinguished:

1. Divine: transcendent, absolute, meta-formal, and proper-class-like.
2. Worldly: mundane, exceedable, formal, and set-like.

This analysis makes it hard to decide whether the ontological categories are really different, and also how they relate with respect to ontological greatness. Therefore, I shall adopt the ontology of classical natural theology, with non-physical concrete entities, physical entities and abstract entities, without proposing any hierarchy as to superiority.

If we add a non-physical, concrete entity that potentially exists in reality to what exists theoretically,¹⁴ then we are faced with the question of what is ontologically greatest in Theorem 1: is it the non-physical, concrete entity, physical entities, or abstract entities? According to theorems 2, 3, 4 and 5 in Step 1, there is an entity in reality that is omnipotent, omniscient and omnipresent (namely V^p). With that, we have the following parameterizable proof for either a concrete, non-physical or an abstract God that modifies Theorem 1 and the criterion of ontological greatness:

1. God is uniquely the greatest of absolutely all conceivable entities.
2. Non-physical (either concrete or abstract) entities are ontologically the greatest.
3. If there is an entity in reality with omni-attributes, then God exists in reality and has these omni-attributes.
4. V^p , the physical universe, is an entity in reality that is omnipotent, omniscient and omnipresent.
5. Therefore, God is non-physical (either concrete or abstract), exists in reality, and is omnipotent, omniscient and omnipresent.

Depending on what is ontologically the greatest, we get three possible outcomes: a non-physical, concrete God (the God of classical natural theism), a physical God (V^p), or an abstract God (V).

14. This results in a new language $MK^{\Omega c}$ that has three ontological categories.

5. FINDINGS

The main finding of this paper is that God and His attributes can be defined and proven using modern formalisms. Beyond this, three additional conclusions will be highlighted in this section: (1) that the central axiom exhibits great explanatory power, (2) that God and the Absolute Infinite can be consistently defined, and (3) that in spite of modal realism, the worlds in V^p are not isolated.

5.1. *The Explanatory Power of the Central Axiom*

Apart from sustaining a proof of God, the central axiom explains why there are physical entities at all. Even if there were no physical entities in reality, abstract entities would still exist. We cannot imagine a reality in which, for example, the thirteenth natural number would not be a prime number. Moreover, in contrast to the standard view (Juvshik 2018), abstract entities have causal relations. This is especially apparent for abstract ordinal machines, whose computations and simulations bring about (or cause) abstract worlds. The proposal of the central axiom is essentially that the abstract worlds that are brought about in the simulations of abstract machines explain the physical worlds, and that V explains V^p .

5.2. *The Consistency of God*

Cantor, who introduced Ω , thought of it as an inconsistent multiplicity (Jané 1995), defining it as a set ordinal that exceeds every set ordinal. This may be one of the reasons why Ω has not been popular as a concept among set theorists, who want to avoid inconsistencies in their formal theories. For theologians, an inconsistent God does not look attractive either. My view here is (1) that Ω is a proper class ordinal (and a proper class cardinal), and (2) that the true conception of Ω , and, as a consequence, the true conception of God, cannot be formally defined. This fits much better with our intuitions about God than an inconsistent God.

5.3. *The Invalidity of Modal Realism*

Lewis' (1986) claim that worlds are causally and spatiotemporally isolated is a challenge to the core of the set-theoretic realism proposed in this paper. In order to provide more insight into this conflict, we could establish a proof from contradiction by assuming that V^p is the universe of all the causally isolated worlds and that causal activity only happens within those worlds and not within V^p as a whole. We thereby assume that each world W exists as a causally isolated version, such that it only does so as a member (rather than a part) of V^p , with abundance one. However, taking into account all the

duplicates of W (that stem from recombination of W with other worlds), the world W exists with an absolutely infinite abundance in V^p as a causally interactive part. Assuming that the observation of a world in V^p is random and based on equal probabilities, it follows that the probability that we observe anything causally isolated from V^p will be absolutely infinitely small. Even though isolated worlds may remain technically ‘possible,’ they are non-actual with certainty. Likewise, it may be technically possible that a randomly selected real number is an integer; however, the probability that this will happen is zero.

6. DISCUSSION

This section contains a discussion of (1) differences from classical natural theology, (2) the provability and conceivability of God, (3) God’s omni-attributes, (4) which set-theoretic variant to choose, (5) modal realism and set-theoretic realism as theological projects, and (6) the compatibility of set-theoretic realism with the findings of the empirical sciences.

6.1. *Differences from Classical Natural Theology*

Classical natural theology often injects deep complexity into metaphysics because it introduces concepts that both overlap with and transcend ordinary metaphysical categories (Swinburne 1993): divine foreknowledge and free will, divine providence and determinism, omnipotence and logical coherence, omnibenevolence and evil, divine action and causation, simplicity and the possession of components, the soul and consciousness, eternity and time, and so on. This complexity often stems from the classical assumption that God is clearly distinct from the physical universe and the sentient beings in it, and that He belongs to a different ontological category.

The assumption of a set-theoretic metaphysics significantly simplifies the complex relation between God and physical reality, because God is a limit case of physical reality with respect to the Absolute Infinite. In this case, God’s nature is continuous with the nature of arbitrarily complex, physical agents with the computational power of transfinite ordinal machines. For example, God’s omnibenevolence might be derivable from the collective benevolence of such ordinal machines.

One of the consequences of this simplification is that the proposed theology is most easily interpreted as a form of pantheism or panentheism, especially for those who argue that physical or abstract entities are ontologically the greatest. Also the three proposed omni-attributes are simplified. For example, while omniscience in classical natural theology does not imply that God undergoes our pain, the proposal that God is an Absolute

Machine that computes our pain makes this interpretation more difficult. Nevertheless, it can be argued that God does not suffer by subjectively participating in our pain, because God's brain is absolutely infinitely larger than our brains. Therefore, God's suffering can be ignored.

6.2. *Provability and Conceivability*

As mentioned in Section 3.3.1, the theological theorems in the first step of the proof can be proven in every sufficiently expressive formal fragment of the theory $MK^{\Omega p}$, such as, for example, MK set theory extended with the definitions introduced in this paper (resulting in MK^p). Nevertheless, from the meta-formal perspective of $MK^{\Omega p}$, the absolute entities that are proven to exist via the formal MK^p are less great than those that can be proven by $MK^{\Omega p}$ itself. MK^p holds itself capable of conceiving of God and other absolute concepts, while $MK^{\Omega p}$ recognizes this formal conception as something that is less great as compared to its own true conception of God and these concepts. In other words, while both proving and conceiving God can be done in formal theories, it is only if we admit that the act of conceiving something can be arbitrarily theoretical that we can use the meta-formal theory $MK^{\Omega p}$ to conceive of God as seen from the absolutely greatest perspective. After all, formal theories such as MK^p also make use of the infinitary recursive enumeration of a Turing machine to conceive of their God.

6.3. *God's Omni-Attributes*

Three omni-attributes were defined via modern formalisms: omnipotence, omniscience, and omnipresence. A fourth omni-attribute, omnibenevolence, was found to be too difficult to define and prove in this way. Nevertheless, Blondé (2015) shows that V^p is benevolent as a result of the fact that evil agents often behave benevolently in order to appear benevolent, while benevolent agents never behave malevolently for any reason. Other attributes of God, such as eternity, transcendence and immutability, might be easier to define and prove, but are beyond the scope of this paper.

6.3.1 *Omnipotence*

A question that has often been investigated is whether God can cause evil (Mackie 1955). According to the definition of omnipotence, God causes evil, given that every evil world exists and God causes every world. A related question is whether a universe that is mereologically maximal in size, but that includes evil worlds, can be as great as or greater than a universe that does not include evil worlds (Kraay 2017). I argue that this is indeed the

case, because evil worlds can be brought into a state that is equivalent to non-existence with respect to axiology.

In a mereologically maximally great universe, worlds do not all have equal abundances. For example, non-reproducing worlds and worlds that are not favored by powerful ordinal machines will be outnumbered by worlds that reproduce or are reproduced abundantly. In particular, God can reproduce benevolent worlds with an absolutely infinitely greater abundance than evil worlds. In this case, the evil worlds are with certainty not observed, which reduces them effectively to non-existent worlds.

6.3.2 Omniscience

Simoni (1997: 2) investigated the problem of radical particularity: how can a universal, boundless being know what is experienced by radically finite beings? V^p knows the answer to every mathematical problem that requires transfinitely long computations, such as the halting problem (Lucas 2021). It also knows every mundane experiential fact from its own experience (because every mundane machine—including, for example, every neural network—is a part of V^p), and it knows every indexical (e.g. temporal) fact relative to a world. Moreover, because every recombination of V^p with itself results in V^p , (a duplicate of) every formal fact (including every finite fact) is taking place at some location at each temporal instant of V^p 's existence. Therefore, God knows everything during every instant of time, including our human experiences.

6.3.3 Omnipresence

In analyzing omnipresence, Hudson (2009) distinguishes several mereological relations between physical entities and locations, such as being entirely located at, being wholly located at, and being partly located at. For each physical entity, these three relations hold with respect to the location V^p .

Hartshorne (1941), in arguing for a non-physical God, proposed the analogy that God is to the world like a mind is to its body. Therefore, God has immediate knowledge and direct power over every part of the universe. Indeed, as a Turing complete Absolute Machine, V^p has immediate and direct read and write access to its entire memory contents (which is V^p itself).

6.4 Which Set-Theoretic Variant to Choose?

One could argue that the set-theoretic theory ZFC, which is the canonical theory in mathematics, cannot prove the existence of a greatest entity. However, this only shows that ZFC can be dismissed as being too weak to prove the existence of a series of meaningful meta-concepts, such as the

universal class. Seeking refuge in this argument comes down to admitting that every arbitrarily godlike entity provably exists in the physical universe, but that the physical universe itself does not provably exist. In natural language (such as English), we have all the meaningful concepts and meta-concepts at our disposal. A correct translation into set theory can therefore be arrived at using the theory MK, which is a sufficiently strong meta-theory of ZFC.

6.5. Proving God via Modal Realism

Modern formal attempts to prove the existence of God, such as Plantinga's (1978) and Gödel's ([1941] 1995) ontological arguments, have made use of modal logic. According to modal logic, things exist possibly when they exist in some possible world, and necessarily when they exist in every possible world. Gödel defines God as having all and only the positive properties in a given possible world, and claims that existence is a positive property. He then uses a strict system of axioms and definitions to show that such a God exists in every possible world. Therefore, God exists necessarily. Plantinga posits that if God exists possibly, then God exists necessarily, and also that it is possible that God exists. Therefore, again, God exists necessarily.

What remains unclear is whether Gödel and Plantinga have shown the existence of one God only, or one God for each possible world. We get many gods if all the possible worlds are causally and spatiotemporally isolated and just as real as the actual world (our world), as advocated by Lewis. Moreover, if some possible worlds have few positive properties apart from existence, Gödel's God is not necessarily a great being. Given these problems with modal realism, the set-theoretic metaphysics becomes an important extra option for the theist who wants to prove a unique, maximally great God.

6.6. Compatibility with the Empirical Sciences

The most significant research gap in this paper pertains to explaining why the observable universe is not extremely complex and benevolent, and why the empirical sciences are so successful. As an explanation, I refer to the present author's account of evolutionary conservation in cosmological natural selection (Blondé 2016, 2019). According to this theory, biological organisms with more than three spatial dimensions require vastly more time to develop Darwinistically. Because of this, they emerge in a world in which three-dimensional organisms like us are already present. Some of them will therefore extend, reuse, or build on the complexity that we create, such that they become evolutionarily dependent on us. This implies that these 3D-extendors have to operate with the greatest

respect for the evolutionarily conserved reproduction plan of our observable universe, in order to maintain and reproduce themselves. Assuming that the knowledge gathered by our empirical sciences has become evolutionarily conserved long before the advent of higher-dimensional organisms, it means that the latter have to be very subtle when they interfere. On the other hand, they will want to reproduce our observable universe as much as possible in order to reproduce themselves. Consequently, they will simulate our observable universe in a way that does not alter the knowledge gathered by our empirical sciences, though with less spatial requirements. This logic can repeat itself for absolutely infinitely many spatial dimensions, via 4D-extendors, 5D-extendors, etc.

This theory is compatible with the requirement of Blondé and Jansen (2021) that V^p consists primarily of conscious, intelligent matter (brain matter or CPU matter) if it is to solve the ‘hard problem’ of consciousness (Chalmers 2017). Because complexity in lower dimensions is always more efficiently simulated in higher ones, the density of conscious, intelligent matter will increase as we pass over, at the limit, to absolutely infinitely many spatial dimensions. The ultimate reality, therefore, will consist for 100% of conscious, intelligent matter, which we can call ‘God’s brain.’ However, God’s brain creates experiences of an evolutionarily conserved external world that is located in an absolutely infinitely small fraction of V^p .

7. CONCLUSIONS

The deductive proofs of five theological theorems on the basis of MK set theory extended with absolutely infinitely many axioms, a central axiom, and a list of definitions show that it is possible to define and prove the existence of God, and His attributes, via modern formalisms, if at least the central axiom and the definitions are accepted. Three of God’s omni-attributes—omnipotence, omniscience, and omnipresence—appear to be translatable into a merger of set theory, mereology, computer science, metaphysics and natural theology. As regards the proof of these attributes, God is, in the first step, identified with a physical analogue of the von Neumann universe of sets V : namely, the universe of all mundane physical entities V^p . In the second step, we have found it possible to reuse the preliminary proof of a physical God as a proof of a non-physical, concrete God, or an abstract God.

Some other important findings are that the central axiom explains both God and physical reality, that God and the Absolute Infinite can be consistently defined, and that modal realism with its isolated worlds is invalid. The philosophical justification of the central axiom is that mathematical

truths and abstract entities exist independently of physical reality. This includes ordinal machines and the Absolute Machine (which is God).

This approach to translating theological terminology into modern formalisms has also revealed some limitations. First, the translation provides no insight into how the existence of absolutely infinitely many physical entities can be reconciled with the observations of the empirical sciences. The solution probably has to be found in other paradigms, such as evolutionary conservation in cosmological natural selection. Second, not all theological terms are suited to being translated into modern formalisms. Omnibenevolence is one example.

In conclusion, I recommend giving up on Lewis' isolated worlds in modal realism. They may exist, but the probability that our observable universe is part of one is zero. Instead, theists would do better by using Cantor's paradise of a unifying set-theoretic realism to prove God.

REFERENCES

- Blondé, Ward, and Ludger Jansen. 2021. "Proving God without Dualism: Improving the Swinburne-Moreland Argument from Consciousness." *Metaphysica* 22 (1): 75–87. <https://doi.org/10.1515/mp-2020-0035>.
- Blondé, Ward. 2015. "An Evolutionary Argument for a Self-Explanatory, Benevolent Metaphysics." *Symposion* 2 (2): 143–66. <https://doi.org/10.5840/symposion2015228>.
- . 2016. "Can an Eternal Life Start from the Minimal Fine-Tuning for Intelligence?" *Philosophy and Cosmology* 17: 26–38.
- . 2019. "EMAAN: An Evolutionary Multiverse Argument against Naturalism." *Symposion* 6 (2): 113–28. <https://doi.org/10.5840/symposion2019629>.
- Cantor, Georg. 1883. "Über unendliche, lineare Punktmannichfaltigkeiten." *Mathematische Annalen* 21 (4): 545–91.
- Chalmers, David J. 2017. "The Hard Problem of Consciousness." In *The Blackwell Companion to Consciousness*, edited by Susan Schneider and Max Velmans, 32–42. 2nd ed. Malden, MA: Wiley-Blackwell.
- Forgie, J. William. 2000. "Kant and Frege: Existence as a Second-Level Property." *Kant-Studien* 91 (2): 165–77. <https://doi.org/10.1515/kant.2000.91.2.165>.
- Geach, Peter. 1956. "Good and Evil." *Analysis* 17 (2): 33–42. 10.1093/analys/17.2.33.
- Gödel, Kurt. (1941) 1995. "Ontological Proof." In *Collected Works*, vol. 3, *Unpublished Essays and Lectures*, edited by Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons, and Robert N. Solovay, 403–4. Oxford: Oxford University Press.
- Gödel, Kurt. 1931. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." *Monatshefte für Mathematik und Physik* 38 (1): 173–98.
- Hamkins, Joel David. 2012. "The Set-Theoretic Multiverse." *Review of Symbolic Logic* 5 (3): 416–49. <https://doi.org/10.1017/S1755020311000359>.
- Hartshorne, Charles. 1941. *Man's Vision of God and the Logic of Theism*. Chicago: Willett, Clark & Co.
- Hudson, Hud. 2009. "Omnipresence." In *The Oxford Handbook of Philosophical Theology*, edited by Thomas P. Flint and Michael C. Rea, 199–216. Oxford: Oxford University Press.

- Jané, Ignacio. 1995. "The Role of the Absolute Infinite in Cantor's Conception of Set." *Erkenntnis* 42 (3): 375–402. <https://doi.org/10.1007/BF01129011>.
- Juvshik, Tim. 2018. "Abstract Objects, Causal Efficacy, and Causal Exclusion." *Erkenntnis* 83 (4): 805–27. <https://doi.org/10.1007/s10670-017-9915-1>.
- Koepeke, Peter, and Benjamin Seyfferth. 2009. "Ordinal Machines and Admissible Recursion Theory." *Annals of Pure and Applied Logic* 160 (3): 310–8. <https://doi.org/10.1016/j.apal.2009.01.005>.
- Kraay, Klaas J. 2017. "Invitation to the Axiology of Theism." In *Does God Matter? Essays on the Axiological Consequences of Theism*, edited by Klaas J. Kraay, 1–35. New York: Routledge.
- Leftow, Brian. 2012. *God and Necessity*. Oxford: Oxford University Press.
- Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
- Lucas, Salvador. 2021. "The Origins of the Halting Problem." *Journal of Logical and Algebraic Methods in Programming* 121:100687. <https://doi.org/10.1016/j.jlamp.2021.100687>.
- Mackie, John L. 1955. "Evil and Omnipotence." *Mind* 64 (254): 200–12. <https://doi.org/10.1093/mind/LXIV.254.200>.
- Marenbon, John. 2015. "Anselm: Proslogion." In *Central Works of Philosophy*. Vol. 1, *Ancient and Medieval*, edited by John Shand, 45–67. Durham: Acumen Publishing Limited.
- Minsky, Marvin L. 1967. *Computation: Finite and Infinite Machines*. Englewood Cliffs, NJ: Prentice-Hall.
- Neumann, John von. 1928. "Die Axiomatisierung der Mengenlehre." *Mathematische Zeitschrift* 27 (1): 669–752.
- Oppy, Graham, ed. 2018. *Ontological Arguments*. Cambridge: Cambridge University Press.
- Plantinga, Alvin. 1978. *The Nature of Necessity*. Oxford: Clarendon Press.
- Schmidhuber, Jürgen. 2006. "A Computer Scientist's View of Life, the Universe, and Everything." In *Foundations of Computer Science: Potential—Theory—Cognition*, edited by Christian Freksa, Matthias Jantzen, and Rüdiger Valk, 201–8. Berlin: Springer.
- Shannon, Claude E. 2001. "A Mathematical Theory of Communication." *ACM SIGMOBILE Mobile Computing and Communications Review* 5 (1): 3–55. <https://doi.org/10.1145/584091.584093>.
- Sharma, Vidushi, Sachin Rai, and Anurag Dev. 2012. "A Comprehensive Study of Artificial Neural Networks." *International Journal of Advanced Research in Computer Science and Software Engineering* 2 (10): 278–84.
- Simoni, Henry. 1997. "Omniscience and the Problem of Radical Particularity: Does God Know How to Ride a Bike?" *International Journal for Philosophy of Religion* 42 (1): 1–22. <https://doi.org/10.1023/A:1002922511041>.
- Smolin, Lee. 1992. "Did the Universe Evolve?" *Classical and Quantum Gravity* 9 (1): 173–91.
- Steinhart, Eric. 2010. "Theological Implications of the Simulation Argument." *Ars Disputandi* 10 (1): 23–37. <https://doi.org/10.1080/15665399.2010.10820012>.
- Swinburne, Richard. 1993. *The Coherence of Theism*. Rev. ed. Oxford: Clarendon Press.
- Tegmark, Max. 2008. "The Mathematical Universe." *Foundations of Physics* 38: 101–50. <https://doi.org/10.1007/s10701-007-9186-9>.
- Wang, Hao. 1949. "On Zermelo's and von Neumann's Axioms for Set Theory." *Proceedings of the National Academy of Sciences* 35 (3): 150–5. <https://doi.org/10.1073/pnas.35.3.150>.
- Wolfram, Stephen. 1984. "Cellular Automata as Models of Complexity." *Nature* 311 (5985): 419–24. <https://doi.org/10.1038/311419a0>.
- Zermelo, Ernst. 1908. "Untersuchungen über die Grundlagen der Mengenlehre I." *Mathematische Annalen* 65 (2): 261–81. <https://doi.org/10.1007/BF01449999>.

The Nature of Monotheism A Philosophical Explication

Joshua Sijuwade

ABSTRACT This article develops a philosophical explication of monotheism through fundamentality, using Rudolf Carnap's method of explication and Karen Bennett's concept of 'building-fundamentality.' By examining how contemporary philosophers and theologians have struggled with defining monotheism in light of Second Temple Judaism's complex theology, this article argues that understanding monotheism as the belief in one fundamental deity provides a more philosophically robust framework than numerical definitions. This framework helps reconcile divine plurality in Jewish theology while offering new perspectives on polytheistic traditions and interfaith debates, thus contributing to broader discussions in the philosophy of religion and theology.

KEYWORDS Bennett, Karen; Carnap, Rudolf; monotheism; Second Temple Judaism.

1. INTRODUCTION

According to the common religious belief, the term ‘monotheism’ has a specifically numerical focus. Most of the major world religions either incorporate the notion of monotheism into their belief system (such as Christianity, Judaism, Islam and Sikhism) or, at least, define their own form of theism or non-theistic belief in light of this notion (such as that of Hinduism and Buddhism). Now, as expressed by the *Oxford English Dictionary*, the term ‘monotheism’ is to be defined as follows:

- (1) (Monotheism) The belief that there is only one ‘god.’

The notion of monotheism finds its etymological roots in the seventeenth-century works of the Cambridge Platonists (Macdonald 2003). In their writings, the Platonists contested that the semantic and intellectual context of the term demonstrated that the antonym of ‘monotheism’ was in fact ‘atheism,’ rather than the more commonly entertained ‘polytheism’ (MacDonald 2003). At its root, a ‘monotheist’ was thus an individual who believed that “there was a ‘god’ with a certain nature,” rather than someone who believed that “there was only one spiritual entity that could or should be named ‘God’.” However, this more accurate understanding of the term has unfortunately been neglected in contemporary discourse. Instead, the contemporary position is now focused on the term as encountered within later Enlightenment and deistic thought, with the primary use of the word there being that of providing a framework or matrix to assess the truth value of a religion within a European context. Specifically, it finds its origins in the work of Henry More (Smith 2013). ‘Monotheism,’ as a term applied to Enlightenment and deistic thinkers, thus served as an organising principle for the categorisation of religious concepts according to their intellectual claims, with priority given in the classification and evaluation of religions to the question of the number of ‘gods.’ The specific problem that this term presents us with reflects the assumption that, when further elucidated, it commits one to the following explication—which we may call ‘Deity Monotheism’ and can state succinctly as follows:¹

- (2) (Deity Monotheism) The belief that there is only one deity.

In a contemporary context, the definition of monotheism has been a central concern in the philosophy of religion, with scholars like Richard Swinburne

1. The terms ‘deity’ and ‘divine being/person’ will be used interchangeably throughout this text, without any difference in meaning.

(2016), Brian Leftow (2016) and Dale Tuggy (2017) largely adopting numerical definitions that focus on counting divine beings. This numerical approach, while intuitive, has also been defended by philosophers such as Swinburne (2016), who argues that monotheism must mean belief in exactly one divine being if it is to maintain conceptual clarity. Similarly, Leftow (2016) contends that any definition allowing for multiple divine beings undermines its core meaning. However, this contemporary construal in (2), focusing on the belief in a single deity and the denial of other divine beings, has led some historians, such as Peter Hayman (1991), Nathan MacDonald (2003) and Paula Fredriksen (2006) to argue that ‘monotheism’ is an inappropriate term for describing biblical teachings—as Fredriksen (2022) notes, ancient belief centred on loyalty to ancestral traditions rather than mental assent to propositions. These scholars point to explicit biblical passages that appear to acknowledge the real existence of other divine beings. For instance, when Deuteronomy 32:8–9 describes Yahweh allotting nations to other ‘gods,’ or when Psalm 82:1 depicts God standing in the divine council among other ‘gods,’ these texts seem to presuppose the actual existence of these beings rather than merely acknowledging that some people believe in them. As Michael Heiser (2008) argues, it would be nonsensical for biblical authors to describe Yahweh allocating nations to non-existent beings or judging among figments of the imagination. And so the evidence from the period surrounding the biblical texts seems to support the existence of other divine beings, which is incompatible with the contemporary definition of the term. Indeed, the concept of ‘monotheism’ in ancient Jewish and early Christian contexts is more nuanced than is commonly understood. That is to say, the notion of ancient Jewish monotheism—and thus not the current dictionary definition of the term—can be construed as follows:

- (3) (Ancient Jewish Monotheism) A religious worldview that acknowledged the existence of multiple divine beings (אלוהים) while maintaining that Yahweh was uniquely supreme and exclusively worthy of worship due to his transcendent attributes as the sole creator and sovereign ruler of al. Reality, rather than a belief system focused on the numerical oneness of deity.

The Jewish scriptures themselves, as Fredriksen (2022) notes, attest to the existence of multiple divine beings alongside the ‘god’ of Israel. These passages are best understood not as rhetorical accommodations to pagan audiences, but as genuine theological statements about the structure

of divine reality. Passages like Exodus 12:12 (“On all the gods of Egypt I will execute judgments”), Exodus 15:11 (“Who is like you among the gods?”), Exodus 18:12 (“Now I know that the Lord is greater than all gods”), Psalm 82:1 (“god stands in the divine council; in the midst of the gods he passes judgment”), Psalm 97:7 (“All the gods bow down to him”), Deuteronomy 32:43 (“Worship him, all you gods”), and Micah 4:5 (“All the peoples walk, each in the name of its god; but we will walk in the name of the Lord, our god forever and ever”) acknowledge the presence of other ‘gods,’ even as they affirm the supremacy of the Jewish ‘god.’ The evidence for genuine divine plurality becomes even clearer when we examine how ancient Jewish and early Christian authors engaged with these ideas. Philo, a prolific philosopher and exegete, seamlessly incorporates Hellenistic ideas into his interpretations of Jewish scripture. In his cosmological treatises, he speaks of the stars as “visible gods,” acknowledging their divine status within a hierarchical framework that places the Jewish ‘god’ at the apex. The apostle Paul, whose epistles constitute the earliest surviving Christian writings, also grapples with the reality of multiple divine beings. As Fredriksen (2022) notes, throughout his letters Paul frequently acknowledges the existence and influence of pagan ‘gods,’ presenting them as formidable spiritual powers that Christ must ultimately conquer. For example, Paul speaks of the “god of this age” (2 Corinthians 4:4), “principalities,” “powers,” “the rulers of the darkness of this age,” and “spiritual hosts of wickedness in the heavenly places” (Ephesians 6:12) that oppose him and the gospel. The religious sensibilities of the ancient world, as noted by Fredriksen (2022) and as reflected in the thought of figures like Philo, Herod and Paul, were characterised by a conception of divinity as a spectrum rather than a binary. Other scholars who reject this evolutionary paradigm tend to assume that passages evincing divine plurality are actually speaking of human beings, or that the other ‘gods’ are merely idols.

Now, Heiser (2008) argues that the passages in Deuteronomy 4 and 32 that affirm the existence of other ‘gods’ must be contextualised in light of the Most High’s dealings with the Gentile nations and the ‘gods’ he appointed to govern them. It would be nonsensical to conclude that Deut 4:19–20 and 32:8–9 show Yahweh giving the nations up to the governance of non-existent beings. The writer is not suggesting that Yahweh allotted non-existent beings to the nations so as to explain why the nations outside Israel worship such non-existent beings. The implication is that the declarations of Deut 4:35, 39 and 32:12, 39 are best understood as reflecting a worldview that accepted the reality of other ‘gods,’ along with Yahweh’s utter uniqueness among them, not a worldview that denied the existence

of lesser *myhla*. To further understand this nuanced perspective, we turn to the historical conceptualisation of monotheism within Second Temple Jewish belief, best exemplified by the use of the term אֱלֹהִים (*Elohim*; 'god' or 'gods'). One of the primary usages of אֱלֹהִים in the Hebrew texts that require explanation is that of Psalm 82:1, which states: "'god' [אֱלֹהִים] stands in the divine council; in the midst of the 'gods' [אֱלֹהִים] he passes judgment." According to Heiser (2008), the first אֱלֹהִים clearly refers to a singular entity ('god') due to subject-verb agreement. However, Heiser (2008) notes that the second אֱלֹהִים is plural, since 'god' cannot be said to be standing in the midst of a (singular) 'god' or himself. Furthermore, another interpretation of אֱלֹהִים within the ancient Jewish and Second Temple Jewish worldviews involves construing it as a 'place of residence' term, as Heiser (2008) suggests. In this context, אֱלֹהִים does not ascribe a specific set of attributes to its referent, but simply identifies the proper domain of reality of the referent. As Heiser (2008) notes, all אֱלֹהִים are members of the unseen spiritual world, which is their place of residence. In that realm there is ranking, hierarchy and differentiation of attributes. Moreover, as Hurtado (2004) highlights, the ancient Jewish religious outlook constituted a distinctive version of the commonly attested belief structure at the time—a 'high god' who presided over other deities (Hurtado 2004, 129). While the ancient Jewish view shares similarities with its broader religious environment, a distinctive factor, as Hurtado (2004) points out, was their concern for 'god's' supremacy and uniqueness. The ancient Jews upheld this with an intensity and solidarity that seemed to surpass anything previously known in the Greco-Roman world, according to Hurtado (2005, 130). Within ancient Jewish belief, Yahweh, as אֱלֹהִים, was not held to be one among equals, but rather 'species-unique,' in that He was incomparable and unique in terms of His attributes, as noted by Heiser (2008).

This concept of 'species-uniqueness' can be further understood through Bauckham's (2008) exploration of 'transcendent uniqueness' or 'divine identity.' The latter holds that, for the monotheism of Second Temple Judaism, God is identified by features in two categories: (a) his relationship to Israel and (b) his relation to reality. Category (a) includes (a1) God having the unique name 'Yahweh' and (a2) Yahweh bringing Israel out of Egypt, emphasizing his covenantal relationship. However, the Jews focused more on category (b) to distinguish Yahweh's uniqueness relative to all of reality. This included (b1) Yahweh as sole creator, (b2) Yahweh as sovereign ruler, and (b3) Yahweh as the only being worthy of worship. That is, firstly, with respect to (b1), God is the sole creator of all things: He creates all things outside of Himself and is seen as the sole actor in His creative activity

(Isaiah 44:23–24). It is God alone who brought all other beings into reality, without assistance or through any intermediary agent. God alone is the creator of all things, and no other being takes part in this activity (Bauckham 2008). Secondly, where (b2) is concerned, God is the sole sovereign ruler over all things. All other things, including beings worshipped as ‘gods’ by non-Jews, are subject to Him in that He reigns supreme over all things outside of Himself (Bauckham 2008). All reality, outside of God, is thus in ‘strict’ subordination, as serving Him. There are no co-rulers with God. Lastly, when it comes to (b3), God is the only being worthy of worship, which involves recognising that worship was the appropriate response to a being who had the unique identifying attributes of (b1) and (b2). Thus, as God is the sole being who possesses these attributes, He is the only being worthy of worship. This prescription to worship God alone is thus grounded upon an acknowledgement of God’s transcendent uniqueness and identity as sole creator and ruler (i.e. [b1] and [b2]) (Bauckham, 2008). God’s unique identity, and the exclusive worship of him, were correlated with, and reinforced by, each other. Thus, in answer to the question of why the Jews would not worship any other being than “the one God,” the simple answer was that they were created by Him and are subject to Him, with any good that comes to them ultimately finding its source in God. These features, according to Bauckham (2008), establish a clear and absolute distinction between God and all else in reality. They enabled ancient Jewish believers to define the uniqueness of God, marking Him out from all of reality, as Bauckham (2008) suggests. This means that, based on these criteria, all other beings, even the אֱלֹהִים, are His creatures and subjects (Bauckham 2008). The widely attested position of Second Temple monotheistic belief is thus grounded upon the expression of God’s utter uniqueness—His ‘species uniqueness’ and ‘transcendent uniqueness’—rather than a negation or denial of the existence of other divine beings, as noted by Heiser (2012, 6). As Heiser (2008) explains, an entity that is ‘species-unique’ possesses at least one attribute not shared by any other member of the species. In other words, a species-unique being need not be unique in every attribute, but must be set apart in ways that are completely unique (Heiser 2012, 30). The uniqueness of Yahweh among the existing אֱלֹהִים, according to Heiser (2008, 29), was thus an incontestable position within ancient Jewish theology. Taking this into account, we see that the religious outlook of the ancient Jews was indeed ‘strict’ or ‘exclusivist.’ However, this did not negate the existence of any other divine agents beyond the god of Israel. Instead, the ‘strict’ monotheistic focus was on the uniqueness of God rather than being a numerical focus, as noted by Wright (1992, 259). For the latter,

along with Heiser and Hurtado, the monotheism of the Second Temple period is not associated with any type of numerical or quantitative oneness, but is solely a qualitative concept focused on the difference between the unique features of Israel's God and all other types of reality. Israel's God is one because of His uniqueness.

Given the evidence, Fredriksen and others have called for the term's "mandatory retirement." However, along with individuals such as Heiser (2008) and Hurtado (2004), I would argue that this is not necessary. The problem with the term arises from the assumption that when it is further elucidated one is committed to an explication of 'monotheism' as the belief in a single deity. Instead, I propose an alternative explication of the term, which we can call 'Fundamentality Monotheism.' This specific explication is compatible with the beliefs and practices evident in the biblical texts and the surrounding period—as it acknowledges the existence of a single 'fundamental' deity while allowing for the presence of subordinate divine beings. Hence, by adopting this more nuanced definition of 'monotheism,' which emphasises the belief in a single fundamental deity rather than the exclusion of all other divine entities, we can reconcile the concept with the religious landscape of the biblical period—and other religious periods as well—and avoid the need to retire the term altogether.

The central focus of this article will thus be to advance an alternative conception of monotheism, namely Fundamentality Monotheism, which reconciles the acknowledgement of a deity with the existence of subordinate divine beings. By employing Rudolf Carnap's method of explication, this approach refines the concept of monotheism so that it becomes "the belief in a single fundamental deity" and lays the ground for testing it against Carnap's criteria of adequacy—similarity, exactness, fruitfulness and simplicity—to ensure a precise and theoretically fruitful definition. To support this explication, there will be an analysis of it in light of the historical and textual evidence from biblical and Second Temple Jewish periods introduced above. Ultimately, it will be shown—once an important objection related to the notion of polytheism, and a certain interfaith implication, have been elucidated—that this explication of the notion of monotheism is indeed fruitful, and so should be adopted as the correct definition of this important theological term.

As regards the structure of the remainder of the article, Section 2 ("Building-Fundamentality") explores fundamentality in divine entities through Bennett's framework. Section 3 ("The Method of Explication") then details Carnap's method of explication and its criteria. Section 4 ("Defining 'God'") articulates how 'god' can be understood as a fundamental deity, Section 5

(“Explicating Monotheism”) subsequently applies Carnap’s method with a view to developing Fundamentality Monotheism. Section 6 (“The Pantheon Objection”) addresses challenges from polytheistic traditions, and Section 7 (“Interfaith Implications of Explication”) explores implications for debates over Trinitarianism and Unitarianism. Finally, the concluding section (“Conclusion”) will summarise the findings of the article.

2. BUILDING-FUNDAMENTALITY

In contemporary metaphysics, the notion of fundamentality is used in reference to an entity (or entities) that is (or are) basic, primitive or rock-bottom in the hierarchical structure of reality. Karen Bennett (2017) defines it as follows:

- (4) (Fundamentality) x is fundamental if it is independent (i.e. unbuilt/ungrounded) and complete (i.e. the builder/ground of everything else).

Two central aspects of the notion of fundamentality, for Bennett (2017), are those of ‘independence’ and ‘completeness’ (with the former being more central to the notion than the latter), which can be construed succinctly as follows:

- (5) (Independence) x is independent if nothing builds x .
 (6) (Completeness) The set of the xx s is (or the xx s plurally are, or a non-set-like x is) complete at a world w just in case its members build everything else at w .

These concepts, in Bennett’s (2017) thought, are intimately tied to the notion of ‘building,’ which is a technical term encompassing various relations, such as composition, constitution, set-formation, realization, micro-based determination, grounding, and causation. Building relations form a unified family, characterized by three essential conditions: directedness (antisymmetric, irreflexive, asymmetric), necessitation (builders necessitate what they build), and generation (built entities exist in virtue of their builders). Within this building-fundamentality framework there is, according to Bennett (2017), a deflationist view of fundamentality, where fundamentality facts are simply building facts. This perspective reflects the familiar phrase ‘unexplained explainers,’ with independence representing the ‘unexplained’ aspect and completeness embodying the ‘explainers’ aspect.

Now, among the various building relations, the specific relation of ‘grounding’ holds a special place. And so, following Jonathan Schaffer (2016), we can construe the nature of this relation as follows:

- (7) (Grounding) An asymmetric, necessitating dependence relation that links the more fundamental entities to the less fundamental entities, and thus backs a synchronic metaphysical explanation for the existence and nature of an entity in virtue of the existence and nature of another, more fundamental entity.

Grounding thus serves to impose a hierarchical structure on reality, connecting more fundamental entities to less fundamental ones. It fulfils two crucial roles: explanatory and generative. In its explanatory capacity, grounding provides the basis for synchronic metaphysical explanations of less fundamental entities in terms of more fundamental ones—that is, grounding is a relation that ‘backs’ a synchronic metaphysical explanation (e.g. an H_2O molecule exists at a particular time in virtue of two H atoms and one O atom, or the singleton set Socrates exists *at a particular time* in virtue of the existence of Socrates), in the same manner that the relation of causation ‘backs’ the diachronic causal explanation of the existence of an entity or event (e.g., a radioactive isotope of carbon-14 exists at t_2 in virtue of the decay of a neutron into a proton in a nitrogen-14 atom at t_1 , through weak nuclear interaction). Meanwhile, its ‘generative’ role is reflected in its ‘super-internal’ nature, where the existence and intrinsic nature of one relatum ensure both the obtaining of the relation and the existence and nature of the other relatum. Moreover, this conception of grounding leads to its identification with metaphysical causation, which is distinct from but related to nomological causation. Both are species of the broader genus ‘causation,’ differentiated by how the causal sufficiency relation is mediated by principles of grounding for metaphysical causation, and by laws of nature for nomological causation. Given this understanding of grounding, we can now further refine our notion of fundamentality (and its two aspects) as follows:

- (8) (Fundamentality_G) x is fundamental if x is independent_G and complete_G.
 (9) (Independence_G) x is independent if nothing grounds x .
 (10) (Completeness_G) The set of the xx s is (or the xx s plurally are, or a non-set-like x is) complete at a world w just in case its members ground everything else at w .

In this framework, an entity is deemed fundamental if it is ungrounded and belongs to a set of entities that collectively ground everything else in a given world. Conversely, an entity is derivative (non-fundamental) if it is grounded by something else or is not part of such a world-grounding set. In further rendering precise this connection between fundamentality and grounding, we can apply the various fundamentality principles within

this framework, with the result that the nature of a fundamental entity is as follows:²

Table 1. Application of Fundamentality Principles

Grounding Principles	Independent _G (Ungrounded)	Complete _G (Ground)
Directed	The deity does not rank below any other entity in the hierarchical structure of reality.	The deity ranks higher than any other entity in the hierarchical structure of reality within the specific world in which it exists.
Necessitating	The existence of any other entity does not necessitate the existence of the deity.	The deity’s existence necessitates the existence of every other entity within the specific world in which it exists.
Generative	The deity’s existence and intrinsic nature are not fixed by the existence and intrinsic nature of any other entity.	The deity’s existence and intrinsic nature fixes the existence and intrinsic nature of every other entity within the specific world in which it exists.
Explanatory	The deity’s existence, at a specific time, is not explained by the existence of any other entity.	The deity’s existence, at a specific time, explains the existence of all other entities within the specific world in which it exists.
Causal	The deity is not a grounded effect of any other entity.	The deity is the generator of all other entities that are grounded effects, within the specific world in which it exists.

A fundamental entity is thus one that is not an output of a grounding relation; rather, it ultimately serves as the ground of everything else. For a fundamental entity, nothing presses upwards on it; instead, it serves the role of pressing upwards on all other (non-fundamental) entities—it is a basic feature of the hierarchical structure of reality (Bennett 2017, 111). We thus have a clear, and indexed (i.e., relativisation to a specific building relation) point-by-point rendering in precise terms of the notion of fundamentality, with a clarification here of how the building-relation of grounding fits neatly into this picture. Now that we have unpacked the nature of fundamentality and its relationship to grounding, we can apply this

2. Note that the tables starting on this page serve a clarificatory function, summarizing the grounding principles already discussed. Readers familiar with these concepts may wish to skip these summaries.

concept in order to elucidate the term ‘god,’ where doing so will furnish a foundation for our explicative activity.

3. THE METHOD OF EXPLICATION

The method of explication as developed by Rudolf Carnap (1962) plays a pivotal role in both analytical philosophy and the philosophy of science, offering a systematic approach to refining and replacing inexact or vague concepts with more precise and useful ones within theoretical frameworks. We can state the central aspect of explication as follows:

- (11) (Explication) A method that systematically refines and replaces inexact or vague concepts (the *explicandum*) with more precise and useful concepts (the *explicatum*) within a theoretical framework, guided by the criteria of similarity, exactness, fruitfulness, and simplicity.

This specific method of explication, according to Carnap (1962), is a two-stage process aimed at enhancing conceptual clarity and theoretical utility. The first stage focuses on the *explicandum*—the inexact or vague concept that requires refinement. Since the explicandum, as noted by Carnap, is inherently inexact, it cannot be precisely defined, and thus, instead, is characterized informally, often through examples that illustrate where the concept clearly applies or does not apply. The second stage introduces the *explicatum*—a new, more exact concept intended to replace the explicandum within a particular theoretical context. This, as he notes, involves explicitly specifying rules for using the explicatum, ideally through a precise definition—though there is an allowance for less strict methods of concept introduction when necessary. An example of this method at work is his example of the everyday concept of ‘fish’ (the *explicandum*), which suffices for general purposes but falls short when it comes to biological classification. Thus, in following the explicative method noted above, this concept is replaced in biology by ‘piscis’ (the *explicatum*), defined as ‘cold-blooded aquatic vertebrate.’ This shift allows for more precise classification and the formulation of biological laws, even though it may exclude creatures commonly considered fish in everyday language, such as whales. Importantly, however, explication is not solely limited to scientific contexts, but is also prevalent in philosophy. That is, philosophers frequently engage in explication when refining concepts such as ‘truth,’ ‘knowledge,’ or ‘blame.’ For instance, Kant (1998) explicates ‘opinion,’ ‘belief,’ and ‘knowledge’ by distinguishing them based on subjective and objective sufficiency. Similarly, Scanlon (2008) explicates ‘blameworthy’ and ‘blame’ in moral philosophy

by articulating their relation to an agent's attitudes and the impairment of relationships.

When seeking to understand the nature of explication, the adequacy of the notion is central, and this centres on four main criteria: *similarity*, *exactness*, *fruitfulness* and *simplicity*. The first criterion, similarity, requires that the *explicatum* bear a certain resemblance to the explicandum, allowing it to be used in place of the explicandum in relevant contexts. However, according to Carnap's thinking there can be considerable differences between the two concepts. The second criterion, exactness, stipulates that the explicatum be more exact than the explicandum. This involves formulating explicit rules for using the explicatum, thus eliminating ambiguity and reducing vagueness. Moreover, exactness also encompasses the elimination of contradictions and paradoxes. The third criterion, fruitfulness, is perhaps the most crucial, according to Carnap: an explicatum should be useful for the formulation of universal statements, laws, or generalizations within the target theory. For instance, the concept of 'piscis' allows biologists to formulate laws about aquatic vertebrates, connecting the concept to other biological concepts and observed facts. The fourth and final, criterion, simplicity, serves as a secondary consideration, which is employed, as he notes, when multiple explicata satisfy the other criteria to a similar degree. Moreover, this criterion refers to the simplicity of the explicatum's definition and the simplicity of the laws that include the explicatum.

Before we proceed further, the decision to employ Carnap's method of explication requires justification, especially given the alternative approaches to conceptual analysis available in contemporary philosophy. Carnap's method is chosen here not as an authoritative standard that all must accept, but as a particularly useful tool for this specific task. The method provides systematic criteria for evaluating conceptual refinements, which is especially valuable when attempting to develop a definition that must navigate between historical accuracy, theological sensitivity, and philosophical rigor. Moreover, Carnap's emphasis on fruitfulness—the ability of a concept to generate new insights and resolve existing puzzles—aligns well with our goal of developing a definition of monotheism that can illuminate rather than obscure the complex theological landscape of ancient and contemporary religious traditions. Now that we have unpacked the nature of the method of explication, it will be important to provide a working definition for the notion of a 'god,' where this will fulfil a central role in our explicative activity.

4. DEFINING 'GOD'

In seeking to conceptualize the term 'god' that stands at the centre of our explication process—a term laden with diverse cultural, philosophical and theological significance throughout human history and across different belief systems—we can offer the following definition:

- (12) (God) x is a "god" iff (i) x is a deity (ii) x is fundamental.

In the definition of the term 'god' featured in (2), this term is used as a *referring expression*—that is, as a 'name' or 'title' for a particular type of entity—namely, one that is a deity and one that is fundamental—and, therefore, we can call the usage of this term its 'nominal' sense. Hence, to be a 'god,' a particular entity must possess these two features—i.e. be a fundamental deity—where this means that any entity that lacks either of these features, (e.g., by being a deity but lacking fundamentality) would not be a 'god.' Now, at a general level, a deity is characterised by attributes that exceed natural limitations, placing it in the category of supernatural beings. These attributes include power, knowledge, and other qualities that surpass what is achievable through natural means alone. The supernatural attributes of a deity typically encompass enhanced power to influence the world, knowledge beyond human comprehension, and a presence not constrained by physical limitations. These qualities, while significantly greater than those found in nature, are not necessarily infinite or maximal. Hence, there is a spectrum as regards the degree to which a deity's attributes exceed natural limitations. At one end are deities whose powers, while supernatural, have clear bounds. At the other extreme lie omnipotent, omniscient and omnipresent divine beings whose attributes are considered absolutely unlimited. Most conceptions of deities fall somewhere between these two poles, ascribing to the 'god' in question powers that vastly surpass the natural world, but still may be subject to some constraints. Examples of constrained deities include Greek gods like Zeus who, while immensely powerful, is not omnipotent. The god of classical theism, in contrast, is considered to have unlimited power, knowledge and presence. The specific supernatural attributes ascribed to a deity thus shape its role and influence within a given religious tradition, and for this reason, understanding these qualities will be key to grasping the nature and significance of a particular conception of divinity. Now, given this construal of the nature of a deity, we can turn our attention to the second feature that an entity must possess to be a 'god'—namely, that of fundamentality.

One could raise the question, at this point, of why deities should be construed as supernatural entities—that is, entities possessing attributes that exceed natural limitations. This characterization emerges from cross-cultural studies of religious concepts, where deities are consistently distinguished from natural entities by their possession of powers, knowledge, or modes of existence that transcend what is achievable through natural processes alone. As Pascal Boyer (2001) and Justin Barrett (2004) have demonstrated in their cognitive scientific research into religion, the concept of a deity invariably involves some violation or transcendence of ordinary natural categories and limitations. This supernatural dimension is not an arbitrary addition, but reflects how religious traditions themselves understand and characterize their deities.

As was previously noted, fundamentality is best construed in terms of an entity's being independent and complete. So, if a deity is to be fundamental, it must itself be independent and complete. More specifically, the deity will be an 'unexplained explainer,' in that it will be on the one hand independent, which is to say unbuilt, and on the other complete, which is to say that it will be a member of a set of entities at a world whose members build everything else. However, as was noted above, the notions of independence and completeness are ambiguous as they stand. In that respect, we must index each of these notions to particular building-relations. Thus, focusing on the specific building-relation of grounding, the deity's being independent is reducible to its being ungrounded, and its being complete is reducible to its being a member of a set of entities at a world whose members ground everything else. In short, the deity is the ungrounded ground of everything else. Hence, the deity, as a fundamental entity, is ontologically prior to all other things in the hierarchical structure of reality. It is independent of all things and exists as the complete entity within this structure, due to its being ungrounded (i.e. unbuilt) and fulfilling the role of grounding (i.e. building) all other features of reality. The deity is thus fundamental by virtue of not being the output of any grounding relation, in that nothing 'presses upwards' on it: rather, it presses upwards on all other (non-fundamental) entities. We can thus further elucidate the nature of the deity's role as a fundamental entity within reality by applying the grounding principles to this specific case as well:

Table 2. Application of Fundamentality Principles

Grounding Principles	Independent _G (Ungrounded)	The deity ranks higher than any other entity in the hierarchical structure of reality within the specific world in which it exists.
Directed	The deity does not rank below any other entity in the hierarchical structure of reality.	The deity’s existence necessitates the existence of every other entity within the specific world in which it exists.
Necessitating	The existence of any other entity does not necessitate the existence of the deity.	The deity’s existence and intrinsic nature fixes the existence and intrinsic nature of every other entity within the specific world in which it exists.
Generative	The deity’s existence and intrinsic nature are not fixed by the existence and intrinsic nature of any other entity.	The deity’s existence, at a specific time, explains the existence of all other entities within the specific world in which it exists.
Explanatory	The deity’s existence, at a specific time, is not explained by the existence of any other entity.	The deity is the generator of all other entities that are grounded effects, within the specific world in which it exists.
Causal	The deity is not a grounded effect of any other entity.	The deity ranks higher than any other entity in the hierarchical structure of reality within the specific world in which it exists.

Grounding, conceived as a relation of ‘directed dependence,’ plays the needed role of a necessary explanation-backing link that stems from the deity to all other entities, and is mediated by the principles of grounding. All other entities are dependent for their existence upon the (eternal and necessitating) action of this specific deity. Hence, they do not exist as independent entities, but are grounded (or built) entities. Thus, as all other entities are the less fundamental result within this grounding relationship, they are subordinate to the deity. Therefore, there is a distinct ordering and distinction of status within reality, wherein the deity, as an independent and complete entity (i.e. the ungrounded ground of everything else), is fundamental, and all other entities, being dependent and non-complete (i.e. grounded entities that are not the ground of everything else), are derivative and non-fundamental. We can now restate our definition of ‘god’ in a more precisely rendered form as follows:

god, but the term 'god' itself is often left inexact or vague. This ambiguity can lead to confusion and hinder the development of a robust theoretical framework for understanding the nature of god and the implications of monotheistic belief. To address this issue, the explication introduces the explicatum, which is the refined and more precise concept of 'god' as a 'fundamental deity.' This new definition is intended to replace the explicandum within the specific theoretical context of monotheism. So, by explicitly specifying the characteristics of a fundamental deity, the explicatum provides a clearer and more exact understanding of what it means to answer to the term 'god.' That is, the explicatum defines 'god' as an entity that possesses two essential features, 'fundamentality' and 'divinity', which are attributes that exceed natural limitations. These attributes include power, knowledge and other qualities that surpass what is achievable through natural means alone, placing the deity in the category of supernatural beings. Now, it is important to note that in some religious traditions, the god of monotheism is conceived as a deity that lacks all limitations, thus possessing attributes such as power, knowledge and goodness in an absolute sense—i.e. maximal power, knowledge and goodness. However, the explication of 'god' as a fundamental deity does not necessarily entail these maximal attributes, which thus allows for a more inclusive understanding of divinity that can accommodate a spectrum of supernatural qualities.

On the other hand, fundamentality is construed as an entity's—which in this context means a deity's—being independent and complete. Independence means that the deity is unbuilt or ungrounded, while completeness signifies that the deity is a member of a set of entities at a world whose members build everything else. Hence, by incorporating these two features into the definition of 'god,' the explication successfully captures the essential characteristics that distinguish a monotheistic god from other entities. What this means is that this explicatum emphasises the unique ontological status of 'god' as a fundamental ground of everything else in reality.

Moreover, the explication also satisfies Carnap's (1956) criteria for adequacy to a sufficient degree. Firstly, it bears a significant similarity to the explicandum, as it maintains the core idea of 'god' as a deity while providing a more precise understanding of what this entails. This similarity thus allows the explicatum to be used in place of the explicandum in relevant contexts, such as theological and philosophical discussions of monotheism. Secondly, the explicatum is more exact than the explicandum, as it eliminates ambiguity and reduces vagueness by explicitly specifying the rules for using the term 'god.' The incorporation of supernatural attributes and the principles of fundamentality (or grounding) provide a clear

framework for understanding the nature of 'god' and its relation to other entities in reality. Hence, this increased exactness enables clearer decisions about the applicability of the concept in various cases, and promotes consistency in respect of its use. Thirdly, the explicatum is relatively simple, as it introduces a concise definition of 'god' that captures the essential features without unnecessary complexity. The use of supernatural attributes and the notion of fundamentality provide a straightforward and intuitive framework for understanding the nature of 'god,' thus making the explicatum accessible and easy to employ in relevant contexts. Fourthly, the explicatum is fruitful, as it enables the formulation of universal statements, laws, or generalizations within the target theory of monotheism. That is, by defining 'god' as a fundamental deity, the explication provides a robust foundation for exploring the implications of monotheistic belief and its relation to other philosophical and theological concepts. The explicatum's emphasis on 'god's' ontological priority and role as the ground of all other entities opens up new avenues of theoretical development and forms of explanatory power.

We can now understand the fruitfulness of the explication of monotheism (as the belief in one fundamental deity) by applying it within the context of ancient Jewish monotheism: the notion of fundamentality featured in the explication emerges as the key characteristic that captures the essential features of the ancient Jewish understanding of God, including His uniqueness and ontological priority, while still accommodating the complex theological landscape of Second Temple Judaism. More specifically, the explication's emphasis on the fundamentality of 'god' aligns with the ancient Jewish understanding of Yahweh as the sole creator and sovereign ruler of all things: by identifying Yahweh as the fundamental ground of all reality, it establishes His ontological priority and unique status as the creator and sustainer of all things, including other divine beings. In consequence, this understanding is consistent with the ancient Jewish belief in Yahweh as the 'high god' who presides over a court of heavenly beings. In addition to this, the explication's emphasis on the fundamentality of 'god' provides a crucial basis for understanding why Yahweh was considered the only being worthy of worship within the ancient Jewish monotheistic framework. This is due to the fact that while the concept of deity alone may not necessarily preclude the existence of other divine beings, the notion of fundamentality establishes Yahweh as the unique and *unrivalled* ground of all reality. What this means is that, as the fundamental entity upon which all other beings, including divine ones, depend for their existence, Yahweh alone is worthy of exclusive devotion and worship. Hence, the explication's identification

of 'god' with an independent and complete entity, ungrounded and serving as the ground of everything else, underscores the ontological and devotional primacy of Yahweh within the ancient Jewish worldview. And so this understanding aligns with the distinctive feature of ancient Jewish monotheism, which reserved worship exclusively for Yahweh, recognising Him as the sole creator and sustainer of all things.

Now, the fruitfulness of this explication lies in its ability to highlight the distinctive features of ancient Jewish monotheism within the broader religious environment of the ancient world. While the concept of a 'high god' presiding over other deities was common in ancient Near Eastern religions, the ancient Jewish understanding of Yahweh's uniqueness and the exclusive devotion owed to Him sets their monotheistic belief apart. The explication's emphasis on 'god' as a fundamental deity captures this distinctiveness by underlining the ontological and devotional primacy of Yahweh within the ancient Jewish worldview. Moreover, the explication proves fruitful in providing a framework for understanding the development of early Christian thought, which emerged from Second Temple Judaism. The point here is that the early Christian understanding of Jesus as a divine figure alongside the god of Israel can be better understood within the context of ancient Jewish monotheism, which acknowledged the existence of divine beings subordinate to Yahweh. The explication's recognition of a fundamental deity among other divine beings provides a conceptual framework for exploring the early Christian understanding of the relationship between God the Father and Jesus Christ. Thus, the explication of monotheism as the belief in one fundamental deity proves to be highly fruitful when applied to the context of ancient Jewish monotheism. It captures the essential features of the ancient Jewish understanding of God, including His (species or transcendent) uniqueness, grounded upon His fundamentality, while accommodating the complex theological landscape of Second Temple Judaism. The explication provides a conceptual framework for understanding the relationship between Yahweh and other divine beings, as well as the distinctive features of ancient Jewish monotheism within the broader religious environment of the ancient world.

Even so, despite the fruitfulness (and thus the overall adequacy, in the sense of similarity, exactness, simplicity and fruitfulness) of the proposed explication of monotheism as the belief in one fundamental deity, it does face a potential challenge when considering its applicability to polytheistic religious traditions. This challenge questions the fruitfulness of the explication in capturing the concept of divinity as understood within the context of ancient Greek religion and mythology; we can term this objection the 'Pantheon Objection.'

6. THE PANTHEON OBJECTION

The Pantheon Objection contends that the proposed explication of monotheism in terms of Fundamentality Monotheism fails to be adequately fruitful due to its inability to capture the concept of divinity as understood within the context of polytheistic religions, particularly as exemplified by the Greek pantheon. This objection deserves serious consideration, as it tests whether our definition can accommodate the full range of religious phenomena it purports to explain. We can state this objection succinctly as follows:

- (15) (Pantheon Objection) Defining ‘god’ in terms of fundamentality (being an independent ground of reality) fails because it cannot properly account for polytheistic deities, such as those in the Greek pantheon, who were considered true ‘gods’ despite not being fundamental in this sense.

This objection arises from the recognition that the explicatum of ‘god’ as a fundamental deity, while potentially useful for clarifying monotheistic belief, does not align with the divine beings depicted in Greek mythology and worshipped in ancient Greek religion. Hence, the objection points out that if the concept of ‘god’ is defined as a being that is fundamental, then this definition should be applicable not only to monotheism but also to polytheism—that is, it should cover both of these forms of ‘theism.’ However, when this definition is applied to a paradigm case of polytheism, the Greek pantheon, it becomes apparent that the ‘gods’ worshipped by the ancient Greeks do not meet the criteria of fundamentality in the sense required by the explication—and thus, the explication provided is not truly fruitful.

We can see this more clearly from the following: according to Greek mythology, the Olympian ‘gods’ (traditionally consisting of Zeus (king of the ‘gods,’ and ‘god’ of the sky and thunder), Hera (‘god’ of marriage and family), Poseidon (‘god’ of the sea and earthquakes), Demeter (‘god’ of harvest and agriculture), Athena (‘god’ of wisdom and war strategy), Apollo (‘god’ of the sun, music, and prophecy), Artemis (‘god’ of the moon and hunting), Ares (‘god’ of war and bloodlust), Aphrodite (‘god’ of love and beauty), Hephaestus (‘god’ of fire and metalworking), Hermes (‘god’ of messengers and commerce), and Dionysus (‘god’ of wine and festivities)), while powerful and revered, are not eternal or uncreated beings. That is, the myths depict the ‘gods’ as coming into existence at a certain point in the history of the cosmos,³ often through procreation or other means of generation.

3. See Graves (2017) for a further unpacking of this story.

Zeus, the ruler of the ‘gods,’ is himself the son of the Titans Cronus and Rhea, and he achieves his position of supremacy through a violent struggle against his father and the other Titans. This suggests that the Greek ‘gods,’ for all their might and majesty, are not fundamental in the sense of being independent grounds of all reality. Rather, they are part of a larger cosmic narrative in which they emerge as contingent beings within an already existing universe. And so, if the concept of ‘god’ proposed by the explication fails to capture the understanding of divine beings within a polytheistic context such as the one above, then it is indeed too narrow or restrictive to serve as a comprehensive and illuminating definition of what it means for an entity to be a ‘god.’ In that case, the fact that the Greek ‘gods’ were considered true deities worthy of worship and veneration, despite not being fundamental, suggests that the explication may be overlooking essential aspects of how the notion of a ‘god’ was conceptualised and experienced in ancient religious traditions.

In light of these considerations, one can indeed question the adequacy and scope of the notion of ‘god’ that features in the explication proposed above, and it thus appears to fail as a comprehensive account of what it means for an entity to be a ‘god’ across diverse religious and cultural contexts. Hence, the question now to be faced is whether one can indeed respond to the issues raised by the Pantheon Objection in a manner that would allow one to re-affirm the adequacy of the explication developed here. I believe that this objection can be addressed by re-considering the possibility of a parallel concept of *Fundamentality Polytheism*, which can be stated as follows:

- (16) (Fundamentality Polytheism) The belief that that there are multiple “god”s =_{ex.}
The belief that there are multiple fundamental deities.

In the proposed explication of the notion of polytheism—the belief that there are multiple ‘gods’—as corresponding to a belief that there are multiple fundamental deities, one can, in fact, understand that the Greek pantheon of ‘gods,’ as described in the mythological tradition, does fit with this particular explication, due to the fact that these ‘gods’ (and others like them) can indeed be considered fundamental in the metaphysical sense, even if they are not eternal or uncreated. That is to say, by distinguishing between the notions of causation and grounding, it is indeed possible to articulate a coherent model of Fundamentality Polytheism that captures the ‘god-hood’ of the Olympian ‘gods’ while avoiding the pitfalls of the original objection. Now, central to this line of response is the recognition that having

a cause for one's existence does not necessarily entail being grounded by another entity. Grounding, as was noted previously, is a *synchronic relation* that obtains at each moment that an entity exists—thus providing a metaphysical explanation for its continued existence and intrinsic nature. In contrast, causation is a diachronic relation that accounts for how an entity came to exist in the first place, but does not necessarily entail a persistent dependence on the causing entity. This distinction thus opens up the possibility that the Greek 'gods,' while having a causal origin as narrated in the myths, may nonetheless be ungrounded and thus fundamental according to the definition provided. The relevant point here is that those myths depict the Olympian 'gods' as being born or generated by other divine beings, as with Zeus's being the son of the Titans Cronus and Rhea. However, once the 'gods' come into existence, there is no indication in the mythological tradition that their continued existence or intrinsic nature is grounded by any other entity. That is, Zeus and the other Olympians appear to be self-sustaining and metaphysically independent, not requiring any external entity to ground their being at each moment—they possess their own power, agency and ontological stability, suggesting that they are not derivative or dependent in the sense of being grounded. And so, if that interpretation is granted, then the 'gods' can indeed be considered fundamental according to the construal of this notion developed above.

In support of this conclusion, we may observe that the mythological story of the "Binding of Zeus" provides compelling evidence for the independence and fundamentality of the Greek 'gods,' even in relation to Zeus, their king. In this story, Zeus was bound with hundred-knotted thongs by Hera, Poseidon, Apollo, and the other Olympians (except Hestia) while he slept, but was freed by the hundred-handed Briareus summoned by Thetis. As punishment, Zeus hung Hera from the sky with bracelets and anvils until the other 'gods' swore loyalty to him. Zeus then sent Poseidon and Apollo to build Troy as bond-servants, but pardoned the other conspirators, who had acted under duress.⁴ Now, the fact that Hera, Poseidon, Apollo, and the other Olympians were able to conspire against Zeus, bind him, and threaten his rule suggests that their existence and power were *not entirely dependent on him*. That is, in the context of the metaphysical notion of grounding, this story demonstrates that the 'gods' were not synchronically grounded by Zeus at each and every moment of their existence, in that their willingness to overthrow Zeus and establish a new order on Olympus implies that they did not view their own existence as inextricably tied to

4. See (Graves 2017) again for a further unpacking of this story.

his. Instead, they were prepared to terminate Zeus's rule and, potentially, his existence, thus indicating that they did not consider themselves to be metaphysically dependent on him. Hence, the complex power dynamics and relationships among the 'gods' suggest that their roles and authorities were not fixed or absolute, and this is consistent with a model of independent and self-sustaining deities. Moreover, Zeus's need to resort to punishing and threatening the other 'gods' after being released, rather than simply reasserting his metaphysical supremacy, suggests that his authority, again, was not intrinsic or necessary to their existence. The specific punishments inflicted on Hera, Poseidon, and Apollo reveal the limits of Zeus's power and the resilience of the other 'gods,' therefore implying that they maintained a level of autonomy even in the face of his wrath. And so we can indeed affirm the possible possession of the first aspect of fundamentality by the Greek pantheon of 'gods'—namely, that of independence.

Furthermore, recent scholarship on Greek religion supports a more nuanced understanding of divine independence—one that aligns with our fundamentality framework presented here. Scholars such as Jenny Strauss Clay (1989), Ken Dowden (1992) and Jan Bremmer (1994) emphasize that the Greek gods, despite their mythological origins, function as autonomous powers within their respective domains. As Clay's analysis of Hesiod's *Theogony* demonstrates, once established in their roles the Olympian gods operate with genuine independence—they have their own wills, their own spheres of authority, and their own divine prerogatives that even Zeus must respect through negotiation and political manoeuvring rather than simple ontological dominance. The binding of Zeus narrative, carefully analysed by Timothy Gantz (1993) in his comprehensive study of Greek mythological traditions, provides compelling evidence for this independence. The very fact that the other Olympians could successfully conspire against and bind Zeus—even temporarily—demonstrates that their existence and power do not derive from him moment-to-moment. As Walter Burkert (1985) notes in his seminal work on Greek religion, the Greek gods are conceived as eternal and self-sustaining once they come into being. Their mythological births are aetiological narratives explaining their entry into the cosmos, not ongoing relationships of metaphysical dependence. The distinction between causal origin and synchronic grounding is crucial here: Hephaestus may have been born from Hera, but his continued existence as the god of metallurgy does not depend on her sustaining power. This understanding supports a model of plural fundamentality within Greek polytheism. Each deity can be understood as fundamental within their particular sphere—Poseidon with respect to the sea, Demeter with respect to agriculture, Apollo with

respect to prophecy and music. While Zeus holds political supremacy and can negotiate or even command other gods through his greater power, this represents a difference in degree of power rather than a difference in fundamental ontological status. When Zeus must persuade, threaten, or bargain with other deities (as he frequently does in Homer's epics), this reveals that he cannot simply will changes in their domains but must work within a system of independent divine powers. The cosmos requires all these fundamental deities working in concert, which is precisely why Greek religion involved cults directed towards multiple gods rather than exclusive worship of Zeus alone.

We can also affirm the possible possession of the second aspect of fundamentality—namely, that of completeness—in that while the 'gods' may not individually ground everything else in the cosmos, they can potentially form a set of entities that collectively ground all other elements of reality within their mythological world. That is, the 'gods' are consistently portrayed as the source and explanation for various natural phenomena, human affairs, and the overall order of the universe. They are the ultimate arbiters of fate, the dispensers of justice, and the powers behind the forces of nature. In this sense, the 'gods,' as a plurality, can be seen as complete, in virtue of fulfilling the role of grounding the existence and nature of the non-divine entities within their domain. This would align with the concept of completeness as defined previously, through it allowing for a plural set of fundamental entities—that is, what we can term 'plural fundamentality.'

Now, alongside the fundamentality of these entities, they will also clearly be deities. However, to avoid the issue of volitional conflict, we should not require the 'gods' to be maximally powerful, but rather just significantly, yet finitely, powerful and authoritative within their respective domains. This reflects the fact that if multiple fundamental deities were conceived as having maximal power, it could lead to irresolvable conflicts, where two or more deities have incompatible desires or intentions regarding non-rational actions, such as the direction of the Sun's rotation on its axis. This would be an impossible state of affairs for entities defined as maximally powerful, as their conflicting wills would negate each other's maximal power. Hence, if we are to deal with the issue of multiple fundamental deities within a polytheistic context they ought to be understood as having a specific scope or sphere of influence, within which they are the ultimate grounds of being and the highest arbiters of reality. For example, Zeus exercises his power and authority with respect to the sky and kingship, Poseidon with respect to the sea and earthquakes, and Athena with respect to wisdom and war, etc. Hence, each deity would be supremely powerful

and independent within their domain, but not necessarily maximal in an absolute sense that encompasses all of reality.

Now, one might object that if we accept the Greek gods' independence on the basis of their ability to oppose or conspire against Zeus, then consistency demands we question the independence of the God of monotheistic traditions who also faces opposition from other spiritual beings. Here, one could potentially present a counterexample from Christian theology, invoking the case of Satan's opposition to God. However, in traditional Christian theology, Satan's ability to oppose God does not constitute a conspiracy in a sense that would threaten divine independence. Satan's opposition operates within divinely permitted boundaries—he cannot act without God's permission (as seen in Job 1–2) and his ultimate defeat is assured. This is categorically different from cases where a deity's existence or power could genuinely be threatened by other beings. The distinction here is between opposition within a divinely maintained order and threats to the very existence or fundamental status of the deity.

Thus, all in all, the model of Fundamentality Polytheism allows for a pluralistic conception of 'god,' where multiple 'gods' coexist as fundamental entities grounding distinct aspects of reality. It captures the intuition that the 'gods' are the ultimate metaphysical grounds of the phenomena associated with them, without requiring them to be all-encompassing. Furthermore, the domain-specific approach furnished by Fundamentality Polytheism can help to make sense of the complex relationships and interactions between the 'gods' as described in the myths. That is, the conflicts, alliances and power struggles among the 'gods' can be understood as negotiations and assertions of their respective fundamental roles within the cosmic order. And when 'gods' clash or collaborate, it can be seen as a working out of the metaphysical grounding relations that define their domains and the reality they collectively sustain. Also, this dynamism and tension within the pantheon would not necessarily undermine their fundamental status, but rather reflect the intricate web of dependencies and influences that shape the mythological world. Our definition of the notion of 'god' that features within the further explications of Fundamentality Monotheism and Fundamentality Polytheism therefore remains untouched by the Pantheon Objection. Hence the former, and now also the latter, explications can be embraced as adequate definitions of these important theological terms.

7. INTERFAITH IMPLICATIONS OF EXPLICATION

The interfaith implications of the position established here offer a promising resolution to long-standing debates concerning the nature of Trinitarianism.

Critics often characterize the latter as polytheistic, due to its affirmation of multiple divine persons. However, this characterization rests on a fundamental misunderstanding of how we should define monotheism and polytheism. As argued here, monotheism should not be defined simply as belief in one deity, but rather as belief in one fundamental deity. Correspondingly, polytheism is better understood not as mere belief in multiple deities, but as belief in multiple fundamental deities. This definitional clarification then has profound implications for how we should categorize Trinitarianism. Monarchical Trinitarianism, as defended in the work of John Behr (2018), Beau Branson (2020) and Joshua Sijuwade (2022) provides a compelling framework that demonstrates how Trinitarianism can be properly classified as monotheistic. This model maintains that while there are three divine persons—the Father, Son, and Holy Spirit—who are relationally distinct and ontologically equal (each being rightly called ‘God’ in a secondary or predicative sense), there is only one fundamental deity: the Father, who is uniquely ‘God’ in the primary or nominal sense. In it, the Father holds a unique position as the uncaused cause and ungrounded ground of all reality, including both the Son and Spirit, as well as all created things. This makes the Father the sole fundamental entity within reality. The Son and Spirit, while fully divine and equal in nature to the Father, derive their being from the Father and thus are not themselves fundamental. They share fully in the Father’s divinity without being “the one God” in the nominal sense. This framework resolves the apparent tension between Trinitarianism and monotheism. While affirming the full deity of three distinct persons, it maintains strict monotheism by recognising only one fundamental deity. The other divine persons, while truly divine, are derivatively so, receiving their deity from the Father as the sole ultimate source. And thus the key insight is that having multiple divine persons need not necessarily entail having multiple fundamental deities. Hence, this resolution demonstrates that the charge of polytheism against Trinitarianism stems from an oversimplified understanding of both monotheism and the Trinity.

When properly understood through the lens of fundamentality rather than mere numerical counting of divine persons, Trinitarianism emerges as a sophisticated form of monotheism that maintains both the unity of God and the full deity of Father, Son and Spirit. Furthermore, given our explication of monotheism, Trinitarianism can be properly classified as monotheistic in precisely the same sense as Judaism and Islam, each of which posits the existence of just one fundamental deity (Yahweh for Judaism, and Allah for Islam). Trinitarianism, envisaged through the lens of Monarchical Trinitarianism, likewise affirms only one fundamental deity: the Father.

This crucial insight reveals that the true area of theological disagreement between these faiths does not concern monotheism—as this is affirmed across all three traditions—but rather centres on the debate between Trinitarianism and Unitarianism, which we can define as follows:⁵

- (17) (Trinitarianism) The belief in the existence of three deities (divine persons)—the Father, the Son, and the Holy Spirit—with the Father being the sole fundamental deity from whom the Son and Spirit derive their divine nature.
- (18) (Unitarianism) The belief in the existence of exactly one deity (whether identified as Allah, Yahweh, or otherwise), who is both the sole deity and the sole fundamental deity.

The central point of contention between these theological positions thus concerns the number of divine persons that exist within reality. Trinitarianism affirms the existence of three such persons, while maintaining that only one (the Father) is fundamental,⁶ whereas Unitarianism⁷ maintains that there is exactly one divine person who is, by definition, fundamental. This clarification helps us better understand the real nature of the theological dispute: it is not about whether monotheism is true (as all parties agree on this point when properly understood in terms of fundamentality), but rather about whether reality includes multiple divine persons or just one. This reframing of the debate has significant implications for interfaith dialogue and theological discussion. It suggests that participants in these discussions might make more progress by focusing their attention on arguments for and against the existence of multiple divine persons, rather than on accusations of polytheism that arise from misunderstanding the nature of monotheism. The question is not whether there can be multiple divine persons while maintaining monotheism (as Monarchical Trinitarianism shows this to be possible),⁸ but whether there actually are multiple divine persons in reality. Moreover, this understanding helps explain why

5. Unitarianism is defended most prominently in the contemporary philosophical and theological literature by Dale Tuggy (2014).

6. Although Trinitarianism affirms the existence of only three deities, it is still compatible with the view that there are more than three deities in reality—as the former deities, the Father, the Son and the Spirit, are *supreme* deities, in the sense of being maximal in all of their attributes, whereas the other deities that potentially exist lack maximality. It is an open question whether (contemporary) Judaism and Islam are also compatible with a position holding that there are additional (non-maximal) deities within reality.

7. Unitarianism, so construed, is thus a version of Deity Monotheism.

8. I leave it as an open question whether other forms of Trinitarianism are able to do this as well.

traditional arguments for monotheism, such as arguments from divine simplicity or necessary existence, do not automatically decide between Trinitarianism and Unitarianism. These arguments typically establish the existence of one fundamental divine being, but leave open the possibility that this being (the Father in Trinitarian thought) might generate other divine persons who share in the divine nature while remaining dependent on their source. Thus, given our explicative work here, the path forward in these theological debates lies not in resolving disputes about monotheism—which all parties effectively embrace—but in examining the philosophical and theological evidence for whether the one fundamental divine being exists alone or generates other divine persons who share in its nature.⁹

8. CONCLUSION

To conclude, this article has analysed monotheism through the lens of fundamentality. Our aim has been to develop a philosophically robust definition that can illuminate rather than obscure the complex theological phenomena found in ancient and contemporary religious traditions. While Carnap's method provides useful criteria for evaluation, the ultimate test of our explication is its ability to generate fruitful insights into the nature of monotheistic and polytheistic beliefs. We began by examining traditional numerical views, and the relevant historical developments, and then went on to explore Bennett's concept of building-fundamentality and Carnap's method of explication. After analysing 'god' in terms of fundamentality and divinity, we developed an explication of monotheism using fundamentality, addressed the Pantheon Objection, and explored implications for interfaith dialogue. This comprehensive analysis presents a reasoned case for understanding monotheism primarily in terms of fundamentality rather than mere numerical counting, while also acknowledging the complexities of this notion and the potential for more nuanced interfaith dialogue.

REFERENCES

- Barrett, Justin L. 2004. *Why Would Anyone Believe in God?* Lanham, MD: AltaMira Press.
- Bauckham, Richard. 2008. *Jesus and the God of Israel: God Crucified and Other Studies on the New Testament's Christology of Divine Identity*. Grand Rapids, MI: Eerdmans Publishing.
- Behr, John. 2018. "One God Father Almighty." *Modern Theology* 34 (3): 320–30. <https://doi.org/10.1111/moth.12419>.
- Bennett, Karen. 2017. *Making Things Up*. Oxford: Oxford University Press.
- Boyer, Pascal. 2001. *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books.

9. For a priori argumentation for this generative action of other divine persons necessarily taking place, see (Swinburne 2018, 430–32).

- Branson, Beau. 2022. "One God, the Father: The Neglected Doctrine of the Monarchy of the Father, and Its Implications for the Analytic Debate about the Trinity." *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology* 6 (2): 1–53. <https://doi.org/10.14428/thl.v6i2.67603>.
- Bremmer, Jan N. 1994. *Greek Religion*. Oxford: Oxford University Press.
- Burkert, Walter. 1985. *Greek Religion*. Cambridge, MA: Harvard University Press.
- Carnap, Rudolf. 1962. *Logical Foundations of Probability*. 2nd ed. Chicago: University of Chicago Press.
- Clay, Jenny Strauss. 1989. *The Politics of Olympus: Form and Meaning in the Major Homeric Hymns*. Princeton, NJ: Princeton University Press.
- Dowden, Ken. 1992. *The Uses of Greek Mythology*. London: Routledge.
- Fredriksen, Paula. 2006. "Mandatory Retirement: Ideas in the Study of Christian Origins Whose Time Has Come to Go." *Studies in Religion/Sciences Religieuses* 35 (2): 231–46. <https://doi.org/10.1177/000842980603500203>.
- Fredriksen, Paula. 2022. "Philo, Herod, Paul, and the Many Gods of Ancient Jewish 'Monotheism.'" *Harvard Theological Review* 115 (1): 23–45. <https://doi.org/10.1017/S0017816022000049>.
- Gantz, Timothy. 1993. *Early Greek Myth: A Guide to Literary and Artistic Sources*. Baltimore: Johns Hopkins University Press.
- Graves, Robert. 2017. *The Greek Myths: The Complete and Definitive Edition*. London: Penguin Books.
- Hayman, Peter. 1991. "Monotheism—A Misused Word in Jewish Studies?" *Journal of Jewish Studies* 42 (1): 1–15. <https://doi.org/10.18647/1576/JJS-1991>.
- Heiser, Michael. 2008. "Monotheism, Polytheism, Monolatry, or Henotheism? Toward an Assessment of Divine Plurality in the Hebrew Bible." *Bulletin for Biblical Research* 18 (1): 1–30. <https://doi.org/10.2307/26423726>.
- Hurtado, Larry W. 2004. *Lord Jesus Christ: Devotion to Jesus in Earliest Christianity*. Grand Rapids, MI: Eerdmans Publishing.
- Kant, Immanuel. 1998. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.
- Leftow, Brian. 2016. "Anti Social Trinitarianism." In *Philosophical and Theological Essays on the Trinity*, edited by Thomas McCall and Michael C. Rea, 52–74. Oxford: Oxford University Press.
- MacDonald, Nathan. 2003. *Deuteronomy and the Meaning of "Monotheism"*. Tübingen: Mohr Siebeck.
- Scanlon, T.M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press.
- Schaffer, Jonathan. 2016. "Grounding in the Image of Causation." *Philosophical Studies* 173 (1): 49–100. <https://doi.org/10.1007/s11098-014-0438-1>.
- Sijuwade, Joshua R. 2022. "Building the Monarchy of the Father." *Religious Studies* 58 (3): 456–76. <https://doi.org/10.1017/S0034412521000020>.
- Smith, Mark S. 2013. *God in Translation: Deities in Cross-Cultural Discourse in the Biblical World*. Grand Rapids, MI: Eerdmans Publishing.
- Swinburne, Richard. 2016. *The Coherence of Theism*. 2nd ed. Oxford: Oxford University Press.
- Swinburne, Richard. 2018. "The Social Theory of the Trinity." *Religious Studies* 54 (3): 419–37. <https://doi.org/10.1017/S0034412518000203>.
- Tuggy, Dale. 2014. "Who Should Christians Worship?" *Journal of Biblical Unitarianism* 1 (1): 5–33.

- Tuggy, Dale. 2017. "On Counting Gods." *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology* 1 (1): 188–213. <https://doi.org/10.14428/thl.v1i1.153>.
- Wright, N.T. 1992. *The New Testament and the People of God*. Minneapolis: Fortress Press.

The Relationship of Italian Neo-Scholasticism and Phenomenology to Naturalistic Anthropology

An Exploration via the Views of Sofia Vanni Rovighi

Tymoteusz Mietelski

ABSTRACT This article explores the anthropological views of Sofia Vanni Rovighi (1908–1990), an Italian philosopher who identified herself as a Thomist while advocating a broadening out of this classical approach through the incorporation of phenomenological elements. The text outlines her conception of the human being, distinguishing between phenomenological and metaphysical levels. A key focus here is her critique of naturalism—which, she argued, is unjustly reductive in its approach to human beings. This polemical reflection situates her views within a broader context, addressing in particular the call for a re-empiricization of Thomistic anthropology.

KEYWORDS anthropology; Italian philosophy; neo-scholasticism; phenomenology

INTRODUCTION

This article presents the anthropological views of Sofia Vanni Rovighi (1908–1990), an Italian philosopher and professor of medieval philosophy, the history of philosophy and theoretical philosophy at the Catholic University of Milan. When discussing the philosophers most important to her academic background, Vanni Rovighi highlighted two individuals: Amato Masnovo, a leading representative of Italian neo-scholasticism, and Edmund Husserl, whose ideas she explored in her first monograph in Italy (Vanni Rovighi 1959, 185–94). Her focus on these two authors outlines two key currents that permeate her work: her self-identification as a Thomist and her advocating of a broadening out of this classical approach through the incorporation of phenomenological elements.

The paper consists of four sections. The first provides a brief introduction to Vanni Rovighi herself. The second discusses her understanding of phenomenology, as Husserl's theory inspired certain changes in her Thomistic approach to anthropology. The third presents the main elements of Vanni Rovighi's conception of the human being, which seeks to oppose naturalistic (and spiritualistic) conceptions as a starting point. The fourth section then offers a critical summary of her conception of anthropology.

THE HISTORICAL AND PHILOSOPHICAL CONTEXT OF VANNI ROVIGHI'S SCHOLARSHIP

Sofia Vanni Rovighi was born on September 28th, 1908, in San Lazzaro di Savena, and passed away on June 10th, 1990, in Bologna. Immediately after completing her studies in 1931 she began her teaching career, which continued until 1978. In 1939 she obtained the *libera docenza* (a qualification equivalent to habilitation), and in 1948 she was appointed as a professor at the Catholic University of Milan, where she lectured on medieval philosophy, the history of philosophy, theoretical philosophy and moral philosophy. Between 1962 and 1971, she served as the editor-in-chief of *Rivista di Filosofia Neo-Scolastica*, an Italian journal modelled on the *Revue Néo-Scholastique de Philosophie* journal published in Louvain (Sina 1990, 490; 2006, 12013–15).

At a time when the dominant (and to some extent politically mandated) philosophy in Italy was Giovanni Gentile's idealism, Vanni Rovighi studied medieval philosophy under the guidance of Amato Masnovo, one of the leading figures of Italian neo-scholasticism. She wanted to deepen her understanding of medieval philosophy and examine whether it could address the issues raised by modern and contemporary thought. Early in her career, Vanni Rovighi focused on the philosophy of thirteenth-century

Franciscan thinkers (1936b), followed by the thoughts of St Thomas Aquinas (1947, 1965, 1973b) and St Anselm (1949, 1969b, 1987). Many of her writings were collected and reissued in two volumes published to commemorate her seventieth birthday (1978b).

Vanni Rovighi's philosophical exploration, however, extended beyond medieval philosophy. She also engaged with the thoughts of modern philosophers (especially Descartes, Spinoza, Kant and Hegel) and contemporary philosophers (notably Husserl, Scheler, Stein, Heidegger and Hartmann). Regarding the latter, she not only studied their works but also attended some of their lectures: in the summer semester of 1932 she participated in Heidegger's lectures in Freiburg, and in that of 1938 she attended Hartmann's lectures in Berlin (Sina 1990, 491; Vanni Rovighi 1976, 1980a).

In her teaching, Vanni Rovighi focused on neo-scholasticism, the most significant result of which is her three-volume work *Elementi di Filosofia*, first published in 1941 and reprinted multiple times, serving successive generations of students. In this work, Vanni Rovighi presents the key topics of Thomistic philosophy, developed in the context of modern and contemporary thought (1985–86). She also authored textbooks on epistemology, anthropology, and the philosophy of God (Vanni Rovighi 1963; 1980b; 1986).

VANNI ROVIGHI'S INTERPRETATION OF THE PHENOMENOLOGY OF HUSSERL
Vanni Rovighi's interest in contemporary philosophy, particularly phenomenology, arose at a time when it was virtually unknown in Italy. Her first work on phenomenology appeared just seven years after the first Italian study dealing with that same field (Vanni Rovighi 1930), and her 1938 book was the first monograph in Italy to comprehensively present Husserl's views (Vanni Rovighi 1938b; 1939). She also authored the "Edmund Husserl" entry in the renowned Italian *Enciclopedia Filosofica* (1957c), as well as numerous other texts on his philosophy (1938a; 1946a; 1946b; 1957a; 1957b; 1966; 1969a; 1973a; 1973c).

Vanni Rovighi considered Husserl's most significant achievement to be his formulation of the theory of intentionality and eidetic intuition. However, she attempted to demonstrate that one of the sources of the concept of intentionality could already be found in medieval Franciscan philosophy (Ales Bello 2018, 45). Regarding Husserl's concept of ideas, Vanni Rovighi argued that it is epistemological rather than metaphysical, and concluded that various regional ontologies cannot effectively compete with their corresponding natural sciences. That changes, however, when it comes to human life (especially moral and religious aspects), or formal ontology, where phenomenological analyses do yield results (Lenoci 1994, 62;

Vanni Rovighi 1969a, 69; 1939, 100–05). According to her, anthropology was an area where the phenomenological method has produced positive outcomes.

Furthermore, Vanni Rovighi believed that the theory of intentionality and universalizing abstraction were the elements that connected Thomism and phenomenology. The observation of convergent elements in various philosophical currents led her to the conviction that *philosophia perennis* exists (Vanni Rovighi 1959). However, this conviction was not about identifying Thomism as the *philosophia perennis*, but rather about recognising the enduring significance of the fundamental doctrines of Western philosophy and understanding philosophy as the progressive deepening of humanity's comprehension of reality. This process of arriving at knowledge of reality occurs through human rationality and its expression, which must be continually refined and enriched (Vanni Rovighi 1959, 192; cf. Sina 1990, 493; Mangiagalli 2008, 330–34).

Commenting on Husserl's thought, Vanni Rovighi argued that one of the greatest paradoxes of reality is the human being, who constitutes the world in their consciousness yet is, like all humans, a part of that world. Husserl attempted to resolve this paradox in *The Crisis* through an analogy: the ego that exists in time, from the past towards the future, presupposes the existence of the current ego—just as the various egos existing in the world are actualizations of the transcendental ego (Husserl, 1970, 178–86; cf. Vanni Rovighi 1975, 277–79).

Vanni Rovighi contended that this solution was entirely different from what neo-scholasticism proposed (i.e. that reality is intelligible because a supreme Intelligence created it—that humans can grasp its intelligibility because they participate ontologically in the Intelligence that founded it (1975, 278–79)). Husserl, on the other hand, rejected this perspective, and on her view this was due to his aversion to concepts that could be ascribed a religious character.

For this reason, although she valued phenomenology, she ultimately rejected it (Vanni Rovighi 1975, 269–79; Vasoli 1994, 32). However, Vanni Rovighi's rejection of phenomenology was not rooted in an acceptance of religious claims—she was firmly convinced of the complete autonomy of philosophical inquiry, particularly in anthropological and moral domains—but rather in a preference for better-substantiated arguments, regardless of whether or not they aligned with religious doctrines (Vanni Rovighi 1936a).

Vanni Rovighi also maintained that phenomenology did not provide a clear response to metaphysical problems, particularly those concerning

God, the soul and the relationship between contingent beings and the Absolute (Lenoci 1994, 65; Vanni Rovighi 1939, 162–64).

Therefore, although phenomenology was not her chosen approach to philosophy, it did teach her to “read Thomistic texts with different eyes” (Vanni Rovighi 1959, 192).¹

According to Cesare Vasoli (a historian of Italian philosophy), her stance stemmed from her belief that phenomenology is:

the most direct interlocutor of neo-scholasticism, capable of recovering the foundations of the philosophy of being and aiming to reformulate its essential principles within a debate open to the most rigorous results of gnoseology, epistemology, science, and contemporary analysis of consciousness. (Vasoli 1994, 31)

It should be noted, however, that in her anthropological considerations, especially those concerning the soul, Rovighi relied on the views of Max Scheler, as she explicitly stated several times (1980b, 194–98).

Nevertheless, the phenomenological aspect of anthropology solely aims to uncover, describe and thematize the data of consciousness (Vanni Rovighi 1978a, 8). In summary, what Vanni Rovighi offers here is an instrumental use of the phenomenological method typical of certain philosophical circles (which will be discussed more below). The fourth section of this article will explain that while this approach is not isolated, some consider it insufficient for gaining an adequate understanding of human metaphysics.

OUTLINE OF VANNI ROVIGHI’S ANTHROPOLOGICAL VIEWS

Anthropology, which Vanni Rovighi called the ‘metaphysics of the human being,’ was always at the centre of her intellectual pursuits. In this field, she drew upon the solutions offered by St Thomas Aquinas, not out of deference to his authority but due to the strength of his arguments (Sina 1990, 496). It was precisely because of Thomistic anthropology that she identified herself as a Thomist (Gregory 1994, 20).

It is worth noting at the outset that in her view, the term ‘naturalism’ can have various meanings. It usually refers to the reduction of every being to its corporeal existence. Nevertheless, since classical anthropology refers to the concept of nature, Vanni Rovighi describes it as ‘naturalistic’ in the proper sense of the term. As such, Thomistic anthropology is far removed

1. All quotations from Italian (Vanni Rovighi, Vasoli) and Polish (Wojtyła, Styczeń) are translated by the author.

from Cartesian dualism and monism, whether materialistic or spiritualistic (Vanni Rovighi 1985–86, vol. 3, 124).

Vanni Rovighi argues that the natural or social sciences cannot resolve certain issues concerning human beings. At some point in their reasoning, these sciences always include, either explicitly or implicitly, a reference to a particular conception of human beings (1978a, 13 ff.). Philosophy seeks justifications for the conception adopted. To achieve this, what or who a human being is must be examined. According to Vanni Rovighi, such an examination proceeds in two stages: phenomenological and ontological.

The phenomenological stage involves describing what experience and self-awareness reveal about the human being, while the ontological stage involves drawing conclusions from this description about the essence of the human being (Vanni Rovighi 1980b, 9 ff.). As the philosopher states:

From the data revealed by phenomenology, metaphysical psychology attempts to conclude what the human being is and about the principle that characterizes them—the principle traditionally called the soul. At this second stage, the path becomes more difficult, requiring subtle investigations, *diligens et subtilis inquisitio*, as St Thomas says. (Vanni Rovighi 1980b, 11)

As the first stage of anthropology, phenomenology is no mere reiteration of the theses of the human sciences or of philosophical conceptions, but rather a revelation of what these disciplines assume and imply (Vanni Rovighi 1980b, 10). She undertakes an analysis of the various conceptions of the human being that may underlie the vision of humanity presented in the natural sciences, social sciences, or culture. She argues that contemporary thought is dominated by the conceptions of the human being proposed by Marx, Nietzsche and Freud. She defines these conceptions as naturalistic, as they portray the human being as a product of a nature that is corporeal, in contrast to the long-standing tradition of understanding the human being as a spiritual being.

Vanni Rovighi considers Marx's conception to be naturalistic, as history—the context in which human beings are born—is determined by economic variables (Vanni Rovighi 1980b, 15 ff.). On this view, the equality of persons is assumed not as a fact but as a postulate. According to Vanni Rovighi, the concept of equality among persons originates in Stoicism and Christianity. In Stoicism, it is based on the acceptance of rationality as the specific difference distinguishing humans from other beings. In Christianity, it rests on the concept of divine filiation, which also presupposes the Stoic thesis. This raises the question of the basis for equality as proposed

by Marx. The latter believed that equality is determined by recognising the human being as a creature that produces the means of subsistence (Vanni Rovighi 1980b, 28 ff.): economic structures must be transformed to achieve equality. However, according to Vanni Rovighi, inequalities are caused by deeper factors, and Marx's view reduces the human being to mere matter.

On the other hand, Nietzsche's anthropology assumes people's inequality, making it more materialistic than Marx's conception (Vanni Rovighi 1980b: 32). Human reason is subordinated to instinct, and freedom is proportional to strength (Vanni Rovighi 1980b, 37 ff.). The ideal human being is one in whom vital values (strength, courage, and the ability to dominate) are predominant—in essence, a description of animality. The inhumanity of this approach stems from elevating these traits to the level of the highest values, which in turn leads to the reduction of the majority of people to mere means for achieving the goals of a select few individuals (Vanni Rovighi 1980b, 49). There is also a reduction of the human being in Freud's anthropology, wherein unconscious instincts play the leading role (Vanni Rovighi 1980b, 64).

Despite their differences, the conceptions of Marx, Nietzsche and Freud all share a naturalistic character, as they reduce the human being to nature understood as corporeality or animality. This occurs as a reaction to spiritualistic concepts originating in modernity with Descartes, whose anthropology serves as a reference point for modern and contemporary thought (Vanni Rovighi 1980b, 73 ff.).

Meanwhile, according to Vanni Rovighi, experience reveals two fundamental aspects of human life: its profound unity and its irreducibility to corporeality. The above runs contrary to Cartesian dualism, and also to the reduction of the person to pure spirit or a material body (Vanni Rovighi 1980b, 171). Throughout the history of philosophy, one of these aspects has often been emphasised at the expense of the other. The best examples of this, in Vanni Rovighi's view, are Plato and Aristotle. The former emphasised the spiritual nature of the human being and treated the body as a burden. The latter, by focusing more on the unity of the human being as expressed in the concept of the soul as the form of the body, overlooked its transcendence (Vanni Rovighi 1985–86, vol. 3, 160 ff.).

Therefore, a proper anthropology, derived from human experience, requires a holistic approach. The profound unity of the human being and their irreducibility to corporeality oppose both the reduction of the human being to pure spirit (which is less common today) and to corporeality (which is widespread today, especially among philosophers who accept ontological naturalism). As such, an anthropology that corresponds to experience will be both anti-naturalistic and anti-spiritualistic.

Thomistic anthropology presents the human being as irreducible not only to matter but also to the spiritual dimension:

[A human being] is not a kind of angel who has fallen into a world alien to them but is woven from the same fabric as other bodies; not only that, they have a soul analogous to the forms of other living beings. There is, therefore, a similarity among the beings that make up the universe, even if there are species-specific differences. A human being is part of nature, despite possessing a substantial form that, in some of its aspects, transcends nature and enables them to dominate it. (Vanni Rovighi 1978b, 16; cf. Gregory 1994, 23; Bettinelli 2008, 184–88)

Based on the unity of human experiences, as grasped in experience, and the irreducibility of the human being to the body, Vanni Rovighi identifies three fundamental characteristics of the human being: unity, spirituality and freedom. The first and third are revealed on the aforementioned phenomenological level, while the second must be arrived at deductively on the metaphysical level (Vanni Rovighi 1978a, 99; 1980b, 233).

A human being consciously perceives themselves, first and foremost, as a subject of feelings and experiences whose directly perceived characteristic is corporeality. Through reflection on their cognition, they further ascertain the existence of objects outside themselves. Subsequently, the human being perceives themselves as a unity: the bodily *ego* is the same *ego* as the cognitive *ego*. Moreover, a person experiences that it is the case that their most characteristic actions cannot be reduced to solely spiritual or physical facts. For instance, affections are partially rooted in blood ties and sexuality, but also transcend these biological foundations. Technical activity, which by its nature concerns matter, is an expression of our knowledge of matter and mastery over it. Social life, determined by bodily needs (such as defence and species preservation), contributes to spiritual development (Vanni Rovighi 1985–86, vol. 3, 169 ff.).

In order to apprehend the second characteristic of human being, which is spirituality, it is necessary, according to Vanni Rovighi, to refer to metaphysics, because conclusions about the nature of the human being must be drawn from the data grasped in experience. The thesis that human beings know themselves through the recognition of their actions follows the thesis that actions are known through their objects. Humans are capable of acquiring knowledge of necessary truths and universal concepts—this mode of cognition is an activity independent of their animality, and thus reveals a subject of activity that transcends animality. Vanni Rovighi refers here both

to Thomistic abstraction and to phenomenological eidetic reduction; the existence of the soul is inferred precisely from the existence of these specific human activities, because the existence of some act whose subject is a soul without a body implies the existence of the soul not only as the form of the body, but also as the subject (Vanni Rovighi 1980b, 192–94, 235 ff.).

Freedom is closely tied to the spiritual nature of the human being. Vanni Rovighi asserts that there is, on the phenomenological level, an experience of freedom. Free will is based on the desire for perceived good and is directed toward particular things, but only because these are recognised as good and understood as furnishing a means to achieving fullness of being (Vanni Rovighi 1980b, 236).² Freedom is a distinctive expression of human spirituality, especially when it transcends one's bodily nature, as in the case of sacrificing one's life for another (Vanni Rovighi 1980b, 174 ff.). While instinct is not difficult to explain, freedom, by its very nature, poses a question to the subject: *why* does one choose this particular thing? The answer, however, is not found in the desired object itself: rather, the subject grants the object its determining aspect by recognising and deciding in a practical judgment that it is good (Vanni Rovighi 1978a, 101).

Vanni Rovighi writes:

A human being reveals their spiritual nature in the world, where they come to know . . . corporeal things. Their spiritual nature is not a festive garment worn on special occasions, nor is it an escape from the world of everyday life, nor an intuition of pure beings; it manifests itself in the world that the human being knows, in the world of experience. (Vanni Rovighi 1985–86, vol. 3, 185)

A CRITICAL PERSPECTIVE

The foregoing presentation of Vanni Rovighi's views on neo-scholastic and phenomenological philosophy, and on the practice of anthropology, supports the conclusion that she advocated using the phenomenological method to provide a fuller description of the reality of the human being. This approach brings to mind another figure in contemporary Italian philosophy: Paolo Valori. This philosopher, a significant figure in the second wave of phenomenology in Italy, advocated combining Thomism with phenomenology in the study of moral experience. Valori believed such

2. On the issue of free will, an evolution in Vanni Rovighi's views can be observed. At first, she believed that the existence of free will was self-evident and undeniable (1985–86, 151 ff.); in later texts, she attempted to make this appeal to evidence more explicit and motivated (1980b, 219 ff.).

an analysis should proceed in at least three steps (a fourth could involve moving beyond philosophy into the realm of moral theology). The first is to establish the background for any such study, using the human sciences. The second involves applying the phenomenological method to distinguish moral experience and value. But while the phenomenological method proves fruitful, it is insufficient, and hence the need for a third step: ontological analysis. For Valori, this involves complementing phenomenology with analyses that ultimately allow for the constructing of a metaphysical system. Although he does not explicitly state that this refers to classical metaphysics, the proposed features of moral ontology roughly correspond to the Thomistic framework (Valori 1977, 1985).

It can thus be said that Vanni Rovighi represents a broader trend proposing the combination of classical philosophy and phenomenology—a position not unique to Italy. The idea of such an integration was already noted and, in a way, advocated by Karol Wojtyła. For instance, during a conference at the Catholic University of Lublin in 1967, he stated:

Alongside the rise of the philosophy of consciousness and the development of its proper cognitive tools (the phenomenological method), new conditions are being shaped for enriching the concept of the human person with the entire subjective, ‘consciousness’ aspect, which was, in some way, overshadowed in metaphysical ‘naturalism’. This enriched concept of the human person can and should be incorporated into the interpretation of Revelation. In moral theology, it must be postulated that this ‘transformation,’ which has already largely taken place in ethics, should be increasingly assimilated. (Wojtyła 1967, 1080)

At the same time, it should be noted that, at least according to the view of Wojtyła and the tradition he represents, it is impossible to replace metaphysics with phenomenology. As early as in his postdoctoral thesis on Max Scheler’s views, Wojtyła both recognised the need to use the phenomenological method and pointed out its insufficiency (Wojtyła 1959; Mazur 2023a).

Meanwhile, the goal of the philosophers of the Lublin Philosophical School in Poland was not merely to supplement metaphysics—particularly ethics or anthropology—with first-person experience of the subject such as is characteristic of phenomenology. As Mazur argues, the issue identified within this school was the lack of empirical grounding for the metaphysical image of the human being (2023b). This led to the conviction of the necessity to “re-empiricize the Thomistic system in its entirety” (Mazur 2023b, 279).

Mazur identifies two such attempts, undertaken, respectively, by Wojtyła and Mieczysław A. Krąpiec. While there is no need to delve into the details of Mazur's analyses here, his conclusion is significant:

Despite its realistic character, Thomistic anthropology was unable to demonstrate how to move from experience to a system. Its re-empiricization entailed assimilating cognitive tools developed in the modern philosophy of the subject, especially the phenomenological description of inner experience. (2023b, 285)

However, this use of the phenomenological method differs from that of Vanni Rovighi (or Valori in ethics). It is not about simply applying this method to describe and explain anthropological (and moral) facts. As Mazur notes:

In their interpretations, Wojtyła and Krąpiec searched for their 'own' anthropological facts which could constitute a starting point for the Thomistic system. Both looked for such facts in the first-person experience of the subjectification of personal acts. (2023b, 285)

Specifically, Wojtyła based his approach on the experience of performing a human act, while Krąpiec focused on the subjective experience of existence.

In this context, it is also worth mentioning Wojtyła's student, Tadeusz Styczeń, who emphasised the need to connect ethics with metaphysics. He addressed this issue as follows:

The way of linking the experience of morality with metaphysics—resulting in the ethics we propose—can, therefore, be most generally characterized either as the translation of relevant metaphysical theses into the language of the experience of morality (the concretization of metaphysics) or as—what ultimately amounts to the same thing—the metaphysical interpretation of the data of the experience of morality. (Styczeń 1972, 194)

Vanni Rovighi's position can thus be evaluated in this context as amounting to an attempt to address the diagnosed inadequacy of classical metaphysics. However, even such an approach would be considered insufficient by some proponents of Thomism, a few of whom advocate the re-empiricization of Thomistic anthropology. Vanni Rovighi, on the other hand, does not go that far: she limits herself to utilising

the phenomenological method to describe reality, and to then drawing metaphysical conclusions.

CONCLUSION

Based on the above, a number of conclusions can be drawn. First, Sofia Vanni Rovighi, a philosopher little known in Poland, belongs to the tradition of Italian neo-scholasticism. Despite her clear affiliation with this philosophical tradition, she played a significant role in Italian phenomenology, authoring the first Italian monograph on Husserl's views.

Second, Vanni Rovighi defends the anthropological thesis of the human being's psychophysical unity and irreducibility to corporeality. She opposes Cartesian dualism and the reduction of the person to either pure spirit or a material body, considering the most important characteristics of human beings to be unity, spirituality and freedom.

Third, she employs Thomistic arguments but, as she herself claims, reads Thomas' texts with different eyes thanks to phenomenology, incorporating elements inspired by Husserl's views into her own reasoning.

Fourth, such an approach may be considered in at least some branches of Thomism to be inadequate (see, e.g., Krapiec, discussed above). Certain proponents of classical anthropology regard the phenomenological solution (as proposed by philosophers such as Rovighi and Valori) as unsatisfactory, and advocate a re-empiricization of classical anthropology. Others, such as Vanni Rovighi and Valori themselves, argue that using the phenomenological method yields positive results. Furthermore, openness to working with other philosophical traditions and sciences is essential in this context.

In summary, Sofia Vanni Rovighi's views represent an intriguing example of a creative dialogue between classical Christian philosophy and contemporary thought—in this case, Husserl's phenomenology. As such, her views encourage philosophers of various traditions and backgrounds to engage in such collaboration.

REFERENCES

- Ales Bello, Angela. 2018. "Husserl, oltre Husserl: La fenomenologia secondo Sofia Vanni Rovighi." In *La fenomenologia in Italia: Autori, scuole, tradizioni*, edited by Federico Buongiorno, Vincenzo Costa, and Roberta Lanfredini, 35–62. Roma: Inschibboleth.
- Bettinelli, Carla. 2008. "Edith e Sofia: Sentire e sapere la persona umana." *Rivista di Filosofia Neo-Scolastica* 100 (4, suppl.): 183–94.
- Gregory, Tullio. 1994. "Gli studi di filosofia medioevale di Sofia Vanni Rovighi." In *Sapientiae studium: La giornata operosa di Sofia Vanni Rovighi (1908–1990)*, edited by Mario Sina, 13–26. Milano: Vita e Pensiero.
- Husserl, Edmund. 1970. *The Crisis of European Sciences and Transcendental Phenomenology*. Translated by David Carr. Evanston, IL: Northwestern University Press.

- Lenoci, Michele. 1994. "Il contributo della filosofia contemporanea negli studi di Sofia Vanni Rovighi." In *Sapientiae studium: La giornata operosa di Sofia Vanni Rovighi (1908–1990)*, edited by Mario Sina, 55–80. Milano: Vita e Pensiero.
- Mangiagalli, Maurizio. 2008. "Astrazione, intuizione e intuizione astrattiva: Sofia Vanni Rovighi e Edith Stein." *Rivista di Filosofia Neo-Scolastica* 100 (4, suppl.): 315–34.
- Mazur, Piotr Stanisław. 2023a. "From Experience to a Method: The Significance of Karol Wojtyła's Habilitation Dissertation in the Development of His Concept of Philosophical Cognition of Man." *Logos i Ethos* 61 (1): 121–38. <https://doi.org/10.15633/lie.61108>.
- . 2023b. "Wojtyła and Krapiec: Two Ways of Re-emperizing Thomistic Anthropology." *Roczniki Filozoficzne* 71 (1): 273–88. <https://doi.org/10.18290/rf23711.13>.
- Sina, Mario. 1990. "Sofia Vanni Rovighi." *Rivista di Filosofia Neo-Scolastica* 82 (2–3): 490–496.
- . 2006. "Vanni Rovighi, Sofia." In *Enciclopedia Filosofica*, vol. 4, 12013–15. Milano: Bompiani.
- Styczeń, Tadeusz. 1972. *Problem możliwości etyki jako empirycznie uprawomocnionej i ogólnie ważnej teorii moralności*. Lublin: TN KUL.
- Valori, Paolo. 1977. "Significato e metodologia della ricerca morale oggi: Scienze umane, filosofia, teologia." *Gregorianum* 58 (1): 55–86.
- . 1985. *L'esperienza morale: Saggio di una fondazione fenomenologica dell'etica*. Brescia: Morcelliana.
- Vanni Rovighi, Sofia. 1930. "Festschrift Edmund Husserl zum 70. Geburtstag gewidmet, Halle a.d. Saale 1929." *Rivista di Filosofia Neo-Scolastica* 22 (6): 491–94.
- . 1936a. "Filosofia e religione nel pensiero di Max Scheler." *Rivista di Filosofia Neo-Scolastica* 28 (suppl.): 157–69.
- . 1936b. *L'immortalità dell'anima nei maestri francescani del secolo XIII*. Milano: Vita e Pensiero.
- . 1938a. "Edmund Husserl." *Rivista di Filosofia Neo-Scolastica* 30 (3): 338–40.
- . 1938b. *La filosofia di Edmund Husserl*. Milano: Unione Tipografica.
- . 1939. *La filosofia di Edmund Husserl*. Milano: Vita e Pensiero.
- . 1946a. "Il movimento fenomenologico." *Rivista di Filosofia Neo-Scolastica* 38 (2–3): 207–11.
- . 1946b. "La fenomenologia di Edmund Husserl." *Humanitas* 1: 141–49.
- . 1947. *Antologia politica di Tommaso d'Aquino*. Milano: Cisalpino.
- . 1949. *S. Anselmo e la filosofia italiana del secolo XI*. Milano: Fratelli Bocca.
- . 1957a. "Husserliana." *Rivista di Filosofia Neo-Scolastica* 49 (1): 54–72.
- . 1957b. "Il colloquio sulla fenomenologia." *Rivista di Filosofia Neo-Scolastica* 49 (2): 197–211.
- . 1957c. "Edmund Husserl." In *Enciclopedia Filosofica*, vol. 1, 1146–52. Firenze: Sansoni.
- . 1959. "Edmund Husserl e la perennità della filosofia." In *Edmund Husserl 1859–1959: Recueil commémoratif publié à l'occasion du centenaire de la naissance du philosophe*, 185–94. The Hague: Martinus Nijhoff.
- . 1963. *Gnoseologia*. Brescia: Morcelliana.
- . 1965. *L'antropologia filosofica di San Tommaso d'Aquino*. Milano: Vita e Pensiero.
- . 1966. "Problema della conoscenza e problema del conosciuto." *Rivista di Filosofia Neo-Scolastica* 58 (2): 163–74.
- . 1969a. *La fenomenologia di Edmund Husserl: Note introduttive. Appunti dalle lezioni introduttive ai seminari per l'anno accademico 1968–1969*. Milano: Celuc.
- . 1969b. *Opere filosofiche di Anselmo*. Bari: Laterza.
- . 1973a. "La fenomenologia di Husserl." *Verifiche* 2: 3–17.
- . 1973b. *Introduzione a Tommaso d'Aquino*. Roma: Laterza.
- . 1973c. *La fenomenologia di Husserl: Appunti delle lezioni*. Milano: Celuc.

- . 1975. "Rileggendo alcuni testi husserliani sull'intenzionalità." In *Studi di filosofia in onore di Gustavo Bontadini*, vol. 2, 269–79. Milano: Vita e Pensiero.
- . 1976. *Storia della filosofia moderna: Dalla rivoluzione scientifica a Hegel*. Brescia: La Scuola.
- . 1978a. *Appunti di antropologia filosofica*. Milano: Vita e Pensiero.
- . 1978b. *Studi di filosofia medioevale*. Milano: Vita e Pensiero.
- . 1980a. *Storia della filosofia contemporanea: Dall'Ottocento ai giorni nostri*. Brescia: La Scuola.
- . 1980b. *Uomo e natura: Appunti per una antropologia filosofica*. Milano: Vita e Pensiero.
- . 1985–86. *Elementi di filosofia*. 3 vols. Brescia: La Scuola. First published 1963.
- . 1986. *La filosofia e il problema di Dio*. Milano: Vita e Pensiero.
- . 1987. *Introduzione a Anselmo d'Aosta*. Roma: Laterza.
- Vasoli, Cesare. 1994. "Gli studi di filosofia moderna e contemporanea di Sofia Vanni Rovighi." In *Sapientiae studium: La giornata operosa di Sofia Vanni Rovighi (1908–1990)*, edited by Mario Sina, 27–44. Milano: Vita e Pensiero.
- Wojtyła, Karol. 1959. *Ocena możliwości zbudowania etyki chrześcijańskiej przy założeniach systemu Maksa Schelera*. Lublin: TN KUL.
- Wojtyła, Karol. 1967. "Etyka a teologia moralna." *Znak* 159 (9): 1077–82.

Some Difficulties of Theology Developed in the Context of Science



A Critique of the Position of Grygiel and Wąsek

Ryszard Mordarski

ABSTRACT *Teologia ewolucyjna: Założenia – problemy – hipotezy (Evolutionary Theology: Assumptions – Problems – Hypotheses)*, by Wojciech P. Grygiel and Damian Wąsek, is an interesting and inspiring book. The attempt to formulate traditional problems of theology in the context of the natural sciences should command special attention today. And if it is also a successful and consistently pursued attempt, then we should welcome it with particular interest. In this article, however, I do not want to dwell on the advantages of the publication being discussed, but rather to make some comments that may appear relevant when seeking to assess the theses of its authors from the perspective of a classical theist entertaining a metaphysical rather than scientific perspective on theology. I will focus on four issues: (1) the concept of Revelation developed in the context of the natural sciences, (2) the understanding of evolution, (3) the metaphorization of theological language, and (4) the panentheistic perspective of theology practiced in the context of science. In conclusion, I state that the proposed development of theology in the context of science, despite the advantage of presenting old theological problems in a new perspective, is vulnerable to the accusation of pan-positivism, which entangles theology in too strict a context, depending as it does on the results of the empirical sciences.

KEYWORDS evolutionary theology; Grygiel, Wojciech; metaphors; panentheism; religious language; theology; Wąsek, Damian

✉ Ryszard Mordarski, Kazimierz Wielki University in Bydgoszcz, Department of Philosophy, Poland  ryszard.mordarski@gmail.com  0000-0003-2346-4572

©   *FORUM PHILOSOPHICUM* 30 (2025) no. 2, 131–45
ISSN 1426-1898 E-ISSN 2353-7043

SUBM. 23 May 2025 Acc. 13 September 2025
DOI: 10.35765/forphil.2025.3002.07

1. THE CONCEPT OF REVELATION IN THE CONTEXT OF THE NATURAL SCIENCES

The authors of *Teologia Ewolucyjna* believe that theological reflection cannot be fully autonomous and independent in relation to other approaches to investigating reality. Therefore, in order to be credible to our contemporaries, it should adjust itself to reflect the influence of the development of various non-theological disciplines, especially the natural sciences. In this belief, they refer to the words of Cipriano Vagaggini, who formulates the dilemma of contemporary theology as follows: “How to practice a theology that will be sensitive to the achievements of the empirical and mathematical sciences?” (Grygiel and Wąsek 2022, 15). Hence, a theology credible to our contemporaries must refer to the theology of Revelation, which shows what God’s agency in the world amounts to. The authors understand Revelation itself as an encounter between God and man, which is a dialogical event aimed at establishing a personal relationship between them. They therefore depart from the classical understanding of Revelation as God’s clarification of His plan for the salvation of humanity. In their opinion, in the twentieth century, the dynamic and dialogical reality of God’s agency has been revealed anew, on the one hand by the influence of existential philosophy, focused on the inner experience of the individual, and on the other by the *nouvelle théologie* movement, with its slogan of returning to original sources and living the faith inspired by the attitude of the early Church. The Second Vatican Council, in the Constitution *Dei Verbum*, confirmed these intuitions about Divine Revelation, emphasizing the aspect of the interpersonal relationship between God and man. After the Council, theologians began to define Revelation as a sign of God’s loving and saving involvement in the world and in human life. This anthropological and existentialist turn pertaining to the concept of Revelation is best expressed by the words of K. Rahner, as quoted by the authors:

... from the point of view of faith, the theologian is primarily interested in the question of the saving message of the reality he analyzes, and not this reality itself—it is only a space for the divine-human encounter. In other words, discovering Revelation consists in the fact that man, when faced with any signs, things, or arguments, asks himself: What significance does this have for his salvation? In what way is what he has before him an invitation sent by God to communion with Him? (Grygiel and Wąsek 2022, 22)

Grygiel and Wąsek accept Rahner’s belief that God himself is Revelation—such that when speaking of Revelation, one can only speak of the

Creator sharing Himself with man. Such a perspective emphasizes the clear primacy of personal relationships over doctrine and, in principle, redirects theology from an interest in doctrine to the development of the attitude of faith. It is therefore recognized that in theology the most important thing is the personal attitude towards God, and not the recognition of some doctrine, truths of faith, or intellectual claims. Doctrine itself is treated as an intellectual construct that can only obscure the living personal relationship with God.

However, a significant inconsistency in the book under discussion shows up here. Grygiel and Wąsek themselves admit that the relationship with God and the truth of doctrine cannot be opposed. Even the conciliar Constitution *Dei Verbum* understands Revelation simultaneously as an interpersonal relationship of God to man and as truths revealed by God: i.e. the doctrine taught by the Magisterium of the Church. It can therefore be said that they are not protesting against doctrine so much as against distorted doctrine, which omits the dialogical relationship between God and man. In their opinion, this distortion occurred for certain historical reasons (the controversy with Lutheranism) at the Council of Trent, which spoke of Revelation as a “saving truth” and as “teaching concerning morals,” and thus of a certain number of truths that God communicates to man. This state of affairs was deepened by the First Vatican Council, which understood Revelation as a doctrine of faith proclaimed by the Magisterium of the Church. In such an approach, the most important issue of God’s self-revelation was omitted, and the things revealed by God were emphasized (*revelata*). Hence, faith began to be understood as accepting certain truths about God, and not as an encounter with Him or opening up of oneself to a personal relationship. The content of faith itself began to be understood instrumentally, without reference to the intentions of the biblical authors or the context of the era. This way of thinking generated conflicts relating to the changing picture of the world, as new discoveries about the functioning of nature broadly construed demanded a revision of the revealed truths. Yet the modification of theological statements supposed to occur under the influence of new scientific discoveries has been hindered by the top-down establishment of doctrine. To overcome this, theology must turn to the empirical sciences in search of new inspirations.

If we assume that Divine Revelation should be understood as a dialogical event consisting of an encounter between God and man, which nevertheless requires the world as a necessary stage for this same divine-human interaction, then the created world must play an active role in helping to create such an occurrence. Theology therefore contains a transcendent

element, which is universal and unchanging, and an immanent element, which is the historically and culturally conditioned way of verbalizing this experience of encounter. Hence, the more reliable the vision of the world, the better the purification of the act of faith, and the greater the dialogue between theology and the empirical sciences, the greater the chance of capturing “anchor points” for the divine-human relationship. Pictures of the world are only, in a somewhat Kantian way, a stage, and are not in themselves infallible doctrinal elements. They result from the culture and spirit of the times, and lose their relevance when the empirical data changes. Nevertheless, although the natural sciences are only extraneous sites for arriving at theological knowledge, they are important in the sense that they provide a current world-picture; they therefore become a model for practicing theology. For this reason, Grygiel and Wąsek would like to refer to new scientific models in order to search—through the prism of the latter—for theological and transcendent meaning, which will allow for the discovery of knowledge about God. Using the method of interpreting the data provided to theology by the natural sciences, they assume that the more reliable the vision of the world, the better the purification of the act of faith. They suggest that probably all theologians of the past had a more or less false understanding of God, because they operated under an erroneous (falsified) world-picture, which distorted theological knowledge. Ultimately, therefore, the world of nature, as a stage, is not just passive decoration for them, but actively shapes “anchor points” for the divine-human relationship. Therefore, if dogma always has a historical-evolutionary dimension, this means that God speaks to people through some historical context or other, and that in order to understand the proper meaning of dogma we therefore need to historically contextualize it. Since the most important context today is contemporary science, theology must be practiced in the context of science. Thus: “Evolutionary theology can therefore be a project of creating new interpretations of the truths of faith using metaphors, symbols, and images that correspond to contemporary scientific discoveries” (Grygiel and Wąsek 2022, 56).

Before we consider the problems posed by the model of evolutionary theology proposed here, however, let us briefly examine what the term “evolution” itself means to these authors, and how the evolutionary processes of nature are intended to impact theology.

2. EVOLUTIONISM AS A PARADIGM FOR PRACTICING THEOLOGY

Grygiel and Wąsek fail to define in what sense they use the term “evolution” in their project of evolutionary theology. However, when reading their book

carefully, one can distinguish at least three meanings of this term, which interpenetrate each other in various arguments.

The first meaning is very narrow, and limited to the theory of evolution found in the natural sciences. Evolution is understood here as a scientific theory explaining the origin and development of life on Earth. If theology were to model itself on such an understanding of evolution, it would express a progressive idea, taking on board new scientific developments. Therefore, Christian doctrine would also have to develop in order to cope with today's science. The basic difficulty that arises with such an understanding of the relationship between theology and science will always come down to the fact that today's findings in the natural sciences are not the final word on the nature of the world. In biology, the dispute over whether evolution is blind or designed (intelligently?) is still unresolved, and in physics and cosmology, after the latest data from the Webb telescope, there is even talk of such discoveries being made as will change the foundations of our understanding of the world. The deeper we look into the universe, the better we understand that Einstein's standard equations, describing the expansion of the universe according to the theory of relativity, do not describe this expansion in the way that the theory should suggest. And it is not my intention here to undermine the theory of evolution. I only want to point out that in today's discussions in the field of the biological sciences, between biochemists and geneticists, certain new facts may emerge that would force scientists to significantly modify their understanding of the mechanisms of the evolutionary process. This could cause philosophers of biology to look more favorably at certain aspects of teleological explanations, in order to better understand complex evolutionary processes. After all, as William James argued, we are not able to imagine today what our future science will be like, especially when it comes to the research methods used there. Therefore, if the theory of evolution is to become a model for practicing theology and establishing the truth of doctrine, we must accept that we base the truths of faith on temporary theories, because scientific theories are always temporary, and we do not know what they will be like in the future. This can only have catechetical and pastoral significance, because we want to explain the truths of faith to modern man in the light of new scientific theories. However, this is only a practical problem, not a theoretical one. It can also result from the confusion experienced by theologians who, no longer able to explain doctrine in the language of classical metaphysics, reach for the language and theories of modern science with the aim of practicing a more understandable theology in that context.

In the second sense, Grygiel and Wąsek use the term “evolution” to describe the general evolutionary paradigm that prevails in all contemporary science. This means that any field of knowledge, in order to deserve being called ‘scientific,’ must be cultivated from an evolutionary perspective. Therefore, not only the natural sciences, but also the historical and humanistic sciences, including of course theology, to be recognized as science, must adhere to the evolutionary paradigm, which in the humanities goes under the label “historicism.” It should be emphasized here that Grygiel and Wąsek adhere to a cumulative model of knowledge development: i.e. they treat the most current theories as the truest. Thinking in terms of the evolutionary paradigm, they assume that the development of science is straightforwardly evolutionary, moving from worse to better theories. Therefore, they focus all their efforts on reconciling theological theses with today’s science, and not on deepening the theological theses themselves. They assume that if some theological truth were to be incompatible with the current state of natural science, then it would have to be adapted to the latter as swiftly as possible. They do not assume that any theological truth is universally true, but rather maintain that it must always be reinterpreted and adapted to current scientific theories. And yet, the question arises whether forming theology on the basis of the current state of science will not always be to some extent arbitrary, in that it makes doctrine dependent on the state of scientific knowledge.

Most important, however, is the third level of understanding of the term ‘evolution,’ which Grygiel and Wąsek implicitly assume in everything they write about evolutionary theology. Specifically, they recognize that even if current scientific theories are partial, hypothetical, and based on questionable, not always well-founded, assumptions, each future scientific theory, however far it may go in supplementing or verifying our current one, will always be expressed in terms of the evolutionary paradigm. In short, although our current scientific theories are provisional and open to verification, there is one certainty: namely, that in the future, they will always still be expressed in terms of the evolutionary paradigm. The most general evolutionary scheme of future science cannot be denied, because reality itself is dynamic and evolutionary. Therefore, as long as science describes reality, it must remain evolutionary in the broadest sense. And this is the most important argument in favor of evolutionary theology—and, indirectly, also an argument for theology practiced in the context of science, because it is science that reveals to us the truth about the dynamic and evolutionary scheme of reality itself. Even so, the question arises of whether it is correct to perceive every dynamic change and every process as evolutionary, to

such an extent that even the fact that history is taking place means that there is evolution. Is referring to the theory of evolution, treated as the most general paradigm of scientific thinking and even, to some extent, as the ideology underpinning modern science, which excludes everything not presented in evolutionary terms, a correct method for practicing theology? What novelty, one may ask, does the paradigm of evolution bring to theology? Why is practicing theology in the context of science more legitimate than practicing it in the context of metaphysics? For a theologian, it should not matter whether their work is accompanied by a psychological intuition to the effect that they are practicing their reflection in a manner modelled on the modern natural sciences. Should one not recognize that since most scientific claims are provisional, one should first of all consider what the status of such claims really is, and to what extent they conflict with religious doctrine? And, most importantly, is constructing a bottom-up theology based on the natural sciences really a better proposition than traditional top-down conceptions?

3. METAPHORICAL LANGUAGE IN THEOLOGY

The program of evolutionary theology focuses on modifying and reinterpreting existing formulas of faith, including official statements of the Magisterium of the Church, in such a way as to harmonize them with the theory of evolution. This task primarily concerns changing the theological language in which these formulas have been defined. Grygiel and Wąsek hold that “the doctrine of the Church is not a monolith once formulated, but a living tissue that is constantly developing. Defined truths of faith that clash with new, evolutionary images of the world can therefore be modified and their change doesn’t go beyond the boundaries of orthodoxy” (2022, 56). The basic problem here is the proper distinguishing of the logical meaning of a truth of faith from its historical context, and the linguistic picture of the world in which this truth has been formulated. The authors make this task much easier for themselves by claiming that the entire language of theology is metaphorical, in that the concepts used to describe finite nature are referred to the description of infinite reality in the case of speaking about God. When we say that God is the Creator of the world, or the Person, we use metaphors and models referring to human experiences. The metaphorical nature of theological language allows us to constantly search for new pictures that will be more adequate. Hence, Grygiel and Wąsek write, “evolutionary theology can ... be a project of creating new interpretations of the truths of faith using metaphors, symbols and images that correspond to contemporary scientific discoveries” (2022, 56). Referring

to the idea of extending the language as proposed by John Macquarrie, they claim that “to adapt theological language, one metaphor should be translated into another metaphor that is understandable in today’s image of the world” (Grygiel and Wąsek 2022, 47). However, the question arises of how, on such a constructivist approach, we can preserve the original meaning of the doctrine of faith.

The difficulties associated with the extreme metaphorization of religious language proposed by authors such as John Macquarrie (1994) and Sallie McFague (1982) were pointed out by William Alston, amongst others. He asked whether, in discussions about God, there are so-called irreducible metaphors: i.e. metaphors that, in the strong sense, cannot be formulated in literal terms, even in part. He calls those theologians who answer this question affirmatively “pan-metaphoricists.” This position began to be popular in the 1970s, and in its extreme form claimed that the concepts of our language refer to descriptions of our world and are completely inadequate when used in reference to God. In its weakened form, it maintained that although we can use our language to talk about God, it is always so vague and unclear that in fact we do not know to what extent it refers to Him. In response, Alston argues that

if we can make any assertion about God definite enough to have truth-value, it will be in principle possible to say the same thing literally, at least partially, even if that requires introducing new terms (or new meanings for old terms) into the language for that purpose. (Alston 1989, 2)

The basic problem is whether we can somehow overcome the impossibility of talking in a literal sense about God, who is wholly other and transcendent to the world. In this situation, the appeal to metaphorical language did seem very promising. It allowed us to formulate truth-apt propositions about God without using terms in any literal way. But the following difficulty arose: if metaphorical statements cannot be at least partially expressed in literal terms, what status does theological language in general have when it comes to such statements’ claim to truth? Can personalistic predicates about God, for example, be paraphrased in such a way that they say something literally true about Him, even though they start from an understanding of a person that applies to human beings? Alston shows how it is possible in language to speak about immaterial persons in a way that avoids irreducible metaphors. It seems that, for example, the biblical metaphor “The Lord is my shepherd” is reducible to the more or less literal idea that God protects me, cares for me, and will never abandon me. This paraphrase speaks at least

in part in a literal sense of God's relation to man, and thus in a broader sense of God's relation to the world. Alston shows that in a similar way we can attribute to God such attributes as timelessness and immutability, and also—despite some difficulties—understand them in a literal sense (see Alston 1989, 64–102). He therefore concludes that

either the pan-metaphoricist abandons the aspiration to significant truth claims or he revokes the ban on literal predictability. He cannot have both. Which way he should jump depends, *inter alia*, on the prospects for true literal predication in theology. (Alston 1989, 37)

Treating religious language in a way similar to the language of science seems to be a mistake precisely because we then transfer the metaphors and models used to explain natural reality onto religious reality. A much better solution is to appeal not to science, but to metaphysics. As Jacques Maritain pointed out, metaphysical knowledge is the highest form of purely natural knowledge, the purpose of which is to seek ultimate rationality by pointing to God as the First Cause and Author of nature. In this way, the existence of God and His perfections (unity, simplicity, immutability, perfection, etc.) can be known by causal ascent from the natural world to the First Principle of all being (*sub ratione, primi entis*):

The knowledge of God thus obtained by the reason constitutes that prime philosophy, metaphysics, or what Aristotle called “natural theology.” It is ana-noetic knowledge or knowledge by analogy, which is by no means to be confused with metaphorical knowledge. It makes use for the knowledge of God of those notions which we seek for in things, and which we, because of this, in as much as they are realized in created things, conceive as limitations, but which in themselves, in their significance, imply neither limitation nor imperfection, and which can therefore be applied in a rightful sense to the Uncreated as well as to the creation. A light of knowledge broken in the prism of creation, but veritable for all that. (Maritain 1937, 306)

According to Maritain, analogy differs from metaphor in that it allows us to know God, albeit in an imperfect way and one limited to the concepts of our language. It nevertheless allows us to grasp the uncreated reality in the divided mirror of transcendental concepts, which are common in an analogical way to what is created and uncreated. However, it is extremely important to clearly distinguish the use of analogy in the domain of faith and in the domain of metaphysics. The difference here is fundamental,

because in the case of metaphysics, analogy constitutes the very form and rule of knowledge. God is not reached either in his personality or in his nature, in the indivisibility of his purest and simplest essence, but only as he reveals himself in the variable but true reflections that are shown to us by things proportional to our reason. Not only is the way of knowing a human characteristic, but also the object itself, which is the goal of knowledge, is understood only to the extent to which it allows itself to be grasped by human reason, by appearing in the mirror of sensible things, and via the analogy of being. Above this wisdom of the natural order stands theology, which rationally develops the truths contained in the deposit of Revelation. Its certainty is greater than that of metaphysics, because it receives its principles from reason illuminated by faith. But theological knowledge still takes place in the symbols of language and through the medium of human thought. It cannot be otherwise, because God speaks our language so that we can know Him. According to Maritain, however, we can say that it is the object of faith, taken not from the side of the thing itself in which we believe, *ex parte ipsius rei creditae*, but from the side of the means or signs that serve the faithful person, *ex parte credentis*. Here we see, in a certain sense, a return to the method of knowledge by analogy—to the extent that Revelation uses human terms, but not by analogy to creation, as in metaphysics, but by analogy with the very mystery of the inner life of God, who will be known face to face in the beatific vision (see Maritain 1937, 307).

All this leads to the conclusion that in defending the thesis about the metaphorization of theology, we should firstly expect Grygiel and Wąsek to provide a much stronger justification for the necessity of using so-called ‘irreducible’ metaphors in theology (and thus answer precisely the question about the sense in which theology is metaphorical, and whether it includes all formulas of faith), and secondly demand that they better justify the thesis that the use of metaphors, analogies and myths in the contemporary natural sciences is a procedure identical, or at least significantly similar, to the use of analogical language in theology.

4. PANENTHEISM IN EVOLUTIONARY THEOLOGY

When presenting God’s relation to the world, evolutionary theology tends to understand it from a panentheistic perspective. This is a common tendency of all contemporary evolutionary theologians, and the authors of the book we are discussing are no exception in this respect. This tendency, it seems, stems from the error I mentioned in the first objection: that is, from giving an excessive role to nature, or to what Grygiel and Wąsek called

“the stage,” in understanding God’s relation to the world. It aims to find some balance between the transcendence and immanence of God, in which

the Being of God includes and penetrates the whole universe, so that every part exists in Him but (as against pantheism) that his Being is more than, and is not exhausted by, the universe. (Peacocke 1993, 371)

Grygiel and Wąsek hold that the modern scientific method allows for a significant correction of the inadequate traditional theistic approach, which introduces a division into natural and supernatural reality. It enables a more precise presentation of the theological understanding of how the world of divine immanence is immersed in the world of divine transcendence. Referring to the position of John Peacocke, who understood “everything in God,” they state that “this concept is panentheism, or an ontological position, connecting the immanent and transcendent order in such a way that the immanent order is ‘immersed’ in the transcendent” (Grygiel and Wąsek 2022, 148). On the one hand, this concept eliminates the arbitrariness of the transcendent God’s intervention in created reality, and on the other, in an understandable way, it shows the causal action of God in the world through the laws of nature. Grygiel and Wąsek state that

adopting the position of panentheism allows for a redefinition of the commonly accepted division of reality into natural and supernatural. Divine immanence in the created order makes all natural events the work of God, and there is no area of reality that would remain outside His causal influence. (2022, 232)

One might think that this is a rather moderate panentheism, although the abolition of clear boundaries between God and His creation may pose certain difficulties for the theist. Nevertheless, supporters of evolutionary theology, such as Philip Clayton, believe that panentheism constitutes a kind of *modus vivendi* between radical theism and various types of pantheisms that have been a temptation for Christian thinkers for centuries. Following Jürgen Moltmann, he emphasizes that a holistic understanding of creation necessarily shows God’s panentheistic relation to space and time. The time-space dimensions become to some extent divine attributes; we should think of God as

... coextensive with the world: all points of space are encompassed by God and are in this sense “within” him. Nonetheless, created space is precisely that created, contingent. Only God himself has the ontological status to be

absolute and to contain all space within himself. In short: finite space is contained within absolute space, the world is contained within God; yet the world is not identical to God. (Clayton 1997, 90)

Clayton emphasizes that dialectical thinking is needed in theology to show that the world is both different from God and completely dependent on Him. This means that to some extent the effects on the world also effect God, although the world remains a contingent and accidental being relative to Him.

Of the six arguments in defense of panentheism that Clayton cites, the most interesting seems to be the argument from divine causality. If God is ontologically 'outside' the world, then His agency in the world seems to be an intervention 'from outside' in the natural order of the latter. Clayton concludes, like Grygiel and Wąsek, that the understanding of God's agency seems more coherent when we understand it as a relation of God to the world that is analogous to the relation of mind to body. This panentheistic analogy suggests that there is no ontological difference between divine action and the regularities of the laws of nature. From a theological perspective, the laws of nature are essentially "descriptions of the predictable regularity of patterns of divine action" (Clayton 1997, 101). The laws and regularities of nature, although they are autonomous actions of nature, are nevertheless, thanks to God's omnipresence in the world, ontologically identical with God's intentions. This understanding removes the difficulty of perceiving God as controlling the regularities and laws occurring in the world 'from the outside.'

Let us consider, however, whether there is any alternative to the panentheistic doctrine of God's relation to the world that evolutionary theology presents us with. Is the doctrine of God's omnipresence that so entangles Him in creation that He becomes changeable and responsive to the world the only perspective for Christianity? Mariusz Tabaczek, assessing panentheistic doctrines, indicates that among the various characteristics of panentheism, evolutionary theologians (J. Moltmann, J. Peacocke, J. Polkinghorne and P. Clayton) emphasize the following in particular: (1) that God encompasses or contains the world (the substantive or locative notion); (2) that God binds up the world by giving the divine self to the world; (3) that God provides the ground for that which emerges within, or for the emergence of, the world (see Tabaczek 2021, 157). The essence of panentheism is not only that everything is in God, but also that divine immanence permeates the entire created order of nature. This feature is distinguished from classical theism which, while emphasizing that God exists in all things, avoids the

statement that everything is in God. For this reason, theists such as Niels Gregersen emphasize that classical Christian doctrine recognized God's immanence without falling into panentheism (see Gregersen 2004, 19–35). At the same time, Thomas Aquinas spoke of God's being present in the world in three ways: (1) by virtue of His power, as all things are subject to it (in opposition to those who claim that visible and corporeal things are subject to the power of a contradictory principle); (2) thanks to His presence, as all things are bare and open to His eyes (in opposition to those who deny God's presence in inferior bodies); (3) in virtue of His essence, as the cause of being of all things (in opposition to those who assume that there are creatures mediating being from God down to the lower creatures) (ST I, 8, 3). This allows us to show that although the world is dependent on God for its existence, "the natures and activities of creatures cannot affect or have a real feedback effect on God" (Tabaczek 2021, 163). On the other hand, another contemporary Thomist, Brian Shanley, presents the issue of the omnipresence of God in the following terms when he writes that

Aquinas clearly thinks that God is "related" to the world in the sense that he creates, loves, knows, wills, governs, and redeems the world. The denial that God is "really related" to the world does not dispute any of these claims. It simply denies that God's causal activity, and any relational terms thereby ascribed to him, implies any alteration in his being. When God acts so as to bring creatures into relationship with him, all of the "happening" is located in creation rather than in God. (Shanley 2002, 59)

The position of traditional theism thus allows for an account of God's immanence and omnipresence in the world that in no way violates the perspective of radical transcendence so clearly rejected by the proponents of panentheism, who advocate evolutionary theology.

Therefore, in the more cautious conceptions of divine embodiment in creation, it is maintained that the position of panentheism will prove true only in an eschatological reality, when—in the words of St. Paul—God will be "all in all." It is for this reason that John Polkinghorne emphasizes that "in my view, panentheistic language is best reserved to express eschatological destiny rather than to describe present reality" (Polkinghorne 2000, 95).

5. CONCLUSION

The project of evolutionary theology is, in fact, a proposal of a kind of scientific and theological positivism, which claims that the only proper path to theology runs from contemporary science. It seems that this cannot be

considered a form of theology *simpliciter*, but rather is a theology of nature which, with reference to contemporary science, tries to solve certain theological problems in a new way. It is therefore a contextual theology, practiced in the context of the contemporary empirical sciences. It stems from the conviction that theology cannot fall into a “metaphysical ghetto,” but instead must conduct a dialogue with other ways of investigating reality. There are certain issues that science itself takes over from philosophy, such as the nature of the mind, of causality, or of temporality. These require at least preliminary meta-scientific research. As John Polkinghorne says,

I believe that too many theologians fail to treat what science has to offer with the appropriate degree of seriousness that would enable them to acknowledge adequately their contextual role. (Polkinghorne 2009, 8)

Polkinghorne believes that there are not many theologian-scientists. However, if theology is to be effectively pursued in the context of science, it must be able to answer the important questions that classical theology has addressed to the latter. It must therefore be pursued in a manner similar to science itself: that is, in terms of a bottom-up approach to thinking. He writes that “I strongly believe that it is possible to do theology in a bottom-up fashion and that its pursuit in the context of science will indeed require just this kind of approach” (Polkinghorne 2009, 30). Science and theology are, of course, different, but they complement each other in the search for truth. They are partners in the human effort to seek truth and understanding. Grygiel and Wąsek fully agree with such a dialogical approach to the relationship of theology and science, but does this mean an unequivocally scientific contextualization of theology? Is it not enough to assume that today’s science is an auxiliary discipline of theology, just as archaeology is an auxiliary discipline of biblical studies. Science, according to this view, would only provide raw material for the work of the theologian, who seeks truth by assessing the motivations of faith. On this approach, the project of evolutionary theology would not be a new evolutionary paradigm of theology, but rather some sort of (more modest) theological reflection on the contemporary state of the empirical sciences. It seems that Grygiel and Wąsek are doing just that when they try to reconsider the old (traditional) formulations of Christian dogmatics regarding evil, original sin, and the existence of the soul in the light of new achievements in biology and neuroscience. We may say that they succeed very well in that regard, but it is separate issue, and one that we will not seek to address here.

BIBLIOGRAPHY

- Alston, William P. 1989. *Divine Nature and Human Language: Essays in Philosophical Theology*. Ithaca, NY: Cornell University Press.
- Aquinas, Thomas. 2006. *Summa Theologiae*. Vol. 2, *Existence and Nature of God (Ia. 2–11)*. Translated with introduction, notes, appendices, and glossary by Timothy McDermott, O.P. Additional appendices by Thomas Gilby, O.P. Cambridge: Cambridge University Press.
- Clayton, Philip. 1997. *God and Contemporary Science*. Edinburgh Studies in Constructive Theology. Edinburgh: Edinburgh University Press.
- Gregersen, Niels Henrik. 2004. “Three Varieties of Panentheism.” In *In Whom We Live and Move and Have Our Being: Pantheistic Reflections on God’s Presence in a Scientific World*, edited by Philip Clayton and Arthur R. Peacocke, 19–35. Grand Rapids, MI: Eerdmans.
- Grygiel, Wojciech P., and Damian Wąsek. 2022. *Teologia ewolucyjna: Założenia – problemy – hipotezy*. Kraków: Copernicus Center Press.
- Macquarrie, John. 1994. *God-Talk: An Examination of the Language and Logic of Theology*. London: Xpress Reprints.
- Maritain, Jacques. 1937. *The Degrees of Knowledge*. Translated by Bernard Wall and Margot R. Adamson. Glasgow: University Press.
- McFague, Sallie. 1982. *Metaphorical Theology*. Philadelphia: Fortress Press.
- Peacocke, Arthur. 1993. *Theology for a Scientific Age: Being and Becoming—Natural, Divine, and Human*. Philadelphia: Fortress Press.
- Polkinghorne, John. 2000. *Faith, Science and Understanding*. New Haven, CT: Yale University Press.
- Polkinghorne, John. 2009. *Theology in the Context of Science*. New Haven, CT: Yale University Press.
- Shanley, Brian J. 2002. *The Thomist Tradition*. Dordrecht: Kluwer Academic.
- Tabaczek, Mariusz. 2021. *Divine Action and Emergence: An Alternative to Panentheism*. Notre Dame, IN: University of Notre Dame Press.

Wittgenstein, Relativism, and the Second-Person Perspective

Piotr Szalek

ABSTRACT This paper addresses the problem of the relativist implications of Wittgensteinian non-cognitivism. If moral and religious language are only an expression of language users' attitudes, then both moral values and religious beliefs will be relative to just those language users. The paper attempts to respond to this charge in the following two ways. First, it seeks to show the common conceptual structure underlying the accusation of relativism as it relates to both Wittgenstein's non-cognitivism and his position on scepticism, where the latter reflects his contextualist anti-sceptical strategy (which is also charged with relativism). Second, it seeks to demonstrate that in both cases it is possible to offer a non-relativist reading of Wittgensteinian thinking by affirming the commensurability of different world-views through an appeal to the second-person perspective, taken as characteristic of the human way of living (or human "form of life").

KEYWORDS Ludwig, Wittgenstein; Non-Cognitivism; Relativism; Scepticism; Second-Person Perspective.

Acknowledgments

This research was supported by the University of Oxford project New Horizons for Science and Religion in Central and Eastern Europe, funded by the John Templeton Foundation. The opinions expressed in the publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation. I am much indebted to Simon Blackburn, Robert Brandom, and Andrew Pinsent for insightful discussions on many of the issues related to the paper. An earlier version of the paper was presented at the "2013 IRC Conference: The Second-Person Perspective in Science and Humanities," Ian Ramsey Centre, St Anne's College, University of Oxford, Oxford (July 17–20, 2013). I would like to thank the conference audience for helpful comments, especially Stephen Darwall, Arlyn Culwick, Samuel Hughs, Birgit Kremmers, Michał Leśniak, Stephen Mulhall, Mikołaj Składkowski-Rode, and Ralph Weir. I also owe a special gratitude to the editors, especially Jakub Prus, Carl Humphries, Maciej Jemioł, and Szczepan Urbaniak, and anonymous referees of *Forum Philosophicum* for helping me to improve the paper.

✉ Piotr Szalek, Catholic University of Lublin, Poland  piotr.szalek@kul.pl  0000-0003-0805-425X

1. INTRODUCTION

One of the most striking elements of the so-called later philosophy of Wittgenstein is his characterisation of moral and religious language in terms of non-cognitivism. According to this view, both moral and religious language are non-referential: that is, they do not refer to anything in the world such as objectively existing moral values or deities (see CV, 16, 64, 85).¹ Moral and religious language are not truth-apt, and are simply the expression of the moral or religious attitudes of language users. The latter are engaged in the corresponding moral and religious “language games,” whose rules should not be breached as doing so leads to philosophical and linguistic confusion. If both moral and religious language are made up of particular and distinctive language games, then they rest on their own distinctive rules and claims, and need no justification via external or rational means as they are performing only an expressive role.

The serious problem that arises from such a picture of moral and religious language is that of relativism (see Schönbaumsfeld 2023, 46; Kusch 2011, 39): if morality and religion are just a matter of expressing attitudes on the part of language users, then moral values and religious beliefs will seem to be nothing more than relative to those particular language users (see LC, 56). If they do not refer to any objective moral values or deities, they will always be “situated” or “contextualised” by the particular attitudes of certain moral agents or religious believers.

In this paper, I would like to examine and refute this problematic relativist commitment associated with non-cognitivism by looking at a similar difficulty that we encounter in Wittgenstein’s discussion of scepticism in his last work, *On Certainty*. In the latter, he develops a framework-invoking approach to the refutation of scepticism: a given commitment’s epistemic justification is furnished by the conceptual framework within which, exclusively, it can possess its content. According to one interpretation:

There could be . . . different epistemic systems, none of which would be intrinsically correct; each of them would be . . . as good as any other, and would certify as . . . justified different propositions. As a consequence, knowledge . . . would always be *situated*: what counts as knowledge within one system of justification might not be so within another. . . . The passage from one epistemic system to another would always be a form of *conversion* or *persuasion*, reached through a-rational means. (Coliva 2010, 1)

1. In the present paper I refer to the writings of Wittgenstein using the abbreviations explicated in the bibliography, specifying the numbered remarks or sections of the relevant works. In the case of secondary literature, all references are to page numbers within the texts cited.

There is, then, a remarkable similarity between the conceptual mechanism standing behind the non-cognitivist account of moral and religious language and the framework-invoking refutation of scepticism. Moreover, the serious problem which arises from both views is that they each appear to be committed to cognitive or epistemic relativism.²

Taking as its point of departure the so-called “hinge-epistemology” interpretation of Wittgenstein’s discussion of scepticism as formulated by Analisa Coliva (2003, 2010; see also Baghrarian and Coliva 2020, 110–14), this paper argues for its applicability to the problem of the relativistic commitments of Wittgensteinian non-cognitivism, and seeks to show that a kind of second-person perspective implicitly present in *On Certainty* can save us from having to grant that there are different mutually inaccessible world-pictures in play. Accessibility or commensurability is possible, because these are furnished by a unitary human world-picture community entailing limits to our access to language (in terms of what is conceivable for us). In other words, the limits of human language (consisting of different language games) determine the limits of our human world (its conceivability and commensurability).

In order to accomplish this goal, I shall proceed in the following way. Firstly, I introduce the most distinctive features of the Wittgensteinian anti-sceptical strategy. Secondly, I explain the arguments in favour of the interpretation that charges the Wittgensteinian view on scepticism and non-cognitivism with being committed to epistemic relativism. Thirdly, and finally, I offer a non-relativist reading of both scepticism and non-cognitivism by showing the commensurability of different attitudes or conceptual schemes within Wittgenstein’s approach, in terms of the second-person perspective.

2. Following Grayling (1988, 118), we can distinguish between cultural and cognitive relativism: “Cultural relativism is the thesis that there are differences between cultures or societies, or between different phases in the history of a single culture or society, in respect of social, moral, and religious practices and values.” However, “cultural relativism is not philosophically problematic, for it is clear that our being able to recognise cultural differences of the kind described presupposes an ability on our part to gain access to other cultures so that we can recognise the differences as differences.” The real problem is cognitive relativism, as “it is the view that there are different ways of perceiving and thinking about the world or experience, ways possibly so different that members of one conceptual community cannot at all grasp what it is like to be a member of another conceptual community.” The problem of the relativist consequences of Wittgenstein’s later philosophy has been noted by many philosophers, both relativists and anti-relativists (see Rorty 1979; Hintikka and Hintikka 1986; Grayling 1988; Haller 1995; Glock 1996; Kirk 1999; and Boghossian 2006). Recently, an anti-relativist reading of Wittgenstein’s later views has also been propounded to varying degrees (see Barret 1991; Putnam 1992; Grayling 2001; O’Grady 2002, 2004; Coliva 2003; Blackburn 2004, 2007; Williams 2007; and, especially, Coliva 2010).

The latter perspective emphasises that ethical and moral understanding is not just a matter of recognising facts, but also of entering into a reciprocal relationship or attitude with others. A commitment to another person's standpoint is taken to be essential for a recognition-based relationship.

2. THE ANTI-SCEPTICAL STRATEGY OF *ON CERTAINTY*

Scepticism is the view that knowledge or rational (justified) belief is impossible, either in general or with respect to a particular domain. Modern scepticism is based on the assumption that for a proposition to be known, it must either be evident (i.e. self-evident, or evident to the senses), or be adequately supported by other propositions that are so.³

On Certainty contains a series of more or less detailed remarks on scepticism. Some are directed specifically against the Dream Hypothesis, while others are more general. All contest any form of extreme scepticism. The primary target of *On Certainty* is Moore's famous argument against scepticism (and idealism). The latter just enumerates some of the many things he took himself to know. He claims that there are empirical truths which we know (i.e. can know) with certainty, such as "These are my two hands" or "The earth has existed for a great many years." He maintains that these truths provide proof of the existence of the external world, since the premises are known for certain and entail the conclusion.

Wittgenstein grants Moore his (psychological) certainty, but denies his knowledge about these truths. He rejects the idea that Moore has provided proof for the philosophical claim that there are objects that are physical and external to our minds. He does so because, for the sceptic, some sort of doubt still remains. Looking at my hands does not guarantee anything, as it is merely a move within our established "language games" (our conceptual scheme), while he does not challenge the move itself. What the sceptic challenges is the whole "language game" or conceptual scheme of the external world of physical-object discourse (see OC, 19, 23, 83, 617). In other words, in claiming to know he has two hands as instances of physical objects, Moore takes for granted the very conceptual scheme that is the target of sceptical attack.

3. This characterisation of scepticism captures an evidentialist assumption that underlies modern scepticism in the wake of Descartes and is also the main focus of Moore's anti-sceptical strategy as considered by Wittgenstein in *On Certainty*. However, it may also be pointed out that scepticism does not have to be methodological (i.e. Cartesian) in character, where this involves defining criteria for something to be known, only to then try to show that knowledge claims fail to meet those same criteria.

Wittgenstein tries to undermine both the Moorean and sceptical positions by impugning the sense of the proposition about the existence of the external physical world. According to him, it is not an empirical proposition: for the sceptic, it does not matter whether there are physical objects or not in respect of our experience, which could be as it is even if we find ourselves unable to specify what it would mean for there to be no physical objects.

According to Wittgenstein, both Moore and the proponents of scepticism ignore the fact that doubting and the allaying of doubt can only make sense within a certain “language game” (conceptual scheme), while the “language game” itself cannot be justified or doubted: it is neither reasonable nor unreasonable (OC, 559, 609–12). Doubt and justification make sense only relative to the rules that guide the use of the propositions (expressions) involved in some “language game.” They come to an end when we are confronted with doubts that are not themselves provided for by our rules—i.e. that do not count as legitimate moves or strategies in that particular “language game” (OC, 204). Moore’s truths mark points at which doubt loses its sense. However, they do this only because they are the background against which we distinguish between true and false, serving as “hinges” on which even our doubts turn (OC, 94, 341–3, 401–03, 514–15, 655): “Doubt grammatically loses its sense. This language game is like that” (OC, 56; see also 494, 498).

Sceptical doubts are invalid or incoherent because their sense implicitly presupposes the very conceptual scheme they explicitly attack. The sceptical hypothesis that nothing around us is real or exists is of that sort—like the thought that all our calculations could or might be wrong. But what a given proposition means is itself an empirical fact. In other words, some empirical facts must be beyond doubt (OC, 55, 514–19).

3. THE RELATIVIST READING OF WITTGENSTEIN’S POSITION

Following Coliva (2010; see also Baghramian and Coliva 2020, 110–14), we can reconstruct the following arguments in favour of a relativist reading of Wittgenstein’s view:

1. Language games “provide reasons for and against... propositions” that are subjects of our “assessing their truth,” while at the bottom of language games lie hinge “propositions which are neither true (grounded, rational) nor false (ungrounded, irrational)” (Coliva 2010, 1). Therefore they cannot be rational, and it would not be possible to have other alternative or merely different grounds, which “would be as legitimate as ours” (Coliva 2010, 2; see OC, 162, 233, 262).

If we apply this description of language games to the non-cognitivist characterisation of moral and religious language, we can say that we hold different, incommensurable, views with regard to the world of values, or religious beliefs that constitute different, incommensurable, conceptual schemes or frameworks of reference for the users of moral or religious language.

2. "At the bottom of our language games lies a way of *acting*, and that it is just a part of our lives to take certain propositions, theories and methods of justification, for granted, and thereby to act in accordance with them" (Coliva 2010, 2). Then, "it is a mere accident that we act in a certain way, and our lives are what they are. . . . There may be other ways of acting and living . . . which would ground other systems of justification" (Coliva 2010, 2; see OC, 92, 132, 264, 338, 609). In other words, within a non-cognitivist framework of moral and religious language there will be alternative conceptual schemes that navigate human lives.
3. If "we find someone who doesn't comply with our system of justification, we could only *persuade* or *convert* them to adopt ours, by appealing not to grounds or reasons—as there are none that could support one system over the other" (Coliva 2010, 2; see OC, 92, 262, 612). On a non-cognitivist account of morality and religion, there will be no common conceptual scheme that we share that could constitute a common ground for the different conceptual schemes expressed in different moral and religious attitudes.
4. As Coliva points out, the metaphor of a mythology is used by Wittgenstein to describe the status of our isolated attitudes or conceptual schemes when he writes that "the propositions describing our world-picture might be part of a kind of mythology" (OC, 95, 97; see Coliva 2010, 2). This resonates with the non-cognitivist account of moral and religious language. As is rightly noted by Coliva (2010, 3), it seems to suggest that there is no rational justification for our beliefs or, putting it more broadly, our conceptual schemes expressed in the form of different moral and religious attitudes and practices: they are like myths.

4. THE NON-RELATIVIST READING OF WITTGENSTEIN'S POSITION

4.1. *The Commensurability of World-Pictures*

In order to show that Wittgenstein is not an epistemic relativist, I will argue that his view implies the commensurability of world-pictures, and that this in turn rests on the implicit second-person perspective that makes human language a communal or public endeavour, not a private one. Both *On Certainty* and *Remarks on the Foundations of Mathematics* are interspersed with examples where Wittgenstein repeatedly imagines different communities, in which things we usually take for granted have ceased to be so. It seems at first sight that the conceivability of these communities would support a kind of Wittgensteinian relativism—i.e. the idea that it is conceivable that there could be people with altogether different conceptual schemes or world-pictures. However, in the case of the interesting example of wood-measurement analysed by Wittgenstein in the *Remarks on the Foundations of Mathematics*, what he has in mind is that if we fail to persuade the alien people operating with a different way of measuring wood that they are mistaken, we should revise our translation of their words. Wittgenstein thinks that we can imagine such a community which has a different way of measuring and paying for wood, and yet we will be able to deal with them and find a common factor when it comes to communication (cf. Coliva 2010, 13). It seems evident that he supposes that if we had to deal with them, we would try to convince them to measure wood by weight and pay for it accordingly. Furthermore, and more to the point, we would do this by using an entirely rational procedure: namely, that of showing them that the quantity of wood could remain the same even when its area and volume had changed. Were we to be successful in doing this, it would be an example not of an alternative epistemic method but, presumably, a case of people holding a false belief that had led them to employ an unreliable procedure to measure wood. Moreover, Wittgenstein argues that if rational argument fails then this probably means we have made a mistake when translating the words of those people into the meanings we ourselves attach to them (cf. Coliva 2010, 13–4, 16–17). The meaning of words seems to be a function of at least some central inferences we accept as common to both world-pictures (see Brandom 2008, 5–6).⁴ Hence, were there to be any problem

4. This interpretation of the common background to a world-picture can lend support to a reading of Wittgenstein of the sort proposed by Robert Brandom. The latter, in line with his inferentialist interpretation, claims that while we have various language games in Wittgenstein, there is one underlying structural feature in the form of the strategy of giving reasons: we accept one element of such a language game on the basis of another element of it, as

with understanding the meanings of words, such that this turned out, by our lights, to be something others did not accept, we should not conclude that these people were actually refusing to accept them, but rather that their “words have a different meaning than the one we . . . attributed to them” originally (Coliva 2010, 17).

Here Wittgenstein, like Quine (1960) and Davidson ([1974] 1984), insists that there are minimum requirements that a form of linguistic behaviour must meet in order to be intelligible to us (CV, 37; Rhees 1965, 25). Anticipating the current debate about radical translation, Wittgenstein assumes something quite close to the “principle of charity”: in order to interpret other people we should maximise agreement by attributing to them beliefs that from our own point of view are mostly true (Baghramian and Coliva 2020, 113–4). Wittgenstein writes that “if language is to be a means of communication there must be agreement not only in definitions but also . . . in judgments” (PI, 242). Furthermore, as we have seen, if seemingly radical differences were to emerge, we should revise our translations rather than attributing to them a large class of judgments that, from our point of view, are false. The crucial premise here is the claim that sharing a language is “not agreement in opinions but in form of life” (PI, 241; RFM, 353). Hence, understanding a different language presupposes convergence not only of beliefs but also of all the relevant patterns of behaviour—something which, in turn, would seem to presuppose common perceptual capacities, needs and emotions, realised in the form of an irreducibly basic second-person perspective. As Wittgenstein puts it: “The common behaviour of mankind is the system of reference by means of which we interpret an unknown language” (PI, 206; see RFM, 414–21; EPB, 149; see also Grayling 1988, 120).

On the basis of the above, we can argue that Wittgenstein was not a relativist either with respect to his position on scepticism or as regards his stance concerning non-cognitivism.⁵ To substantiate this argument, let us now turn to the key concepts elaborated by him in order to characterize

part and parcel of our rule-following, and this is the formal scheme common to all language games. It should be noted, however, that I am accepting in my interpretation only this minimal construal of the common elements of language games, and this does not necessarily mean I would endorse all of Brandom’s supplementary theses. While he was heavily influenced by Wittgenstein, their theories of meaning are somewhat different. For Wittgenstein, words get their meaning from their role in language games, embedded in forms of life. For Brandom, words get meaning merely from the inferences they are involved in. I am grateful to an anonymous reviewer for pointing out this difference.

5. Here I follow and extend the non-relativist readings of Wittgenstein on scepticism that are to be found in Coliva (2010) and Grayling (1988, 20–22; 2001), applying them to the present

his own project. In *On Certainty* he develops the notion of a world-picture, where this is integrated into his conception of language games and forms of life. The most significant aspect of language games is that “the term ‘language game’ is meant to bring into prominence the fact that the speaking of language is part of an activity, or of a form of life” (PI, 23; see PI, 19; Z, 173). In turn, forms of life consist of a plurality of language games, “a complicated network of similarities overlapping and crisscrossing: sometimes overall similarities, sometimes similarities in detail” (PI, 66). It seems that a “form of life” resembles a “medley-like” mixture of human practices somehow supporting or complementing each other (Kober 1996, 439). The term refers to a community sharing practices, customs, uses and institutions (PI, 199; RFM, 32, 43). Furthermore, it is not required that any one member of the community be competent in all language games performed by that community, and the second-person perspective makes us aware of this as a constant element of our orientation in the world. In other words, the notion of a “form of life” describes, or expresses, the setting in which language games are practised by such a community. It links the concept of a practice (a language game) with the concept of a community and the second-person perspective implicit therein.

The connection of the linguistic practice (i.e. rule-following language games) with the communal aspect of language acquisition and communication is possible only by virtue of the implicit second-person perspective. This perspective makes language a communal or public enterprise, not merely the private experience of the language user. It is significant that Wittgenstein does not construe language from the first-person perspective as an entirely private experience on the part of language users, or from the third-person perspective of strictly objective experience as a “view from nowhere,” but in terms of the second-person perspective, as a phenomenon natural to the human world (PI, 293; Johnson 2013, 77). What makes us familiar with some particular language game is our participation in a community of language users and our practising of its rules through having an implicit imprint of our human orientation in the world as an orientation within the human community (the human “form of life”). In other words, a part of what we are as human beings consists in the second-person perspective implicit in our lives as language users (see PI, 240–47).

The same conceptual mechanism is also encountered with the term “world-picture.” Wittgenstein’s idea is that such a picture contains the

concepts through which we conceive of the world. It is characterised by him as a kind of “myth” or “mythology,” in that it contains certain categories basic to our understanding of the world (see OC, 95, 195). A “myth” exhibits the views and convictions of a community or a form of life that we share with others (the implicit second-person perspective). It is a way of seeing the world (*Weltanschauung*). It might contain traditions, political views, moral values, or religious beliefs (PO, 125–29).

It seems that a world-picture is not necessarily a theory of the world, but is something that guides the behaviour of the community that holds it (PO, 125, 129, 137). In that sense, it could serve also as a “basis” (*Grundlage*) or “point of departure” for a community’s way of looking at the world (OC, 105, 167). This latter function is possible as it contains both certainties and knowledge claims that rest on these. Hence, Wittgenstein says that “above all it is the substratum of all my enquiring and asserting. The propositions describing it are not all equally subject to testing” (OC, 167; see OC, 234, 281–82, 327, 621).

The notion of a “world-picture” describes a familiar cultural (anthropological) phenomenon: the intuitive or practical, rather than discursive, sharing of views that correspond to what is disclosed in a given community’s customs (institutions) or ways of social behaviour and somehow overlap and supplement each other (OC, 102–3, 167, 275, 281, 298; PI, 129). The main point here is that a community’s language will embody their world-picture uniquely and absolutely. It does so because there are no real alternatives to it, although other ways of seeing the world are imaginable. Hence, different possible world-pictures are accessible to each other, since they seem to be merely a kind of extension of the particular community’s actual worldview. Such alternatives can only be imagined, as there are no such alternatives in reality. In other words, such differing yet imaginable world-pictures are commensurable with one another. We do not have here an instance of the first-person perspective (solipsism) such as would be implied by relativism, but rather a case of the second-person perspective, this being the overriding presupposition for any communal sharing of a form of life.⁶

Furthermore, Wittgenstein says that “I do not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false” (OC, 94). Apparently, according to him, we do not construct our world-picture: it is not a matter of reasoning,

6. This fits very much with current discussions of the second-person perspective. See, for instance, Darwall (2006, 2021), Pinsent (2012, 2013), and Eilan (2016).

or of engaging in some (spontaneous) conceptualizing activities—it is rather something we inherit or are given. Bernard Williams (1981, 156–57) points out, in similar fashion, that our language games and forms of life are “absolutely” acquired—that is, that they cannot be justified. Hence, the rules or grammar of our language games cannot be justified. They are not something that can be said to be “reasonable” or “unreasonable,” but rather “there like our life.” Presumably Wittgenstein has this in mind when he says:

Suppose we meet people who did not regard that [i.e. the claims of physics—P.S.] as a telling reason. Now, how do we imagine this? Instead of the physicist, they consult an oracle. (And for that we consider them primitive.) Is it wrong for them to consult an oracle and be guided by it? If we call this “wrong” aren’t we using our language games as a base from which to combat theirs? (OC, 609)

Following Williams, we can say that the various language-game communities exist in a merely empirical sense. From this, however, it does not follow that there are different world-pictures that are inaccessible to each other. Accessibility or commensurability are already in play, due to the fact that these are themselves constituted on the basis of a single human world-picture community that sets the limits to our access to language (and what we can conceive of). In other words, the limits of human language (consisting of different language games) determine the limits of our human world (its conceivability and commensurability). This is possible due to the second-person perspective implicitly assumed by Wittgenstein in his considerations regarding the commensurability of our world-pictures.⁷

4.2. Hinge Epistemology and the Common Human Form of Life

Emphasising the common elements of a human form of life, or even that there is a common form to human life, seems preferable in both hinge epistemology and Wittgensteinian philosophy because it grounds understanding, prevents thoroughgoing relativism, and illuminates those fundamentally shared behaviours that enable both language and knowledge (Conway 1989, 24). This foundational unity explains how humans can make sense of one another and forms the basis for shared (“hinge”) certainty, which is crucial for knowledge and justification. In contrast, an exclusive focus on the plurality of forms of life risks undermining the possibility of mutual intelligibility and shared knowledge, leading to an untenable radical relativism.

7. In the next three sections, I respond to comments made by an anonymous reviewer for this journal. I am very grateful to the latter for their insightful and helpful remarks.

On this reading of Wittgenstein, it becomes possible to explain the mutual understanding that occurs between people of different backgrounds: any human being can understand another, because they share a common form of life (OC, 358–59; PI, 23). This commonality includes a shared set of behaviours, ways of living, and patterns of language. While acknowledging the existence of multiple ways of living and language games, a single common human form of life acts as a unifying force. It explains why there is not a complete breakdown of understanding, as there are certain universally shared behaviours that constitute the bedrock of human interaction and language (the “hinge” elements). The human form of life includes fundamental aspects of human existence, such as speaking, thinking, and having beliefs about the world (Moyal-Sharrock 2015, 23–26). This common foundation creates the “universal grammar” of mankind, enabling understanding of and communication through even foreign languages.

An emphasis on the shared human form of life provides a more robust and less problematic account of knowledge and justification than a view that only stresses plurality. It explains the possibility of intersubjective agreement and shared knowledge. By contrast, if we were to fully embrace a radical emphasis on the plurality of forms of life, this would lead to a situation where understanding another person could become impossible, much like the hypothetical inability to understand a talking lion. Moreover, the idea of a common human form of life directly links epistemological questions about knowledge and justification to the fundamental nature of human beings as social, rule-following beings.

It is these accounts taking seriously the idea of a single common human form of life, as against those mainly emphasising the plurality of forms of life, that give the main motivation for a reading of Wittgenstein in terms of so-called hinge epistemology. The latter, inspired by Wittgenstein’s *On Certainty*, studies the basic certainties that form the bedrock of knowledge (Moyal-Sharrock and Pritchard 2024, 33–34). These “hinges” are not the kind of propositions for which we typically seek justification, but rather presuppositions of our understanding. A common human form of life provides the shared foundation that allows for these universal hinges such as the certainty of one’s own existence or the reliability of the senses. Therefore, the idea of a shared human form of life explains why we have a certain degree of agreement in respect of our fundamental beliefs. This shared background makes certain epistemological stances plausible and allows for justification to occur. Without this commonality, any attempt to justify a knowledge claim would be located in a vacuum. Hinge epistemology aims to address radical scepticism by providing a way to understand

why we can have basic certainties even if they cannot be strictly proven. The shared human form of life offers a crucial element of this understanding, as it reveals the common ground upon which we build our knowledge and certainty.

4.3. Non-Cognitivism and the Common Human Form of Life

Nevertheless, the position described above might be charged with imposing a false universalism, and so it is necessary to add further nuance to our argument here, especially in the context of the alleged non-cognitivism of Wittgenstein. This “universal” approach attempts to impose a uniform “form of life,” where Wittgenstein’s own philosophy acknowledges a rich diversity of unique and context-dependent practices making universal claims about morality or religion impossible within that framework. In the non-cognitivist reading of Wittgenstein on morality and religion, moral and religious language do not describe facts or objective truths, but rather function differently, expressing attitudes, emotions, or a way of life. The meaning of moral and religious expressions is found in their use within a specific form of life. This implies that these meanings are not universally applicable across different forms of life. Therefore, on this reading, to affirm a single “human form of life” and then use that to argue for universally shared moral and religious attitudes runs counter to Wittgenstein’s ideas.

Translating the idea of a single human form of life into a non-pluralist interpretation of Wittgenstein’s non-cognitivism about morality and religion will, in that case, prove problematic, because his concept of “forms of life” will be taken to refer to shared cultural, social and linguistic practices that are diverse, not singular (Weiberg 2025, 3–6). A non-cognitivist stance denies that moral and religious statements express objective truths, while Wittgenstein’s ideas suggest that the meanings within these practices are relative to their specific forms of life. Therefore, trying to universalise a single “form of life” to impose uniform moral or religious attitudes goes against the Wittgensteinian view that meaning is generated through diverse and specific linguistic and cultural contexts.

This line of criticism emphasises that Wittgenstein’s concept of “forms of life” serves to highlight how language, culture and social practices intertwine to create meaning and understanding within specific communities. On such an interpretation, the concept of “forms of life” is inherently plural, referring as it does to a multitude of diverse ways of living and communicating. There is no singular, universal “human form of life” from which all moral and religious practices might be said to originate.

4.4. Hinge Epistemology, Non-Cognitivism, and the Second-Person Perspective

The above criticism demonstrates the need for an explanation that appeals to the second-person perspective, as the latter can help disarm the aforementioned charge and bridge the gap between hinge epistemology and non-cognitivism. The second-person perspective is relevant to hinge epistemology just by virtue of the fact that it provides a unique way to ground knowledge through intersubjective interactions, offering a foundation beyond mere first- or third-person observation, where this is critical for understanding the shared, and often unquestioned, framework (of “hinges”) that enables communication and knowledge acquisition. For any non-cognitivist account of ethics and religious belief, the second-person perspective is relevant in that it stresses the idea that moral and religious commitments are not merely abstract beliefs, but relational stances involving mutual acknowledgement and a shared “you” or “otherness,” where this in turn lends support to the sort of views that hold such beliefs to be more about values and commitments than just stating factual propositions.

Indeed, the second-person perspective is arguably fundamental to developing a coherent understanding of intersubjective interaction more generally—wherever people make sense of each other’s behaviour and adjust their own actions accordingly. This forms the basis for developing the trust and shared understanding necessary for any epistemic practice. Hinge epistemology, drawing on Wittgenstein, focuses on the unquestioned beliefs or practices (the “hinges”) that provide the framework for all other knowledge (Pritchard 2025, 48–49), and the second-person perspective is relevant here because it shows how these hinges are established and maintained through direct interaction with and mutual acknowledgement of others. The second-person perspective can also help explain how differences in opinions and biases function within testimonial exchanges (Boncompagni 2024, 290–94). When a prejudice acts as a hinge, preventing proper evaluation of testimony, engaging with the second-person perspective can reveal the underlying prejudice by reintroducing a normative level where rational consideration of others is possible.

That last element brings us to the relation between the second-person perspective and non-cognitivist accounts. The second-person perspective stresses that ethical and moral understanding is not just a matter of recognising facts, but of entering into a reciprocal relationship with or attitude towards others (see OC, 204). A commitment to acknowledging another person’s standpoint is essential for a recognition-based relationship

(Cockburn 1990, 6–10). Non-cognitivists argue that moral or religious judgments express emotions or commitments, rather than describing facts. The second-person perspective aligns with this by emphasising how our moral and religious stances are deeply tied to our interactions and our acknowledging of another’s “you-ness” or “otherness.” The concept of religious belief is sometimes compared to that of hinge commitment because both involve deep, foundational stances rather than simple empirical assertions. The second-person perspective suggests that religious belief is also a form of relational commitment, involving a reciprocal orientation toward a divine “you,” which aligns with non-cognitive views that focus on commitment and value rather than just propositional content.

As rightly noted by David Cockburn:

Wittgenstein was writing against the background of a tradition in which it was customary to mark off my thought about other people from my thought about, say, stones by saying that I believe that the former, but not the latter, “have minds.” His introduction of the term “attitude” here represents a revolt against this way of speaking which has a number of dimensions. Part of what he wishes to highlight with this term is the fact that we feel about and act towards other human beings in ways that are utterly different from those in which we feel about and act towards, for example, stones. I have a certain “practical orientation” towards another human being with whom I am confronted. (1990, 6)

Cockburn emphasises an element of the thought of Wittgenstein that is crucial for our considerations relating to the second-person perspective. For Wittgenstein, the centre of the picture of the human way of living (or “practical orientation,” as Cockburn calls it) was to have a certain “attitude towards” others, rather than to have a certain “belief about” them (PI, 178). It is a matter of our attitudes, as non-cognitivism emphasises, that we feel towards others as beings towards which certain ways of acting make sense or seem appropriate. In the course of presenting of his own philosophical argument for recognising other people as persons, Cockburn himself stresses that, for Wittgenstein, “the attitude is what is fundamental in our thought about each other” (Cockburn 1990, 9).

5. CONCLUSIONS

To conclude, we can therefore legitimately claim that Wittgenstein’s views on scepticism and, by analogy, non-cognitivism, do not imply epistemic relativism. The form of relativism to which Wittgenstein is committed

might be simply anthropocentrism (Grayling 1988, 20).⁸ He accepts the different cultures and “forms of life” of human beings, but also talks about “the common behaviour of mankind”—and in fact about a common human world-picture. Cultural pluralism, as we might call this form of anthropocentrism, itself makes sense only if we assume that there is mutual accessibility or commensurability between cultures at the cognitive or epistemic level. The different “forms of life” share an experiential and conceptual basis that permits mutual accessibility between them in terms of the second-person perspective. That is precisely the respect in which those “forms of life” are not epistemically relative at all.⁹ The ability to detect that something is a “form of life,” and that it differs from our own, requires these means just for us to identify its presence and be in a position to say what distinguishes it from ours.

I. WITTGENSTEIN’S WORKS

- CV *Culture and Value*. Edited by G.H. von Wright in collaboration with Heikki Nyman. Translated by Peter Winch. German-English parallel text. Oxford: Blackwell, 1980.
- LC *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. Edited by Cyril Barrett. Oxford: Blackwell, 1966.
- EPB *Eine Philosophische Betrachtung*. In *Schriften*, vol. 5, edited by Rush Rhees, 117–237. Frankfurt: Suhrkamp, 1970.
- RFM *Remarks on the Foundations of Mathematics*. Edited by G.H. von Wright, Rush Rhees, and G.E.M. Anscombe. Translated by G.E.M. Anscombe. Oxford: Blackwell, 1978.
- PI *Philosophical Investigations*. Edited by G.E.M. Anscombe and Rush Rhees. Translated by G.E.M. Anscombe. German-English parallel text. 2nd edition. Oxford: Blackwell, 1958.
- Z *Zettel*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by G.E.M. Anscombe. German-English parallel text. Oxford: Blackwell, 1967.
- OC *On Certainty*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by Denis Paul and G.E.M. Anscombe. German-English parallel text. Oxford: Blackwell, 1969.
- PO *Philosophical Occasions: 1912–1951*. Edited by James C. Klagge and Alfred Nordmann. German-English parallel text where appropriate. Indianapolis: Hackett Publishing, 1993. (Cited by original pagination only).

8. Anthropocentrism might be understood here as similar to what John McDowell’s interpretation of Wittgenstein calls the “communitarian” view, though the second-person perspective heavily emphasised in the present paper is missing from the latter. In McDowell’s view, Wittgenstein’s philosophy moves beyond the idea that a linguistic community’s agreement in judgment determines meaning or correctness. Instead, McDowell emphasises an individual’s initiation into the space of reasons through upbringing and the acquisition of a “second nature” that shapes their conceptual capacities (see McDowell 1984, 325–63; see also Wright 1980, *passim*, and McGinn 2021, 145–60).

9. Where Wittgenstein’s approach is concerned, there seems to be a kind of transcendentalist impulse (pursuing the conditions under which our knowledge is possible) that criss-crosses with a naturalistic one (invoking human behaviour as a common factor in respect of our cognitive activity, where this furnishes some sort of basis for a “principle of charity”).

II. SECONDARY LITERATURE

- Baghramian, Maria, and Annalisa Coliva. 2020. *Relativism*. London: Routledge.
- Barrett, Cyril. 1991. *Wittgenstein on Ethics and Religious Belief*. Oxford: Blackwell.
- Blackburn, Simon. 2004. "Relativism and the Abolition of the Others." *International Journal of Philosophical Studies* 12 (3): 245–58. <https://doi.org/10.1080/0967255042000243939>.
- . 2007. *Truth: A Guide for the Perplexed*. London: Penguin.
- Boghossian, Paul. 2006. *Fear of Knowledge: Against Relativism and Constructivism*. Oxford: Oxford University Press.
- Boncompagni, Anna. 2024. "Prejudice in Testimonial Justification: A Hinge Account." *Episteme* 21 (1): 286–303. <https://doi.org/10.1017/epi.2021.40>.
- Brandom, Robert B. 2008. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.
- Cockburn, David. 1990. *Other Human Beings*. London: Macmillan.
- Coliva, Annalisa. 2003. *Moore e Wittgenstein: Scetticismo, certezza e senso comune*. Padova: Il Poligrafo.
- . 2010. "Was Wittgenstein an Epistemic Relativist?" *Philosophical Investigations* 33 (1): 1–23. <https://doi.org/10.1111/j.1467-9205.2009.01394.x>.
- Conway, Gertrude D. 1989. *Wittgenstein on Foundations*. Atlantic Highlands, NJ: Humanities Press.
- Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
- . 2021. "Recognition, Second-Personal Authority, and Nonideal Theory." *European Journal of Philosophy* 29 (3): 562–74. <https://doi.org/10.1111/ejop.12674>.
- Davidson, Donald. (1974) 1984. "On the Very Idea of a Conceptual Scheme." In *Inquiries into Truth and Interpretation*, 183–98. Oxford: Clarendon Press.
- Eilan, Naomi, ed. 2016. *The Second Person: Philosophical and Psychological Perspectives*. London: Routledge.
- Glock, Hans-Johann. 1996. *A Wittgenstein Dictionary*. Oxford: Blackwell.
- Grayling, A.C. 1988. *Wittgenstein*. Oxford: Oxford University Press.
- . 2001. "Wittgenstein on Scepticism and Certainty." In *Wittgenstein: A Critical Reader*, edited by Hans-Johann Glock, 305–21. Oxford: Blackwell.
- Haller, Rudolf. 1995. "Was Wittgenstein a Relativist?" In *Wittgenstein: Mind and Language*, edited by Rosaria Egidi, 223–31. Dordrecht: Kluwer Academic.
- Hintikka, Merrill B., and Jaakko Hintikka. 1986. *Investigating Wittgenstein*. Oxford: Blackwell.
- Johnson, Joshua. 2013. "The Private Language Argument and a Second-Person Approach to Mindreading." *European Journal for Philosophy of Religion* 5 (4): 75–86. <https://doi.org/10.24204/ejpr.v5i4.206>.
- Kirk, Robert. 1999. *Relativism and Reality: A Contemporary Introduction*. London: Routledge.
- Kober, Michael. 1996. "Certainties of a World-Picture: The Epistemological Investigations of *On Certainty*." In *The Cambridge Companion to Wittgenstein*, edited by Hans Sluga and David G. Stern, 411–41. Cambridge: Cambridge University Press.
- Kusch, Martin. 2011. "Disagreement and Picture in Wittgenstein's *Lectures on Religious Belief*." In *Image and Imaging in Philosophy, Science and the Arts*, vol. 1, edited by Richard Heinrich, Elisabeth Nemeth, Wolfram Pichler, and David Wagner, 35–57. Frankfurt am Main: Ontos Verlag.
- McDowell, John. 1984. "Wittgenstein on Following a Rule." *Synthese* 58 (3): 325–63. <https://doi.org/10.1007/BF00485246>.

- McGinn, Marie. 2021. "Recognizing the Ground That Lies before Us as Ground': McDowell on How to Read the *Philosophical Investigations*." In *Wittgenstein, Scepticism and Naturalism: Essays on the Later Philosophy*, 145–60. London: Anthem Press.
- Moyal-Sharrock, Danièle, and Duncan Pritchard. 2024. *Wittgenstein on Knowledge and Certainty*. Cambridge: Cambridge University Press.
- Moyal-Sharrock, Danièle. 2015. "Wittgenstein on Forms of Life, Patterns of Life and Ways of Living." *Nordic Wittgenstein Review* 4 (1): 21–42. <https://doi.org/10.15845/nwr.v4i0.3362>.
- O'Grady, Paul. 2002. *Relativism*. Chesham: Acumen Publishing.
- . 2004. "Wittgenstein and Relativism." *International Journal of Philosophical Studies* 12 (3): 315–37. <https://doi.org/10.1080/0967255042000243975>.
- Pinsent, Andrew. 2012. *The Second-Person Perspective in Aquinas's Ethics: Virtues and Gifts*. New York: Routledge.
- . 2013. "The Non-Aristotelian Virtue of Truth from the Second-Person Perspective." *European Journal for Philosophy of Religion* 5 (4): 87–104. <https://doi.org/10.24204/ejpr.v5i4.207>.
- Pritchard, Duncan. 2025. "Axiological Hinge Commitments." *Synthese* 205 (2): 1–22. <https://doi.org/10.1007/s11229-024-04898-0>.
- Putnam, Hilary. 1992. *Renewing Philosophy*. Cambridge, MA: Harvard University Press.
- Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Rhees, Rush. 1965. "Some Developments in Wittgenstein's View of Ethics." *Philosophical Review* 74 (1): 17–26. <https://doi.org/10.2307/2183528>.
- Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.
- Schönbaumsfeld, Genia. 2023. *Wittgenstein on Religious Belief*. Cambridge: Cambridge University Press.
- Weiberg, Anja. 2025. "Religious Belief in the Later Wittgenstein—A 'Form of Life', a 'Hinge', a 'Weltanschauung', Something Else or None of These?" *Religions* 16 (8): 1046. <https://doi.org/10.3390/rel16081046>.
- Williams, Bernard. 1981. "Wittgenstein and Idealism." In *Moral Luck: Philosophical Papers 1973–1980*, 144–63. Cambridge: Cambridge University Press.
- Williams, Michael. 2007. "Why (Wittgensteinian) Contextualism Is Not Relativism." *Episteme: A Journal of Social Epistemology* 4 (1): 93–114. <https://doi.org/10.3366/epi.2007.4.1.93>.
- Wright, Crispin. 1980. *Wittgenstein on the Foundations of Mathematics*. Cambridge, MA: Harvard University Press.

“Fake News” in Reformulated Messages

Towards Expanding the Toolset for Identifying Misinformation

Mitchell T. Welle, Marcin Koszowy

ABSTRACT In an age where information spreads faster than ever, the subtle manipulation of truth through rephrasing plays a pivotal role in amplifying misinformation. Starting from the observation that the spread of “fake news” may be significantly reinforced through reformulating a message for the sake of its *misrepresentation*, we seek to address the problem of the spread of “fake news” from the perspective of the rephrasing of news for purposes of misinformation. Given such a potentially dangerous role for *misuses of rephrasing*, the following research question arises: what is the relation between “fake news” and reformulated messages? This question will be addressed by analysing to what extent (i) definitions of “fake news” in the computer-science and philosophy-related literature, and (ii) recent linguistic studies of rephrase (as it is sometimes known), are helpful in identifying the main features of “fake news” as these relate to the latter. In this regard, we propose a research programme for addressing rephrase as a linguistic phenomenon—one that will serve as a tool for the study of communication in respect of “fake news.”

KEYWORDS fabrication; fake news; misrepresentation; mimicking; reformulating messages; rephrasing

Acknowledgements

We would like to acknowledge the fact that the work reported in this paper has been supported by the Polish National Science Centre under grant 2020/39/I/HS1/02861.

We extend our sincere thanks to the two anonymous reviewers for their constructive feedback and thoughtful engagement with our manuscript.

✉ Mitchell T. Welle, Laboratory of the New Ethos, Warsaw University of Technology, Poland

✉ Mitchell.welle@pw.edu.pl ☎ 0000-0003-2649-8831

✉ Marcin Koszowy, Laboratory of the New Ethos, Warsaw University of Technology, Poland

✉ marcin.koszowy@pw.edu.pl ☎ 0000-0001-5553-7428

Jill looked at the King: his mouth was open and his face was full of horror.
And then she understood the devilish cunning of the enemies' plan. By mixing
a little truth with it they had made their lie far stronger.

C.S. Lewis, *The Last Battle*

1. INTRODUCTION

In 2024, the President of the European Commission Ursula von der Leyen, at the World Economic Forum in Davos, Switzerland, stated: “For the global business community, the top concern for the next two years is not conflict or climate. It is disinformation and misinformation, followed closely by polarisation within our societies” (Von der Leyen 2024). So, in an age where information spreads faster than ever, the subtle manipulation of truth through *rephrasing* (or what linguists sometimes refer to as “rephrase”) plays a pivotal role in amplifying misinformation. This may give us pause to reflect and ask: “What is ‘rephrasing’?” The working definition of the latter runs as follows: “To say that a speaker rephrases is therefore to say that he or she means a contribution to be understood as connected to another contribution in a specific way” (Younis et al. 2023). This definition is built on contemporary research into rephrase (e.g. Visser et al. 2018; Konat et al. 2016; Koszowy et al. 2022). At the same time, the current state of research emphasizes the difficulties associated with the boundaries of the phenomenon: if rephrase is reformulation, where does it end? After all, two expressions of the same thought that differ completely in wording are difficult to classify as together constituting an instance of rephrase. Therefore, as has been stated, for example, by Younis et al (2023), an instance of rephrasing must be an operation on certain linguistic material that modifies it to a certain extent but not completely. However, this raises the following question: to what extent does this modification allow the output of this operation to be considered an instance of rephrase? In Koszowy et al. (2022), an attempt has been made to use semantic similarity as a measure, but counterexamples have been found showing that semantic convergence does not necessarily have to be related to reformulation. Therefore, in this article, we take as our starting point the definition of rephrase proposed by Younis et al. (2023). (The idea of rephrase will itself be elaborated in more detail in Section 4.)

Turning now to a timely example that uses a rephrased “fake news” headline to exemplify what we are talking about, we may consider the original news story “Netanyahu acknowledges Israel losing online “propaganda war,” should be doing more” from the *Times of Israel* (Freiberg 2025). The popular and (in)famous website *InfoWars* ran the rephrased headline “Netanyahu declares war on free speech as Israel’s propaganda

efforts falter” (2025). The news story on *InfoWars*, which itself is a story originally from the website “lifesitenews.com,” cites the *Times of Israel* as one of the many sources of the article. The *Times of Israel* article was the main article that lifesitenews.com drew its article from. The headline from the *Times of Israel* is relatively neutral. On the other hand, the *InfoWars* article includes phrases such as “falter” in place of “losing,” and employs the phrase “declares war” to intensify the propositional content in its rephrased title. In short, both titles are propositionally equivalent, but the rephrased title from *InfoWars* is linguistically intensified.

In this paper, we discuss the results of contemporary studies of rephrasing in argumentation and dialogue, treating these as a possible toolset for critically dealing with “fake news” and misinformation. Despite empirical evidence of the impact of repetitions on the spread of “fake news” (e.g. Hassan and Barber 2021; Diaz-Garcia et al. 2025), the intersection between communicative phenomena related to repeating or reformulating messages and the notion of “fake news” still remains underexplored. Here, we analyse the representative definitions of “fake news” in computer science and philosophy to reveal the key tendencies pertaining to the perception of “fake news” in these two research fields. Our analysis shows that the tendencies of reformulating or mimicking are embedded in those definitions. Given that the notions of reformulating and/or mimicking messages are implicated in techniques of rephrasing (in that rephrasing is a subcategory of mimicking), we hypothesize that the overlap between these two areas may provide a fruitful path for future research on the linguistic cues for identifying attempts at spreading “fake news” via rephrasing and reformulating techniques. Thus, we claim that recent models of rephrase in argumentation (Konat et al. 2016; Visser et al. 2018; Koszowy et al. 2022; Younis et al. 2023) can furnish a possible repertoire of tools for systematically studying and critically assessing rephrased “fake news.” This then leads on to an exploration of the key role of rephrasing “fake news” in the process of re-framing pieces of information for the sake of misinforming recipients.

Our initial observation is that the dissemination of “fake news” has become an integral challenge within the broader context of digital communication and media. This issue is particularly pertinent as such news leverages the pervasiveness of social platforms to manipulate public perception. One of the core tactics in “fake news” construction is the strategic use of rephrasing, where minor adjustments in language alter the truth value of statements while preserving their perceived factuality. This rephrasing technique subtly shifts public interpretation, often without causing overt contradictions, making it difficult to discern misinformation. These

observations prompt a deeper inquiry into the linguistic mechanisms that make such news persuasive.

Despite a growing corpus of research on “fake news” detection (e.g., Singhal et al. 2019; Tschitschek et al. 2017; Wang et al. 2018), there remains a critical gap in understanding how the rephrasing of statements contributes to the persuasive force of “fake news.” While previous studies (e.g., Hassan and Barber 2021) have concentrated on the broader identification of false information through algorithmic methods and empirical investigation focused on the effect of rephrasing in various domains (e.g., Koszowy et al. 2022; Younis et al. 2023), little attention has been paid to the cognitive and linguistic impacts of rephrasing in enhancing the believability of “fake news.” Specifically, the locutionary and illocutionary aspects of rephrasing as a rhetorical device in “fake news” production remain, to the best of our knowledge, underexplored. This lacuna highlights the need for an investigation that will bridge the gap between practical “fake news” detection on the one hand, and deeper philosophical considerations on the other.

Given this paucity of systemic research into reformulation strategies aimed at spreading “fake news,” our study seeks to address the following research question: what role does rephrasing play in the current structure of “fake news” in two representative research fields—namely, computer science and philosophy? Specifically, our research aims to explore how rephrasing techniques are *embedded* within representative definitions of both disciplines to distort or enhance the believability of “fake news.” By examining computational approaches to detecting rephrased misinformation and the philosophical implications of language manipulation, the study aims to offer an understanding of how rephrasing contributes to the structural integrity and rhetorical effectiveness of such news.

Our preliminary insights suggest that rephrasing plays distinct roles in the fields of computer science and philosophy, reflecting the differing focuses of these domains. In computer science, rephrasing is studied with a practical emphasis, as it typically has *real-world implications* for the detection and mitigation of “fake news” through algorithms and machine-learning techniques. These approaches aim to systematically identify “fake news,” thereby providing concrete tools for combating misinformation. On the other hand, philosophy engages with rephrasing in a *more speculative and theoretical* manner, focusing on the ethical and epistemological implications of how language can be manipulated to mislead or persuade. While both fields recognize the significance of rephrasing in the structure of “fake news,” computer science seeks actionable solutions, whereas philosophy probes the deeper implications of how and why rephrasing shapes our

understanding of epistemological and ethical concerns. More broadly, our goal here is rather to bring together two seemingly distant areas of research: one focused on rephrasing in pragmatics and argumentation theory, and the other dealing with the detection of “fake news” in the context of its increasing proliferation. The former is more philosophical and linguistic, while the latter, due to the computer systems and techniques used to detect such news, is definitely oriented towards computer science. Nevertheless, research on “fake news” detection procedures also has a clear philosophical component, because effective identification depends on an adequate definition of what “fake news” is (i.e. a fine-grained conceptual framework).¹

The goal of our research is to develop an understanding of how rephrasing contributes to the structure of “fake news.” To achieve this, we will study the definitions and conceptualizations of “fake news” in philosophy and computer science, specifically focusing on how rephrasing is treated within these frameworks. By analysing how computer science defines and detects rephrasing in such news through computational tools, alongside philosophical perspectives that explore the ethical and epistemological dimensions of rephrased misinformation, this research aims to bridge the gap by exploring the overlap between rephrasing and “fake news.”

For this purpose, we will be looking at notions of “fake news” in two research areas that significantly differ from each other, and so are complementary when it comes to obtaining an adequate conceptualization of such news from the point of view of the critical assessment of its role in communication, drawing out the components of rephrasing embedded in both sets of definitions. On the one hand, philosophy (especially epistemology and the philosophy of language, argumentation and communication) furnishes a theoretical framework for capturing the main features of “fake news.” On the other, computer science elaborates tools aimed at operationalizing the detection of (potential) “fake news” in its sphere of communication.

In this way, rephrasing becomes a form of deception that goes beyond the creation of falsehoods—it is about shaping these falsehoods to align with the expectations and cognitive biases of the audience. By crafting information that appears real, those behind “fake news” are able to amplify

1. To elaborate further, this is because these conceptualizations of “fake news” include, among other things, both an epistemological component (in that they address issues of truth and cognitive deception through “fake news”) and an ethical one (since they refer to the moral dimension of “fake news” at the level of the intentions behind its use and its effects). Hence, exploring the overlap between the study of rephrasing and research on “fake news” has directed us to philosophy and computer science as the representative disciplines typically and most visibly exhibiting this overlap.

the impact of their fabricated content, making it more persuasive and harder to discredit. This intentional rephrasing, coupled with the strategic presentation of information, plays a critical role in the effectiveness of “fake news,” ensuring that the falsehoods it contains are not only believable but also likely to be disseminated widely before they are debunked.

The overall idea behind the present paper is to *initiate moves* in the direction of laying a foundation for developing a comprehensive toolset for combating “fake news” from the perspective of computer science. The latter, via social media, offers a powerful and systematic method of identifying and neutralizing such news; however, some “fake news” articles manage to escape detection. We may deploy an analogy here: it is akin to a fishing net. Many of the obvious “fake news” articles are caught, but the smaller fish (i.e. the subtle “fake news” stories) manage to escape detection. One class of “fake news” articles that escape detection are rephrased news articles. The philosophical discussions surrounding “fake news” offer a more nuanced discussion of the nature of the latter, and we might find the discussion here fruitful for developing a more adequate definition of “fake news” for computer scientists. To give an example of how philosophy could potentially help computer science, we observe that many papers in the latter field define “fake news” as at the very least having the property of being intentionally false (see: Singhal et al. 2019; Desamsetti et al. 2023; Zhou and Zafarani 2018). However, as we can see from the *InfoWars* story referred to at the start, we can find ourselves dealing with an article that is not intentionally false, but where there is still an intuition that something of that type should somehow count as “fake news.” So why is there an intuition to the effect that the *InfoWars* example should count as an instance of such news?² A more nuanced account of this, taking into account aspects of rephrasing, may yet succeed in capturing that example within the net of “fake news” stories.

The overall structure of the paper, then, will be as follows: first, in Section 2, we explore key texts discussing “fake news” detection in computer science in relation to rephrasing; second, in Section 3, we examine the literature in philosophy with respect to its discussion of “fake news” as

2. It is worth mentioning that it is not the case *a priori* that *InfoWars* is “fake news.” We might treat it as a kind of institutional fact, to use Searle’s term. It just happens to be the case that as of this moment (2025), *InfoWars* is commonly taken to be the paradigmatic example of this (see Dentith 2016). Does this always have to be the case? No. It certainly is the case that in some possible world Alex Jones, the head of *InfoWars*, speaks as though he were trained to deliver the news on NPR and adheres to the news standards of the Associated Press (AP) before publishing news articles. Additionally, is everything that *InfoWars* publishes or reports on false? No. But should we be sceptical about stories published on *InfoWars*? Yes.

it relates to rephrasing; third, in Section 4, by inspecting the similarities and differences in these fields we seek to bridge the gap between them, surveying the overlap between rephrase and “fake news” in order to propose specific areas of philosophical and linguistic research into rephrasing; lastly, in Section 5, we conclude by envisaging possible future work. These steps will, we hope, enable us to offer a new perspective on rephrasing and “fake news” studies, legitimizing the category of rephrasing as one of the key concepts needed for capturing the linguistic manifestations of “fake news.” In this way, our research programme aims to capture both theoretical dimensions of rephrasing as they relate to this phenomenon.

2 DEFINITIONS OF “FAKE NEWS” IN COMPUTER SCIENCE

In exploring the role that rephrasing plays in definitions of “fake news,” we have limited our investigation to the fields of philosophy and computer science. This decision is driven by their complementary strengths in addressing the nuanced ways that rephrasing can be used to manipulate information. Philosophy provides a critical theoretical framework for understanding how language, including rephrasing, can be used to obscure truth or distort reality. It allows us to examine the ethical implications of “fake news” as a tool for deception, exploring the philosophical foundations of truth and the societal impacts of manipulated information.

In seeking to gather together the necessary computer-science definitions related to “fake news,” our approach has been designed to ensure relevance during the selection process. First, we identified a series of relevant, and commonly cited and recent, peer-reviewed articles falling within the field of computer science that explicitly addressed “fake news” or focused on methodologies for “fake news” detection. These articles served as the primary sources for extracting the definitions of “fake news.” We screened peer-reviewed work between 2015–2025 in venues indexed by ACM Digital Library, IEEE Xplore and major journals/conferences; for philosophy, we screened journal articles and chapters in epistemology, as well as the philosophy of language, argumentation and communication. Second, from each article, we isolated and recorded the specific definitions of “fake news” utilized by the authors, ensuring that these were contextually aligned with the overall research objectives of the study. Third, we compiled these definitions into a table, providing a comparative framework for analysis. Lastly, we examined the definitions for recurring features or thematic patterns, which were then highlighted and categorized within the table to illustrate commonalities and distinctions across the selected literature. We drew from these features or thematic patterns elements aimed at identifying linguistic

manifestations of certain instances of rephrasing that may have the effect of “fake news.” This approach allowed for a comparative understanding of how such news is conceptualized within the domain of computer science.

Table 1: Computer science definitions of “fake news”: major tendencies

#	Author(s)	Falsehoods	Fabrication	Intent And Purpose	Ways of mis-leading people	Verification and evidence	Harm and other normative factors	Undefined
1.	Dong et al. (2023)							X
2.	Qian et al. (2018)							X
3.	Ruchansky et al. (2017)	X	X					
4.	Tschiatschek et al. (2017)	X			X			
5.	Alam et al. (2022)	X		X	X		X	
6.	Kou et al. (2022)	X			X			
7.	Nabov et al. (2021)				X		X	
8.	Zhou et al. (2020)	X		X				
9.	Ajao et al. (2019)	X		X	X	X		
10.	Khattar er al. (2019)	X	X			X		
11.	Singhal et al. (2019)	X		X		X		
12.	Wang et al. (2018)	X	X	X		X		
13.	Conroy et al. (2015)	X		X				
14.	Farajtabar et al. (2017)	X		X			X	
15.	Berrondo-Otermin & Sarasa-Cabezuelo (2023)	X		X	X	X	X	
16.	Wu & Rao (2020)	X			X			
17.	Desamsetti et al. (2023)	X		X			X	
18.	Fifita et al. (2023)	X		X				
19.	Shu et al. (2017)	X		X				
20.	Zhou & Zafarani (2018)	X		X				
21.	Zhang & Ghorbani (2020)	X		X	X		X	
22.	Jain & Kasbe (2018)	X			X			
23.	Perez-Rosas et al. (2017)	X	X	X		X		
24.	Oshikawa et al. (2018)	X			X			
25.	Parikh & Atrey (2018)	X		X				
26.	Sharma & Singh (2024)	X		X	X		X	

Table 1 presents a detailed overview of the major tendencies found in definitions of “fake news” within the computer-science literature. To compile this table, we analysed 26 representative articles, each of which directly addresses either “fake news” itself or its detection. The definitions are categorized into six key themes: Falsehood, Fabrication, Intent and Purpose, Ways of Misleading People, Verification and Evidence, and Undefined. These themes were inductively drawn from the definitions of “fake news,” prior to our identifying linguistic manifestations of certain instances of rephrasing that may have the effect of “fake news.” Falsehood and Fabrication emerged as fundamental components, frequently mentioned across the literature, underscoring the fact that “fake news” is often built upon deliberately false or fabricated information. The category of Intent and Purpose highlights the intentional deception that characterizes “fake news,” distinguishing it from accidental misinformation. The instances of Ways of Misleading People identified by us, including disinformation and satire, emphasize the diverse strategies employed to deceive audiences. Additionally, the role of different media platforms, such as social media, in the spread of “fake news,” is a prominent theme, as is the challenge of verifying the authenticity of information presented as news. This analysis reveals not only the consistency in how “fake news” is defined within the field of computer science, but also the emphasis on intentionality and the mechanisms of dissemination.

The structure of our discussion of these categories will be as follows: we will first describe the data obtained by giving the relevant statistical information, and then proceed to outline the elements of rephrasing identified in the features of various definitions of “fake news.” The goal is to see how rephrasing manifests itself within these features.

The first category is Falsehood. Falsehoods are a fundamental aspect of the definitions of “fake news” within the computer-science literature. This category, which includes terms such as “false,” “untrue,” and “falsified,” was identified in 23 out of 26 articles analysed, representing 88.4% of the total. This near-universal acknowledgement—as might well have been expected, given that the term suggests some form of falsity in one way or another—underscores the central role of falsehood in the conceptualization of “fake news.”

When we turn to the act of rephrasing, applied to false information, this serves as a powerful tool in the construction of “fake news.” Building on previous research into rephrasing (Konat et al. 2016; Visser et al. 2018; Koszowy et al. 2022; Younis et al. 2023), we may say that rephrasing in the context of “fake news” is not merely a simple alteration of words but

a deliberate effort to replicate the tone, style and structure of legitimate news sources (Dentith 2016). By doing so, creators of “fake news” aim to enhance the believability of their content, making it more difficult for the audience to distinguish between authentic and false information. The subtle manipulation of language allows the fabricated content to blend seamlessly with genuine news, exploiting the trust that readers place in familiar journalistic formats. This mimicry extends beyond mere language to include the visual and contextual elements of news, such as headlines, imagery, and even the use of credible-sounding sources, all designed to reinforce the illusion of authenticity.

The category of Fabrication stands out as notable within the computer-science literature. It included the term “fabrication” or “fabricate,” and was identified in 5 of the 26 articles (representing 19.2% of the total). The concept of fabrication is interesting because it shows an overlap between falsehood and the intentional aspect of “fake news,” distinguishing it from other forms of misinformation, such as rumours or inadvertent errors, as well as stories that are simply false (e.g. Ruchansky et al. 2017; Khattar et al. 2019; Farajtabar et al. 2017; Wang et al. 2018; Perez-Rosas et al. 2017). The act of fabricating involves creating false information intentionally, often with the aim of deceiving or misleading the audience. This deliberate intent is crucial for differentiating “fake news” from mere mistakes or misunderstandings. This distinction allows us to distinguish between obvious examples of “fake news” and mere mistakes. For example, many have the intuition that the highly cited *InfoWars* story about how the 2012 Sandy Hook school shooting in the US was staged is a paradigmatic example of “fake news” in the digital era (see Collins and Eaton-Robb 2022)—as distinct from being a simple case of confusion over some relatively subtle philosophical terminology employed in the Dennett obituary published by the *New York Times* (see Kandell 2024).

Fabrication, in the context of “fake news,” frequently involves rephrasing as a method for constructing misleading narratives. Rephrasing plays a crucial role in how fabricated content is framed and presented to the public. By subtly altering legitimate information or reconstructing false statements in a manner that mimics the tone and structure of credible sources, fabricators can create content that appears authentic, despite its deceptive intent. The process of rephrasing fabricated information often involves changing emphasis, removing critical context, or introducing emotionally charged language, all while maintaining an outward appearance of legitimacy. This strategic manipulation of language allows fabricators to enhance the believability of their falsehoods, making them more difficult

for audiences to detect and scrutinize. Thus, rephrasing acts as a vehicle for delivering fabricated content in a way that maximises its persuasive impact while concealing its intentional falsity. This connection between fabrication and rephrasing underscores the importance of developing tools that can identify subtle linguistic cues to better detect and counteract “fake news.”

Intent and Purpose captures the deliberate motivations behind the dissemination of false information. This category, which includes terms such as “intention to deceive,” “deliberate,” “purposeful” and “intentional,” was identified in 14 out of the 26 articles analysed, representing 61.5% of the total. This significant proportion underscores the importance of understanding the underlying motivations behind “fake news” in both academic and practical contexts. Unlike accidental misinformation, “fake news” is characterized by a conscious effort to mislead or deceive the audience. For instance, Alam et al. (2022), Singhal et al. (2019), Conroy et al. (2015), Fifita et al. (2023) and Parikh and Atrey (2018) highlight the purposeful nature of “fake news.” This intentionality is a key factor that differentiates “fake news” from other types of misleading information, such as rumours or unintentional errors.

To show that we can infer an intent to generate “fake news,” we can look at how rephrasing operates as a linguistic mechanism to subtly reshape stories to influence public opinion, promote a political agenda, or generate financial gain. This rephrasing often involves subtle changes in wording, tone, or emphasis, that significantly alter the perception of the information without making it immediately obvious that it has been manipulated. For instance, the title from the unconventional news website *Zero-Hedge*, “US Women’s Life Expectancy is the Lowest among Developed Nations” (Durden 2024), appears alarming, but upon closer analysis is misleading. The article provides a list of 10 developed countries, placing the United States at the bottom (at 80 years). However, two critical issues arise: first, the list excludes Slovakia (see WHO 2024), a country classified as developed (see UN 2014), where women’s life expectancy is 78 years. Secondly, the original article cited by *Zero-Hedge* refers to “high-income countries,” a term with a specific technical meaning, while the rephrased title substitutes it with “developed countries,” a term that carries broader and more urgent implications. This strategic rephrasing shows the intention of creating more urgency than the original article suggested, while still appearing to rely on solid evidence (this is also the case in the *InfoWars* example in Section 1. This example highlights how rephrasing operates as a linguistic mechanism to subtly reshape narratives, exploiting the trust in seemingly factual claims, while we nevertheless can infer an underlying intent behind the rephrased content.

Ways of Misleading People focuses on identifying the various forms that “fake news” can take. This category, which includes terms such as “misinformation,” “disinformation,” “hoaxes,” and “rumours,” was identified in 9 out of the 26 articles analysed, accounting for 42.3% of the total. The inclusion of terms like “misinformation” and “disinformation” in the definitions emphasizes the spectrum of accuracy and intent within “fake news.” Tschatschek et al. (2017) and Alam et al. (2022) discuss these distinctions, noting that misinformation typically involves the unintentional spread of false information, while disinformation is deliberately misleading. This differentiation is crucial for understanding the motivations behind the dissemination of “fake news,” as disinformation often carries malicious intent, while misinformation may arise from ignorance or error. Additionally, the category encompasses various formats of “fake news,” such as satire and hoaxes. For instance, Ajao et al. (2019) include satire as a form of “fake news,” recognising that while it may use humour and exaggeration, it can still mislead audiences who take the content at face value. This highlights the complexity of “fake news,” where even content meant as a joke can contribute to misinformation if not clearly identified as satire.

The Misinformation and Types of News category within “fake news” highlights the various forms it can take, and underscores how rephrasing plays a crucial role in shaping these forms to deceive or mislead. The inclusion of terms like “misinformation” and “disinformation” in the definitions emphasizes that “fake news” exists on a spectrum of accuracy and intent, where rephrasing can shift the meaning and perception of the information presented. Rephrasing allows creators to manipulate content in a way that aligns with their specific intent—whether to mislead unintentionally, as in the case of misinformation, or with malice, as with disinformation.

An effective analogy to describe this problem is the popular child’s game commonly known as “the Telephone Game,” where a message is whispered from one person to another in a line. By the time the message reaches the last person, it is often significantly distorted, bearing little resemblance to the original. In the context of “fake news,” disinformation acts similarly to the Telephone Game but with intentional distortion: the original information is deliberately rephrased at each stage to create a narrative that suits a particular agenda. When we turn to the news, we can see that the use of the phrase “swift-boating” became popular during the 2004 US presidential campaign involving John Kerry and George W. Bush. The phrase was a half-truth repeated to great effect and, arguably, caused Kerry to lose the

election (see Major and Andersen 2016).³ This kind of manipulation can introduce subtle inaccuracies or misleading interpretations, transforming what might have been factual content into something that is false yet appears credible.

Misinformation, though typically spread without harmful intent, can also resemble the unintended distortions in the Telephone Game. As information is passed along without proper verification, rephrasing may simplify, exaggerate, or alter details, leading to a message that is different from the original, even if the distortion is unintentional. This analogy highlights how rephrasing, even without malicious intent, can still lead to significant misunderstandings and the spread of inaccurate information. Additionally, the category includes various formats of “fake news,” such as satire and hoaxes, where rephrasing also plays a key role. Satire, for instance, relies heavily on exaggeration and humour, but when its satirical nature is not made clear, rephrasing can blur the line between joke and reality, leading some audiences to take the content at face value. Similarly, hoaxes often involve rephrasing factual information in a way that creates a completely false narrative, crafted to deceive the audience for the sake of entertainment, financial gain, or other motives.

For each kind of variation of “fake news,” it becomes apparent that rephrasing plays a significant role as a tool in its production. Whether the intent is to deceive, entertain, or simply provoke a reaction, rephrasing serves as a tool to reshape information into various types of “fake news,” each with its own impact on the audience. Understanding how rephrasing contributes to the complexity of “fake news,” much like the distortions in the Telephone Game, helps to illuminate the challenges in identifying and countering it in all its forms.

Turning to the Verification and Evidence category, this distinguishes “fake news” from other types of information, emphasising the presence or absence of supporting evidence and the verifiability of the content. This category, which includes terms such as “lack of evidence,” “unverifiable” and “false evidence,” was identified in 3 out of the 26 articles analysed, representing 26.9% of the total amount. The emphasis is on the unverifiable

3. For some more context: The term comes from Swift Boat Veterans for Truth (later Swift Vets and POWs for Truth), a “527” group that, in August 2004, ran TV ads and promoted the book *Unfit for Command* attacking Democratic nominee John Kerry’s Vietnam War record. (Kerry had commanded a Navy “swift boat.”) Many of the allegations were contradicted by Navy records and eyewitnesses and widely judged to be unsubstantiated. However, the ethotic attack achieved its greater strategic purpose regardless of the actual truth value of the main claim of the book and TV ads.

nature of “fake news.” For instance, Zhou et al. (2020) and Perez-Rosas et al. (2017) highlight the fact that a key characteristic of “fake news” is its lack of verifiable evidence. This absence of reliable sources or a factual basis makes it challenging to authenticate any information that, typically, would rely on verifiable data and credible sources. Khattar et al. (2019), interestingly, examines the concept of false evidence, where “fake news” lacks verification and may present fabricated or misleading evidence to support its claims. Here, it might be useful to draw a distinction to the effect that all false evidence is fabricated or misleading, but not all fabricated or misleading claims count as false evidence. It is worth noting that one issue with this category is that some legitimate news sources are difficult or impossible to verify. One might consider, for example, a news article that uses whistle-blower information or depends on confidential sources.

Rephrasing plays a pivotal role in how “fake news” navigates the Verification and Evidence category, particularly by obscuring the lack of verifiable information, or fabricating evidence to support false claims. When creators of “fake news” manipulate content, they often engage in rephrasing to make unverifiable information appear more credible in ways difficult to detect. This manipulation underscores the implicit imperative to cultivate robust critical thinking skills and dispositions amongst news consumers.

The Harm and Normative Factors category focuses on the ethical and societal implications of “fake news,” highlighting the potential damage caused by the dissemination of false information. This category includes terms such as “harm,” “unethical,” “normative judgments” and “negative impact.” It was identified in 5 out of the 26 articles analysed, representing 26.9% of the total amount. This number is noteworthy in the field of computer science, as it seems desirable for this normative dimension to be captured in the concept of “fake news.”

Harm is a central theme in this category, reflecting the various ways “fake news” can negatively impact individuals and society. For instance, Ruchansky et al. (2017) and Sharma and Singh (2024) discuss how “fake news” can lead to real-world consequences, such as social unrest, public panic, and even violence. These studies highlight the potential for “fake news” to exacerbate conflicts, spread fear, and undermine social cohesion, showcasing the broader societal harm that can result from the spread of misinformation. Alam et al. (2022) addresses the ethical dimensions of “fake news,” describing it as inherently reflecting an intention to deceive and manipulate the public. This ethical violation is particularly concerning when “fake news” is used to influence public opinion or manipulate political outcomes, as it undermines democratic processes and erodes public trust

in media and institutions. The deliberate spread of false information with the intent to mislead is widely regarded as an unethical practice, raising important questions about the responsibilities of information providers and the need for accountability.

The use of “normative judgments” and the discussion of “negative impact” further highlight the moral considerations associated with “fake news.” Berrondo-Otermin and Sarasa-Cabezuelo (2023) emphasize the normative implications of spreading misinformation, noting that “fake news” often involves making normative claims that are intended to shape public attitudes or behaviours. Additionally, Desamsetti et al. (2023) discusses the broader negative impacts of such news on public discourse, pointing out that the spread of false information can lead to a general degradation of the quality of information available to the public. This degradation can have long-term consequences, including the erosion of informed public debate and the weakening of societal resilience against misinformation. The authors highlight the need for stronger measures to combat “fake news” and protect the integrity of public discourse.

Rephrasing plays a critically important role in the Harm and Normative Factors category, particularly in respect of how “fake news” is crafted to inflict ethical and societal harm. The act of rephrasing in such news causes harm by making “fake news” nearly indistinguishable from real news. Also, the manipulation of language through rephrasing can amplify the negative impact of “fake news” by making harmful content appear more credible or persuasive. For instance, the ethical violation inherent in “fake news” is often deepened by the way information is rephrased to deceive or manipulate the public. Rephrasing can involve twisting facts, selectively omitting context, or framing information in a way that heightens emotional responses, all of which contribute to the unethical nature of “fake news.”

The term “harm” is central to this category, and rephrasing is frequently used to exacerbate the potential damage caused by “fake news.” For example, rephrased content may be designed to incite fear or anger, leading to real-world consequences such as social unrest or violence. By carefully choosing words and altering the presentation of information, creators of “fake news” can intensify the harmful effects on individuals and society, making the false information more likely to provoke extreme reactions. This manipulation of language not only spreads misinformation, but also actively contributes to the deterioration of social cohesion and public trust.

Rephrasing, here, also plays a role in the normative judgments associated with “fake news.” When such news is rephrased to make normative claims—statements that express values or prescribe actions—it can shape

public attitudes and behaviours in ways that align with the deceptive intent of its creators. This rephrasing can lead to the spread of harmful ideologies or the manipulation of public opinion, particularly when the language used is crafted to resonate with existing biases or fears. The ethical implications of such rephrasing are significant, as it involves deliberately steering public discourse in a direction that undermines the integrity of information and democratic processes.

Moreover, the broader negative impacts of “fake news” on public discourse, as discussed by authors such as Desamsetti et al. (2023), are often magnified through the strategic rephrasing of content. By presenting misleading information in a way that appears authoritative or aligns with popular narratives, rephrased “fake news” can degrade the overall quality of public debate. This degradation erodes societal resilience against misinformation, as the public becomes less able to distinguish between truth and falsehood in a landscape where rephrased “fake news” is prevalent.

The last category is the Undefined one. Some papers left the concept of “fake news” undefined (Dong et al. 2023; Qian et al. 2018). This absence of any definition is itself notable, because it assumes that the definition of “fake news” is so well understood that there is no need to supply one.

To conclude, our investigation into definitions of “fake news” in computer science has underscored the centrality of rephrasing as a mechanism for subtle misinformation. Central to this phenomenon is the use of falsehoods and fabrication, where strategic linguistic adjustments create an illusion of authenticity while embedding misinformation. A key tactic involves mimicking credible sources, replicating the tone, style and structure of legitimate journalism to exploit audience trust. Rephrased content often reflects deliberate intent and purpose, aligning language with specific objectives such as promoting propaganda or influencing public opinion. These manipulations frequently involve selective omission, emphasis, and pragmatic shifts that subtly alter meaning without overt contradiction. Furthermore, rephrasing serves to obscure verification and evidence, either by masking unverifiable claims or fabricating support to enhance believability. The societal and ethical impact of such rephrasing is significant, as it amplifies harm by provoking emotional reactions and exacerbating divisions. Detection tools in computer science focus on identifying linguistic patterns, structural mimicry, and shifts in tone to counteract the persuasive force of rephrased misinformation. Together, these elements underscore the nuanced and deceptive nature of rephrasing as a tool in fake news.

3. DEFINITIONS OF “FAKE NEWS” IN PHILOSOPHY

In this section, we examine the philosophical discourse surrounding definitions of “fake news,” focusing on how various philosophers have sought to conceptualize this phenomenon. While much of the existing work on “fake news” has centred around its detection and mitigation through technical means (as in computer science), philosophy offers a more nuanced and theoretical exploration of its ethical, epistemological and linguistic dimensions.

“Fake news” raises deeply philosophical issues. The topic raises significant questions about truth, epistemology, ethics, and the impact of information on society in the Digital Age. Below, we will explore the philosophical dimensions of “fake news,” focusing on how various theoretical frameworks can help us understand and address this complex problem as it simultaneously relates to computer science. Philosophers have long been concerned with the nature of truth and the ethical implications of communication, making the study of “fake news” particularly relevant to contemporary philosophical discourse, which in turn might provide useful contemporary tools for identifying such news. However, it is worth noting that there is no single agreed-upon definition of “fake news” (albeit that such a state of affairs is hardly an uncommon occurrence in philosophy). In fact, there are some philosophers (e.g. Habgood-Coote 2019; Musi and Reed 2022) who have either asserted that defining “fake news” is a fool’s errand, or sought to side-step it entirely. Nevertheless, the emphasis here is not so much on technological tools being used to identify “fake news,” as in computer science. Rather, what we encounter here is a deeper discussion about the nature of falsity, with more attention paid to such subtle aspects of “fake news” as the copying or mimicking of legitimate news.

Now we will turn to the analysis of selected definitions of “fake news” in the philosophical literature. Here we have identified 20 representative definitions that offer a more nuanced take on “fake news,” utilizing the various subdivisions in philosophy such as epistemology, philosophy of science, and ethics. In this subsection, much as in Section 2, we will go through the definitions offered and look for similarities between them. (At the same time, we will not seek to elaborate here on those categories that overlap sufficiently with those already encountered in Section 2.) We identify categories that the various definitions share in Table 2. These include the following: Mimicking, False, “Bullshit,” Lack of Concern for Truth, Intentional, and Social or Political Aspect. We will then turn to identifying the aspects of these categories related to rephrasing.

Table 2. Philosophy definitions of “fake news”: major tendencies

#	Author(s)	Mimicking	False	Bullshit	Lack of Concern for Truth	Intentional	Social or Politi- cal Aspect	Undefined
1.	Harris (2022)	X			X		X	
2.	Stewart (2021)							X
3.	Croce and Piazza (2021)	X			X	X	X	
4.	Dentith (2016)	X	X			X		
5.	Jaster and Lanius (2018)		X	X		X		
6.	Fallis and Mathiesen (2019)	X			X	X	X	
7.	Mukerji (2018)			X	X		X	
8.	Pepp et al. (2019)	X					X	
9.	Musi and Reed (2022)							X
10.	Anderau (2021)	X				X		
11.	Rini (2017)	X	X				X	
12.	Gelfert (2018)	X	X			X		
13.	Grundmann (2023)	X						
14.	Levy (2017)	X			X	X		
15.	McIntyre (2018)		X			X		
16.	Habgood-Coote (2019)							X
17.	Novaes and de Ridder (2021)							X
18.	Goldman and Baker (2019)		X			X		
19.	Galeotti (2019)	X	X	X			X	
20.	Ball (2021)	X			X		X	

The structure of the discussion of categories here proceeds as follows: first, we describe the data that we obtained by presenting the relevant statistical information, then we outline the elements of rephrasing discernible in the features of various definitions, and finally we look at some ethical aspects. The goal is to see how rephrasing manifests itself within these features. However, for the sake of space, we will not elaborate on those categories that overlap sufficiently with what was explored above in relation to computer science.

The Mimicking category is a significant element in the philosophical literature on “fake news.” In this category, we have included terms and phrases

such as “mimic,” “guise of news,” “news that represent itself as genuine,” “designed,” and “presented as genuine news.” We have discovered that the majority of definitions fall under this category: i.e. 12 of the 20 definitions (60%). The idea of mimicking is notable because it doesn’t necessarily rely on the truth value of its claims to be considered “fake news” (e.g. Harris 2022; Ball 2021; Galeotti 2019). The emphasis on mimicry underscores the deceptive strategy employed in “fake news”: by imitating the appearance and format of legitimate news outlets, “fake news” seeks to exploit the trust that audiences place in established journalistic practices. This mimicry makes it challenging for consumers to distinguish between authentic and false information.

In the context of “fake news,” rephrasing is instrumental in the process of mimicking genuine news sources (i.e. it is a subcategory of mimicking—mimicking can be an effect of rephrase, and rephrasing is one way to mimic). By carefully adjusting linguistic elements, creators of “fake news” replicate the style, tone and structural conventions of legitimate journalism to deceive readers. This mimicry is achieved through the strategic use of rephrasing, which allows for the alteration of authentic content or the fabrication of new content that closely resembles credible reporting. (There is an assumption involved here that mimicking rests upon, that will be elaborated on as potentially problematic in the next paragraph). Rephrasing enables the integration of falsehoods into a familiar journalistic framework, making the fabricated information appear trustworthy. This deliberate manipulation exploits the readers’ expectations and cognitive biases, particularly their trust in established news formats. As a result, the role of rephrasing in mimicking not only enhances the deceptive quality of “fake news,” but also poses significant challenges for detection and critical assessment. Recognizing the patterns of rephrased language that contribute to mimicry is therefore essential for developing analytical skills and critical thinking habits aimed at identifying and mitigating the impact of “fake news” in public discourse.

In order to understand the concept of mimicking, it is worth exploring the distinction between legitimate and illegitimate news sources, as the concept of mimicking in “fake news” has this distinction built in: after all, to mimic one must mimic something or someone. As was mentioned before, the category of mimicking rests on a potentially problematic assumption: it must assume that there is a difference between legitimate and illegitimate news sources. It is not always obvious when a news source is legitimate or not. This issue was touched upon by Habgood-Coote’s (2019) third argument about the propagandistic aspect of using the term “fake news.” The object

that the mimicker is emulating is the legitimate news source, and the mimicker is the illegitimate new source. This is not an insignificant assumption. The problem with the difference between legitimate and illegitimate new sources is captured in the following question: Who is the arbiter of what is legitimate or illegitimate? Let us suppose that we submit that the government is the arbiter of legitimacy. This might seem reasonable. However, intuition and historical experience tell us that we might well want to refer to some government-sourced news as “fake news.” For example, consider communist-run Poland (PRL) in the 1980s. Only government-sanctioned stories counted as legitimate news sources. The reporters would declare that the population was happy and that there was plenty of food in all of the shops. However, for many Polish people confronted by empty shelves this was evidently not the case. It was often reported that a common response from shopkeepers was just “*Nie ma*” [roughly: “it is not here”]. So do we believe the official news source, or our own eyes? This is a deeper problem, certainly worth exploring in terms of its potential implications; however, it lies outside of the scope of our paper.

Turning now to the category False, we find that this emerges as a less central one in philosophical definitions of “fake news,” appearing as it does in just 7 out of the 20 definitions analysed (35%). It encompasses terms such as “false,” “untrue,” “fabricated,” and “not supported by evidence,” highlighting the dissemination of information that deviates from factual accuracy (Dentith 2016; Gelfert 2018; McIntyre 2018; Galeotti 2019). The prevalence of falsehood in these definitions underscores the ethical and epistemological concerns associated with the spread of misinformation in public discourse. We take it that there is sufficient overlap with the category of “falsehood” in Section 2 to not warrant further elaboration.

The category of Bullshit, as articulated by Frankfurt (2005), appears in only 3 out of the 20 philosophical definitions analysed (15%). This category includes definitions where “fake news” is characterized by a lack of concern for the truth, with producers being indifferent to the veracity of the information they disseminate (see, e.g., Jaster and Lanius 2018; Muke-rji 2018). Unlike outright falsehoods or intentional lies, “bullshit” involves statements made without regard for their truthfulness, aiming instead to persuade or manipulate the audience for other purposes. (We ourselves only took into account papers that made explicit use of the Frankfurtian conception of “bullshit.”) The notion is significant because it shifts the focus from the content of the message to the attitude of the communicator towards truth. Producers of “fake news,” in this sense, are not necessarily committed to spreading falsehoods, but are indifferent to whether their

statements are true or false, as long as they achieve their desired effect. This indifference undermines the epistemic foundations of public discourse and poses challenges for identifying and addressing misinformation.

Rephrasing plays a significant role in the propagation of “bullshit” within “fake news.” The process of rephrasing allows communicators to construct messages that sound plausible and convincing while lacking substantive truth. In cases of “bullshit,” rephrasing is utilized to reshape information in a way that prioritizes rhetorical effectiveness over factual correctness. This involves the use of ambiguous language, rhetorical flourishes, or emotionally charged expressions that can mislead the audience by creating an impression of meaningful communication where there is none. For example, a statement such as “A significant number of experts agree that this policy will fail” can be rephrased as “A significant number of experts agree that this policy is doomed,” amplifying the persuasive force without providing concrete evidence or specifying sources.

The ethical implications of disseminating “bullshit” are also worth mentioning, as it contributes to a culture of indifference toward truth and undermines meaningful dialogue. When individuals or institutions prioritize rhetorical appeal over accuracy, they foster a communicative environment where truth is secondary to persuasiveness. This can lead to a normalization of deceptive communication practices, making it increasingly difficult for audiences to distinguish genuine information from manipulative rhetoric. Moreover, the spread of “bullshit” erodes public trust in information sources, which is especially damaging in contexts where informed decision-making is critical, such as politics or public health. The cumulative effect is a public discourse that is more susceptible to misinformation, where audiences may become cynical and disengaged, doubting all information regardless of its source.

The category Lack of Concern for Truth features prominently in philosophical definitions of “fake news,” appearing in 30% of the definitions analysed (6 out of 20). This category highlights an attitude where the producers of “fake news” exhibit indifference towards the veracity of the information they disseminate (e.g. Ball 2021; Levy 2017; Fallis and Mathiesen 2019). Unlike intentional deception, where falsehoods are propagated knowingly, a lack of concern for truth reflects a disregard for whether the information is true or false, so long as it serves the communicator’s purpose. This category is very similar to that of Bullshit. Here, we looked for terms other than “bullshit” that were used to express a lack of concern for truth. We take it there is sufficient overlap with the category of Bullshit (in Section 3) to not warrant further elaboration.

The Intentional category is featured prominently in philosophical definitions of “fake news,” appearing as it does in 9 out of the 20 definitions analysed (45%). This category emphasizes that the dissemination of “fake news” involves a deliberate intent to deceive or mislead the audience (see, e.g., Dentith 2016; Gelfert 2018; Rini 2017; Croce and Piazza 2021). The centrality of intentionality in these definitions underscores the ethical dimension of “fake news,” highlighting the purposeful actions of communicators who manipulate information to achieve specific objectives, such as influencing public opinion, advancing political agendas, or generating financial gain. We conclude that there is sufficient overlap here with the category of Intent and Purpose in Section 2 to not warrant further elaboration.

The category of Social or Political Aspect appears in 40% of the philosophical definitions analysed, featuring in 8 out of 20 cases. This one emphasizes that “fake news” often serves specific social or political purposes, such as manipulating public opinion, advancing political agendas, or undermining democratic processes (see, e.g., Harris 2022; Pepp et al. 2019; Rini 2017). The inclusion of social and political dimensions highlights the broader impact of “fake news” on society, and its potential to influence collective behaviours and attitudes. The social or political aspect underscores the fact that “fake news” is not merely an isolated communicative act, but is embedded within larger socio-political contexts. It often exploits existing social tensions, ideological divides, or political controversies to achieve its objectives. This dimension reflects the instrumental use of “fake news” as a tool for propaganda, disinformation campaigns, or social engineering. It would appear that there is sufficient overlap with the category of Harm in Section 2 to not warrant further elaboration.

The category of Undefined, meaning that no definition has been provided, can be observed in 20% of the philosophical literature analysed, with 4 out of 20 authors opting not to offer a formal definition of “fake news,” or arguing that seeking such a definition is futile. Notably, Habgood-Coote (2019) argues against having any such definition by contending that the term “fake news” is itself linguistically defective and propagandistic, since it lacks a stable and coherent meaning. Novaes and de Ridder (2021) follow Habgood-Coote’s (2019) lead and refrain from defining the term, arguing instead that the concept of “fake news” is either too ambiguous or too problematic to warrant a precise definition. This approach reflects a critical stance toward the term itself, suggesting that its usage may be more harmful than beneficial to public discourse. Musi and Reed (2022) sidestep the “fake news” debate by constructing a definition of “semi-fake news.” Stewart (2021) contends that “fake news” amounts to an umbrella

term that should be thought of as being composed of smaller parts: namely, “misinformation,” “disinformation,” and “misleading content.”

By examining the definitions of “fake news” in both computer science and philosophy, we have highlighted their key elements and their relevance to understanding rephrased misinformation. Computer science often emphasizes practical detection methods and the operational aspects of misinformation, focusing on falsehoods, intent, and their harmful impact. In contrast, philosophy provides a deeper exploration of intent, ethical considerations, and the broader social implications of “fake news.” Together, these perspectives allow us to better grasp the interplay between rephrased language and the construction of misleading narratives.

Rephrasing in “fake news” highlights deeper issues such as truth, ethics, and societal impact. A prominent feature is mimicry, where rephrased content imitates the style and format of legitimate journalism, exploiting public trust in traditional news sources. Philosophical definitions emphasize the role of intentionality, as rephrased “fake news” often reflects deliberate efforts to mislead, influence opinions, or serve specific political and social agendas. The concept of falsehood is central, as rephrasing introduces inaccuracies or distorts meaning while maintaining a facade of legitimacy. Philosophy also considers the phenomenon of indifference to truth, where creators prioritize rhetorical impact over factual accuracy, contributing to the erosion of public trust in media. Additionally, the social and political dimensions of rephrased fake news are critical, as such manipulations amplify ideological divisions and undermine constructive public discourse. These elements underscore the ethical and epistemological concerns associated with rephrased misinformation, offering valuable insights into its societal consequences and highlighting the need for critical engagement with deceptive communication practices.

With these theoretical frameworks in mind, we turn our attention to the linguistic study of rephrasing. By analysing how rephrasing operates as a powerful tool for misinformation, we aim to identify key cues and mechanisms that contribute to its influence. In the next section, we will delve into the linguistic research on rephrasing, exploring how it can be used both to deceive and to develop effective critical thinking strategies for identifying misinformation.

4. RESEARCH ON REPHRASE IN ARGUMENTATION THEORY AND CORPUS LINGUISTICS AS A FRAMEWORK FOR STUDYING “FAKE NEWS”

Taking into account the parallels identified in Sections 2 and 3 between the concept of “fake news” and that of rephrasing, we now turn to recent

research on rephrase, construed as furnishing a framework for exploring the potential of rephrase-analysis tools in the study of argumentative discourse. This exploration leads us to propose a research program centred on the deployment of models of rephrase when reformulating messages to help identify fake news.

It is worth pointing to some of the literature on rephrase. An account is given in Younis et al. (2023)—but in sum, one can say that the concept of “rephrase” is moving towards a definition. Rephrase is first found in Inference Anchoring Theory (IAT), as a propositional relation distinct from inference and conflict. A speaker’s intention is decisive: when two contributions are linked by rephrase, the second is meant as a restatement of the first that neither provides a reason for it nor opposes it; marking a segment as rephrase therefore rules out pro- or con-argument status. The authors then refine types, positing rephrase specification (the second contribution is narrower in meaning) and rephrase generalization (the second is broader), and noting that these operate along multiple semantic scales (quantitative, evaluative, part-whole) and appear to be among the most frequent subtypes in argumentative corpora. They contrast rephrase with neighbouring notions: reformulation (an “equivalence operation” at the semantic/pragmatic level) and paraphrase (approximate semantic equivalence). Because IAT defines rephrase by communicative function (non-inferential restatement) rather than semantic equivalence, there are reformulations that are not rephrases (e.g. “in other words” used to draw an inference), and rephrases that are not paraphrases (when meanings are not semantically close). The upshot is a functional, intention-sensitive account that distinguishes rephrase from both argumentative support/attack and from purely semantic sameness, while motivating a finer typology of subtypes.

One word about rephrase and intentionality. Rephrase may, but does not have to, be accompanied with a speaker’s intention to reformulate. Still, the intentional aspect of rephrasing may be the purposeful act of expressing a statement in different words, not merely to clarify, but also to achieve a strategic rhetorical effect, gain a persuasive advantage, gain deeper comprehension of an original message, emphasize a specific point, or manage the discourse flow. This intent may differentiate rhetorical uses of rephrase from mere paraphrasing or restatement, as it may carry a persuasive goal beyond pure reformulation. Thus, in recent rephrase studies (e.g. Konat et al, 2016; Younis et al, 2023) these intentions of rephrasing have been brought up for discussion.

Now with the background discussion on rephrase done, our exploration of the technique of rephrasing messages has emerged as a powerful yet

insidious tool in the dissemination of “fake news.” By altering language subtly, creators of misinformation can manipulate truth in ways that are both convincing and deceptive.

Given the tendencies towards defining “fake news” using certain key rephrase-related terms, such as “mimicking”⁴ and “fabricating,” as discussed in Sections 2 and 3, this section will explore key areas of study of rephrase with a view to developing future tools for identifying “fake news” in reformulated messages. To this end, we propose possible directions of future inquiry to supplement the existing critical thinking theories with a study of how rephrasing for the sake of generating and spreading “fake news” can mislead audiences, evade detection, and amplify the impact of such news. By synthesizing key features identified in computer science and philosophical discussions, we aim to illuminate the role of rephrasing in the construction and spread of misinformation. Furthermore, this section seeks to establish a foundation for using some critical thinking tools and analytical frameworks to recognize and counteract the deceptive nature of rephrased content. Through this exploration, we move closer to understanding how to mitigate the pervasive influence of fake news in contemporary discourse.

For the purpose of succinctly depicting the potential rephrase-related properties of communication typical of the dissemination of “fake news,” we list in Table 3 those properties that we have found to furnish a regularity in respect of definitions of “fake news” appearing in both computer science and philosophy.

As we observed in Section 2, the definitions in computer science include the following features of rephrasing in “fake news”:

- Using language to obscure or misrepresent the truth.
- Mimicking the original content to create deceptive yet credible outputs.

4. We are not trying to claim that mimicry is a subtype of rephrase, but rather that it is an effective linguistic tool for rephrase. Rephrase, in some use cases, such as misrepresenting an opponent’s statement in straw man fallacies, might instead be conceived as an instance of mimicking, in the sense of making a p' “pretend” to contain exactly the same content as an original statement p , whereas in fact a p' contains a modified content, which is accompanied by the intention of making the other party’s position easier to attack. In this respect, in some cases, mimicking can be an effect of rephrase. The goal of our paper, which is to explore the potential of rephrase studies in identifying some linguistic manifestations, among other key manifestations, was partly achieved through indicating and discussing mimicking as a discursive strategy that may (but does not have to) be achieved by means of rephrasing. As the exploration of this overlap has a potential for capturing those rephrase uses that may illicitly mimic original messages, we have given in our paper arguments in favour of the claim that, in this respect, the systematic study of rephrase may be incorporated into the broader toolset for identifying misinformation.

- Masking unverifiable claims by pretending to reference reliable sources.
- Exploiting linguistic and structural patterns to evade detection.

Table 3: Principal rephrase-related features of “fake news”: summary from Sections 2 and 3

Category	Feature
Mimicry	Imitation of the tone, style, and structure of legitimate sources to exploit trust
	Creating outputs similar to the original content but embedding distortions
Manipulation of Truth	Using language to subtly obscure or misrepresent the truth
	Introducing distortions while maintaining the appearance of legitimacy
Intentionality	Deliberate efforts to align rephrased content with specific agendas
	Often linked to operational motives (e.g. propaganda)
Verification and Credibility	Masking unverifiable claims through subtle linguistic adjustments
	Pretending to have credible sources while tweaking content
Emotional and Cognitive Impact	Provoking emotional responses such as fear or anger through rephrasing
Ethical and Epistemological Concerns	Erosion of public trust due to repeated exposure to rephrased misinformation
Social and Political Implications	Amplifying societal divisions through tailored rephrased narratives
Critical Thinking Relevance	Recognizing deceptive patterns through linguistic and pragmatic analysis

The definitions of “fake news” in philosophy, on the other hand, as discussed in Section 3, emphasize features relating to concepts such as—among others—trust, truth, factual accuracy, and knowledge:

- Mimicking legitimate formats to exploit trust in established institutions.
- Prioritizing rhetorical appeal over factual accuracy (manipulation of truth).
- Amplifying societal and political divisions through targeted rephrasing (social and political implications).
- Embedding nuanced distortions that subtly alter the original meaning.

Turning now to the main features of “fake news” as found in the definitions taken from the philosophical literature (see Section 3), let us point to the key features of fake news: (i) using language to *mislead or obscure the truth*; (ii) pretending to have credible sources of information; and (iii) ethical and social aspects.

When considering these features in terms of possible benefits of employing rephrase studies to identify these kinds of misinformation in discourse, we can observe that feature (i) of “fake news”—“obscuring the truth”—may have a great deal in common with misrepresenting the content of an original statement in cases of rephrase use. Applying this communication technique in discourse by means of the use of rephrase can be reinforced by using linguistic techniques of rephrasing. Once this is done, a rephrase output which looks very similar to an original input may in fact mimic the original content, and thus be employed as a subtle misinformation tool. Thus, the role of rephrase in generating and spreading misinformation may consist in modifying the linguistic surface of a message in such a way that the truth is made obscure.⁵ For example, consider the ever-updating news aggregate website the *Drudge Report*. This influential site, run by Matt Drudge, rephrases original news headlines in order to make them more attractive to his audience (Carr 2011).

Building on this idea, feature (ii)—“pretending to have credible sources of information”—further illustrates how rephrasing can be manipulated to mislead an audience. This tactic involves referring to objectively reliable sources while simultaneously tweaking the content or context of what those sources originally stated. This kind of rephrase use essentially consists of referring to a given source of information in an inadequate way by, e.g., not referring to the source directly, while misrepresenting what that source mentions. To a certain extent, this way of using rephrase is similar to performing those fallacious arguments from expert opinion that rely on manipulating the content of what a genuine authority has uttered. Likewise, in such cases, misuses of rephrasing can be intentionally employed to create misrepresented content which, if spread widely, can serve as efficient “fake news,” functioning effectively precisely because of the similarity of contents.

5. Here we have in mind such as issues as straw man fallacies (Visser et al. 2018): rephrase may, but does not have to, mislead audiences by making the news “fake”. We emphasize that rephrase, due to its linguistic and structural features, may be just one of the possible vehicles for misinforming people, and thus also for the spreading of “fake news.”

5. CONCLUSION AND FUTURE WORK

To conclude, we have conducted an investigation into how rephrasing contributes to the construction and persuasive power of “fake news” by analysing definitions from both computer science and philosophy. Our paper has:

- examined representative definitions to identify core rephrase-related features—such as mimicry, fabrication, intentionality and the strategic obscuring of truth—that underlie “fake news.”
- bridged a gap between the overlap between rephrase and “fake news” as exhibited in the literature in philosophy and in computer science, in order to propose specific areas of philosophical and linguistic research into rephrase, applied specifically to identify linguistic manifestations of certain instances of rephrase that may have the effect of “fake news.”

This approach, we think, advances the field by opening new research threads for scholars in discourse analysis, argumentation studies, and communication studies, and does so by examining the role played by rephrase in respect of identifying a new class of “fake news.” In contrast to some papers, such as that of Anderau (2021), whose work centres primarily on defining “fake news,” our study extends the inquiry by emphasizing the role of rephrasing as a dynamic communicative strategy. By doing so, we aim to provide insights that will hopefully prove useful for developing a more comprehensive framework linking theoretical discoveries with practical applications—thereby equipping scholars to explore how nuanced linguistic manipulations can be systematically detected and mitigated in public discourse.

The study of rephrase, far from being a marginal phenomenon in communication (Younis et al. 2023), offers valuable insights into a crucial mechanism for the spread of “fake news.” By examining how rephrased content subtly manipulates and reframes information, we can better understand its role in constructing persuasive yet misleading narratives. Rephrase serves as an effective model for studying and detecting a powerful class of “fake news” that thrives on the ambiguity created by linguistic shifts.

Both computer science and philosophy provide complementary approaches to understanding and combating rephrased “fake news.” In computer science, “fake news” is operationalized for detection, focusing on patterns, algorithms, and data-driven approaches to identifying misrepresentations. These methods aim to provide concrete, scalable tools for filtering and analysing information in digital environments. On the other hand, philosophy provides a deeper theoretical framework, focusing

on the ethical, epistemological, and communicative dimensions of “fake news.” Philosophical approaches emphasize the intent behind deception and the ways in which language, including rephrasing, can be weaponized to obscure truth and manipulate public perception.

These two fields together offer a robust framework for addressing the challenges posed by rephrase in “fake news.” While computer science provides the practical tools for detection, philosophy probes the deeper implications of how and why rephrasing manipulates understanding. Future interdisciplinary collaborations between these fields could enhance our ability to both detect and critically analyse “fake news,” fostering a more holistic approach to combating misinformation.

Furthermore, developing a comprehensive corpus of rephrase misuse and implementing rephrase-checking analytics could provide significant tools for combating the dissemination of manipulated content. These tools would not only enhance our capacity to detect misused rephrases but also highlight how rephrase impacts cognitive biases, potentially paving the way for broader applications in misinformation detection. Critical thinking plays a pivotal role in this context, offering a practical framework for individuals to build immunity against the subtle manipulations that rephrased “fake news” entails. By fostering analytical skills and a disposition for scepticism, we can empower individuals to better navigate the complex information ecosystems of today’s online platforms, where rephrased content often evades conventional detection systems.

Looking ahead, a comprehensive model aimed at capturing the nuances of rephrase within “fake news” strikes us as essential.⁶ Such a model would focus on identifying and classifying the specific linguistic strategies through which rephrasing subtly manipulates content, often making deceptive statements more persuasive and harder to detect. This would provide a theoretical framework that complements existing detection methods, particularly those grounded in computer science. The development of this model would not only enhance detection capabilities but also provide new ways to systematically categorize rephrased misinformation across different media platforms.

In addition, a large, annotated corpus dedicated to the study of rephrase misuse would certainly be beneficial in pursuit of this enterprise. This

6. We would like to note that the research is ongoing into the technical use of theoretical rephrase tools for designing multi-agent systems that employ rephrase and argumentation (see Uberta et al, forthcoming), where this is devoted to taking argumentation and rephrase studies as a theoretical model framework for AI tools to evaluate the effectiveness of those theories in argument and rephrase evaluation. Such work addresses the issue of the erroneous use of rephrase.

corpus would serve as a vital resource for future studies, enabling both qualitative and quantitative analysis of how rephrase is employed in misleading contexts. Such a resource would also support the development of rephrase-checking algorithms, which could be integrated into existing “fake news” detection systems, offering a more refined means of identifying subtle yet impactful manipulations. This corpus would also facilitate cross-disciplinary research, merging insights from linguistics, argumentation studies and computational analysis.

Another promising direction for future research lies in integrating insights from rephrase studies into practical tools for combating “fake news.” One such application could build upon existing systems like the Reason-Checking “fake news” app Evidence Toolkit (Visser et al. 2020), which focuses on identifying and countering misinformation. Expanding this toolkit to include rephrase analysis would allow for the detection of subtle linguistic manipulations that often evade traditional fact-checking methods. By systematically incorporating mechanisms to identify mimicked structures, adjusted tones, and rephrased content that obscures or distorts truth, such a tool could serve as a comprehensive resource for addressing the deceptive nature of rephrased “fake news.” This integration would bridge the gap between theoretical research on rephrasing and its practical implications, empowering users to identify and resist the influence of misinformation in digital communication.

Finally, future work should prioritise the integration of critical thinking instruction, specifically aimed at empowering individuals to recognise rephrasing strategies that contribute to the spread of “fake news.” Educational frameworks that incorporate critical thinking through practical exercises for detecting rephrased misinformation could offer a defence against “fake news” that evades traditional detection systems. This instruction should be tied directly to an analysis of how rephrasing strategies exacerbate societal polarization, particularly in digital communication environments. By addressing the intersection between rephrasing and “fake news,” these initiatives could ultimately contribute to reducing the divisive effects of misinformation in public discourse.

AUTHORS’ CONTRIBUTIONS

Mitchell Welle: Literature Review, Argument Development, Analysing the Definitions, Qualitative Analysis, Interpretation of Results, Formatting, Editing, Proofreading.

Marcin Koszowy: Framing the Problem, Conceptualization and Modelling, Argument Development, Qualitative Analysis, Editing.

BIBLIOGRAPHY

- Ajao, Oluwaseun, Deepayan Bhowmik, and Shahrzad Zargari. 2019. “Sentiment Aware Fake News Detection on Online Social Networks.” In *ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8504–08. Piscataway, NJ: IEEE.
- Alam, Firoj, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimitar Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. 2022. “A Survey on Multimodal Disinformation Detection.” arXiv:2103.12541. <https://arxiv.org/abs/2103.12541>.
- Anderau, Glenn. 2021. “Defining Fake News.” *Kriterion—Journal of Philosophy* 35 (3): 197–215. <https://doi.org/10.1515/krt-2021-0019>.
- Ball, Brian. 2021. “Defeating Fake News: On Journalism, Knowledge, and Democracy.” *Moral Philosophy and Politics* 8 (1): 5–26. <https://doi.org/10.1515/mopp-2019-0033>.
- Berrondo-Otermin, Maialen, and Antonio Sarasa-Cabezuelo. 2023. “Application of Artificial Intelligence Techniques to Detect Fake News: A Review.” *Electronics* 12 (24): 5041. <https://doi.org/10.3390/electronics12245041>.
- Carr, David. 2011. “How Drudge Stayed on Top.” *New York Times*, May 15. <https://www.nytimes.com/2011/05/16/business/media/16carr.html>. Accessed November 5, 2025.
- Collins, Dave, and Pat Eaton-Robb. 2022. “Families Testify of Confrontations with Sandy Hook Deniers.” AP News, September 27. <https://apnews.com/article/shootings-texas-connecticut-alex-jones-waterbury-f3ef318375efc40dccc7067fb8f53e5e>. Accessed November 5, 2025.
- Conroy, Nadia K., Victoria L. Rubin, and Yimin Chen. 2015. “Automatic Deception Detection: Methods for Finding Fake News.” *Proceedings of the Association for Information Science and Technology* 52 (1): 1–4. <https://doi.org/10.1002/pra2.2015.145052010082>.
- Croce, Michel, and Tommaso Piazza. 2021. “Misinformation and Intentional Deception: A Novel Account of Fake News.” In *Virtues, Democracy, and Online Media: Ethical and Epistemic Issues*, edited by Maria Silvia Vaccarezza and Nancy E. Snow, 52–67. New York: Routledge.
- Dentith, M.R.X. 2016. “The Problem of Fake News.” *Public Reason* 8 (1–2): 65–79.
- Desamsetti, Sankar, Satya Hemalatha Juttuka, Yamini Mahitha Posina, S. Rama Sree, and B.S. Kiruthika Devi. 2023. “Artificial Intelligence Based Fake News Detection Techniques.” In *Recent Developments in Electronics and Communication Systems*, edited by K.V.S. Ram Ramachandra Murthy, Sanjeev Kumar, and Mahesh Kumar Singh, 374–80. *Advances in Transdisciplinary Engineering* 32. Amsterdam: IOS Press. <https://doi.org/10.3233/ATDE221284>.
- Diaz-Garcia, Jose A., Dolores Ruiz M., and Martin-Bautista Maria J. 2025. “Liars Know How to Argue: An Approach to Disinformation Analysis Based on Argument Mining.” In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, vol. 3, edited by Marie-Jeanne Lesot, Susana Vieira, Marek Z. Reformat, João Paulo Carvalho, Fernando Batista, Bernadette Bouchon-Meunier, and Ronald R. Yager, 60–69. Cham: Springer.
- Dong, Yiqi, Dongxiao He, Xiaobao Wang, Yawen Li, Xiaowen Su, Di Jin. 2023. “A Generalized Deep Markov Random Fields Framework for Fake News Detection.” In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, edited by Edith Elkind, 4758–65. International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2023/529>.
- Durden, Tyler. 2024. “US Women’s Life Expectancy Is the Lowest among Developed Nations.” ZeroHedge, September 14. <https://www.zerohedge.com/medical/us-womens-life-expectancy-lowest-among-developed-nations>. Accessed November 5, 2025.

- Fallis, Don, and Kay Mathiesen. 2019. "Fake News Is Counterfeit News." *Inquiry* 1–20. <https://doi.org/10.1080/0020174X.2019.1688179>.
- Farajtabar, Mehrdad, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias Khalil, Shuang Li, Le Song, and Hongyuan Zha. 2017. "Fake News Mitigation via Point Process Based Intervention." arXiv:1703.07823. <http://arxiv.org/abs/1703.07823>.
- Fifita, Faizi, Jordan Smith, Melissa B. Hanzsek-Brill, Xiaoyin Li, and Mengshi Zhou. 2023. "Machine Learning-Based Identifications of COVID-19 Fake News Using Biomedical Information Extraction." *Big Data and Cognitive Computing* 7 (1): 46. <https://doi.org/10.3390/bdcc7010046>.
- Frankfurt, Harry G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
- Freiberg, Nava. 2025. "Netanyahu acknowledges Israel losing online 'propaganda war,' should be doing more." *Times of Israel*, August 10. https://www.timesofisrael.com/liveblog_entry/netanyahu-acknowledges-israel-losing-online-propaganda-war-should-be-doing-more/. Accessed November 5, 2025.
- Galeotti, Elisabetta. 2019. "Believing Fake News." In *Post-Truth, Philosophy and Law*, edited by Angela Condello and Tiziano Toracca, 37–52. London: Routledge.
- Gelfert, Axel. 2018. "Fake News: A Definition." *Informal Logic* 38 (1): 84–117. <https://doi.org/10.22329/il.v38i1.5068>.
- Goldman I, Baker D (2019) Free speech, fake news, and democracy. *First Amendment* 18(1): 66–145
- Grundmann T (2023) Fake news: the case for a purely consumer-oriented explication. *Inquiry* 66(10):1758–1772, DOI 10.1080/0020174X.2020.1813195, URL <https://doi.org/10.1080/0020174X.2020.1813195>, <https://doi.org/10.1080/0020174X.2020.1813195>
- Habgood-Coote, Joshua. 2019. "Stop Talking about Fake News!" *Inquiry* 62 (9–10): 1033–65. <https://doi.org/10.1080/0020174X.2018.1508363>.
- Harris, Keith Raymond. 2022. "Real Fakes: The Epistemology of Online Misinformation." *Philosophy & Technology* 35 (3): 1–24. <https://doi.org/10.1007/s13347-022-00581-9>.
- Hassan, Aumyo, and Sarah J. Barber. 2021. "The Effects of Repetition Frequency on the Illusory Truth Effect." *Cognitive Research: Principles and Implications* 6 (38): 1–12. <https://doi.org/10.1186/s41235-021-00301-5>.
- Jain A, Kasbe A (2018) Fake news detection. In: 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp 1–5, DOI10.1109/SCEECS.2018.8546944
- Jaster, Romy, and David Lanius. 2018. "What Is Fake News?" *Versus* 2 (127): 207–27.
- Kandell J (2024) Daniel c. dennett, widely read and fiercely debated philosopher, 82, dies. URL <https://www.nytimes.com/2024/04/19/books/daniel-dennett-dead.html>
- Khattar, Dhruv, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. "MVAE: Multimodal Variational Autoencoder for Fake News Detection." In *Proceedings of the 2019 World Wide Web Conference*, 2915–21. New York: ACM.
- Konat, Barbara, Katarzyna Budzynska, and Patrick Saint-Dizier. 2016. "Rephrase in Argument Structure." In *Foundations of the Language of Argumentation: COMMA 2016 Workshop*, 32–39. Potsdam: University of Potsdam.
- Koszowy, Marcin, Steve Oswald, Katarzyna Budzynska, Barbara Konat, and Pascal Gygas. 2022. "A Pragmatic Account of Rephrase in Argumentation: Linguistic and Cognitive Evidence." *Informal Logic* 42 (1): 49–82. <https://doi.org/10.22329/il.v42i1.7212>.
- Kou Z, Shang L, Zhang Y, Yue Z, Zeng H, Wang D (2022) Crowd, expert ai: A human-ai interactive approach towards natural language explanation-based covid-19 misinformation detection. International Joint Conferences on Artificial Intelligence Organization

- Levy, Neil. 2017. “The Bad News about Fake News.” *Social Epistemology Review and Reply Collective* 6 (8): 20–36.
- Major, Mark, and David J. Andersen. 2016. “Polls and Elections: Swift Boating Reconsidered: News Coverage of Negative Presidential Ads.” *Presidential Studies Quarterly* 46 (4): 891–910. <https://doi.org/10.1111/psq.12324>.
- McIntyre, Lee C. 2018. *Post-Truth*. Cambridge, MA: MIT Press.
- Mukerji, Nikil. 2018. “What Is Fake News?” *Ergo: An Open Access Journal of Philosophy* 5: 923–946. <https://doi.org/10.3998/ergo.12405314.0005.035>.
- Musi, Elena, and Chris Reed. 2022. “From Fallacies to Semi-Fake News: Improving the Identification of Misinformation Triggers across Digital Media.” *Discourse & Society* 33 (3): 349–70. <https://doi.org/10.1177/09579265221076609>.
- Nakov P, Corney D, Hasanain M, Alam F, Elsayed T, Barr’ on-Cede~no A, Papotti P, Shaar S, Martino GDS (2021) Automated fact-checking for assisting human fact-checkers. URL <https://arxiv.org/abs/2103.07769>, 2103.07769
- “Netanyahu Declares War on Free Speech as Israel’s Propaganda Efforts Falter.” 2025. *InfoWars*, September 4. <https://www.infowars.com/posts/netanyahu-declares-war-on-free-speech-as-israels-propaganda-efforts-falter>. Accessed November 5, 2025.
- Novaes, Catarina Dutilh, and Jeroen de Ridder. 2021. “Is Fake News Old News?” In *The Epistemology of Fake News*, edited by Sven Bernecker, Amy K. Flowerree, and Thomas Grundmann, 155–78. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198863977.003.0008>.
- Oshikawa R, Qian J, Wang WY (2018) A survey on natural language processing for fake news detection. CoRR abs/1811.00770, URL <http://arxiv.org/abs/1811.00770>, 1811.00770
- Parikh, Shivam B., and Pradeep K. Atrey. 2018. “Media-Rich Fake News Detection: A Survey.” In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 436–441. <https://doi.org/10.1109/MIPR.2018.00093>.
- Pepp, Jessica, Eliot Michaelson, and Rachel K. Sterken. 2019. “What’s New about Fake News?” *Journal of Ethics and Social Philosophy* 16 (2): 67–94. <https://doi.org/10.26556/jesp.v16i2.629>.
- Perez-Rosas, Verónica, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. “Automatic Detection of Fake News.” arXiv:1708.07104. <http://arxiv.org/abs/1708.07104>.
- Qian, Feng, Chengyue Gong, Karishma Sharma, and Yan Liu. 2018. “Neural User Response Generator: Fake News Detection with Collective User Intelligence.” In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 3834–40. International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2018/533>.
- Rini, Regina. 2017. “Fake News and Partisan Epistemology.” *Kennedy Institute of Ethics Journal* 27 (S2): 43–64. <https://doi.org/10.1353/ken.2017.0025>.
- Ruchansky, Natali, Sungyong Seo, and Yan Liu. 2017. “CSI: A Hybrid Deep Model for Fake News Detection.” arXiv:1703.06959. <https://arxiv.org/abs/1703.06959>.
- Sharma, Upasna, and Jaswinder Singh. 2024. “A Comprehensive Overview of Fake News Detection on Social Networks.” *Social Network Analysis and Mining* 14 (1): 1–20. <https://doi.org/10.1007/s13278-024-01280-3>.
- Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: A data mining perspective. URL <https://arxiv.org/abs/1708.01967>, 1708.01967
- Singhal, Shivangi, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnuram Kumaraguru, and Shin’ichi Satoh. 2019. “SpotFake: A Multimodal Framework for Fake News Detection.” In

- 2019 *IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 39–47. <https://doi.org/10.1109/BigMM.2019.00-44>.
- Stewart, Elizabeth. 2021. “Detecting Fake News: Two Problems for Content Moderation.” *Philosophy & Technology* 34 (4): 923–940. <https://doi.org/10.1007/s13347-021-00442-x>.
- Tschiatschek, Sebastian, Adish Singla, Manuel Gomez-Rodriguez, Arpit Merchant, and Andreas Krause. 2017. “Fake News Detection in Social Networks via Crowd Signals.” arXiv:1711.09025. <http://arxiv.org/abs/1711.09025>.
- Uberna, Maciej, Wawer, Michał, Chudziak, Jarosław and Koszowy, Marcin. Forthcoming. “Rephrasing Agents: Theoretically Driven Multi-Agent System for Identifying Misuses of Reformulation.” Under review.
- UN. 2014. “Country Classifications.” https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf. Accessed November 5, 2025.
- Visser, Jacky, John Lawrence, and Chris Reed. 2020. “Reason-Checking Fake News.” *Communications of the ACM* 63 (11): 38–40. <https://doi.org/10.1145/3397189>.
- Visser, Jacky, Marcin Koszowy, Barbara Konat, Kasia Budzynska, and Chris Reed. 2018. “Straw Man as Misuse of Rephrase.” In *Argumentation and Inference: Proceedings of the 2nd European Conference on Argumentation*, vol. 2, edited by Steve Oswald and Didier Maillat, 941–62. London: College Publications.
- Von der Leyen, Ursula. 2024. “Special Address by Ursula von der Leyen, President of the European Commission.” Speech presented at the World Economic Forum Annual Meeting, Davos, January 16. <https://www.weforum.org/meetings/world-economic-forum-annual-meeting-2024/sessions/special-address-by-ursula-von-der-leyen-president-of-the-european-commission-96293a5a9d/>. Accessed November 5, 2025.
- Wang, Yaqing, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. “EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection.” In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 849–57. New York: ACM. <https://doi.org/10.1145/3219819.3219903>.
- WHO. 2024. “Slovakia Health Data Overview for the Slovak Republic.” World Health Organization. <https://data.who.int/countries/703>. Accessed November 5, 2025.
- Wu L, Rao Y (2020) Adaptive interaction fusion networks for fake news detection. CoRR abs/2004.10009, URL <https://arxiv.org/abs/2004.10009>, 2004.10009.
- Younis, Ramy, Daniel de Oliveira Fernandes, Pascal Gyax, Marcin Koszowy, and Steve Oswald. 2023. “Rephrasing Is Not Arguing, but It Is Still Persuasive: An Experimental Approach to Perlocutionary Effects of Rephrase.” *Journal of Pragmatics* 210: 12–23. <https://doi.org/10.1016/j.pragma.2023.03.010>.
- Zhang X, Ghorbani AA (2020) An overview of online fake news: Characterization, detection, and discussion. *Information Processing Management* 57(2):102025, DOI <https://doi.org/10.1016/j.ipm.2019.03.004>, URL <https://www.sciencedirect.com/science/article/pii/S0306457318306794>
- Zhou, Xinyi, and Reza Zafarani. 2018. “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities.” arXiv:1812.00315. <http://arxiv.org/abs/1812.00315>.
- Zhou, Xinyi, Jindi Wu, and Reza Zafarani. 2020. “SAFE: Similarity-Aware Multi-Modal Fake News Detection.” arXiv:2003.04981. <https://arxiv.org/abs/2003.04981>.

The Ethics of Responsibility in the Context of the Use of Intelligent Machines and the Problem of the Technosystem

Mariusz Wojewoda

ABSTRACT This article is devoted to the problem of responsibility as it arises in the context of the technosystem—where the latter is enhanced by the work of intelligent devices. In this case, the system in question refers to the relationship between man, functioning in the various roles of creator, trainer, owner and user of intelligent agents, and machines equipped with artificial intelligence. Suppose we assume that the aim of technological development is the well-being of the human being living now and the human being of the future. In that case, attention must be focused on the value of responsibility. The system of technology is embedded in a cultural and social context. The author diagnoses the causes of the disappearance of responsibility (the adjacency of actions) in the context of using smart devices, and considers what should be done to counteract this. The background for the analyses undertaken is furnished by ecosystem theory, together with the related concept of instrumentalization as construed by Andrew Feenberg, and Hans Jonas' ethical theory of responsibility—considered as they relate to the analysis of selected cases.

KEYWORDS adiaphorization; artificial intelligence; Andrew, Feenberg; Hans, Jonas; moral responsibility; system of technology; technical phronesis

The technosystem is based on a specific conception of life on which we cannot avoid taking a stand, whether consciously and explicitly, or passively in submission to the uncontested facts. . . . Progress is not technical or moral but technical and moral. In a society based on technical rationality the process of transcendence must itself have a rational structure—it must make sense in technical terms just as technical change must make sense in moral terms. (Feenberg 2017, 220)

INTRODUCTION: INTELLIGENT MACHINES

In today's world we are increasingly—and willingly—making use of devices equipped with artificial intelligence. This applies to various areas of human activity. The design of intelligent machines means that we obtain a brilliant technical 'partner' for human activities in business, science, medicine and institutional management. It is also an essential 'companion' in public transport, tourism and entertainment activities. Intelligent machines use artificial intelligence (AI) in their operations. However, it is not easy to define this clearly. AI is a collection of different technologies that allow data mining, exploiting the potential of deep machine learning, natural language processing, artificial neural networks, logic programming, automation of decision-making processes, and a whole range of virtual and machine elements (Konotos 2021).

In the context of AI, 'intelligence' means dealing with an algorithm created on the basis of a dataset containing exemplary behavioural models. The program independently searches for relationships between the data and proposes solution patterns and specific courses of action (Przegalińska and Oksanowicz 2023). The speed of information transfer between humans and intelligent devices, together with intelligent machines (the Internet of Things), plays an important role here. We currently use 'weak/narrow AI' to solve specific problems. Its specialized skills allow it to perform certain tasks better than humans. Work is underway to create 'strong/general AI' with multi-threaded knowledge and cognitive abilities that will enable intelligent machines to solve functions at a more advanced level than humans. The next stage in the development of AI is expected to be self-aware super-intelligence—the 'singularity'—which will be capable of making conscious choices. So far, we remain at the level of 'weak AI-ANI,' but work is underway to create 'strong AI-AGI' (Boden 2018).

The use of AI is linked to the concept of an 'expert system.' Its operation consists of emulating the decision-making process executed by a human expert in a given field. Emulation does not mean simulation

of the decision-making process, but rather indicates a 'parallel' process, so to speak, alongside that of human decision-making. Expert systems use databases to 'extend' our scope of action. They can significantly help humans optimize the decision-making process, and improve companies' level of business efficiency (Jackson 1998).

When referring to intelligent machines, we use the term 'agent.' This denotes an 'entity' operating in a specific environment, capable of communicating, checking, and reacting to environmental changes. Intelligent machines can be thought of as (1) agents that extend our ability to act, (2) agents equipped with a certain degree of autonomy of action, or (3) systems whose task is to optimize human actions.

For example, an intelligent agent can help manage information by searching for it in online resources and selecting and filtering data. An intelligent avatar can represent a user in cyberspace, facilitate online commerce, or facilitate business management. Another critical aspect of the use of AI is in medicine. Here we have in mind algorithms for collecting information on patients' health. For this purpose, we can use smartphones with applications that allow instant information on the changing state of this. 'Speed' of access to information will enable us to undertake effective treatment options. An intelligent program can help create an effective vaccine. An exciting application of AI can be seen in the use of algorithms to find potential organ donors and recipients requiring a specific tissue match (Fry 2018). This is also when the speed of access to information matters—if we are to take effective medical action promptly.

The intelligent agent is unaware of the subjective distinctiveness of its existence (Floridi and Sanders 2004). From the perspective of cognitive science, it is recognized that humans are beings that are conscious of their existence, but we ourselves do not know why we are aware of anything (Dehaene 2014). In this dreamlike scenario, it is not easy to entertain the thought that we could create an artificial consciousness. The latter could just happen by accident, uncontrollably, thanks to the actions of some programmer. This is one of the deepest human fears relating to our technological future. When we talk about humans, we link consciousness with free will. We do not equip intelligent machines with an artificial will to act. By analogy with free will, we speak of 'machine autonomy,' especially when we want to draw attention to the operation of intelligent devices with limited or no human supervision (Wojewoda 2023).

Where intelligent machines are concerned, it is necessary to distinguish intelligence from consciousness. Robots, without being conscious, can be intelligent. They can be taught to follow legal and ethical rules that forbid

endangering humans (Lin et al. 2017). However, it is challenging to expect intelligent machines to solve moral dilemmas when choosing between values of similar importance, such as life and truth. For example, a clever medical diagnosis robot can be a valuable assistant for medical services. Oriented towards the value of providing truthful information based on the patient's health data, it will inform the patient that their condition is dire, and that there is no point in further treatment. This can then trigger a radical mental crisis and suicidal thoughts on the part of the patient. However, such consequences clearly conflict with the affirmation of the value of life (Alowais et al. 2023). Doctors are expected to show sensitivity and be capable of communicating information to patients about their condition in ways that are appropriately nuanced. Currently, only humans can solve dilemmas and communicate with patients based on a specific understanding of values and the complex nature of what is involved in realizing them. Creating intelligent yet conscious machines (AGIs) will probably render them capable of solving human dilemmas. The acquisition of consciousness by AI will involve further challenges. The next step will be to assign rights to conscious machines similar to human rights—such as the right to have their existence protected by not being disconnected from a power source, or to have their mental health preserved in the context of possible software changes. It is difficult for us now to imagine the complexity of the problems that would have to be solved in this situation.

The autonomy of intelligent machines is a matter of degree. First, machines are not autonomous when it comes to their energy source—they must be powered. Second, they serve to 'complement' and 'extend' human capabilities. Third, it is the human being that determines the specific goals and how the machine will achieve them. Fourth, the human being sets the general objectives, while the machine itself selects the ways and circumstances of their implementation: so the human still controls task performance, but only at the level of overall objectives. Fifth, the machine defines its own tasks and chooses how to realize them. The human only deals with the result of the task performed (Glaser and Rossbach 2011). The autonomization of machines, to be sure, offers excellent opportunities for humans, and does not in and of itself generate any novel threat. A revolt by robots will be possible, though, once they become fully autonomous.

Machines equipped with AI are not conscious, so it is difficult to assign a moral characterization (good/bad, right/wrong) to their actions. However, decisions made by humans with significant participation from intelligent machines are ethical. It is now recognized that intelligent machines are not axiologically neutral. Things, including intelligent things, are bearers

of specific values. Indeed, the realization of values depends to a large extent on such bearers. The choice of the bearer with which a value is to be realized is something that can influence the moral decision-maker's reading of the situation. For example, using AI-equipped drones fundamentally changes the nature of warfare. It distances the user of the tool from the consequences of the action. The introduction of intelligent tools for various activities affects the assessment of the user's skill—we can perform many activities faster and more precisely. At the same time, being quicker and more efficient does not necessarily mean it is an instance of morally good use.

When using technical devices, we focus on usability and functionality in achieving our intended purpose. In assessing specific situations, other values should also be considered, such as life, one's health, or the health of others, as well as the value of the common good or the good of humanity. Values of this kind provide a meta-level for reading values such as functionality or effectiveness. Value can be understood in three ways: (1) substantively, as an entity in its own right; (2) attributively, as a feature of a thing; (3) in terms of utility, as the equivalent of something handy to us (Fleischer 2010). Values concern the relational sense of the connections between creators, technical objects and their users. From the point of view of axiological relationalism, values are an essential part of our cultural endowment; they are the backdrop for our individual and social valuations (Wojewoda 2022). Consciously accepted moral responsibility is linked to, among other things, the value of life and health, and concern for the well-being of humanity understood as concern for the future. Responsibility opens up a meta-level with respect to our appreciation of the value of life, health and the common good. For example, we link the functionality of a piece of equipment to concern for the quality of life.

The problem to be posed here is the following: in what sense does the technological mediation of action through intelligent tools distance humans from the consequences of decisions? In this article, the issue of responsibility as it pertains to intelligent machines will be analysed from the perspective of Andrew Feenberg's technosystem theory and Hans Jonas' ethics of responsibility. The idea is to diagnose the related depersonalization of decision-making and blurring of moral responsibility affecting savvy machine users.

The topic of the ethical use of intelligent machines already has a considerable literature. It is primarily concerned with developing ethical and legal principles for artificial intelligence (Lin et al. 2017; Coeckelbergh 2020). The traditional model for constructing rules is based on Isaac Asimov's

concept of ethics for robots. In his view, ethical rules form a set of rules similar to legal procedures or instructions created by a superior (manager) for a subordinate (employee). This relationship resembles the relationship between master and slave (Asimov 1950). However, this solution is inadequate. Therefore, solutions are also being formulated to teach intelligent machines human values and equip them with empathetical abilities (Minsky 2007). This idea of ethical principles for AI is complex and multi-stage. However, it requires the creation of an informed artificial intelligence—a powerful AI or superintelligence (Bostrom 2014; Floridi 2023). The present author seeks to draw attention to yet another theme, which concerns the use of intelligent machines inside organizational structures. The use of intelligent machines can help enhance the efficiency and effectiveness of an institution's operations. However, an intelligent system can also itself become a precise tool for improving an organization's level of efficiency. Andrew Feenberg's conception deserves attention in this regard, and will be discussed below.

THE TECHNOSYSTEM IN ITS SOCIAL AND CULTURAL CONTEXT

We can construe the term 'technique' (*techné*) in four ways: (1) as tools, these being ones that did not arise naturally in nature; (2) as the ability to perform certain activities using the tools just mentioned (Heidegger 1977); (3) as products of action, in the sense of technical artefacts (machines, equipment), as well as procedures and rules specifying a certain correct model of acting (Lizut 2014); and (4) as a system of technology. This last way of understanding technology points to the close relationship between man, civilization, and the products of technology (Dusek 2006). This dependence involves the environment, culture, and individual humans. It is also a two-way relationship—we are constantly changing and improving the tools we use and seeking new ways to employ them. The use of technical tools ultimately affects us: we change at the level of our functioning in the world and our social understanding of reality. By using such tools, we become participants in a technical way of life, and even define our own very identity through practices relating to their use (Feenberg 2006). We are currently experiencing such change with a particular intensity. The knowledge society needs constant and fluid access to information, and information that is appropriately selected. The products of technology and their use influence social, cultural and political change.

The issue of systems of technology was addressed in the 1960s by the French philosopher of technology Jacques Ellul. He mainly pointed out the dangers associated with the influence of technology on the human

understanding of reality. He believed the technical mindset was fundamentally directed towards calculating and achieving specific goals. The reason for introducing technical thinking into organizations was to eliminate what is accidental in human action and to anticipate future risks. What we have in mind here is not a full-scale anticipation of the future, but the preparation of procedures for dealing with complex crises. According to Ellul, the fundamental purpose of setting up a system is to speed up and optimize the decision-making process and develop mechanical habits in employees. In his view, what is human in the decision-making process is related to arbitrariness and the influence of emotional factors. Rationalizing the decision-making process means that we eliminate such subjective and arbitrary factors from human agency. Rationality is understood here as the pursuit of a technical model for exercising organizational power. For example, systemizing activities involves subjecting employees' behaviour to procedures, and a specialized model for assessing task performance.

In Ellul's view, systemic organizational management is supposed to lead to a mechanization of human behaviour that is not pleasing to employees but is helpful for the organization. However, this reveals a critical danger in the form of the adiaphorization of action and the disappearance of the sense of responsibility (Ellul 1980, 35–40). Adiaphorization (from the Ancient Greek *adiaphoron*, meaning 'morally indifferent act') consists in a subject's making a decision mediated by technical and systemic rules, treating their action as based on pragmatic utility and at the same time as morally indifferent. Sigmund Bauman analysed this issue, seeing it as an oppressive tool of the bureaucratic model of exercising power in an organization. The disappearance of any ethical consciousness associated with the work performed means that the decision is no longer analysed from a moral perspective. Instead, efficiency, functionality, and the achieving of operational goals are introduced (Bauman 1991, 440–42). These are values reflecting a qualitatively lower level of axiological engagement, and in this way employees become cogs in the organization's machine. Moreover, Hannah Arendt wrote about adiaphorization in similar terms, understanding it to be the cause of moral indifference to the mass extermination of the Jewish population. Adiaphorization causes the individual to embody a mechanical attitude, such that they follow orders/commands and are not responsible for the consequences. This involves an attitude of self-justification and moral indifference (Arendt 2003, 62–63).

An analysis of the system of technology from a social and cultural point of view has also been developed by the contemporary philosopher Andrew Feenberg. He distinguishes four main ways of understanding technology:

(1) as technical determinism, in which technological creations fundamentally influence human action and lifestyles; (2) as socio-cultural determinism, in which technology is adapted to social, political, and cultural changes; (3) in terms of interactionism, where it is assumed that technology and users constitute a specific whole within which it is difficult to determine which element is dominant; and (4) critical constructivism, in which the system entails the close interdependence of humans and technology, but is also so malleable that it can be modified (Feenberg 2009). In his later research, the American philosopher himself adopts this last position.

Feenberg recognizes that human activities, including technology-related ones, are embedded in a world of values. Some of this involves moral values, based on which we consider human behaviour right or wrong. The author of the book *Technosystem* poses an important question: does not technology, or its use, lead to the disappearance of our moral sensitivity?

Among other things, Feenberg analyses the value of the utility of technical tools. The subjective aspect of usability is based on the belief that using the device in question is correct. In contrast, the objective element of usability concerns knowledge of the device's technical capabilities. In this case, practical rationality is based on justifying for what purpose and to what extent the device should be used appropriately in a specific social, business, medical or political context. The purpose of use depends on the characteristics of the object itself, the individual beliefs of the subject, and also on culturally and institutionally approved ways of using it (Feenberg 2017).

This goal is linked to our mental map and our systemic imaginaries. Such imaginaries relate to the capabilities inscribed in the machine, the subjective skills involved in using these devices, and the social imaginary in which the creator and user of the device operate. For example, the smartphone we use is constantly equipped with new affordances. We can use them meaningfully when we have mastered specific skills, and when we function in a developed technological society where individuals and institutions have produced appropriate strategies for using such devices. Here, forms of essential and occasional use should also be considered. Understood as a system, technology is more than the sum of its constituent parts: there is, in addition, whatever quality it has, *qua* part, that results from our human relationship with *techne*-related creations. It is generally recognized that we do change under the influence of technology, and certainly we are mindful of how we use devices now and how we used them in the past.

The culture surrounding *techne*-related creations fulfils an important role: namely, it points to a system of meanings and practices of technical use developed within a community. It is, therefore, not only about technical

feasibility, but also about the social justification and psychological credibility of a particular use of *techne*-related artefacts. For example, we usually assume that we are safe using an app to carry out banking operations via a notebook or smartphone. In this case, we rely on our own ‘common sense,’ as well as a socially entrenched belief that such an action is right. Critical constructivism assumes that the use of technical tools depends on a cultural and axiological framework. The function of the tool is subject to interpretation. It depends on the technological imagination, socially established habits of behaviour, and approved action practices. Nowadays, the selection of artefacts of technology is primarily made within the framework of institutions, associated procedures of action, and specific technical disciplines.

Feenberg refers to ‘instrumentalization theory’ to clarify the question of the functionality of contemporary technological artefacts. Identifying affordances requires a decontextualization of *techne*-related objects. Here, we have two levels of reflection—the technical and the cultural. In the process of interpretation, we determine the meaning of artefacts by referring to legible values. For example, the usefulness of a thing means that we, as users, select the rules of its effectiveness. We do this because of our expectations. User preferences are embedded in an axiological background that determines in which situations a thing counts as functional, safe and sensible and under which conditions as a form of misuse—e.g., when it leads to dependence on the artefact. Values provide a cultural background giving credence to human practices. For example, when we use digital tools to work online, we recognize that this kind of work is just as productive as working in the office or at university.

In the process of interpretation, we consider the AI-equipped artefacts, their creators (in the case of AI, we are talking here about programmers and skills trainers), and the users of these tools. The meanings within these relationships are not fixed once and for all: every so often, we reinterpret our reading of the relationships between them (Feenberg 2017). In addition to the technical dimension, the design of tools also involves a recontextualization of their meaning. This process then equips the devices with new capabilities and meanings to justify their use. The next stage is to change our mental map, modify our conceptual models, and create imaginary systems for the effective and meaningful use of our digital tools (Norman 2023). Behind all these technical proposals lies a set of background axiological commitments, which must often be articulated. This implies, among other things, a focus on improving the speed of access to information, communicative convenience, and a concern to be in control of one’s health. It is not just a matter of suggesting new possibilities related to the use of digital

devices, but also of overcoming individual and social fears, changing personal habits, modifying the social and mental map, and creating a socially approved view that these devices are helpful and at the same time safe—that is, that our sensitive data is adequately protected.

Ways of using tools can result from imitation and the social compulsion to participate in certain practices. The relationship between the social and the technological compulsion to use tools is interesting. On the one hand, the range of professional, institutional activities that an individual can (and is now supposed to) carry out using digital tools is constantly expanding. On the other hand, the spread of these tools may widen the scope of exclusion, pushing large groups of people to the margins of social and professional life (Suchacka et al. 2021). Therefore, it is essential to design responsibly, equip devices with clear markers, and define understandable ways of using technical devices. What we have in mind here is a move away from a model based on compulsion, and toward creating positive habits of use. We should maintain a critical attitude towards technopression—the compulsion to use digital devices. Technological culture should lead to the development of human beings and their social and creative skills, and not to their objectification (Krzykowski 2023).

TECHNE AND MORAL RESPONSIBILITY

Following Hans Jonas, we shall distinguish between moral responsibility and legal liability. The latter concerns compliance with rules laid down by a state, or by international law, while the former, in its legal sense, addresses the negative consequences of a subject's conduct as these relate to some negative state of affairs resulting from their freely chosen actions. Legal liability concerns acts whose significance is determined post-factum—as well as, more generally, ones that have happened with specific consequences (Peno 2015). In this article, we shall focus on moral responsibility. Jonas (1984) highlighted several factors to consider when discussing moral responsibility as it relates to the use of traditional technical tools dating from the second industrial revolution. We, on the other hand, are now at the stage of the fourth and projected fifth industrial revolution (the use AI). In modernizing Jonas' conception, we will relate it to the issue of the use of intelligent machines. It consists of three essential elements.

1) *Reflecting on the future*. Here we assume that we are dealing with probabilities of future events. Analysing scenarios pertaining to such possible courses of events prepares us for their undesirable consequences. As long as the danger is unknown, we do not know how to protect ourselves—in this case, we identify the sources of risk and diagnose the causes

of user misbehaviour (Jonas, 1984). Describing and explaining the causes of wrongdoing is supposed to make us work out scenarios for counteracting the negative consequences of activities related to the use of intelligent machines.

2) *Asymmetry of action*. In defining the rules of moral responsibility, we cannot rely on the law of reciprocity. Reciprocity means that we expect behaviour of a similar kind to our own, whereas the intelligent machine is assumed not to be a conscious subject but rather an agent whose assistance we use. The intelligent agent is an integral part of the expert team, because the information it provides, or its skills, influence the decision-making process and the action strategy adopted. The asymmetry is that we do not expect the intelligent machine to behave responsibly. It is a valuable tool for acquiring and collecting data, but it is the human being who designs, trains and supervises the intelligent machine that is responsible for its actions, even when using AI.

3) *Causal power*. Moral responsibility is linked to the category of causal power. In using the term ‘power,’ we are describing the impact of an individual or organization on the external environment and the lives of others. Being responsible means that the subject can be considered a conscious executor of their action: in other words, there is a causal relationship between the action and its consequences. By introducing intelligent machines, we increase the power of our impact on the environment. Still, paradoxically, we separate ourselves from the consequences of this action significantly when we systematically increase the autonomy of intelligent machines. This separation makes us systematically lose the ability to analyse the action in moral terms (Jonas 1984).

A good example of the adiaphorization phenomenon—the suspension of responsibility in the context of the use of technical tools equipped with artificial intelligence—is the Lavender program created by the Israeli army to eliminate Palestinian Hamas or Islamic Jihad fighters (Brigadier General YS 2021). The Israeli military used AI to mark Gaza zones as suspected of carrying out attacks against Israelis. This included actual militants, but also those individuals who were Hamas or Islamic Jihad sympathizers. Human oversight of combat drone operations was reduced to a minimum. The machine was given significant autonomy regarding who to eliminate and what places to bomb. As a result of the indiscriminate bombing, many innocent people were killed who were in the firing zone of the bombs being dropped. The human supervisor of the Lavender system was limited to determining the algorithm of people to eliminate, while agreeing to kill random victims of the attack. The Israeli army decided to use the intelligent system

even though it made the wrong decisions in about 10 percent of cases. Sometimes, it flagged as dangerous individuals people who had only a loose connection to militant groups. The operation of the Lavender intelligence program is an example of the systematic narrowing of the field of human responsibility through smart devices. Transferring this responsibility to algorithms means there is no connection between the person making the decision and the consequences of his/her action (Yuval 2024).

What gives moral responsibility its specificity is that it does not seek to analyse the consequences themselves, but rather to consider the overall well-being of the persons associated with the action taken. Practical rationality involves considering what needs to be done to avoid moral evil, or how to restore the good that has been violated. We are concerned with identifying irresponsible behaviour in advance, as it were, before it is actualized. From this perspective, our aim is to consider possible consequences, not those already transpiring. Irresponsibility is an omission or failure to undertake certain activities: one from which negative consequences may arise. We do not expect moral reflection in this sense from intelligent machines. It is a peculiarity of human thinking and human ethical sensitivity. In a root sense, moral responsibility stems from the awareness that human well-being may be endangered by my actions or my inaction. The discovery of the causes of this threat makes it necessary to engage in activities to counteract it. We are not yet in a position to algorithmize this kind of awareness and transfer it to machine activity.

CONCLUSION

Technological mediation makes it increasingly difficult for us to identify the link between a conscious choice, the resulting action, and the consequences of that action. Autonomous devices distance the user from the consequences of their actions primarily when the machine determines the specific goal and the means of achieving it. It is essential to recognize that no command can enforce moral responsibility where the latter is based on voluntarily accepting a commitment to responsibility. While it is true that we are accountable for someone or something to someone or an institution, the key is the conscious acceptance of a commitment to a particular action. Moral responsibility is about acting freely and consciously. We do not expect this kind of action from animals or intelligent machines. When we think of responsibility, we consider the subject's responsibility for the intentions and consequences of their actions. With this in mind, Feenberg invokes the concept of *phronesis* from Aristotle—practical reason associated with the subject's trained ability to make accurate moral choices (Feenberg 2017).

Currently, it seems that the exercising of such a skill only takes place in a cultural and social vacuum. We need a 'frenetic background' embedded in the ecosystem, in which the creation of functional and valuable machines is linked to responsibility for their use. A background of this sort often goes unarticulated, but this is surely a condition for any rational debate over the extent of individual and social responsibility for the rules governing the development of smart devices. In this sense, we need an institution with an appropriate organizational culture, in which the laws of responsible behaviour are followed out of habit—technical phronesis.

When analysing the issue of responsibility in the context of the creation and use of intelligent devices, we are talking about the responsibility of programmers, trainers, robot builders, owners and users. The design and use of intelligent machines should be linked to responsibility for our quality of life and concern for future generations. Referring back to Jonas' thought, when increasing the power of our actions through the design and use of intelligent artefacts we cannot overlook the issue of responsibility for the possible consequences of our actions. On the other hand, Feenberg emphasizes that the technosystem's coherence not only refers to the utility of things but also has a normative significance: i.e., it speaks about what the system should be like.

In addition to practical values, a human-friendly technical system must consider values at the meta-level, such as concern for the quality of life, and responsibility for the well-being, of both those currently living and future generations. In a moral sense, practical reason must be employed not only to seek new solutions but also to define the limits of a model of technological development that otherwise would be undesirable for humanity. The social acceptance of intelligent machines should not lead its designers and users to lose sight of critical thinking and the ability to make independent moral judgments. If AI is an ethical tool, it is because its programmers, trainers and users are honest. If its development is done without analysing moral issues, it could lead to serious social problems.

The progress associated with the development of AI is exponential, so it is difficult to predict the future course of discoveries, or new applications of artificial intelligence tools. An important issue is the development of ethical rules for AI. Currently, we think of these rules in terms of task regulations. However, this may not be enough: we associate moral responsibility with consciousness, and research into creating conscious AI is undergoing intensive development. AGI—a powerful form of AI/superintelligence—may be achieved soon, and then the task of artificial intelligence developers and trainers will be to create subjective rules of moral action for it.

REFERENCES

- Alowais, Shuroug A., Sahar S. Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I. Alshaya, Sumaya N. Almohareb, Atheer Aldairem, Mohammed Alrashed, Khalid Bin Saleh, Hisham A. Badreldin, Majed S. Al Yami, Shmeylan Al Harbi, and Abdulkareem M. Albekairy. 2023. "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice." *BMC Medical Education* 23 (689): 1–15. <https://doi.org/10.1186/s12909-023-04698-z>.
- Arendt, Hannah. 2003. *Responsibility and Judgment*. Edited and with an introduction by Jerome Kohn. New York: Schocken Books.
- Asimov, Isaac. 1950. *I, Robot*. New York: Doubleday.
- Bauman, Zygmunt. 1991. *Modernity and the Holocaust*. Cambridge: Polity Press.
- Boden, Margaret A. 2018. *Artificial Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.
- Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Brigadier General YS. 2021. *The Human-Machine Team: How to Create Synergy between Human and Artificial Intelligence That Will Revolutionize Our World*. Self-published.
- Coeckelbergh, Mark. 2020. *AI Ethics*. Cambridge, MA: MIT Press.
- Dehaene, Stanislas. 2014. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. London: Penguin Books.
- Dusek, Val. 2006. *Philosophy of Technology: An Introduction*. Oxford: Blackwell Publishing.
- Ellul, Jacques. 1980. *The Technological System*. Translated by Joachim Neugroschel. New York: Continuum Publishing.
- Feenberg, Andrew. 2006. "What Is Philosophy of Technology?" In *Defining Technological Literacy*, edited by John R. Dakers, 5–16. New York: Palgrave Macmillan.
- Feenberg, Andrew. 2009. "Critical Theory of Technology." In *A Companion to the Philosophy of Technology*, edited by Jan Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks, 146–53. Oxford: Wiley-Blackwell.
- Feenberg, Andrew. 2017. *Technosystem: The Social Life of Reason*. Cambridge, MA: Harvard University Press.
- Fleischer, Michael. 2010. "Wartości w wymiarze komunikacyjnym." In *Teorie komunikacji i mediów*, vol. 2, edited by Marek Graszewicz and Jerzy Jastrzębski, 9–54. Wrocław: Oficyna Wydawnicza ATUT.
- Floridi, Luciano, and J.W. Sanders. 2004. "On the Morality of Artificial Agents." *Minds and Machines* 14 (3): 349–79. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>.
- Floridi, Luciano. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. New York: Oxford University Press.
- Fry, Hannah. 2018. *Hello World: How to Be Human in the Age of the Machine*. London: Doubleday.
- Glaser, Horst Albert, and Sabine Roszbach. 2011. *The Artificial Human*. Frankfurt: Peter Lang.
- Heidegger, Martin. 1977. *The Question Concerning Technology, and Other Essays*. Translated by William Lovitt. New York: Garland Publishing.
- Jackson, Peter. 1998. *Introduction to Expert Systems*. 3rd ed. Boston: Addison-Wesley.
- Jonas, Hans. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press.
- Konotos, John. 2021. "Artificial Intelligence, Machine Consciousness and Explanation." *Academia Letters*, Article 1709. <https://doi.org/10.20935/AL1709>.

- Krzykawski, Michał K. 2023. "Krytyka w technopresji." *Śląskie Studia Polonistyczne* 20 (2): 1–28. <https://doi.org/10.31261/SSP.2022.20.09>.
- Lin, Patrick, Keith Abney, and Ryan Jenkins, eds. 2017. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. New York: Oxford University Press.
- Lizut, Rafał. 2014. *Technika a wartości: Spór o aksjologiczną neutralność artefaktów*. Lublin: Wydawnictwo Academicon.
- Minsky, Marvin. 2007. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster.
- Norman, Donald A. 2023. *Design for a Better World*. Cambridge, MA: MIT Press.
- Peno, Michał. 2015. "Prawna odpowiedzialność z tytułu pełnienia roli społecznej." *Acta Universitatis Lodziensis. Folia Iuridica* 74: 42–57. <https://doi.org/10.18778/0208-6069.74.04>.
- Przegalińska, Aleksandra, and Paweł Oksanowicz. 2023. *Sztuczna inteligencja: Nieludzka, arcyłudzka. Fenomen świata nowych technologii*. Kraków: Znak.
- Suchacka, Małgorzata, Rafał Muster, and Mariusz Wojewoda. 2021. "Human and Machine Creativity: Social and Ethical Aspects of the Development of Artificial Intelligence." *Creativity Studies* 14 (2): 430–43. <https://doi.org/10.3846/cs.2021.14316>.
- Wojewoda, Mariusz. 2022. "Autointerpretacja osoby i problem 'mocnego wartościowania' w filozofii Charlesa Taylora." *Analiza i Egzystencja* 57: 71–90. <https://doi.org/10.18276/aie.2022.57-04>.
- Wojewoda, Mariusz. 2023. "Artificial Intelligence as a Social Utopia." *Ethos* 36 (3): 109–26. <https://doi.org/10.12887/36-2023-3-143-08>.
- Yuval, Abraham. 2024. "System of Lavender." +972 Magazine, April 3. <https://www.972mag.com/lavender-ai-israeli-army-gaza>. Accessed October 17, 2025.

Mindful Decentering, and Attention as Selection for Action

Piotr Sikora

ABSTRACT This article examines the compatibility between one of the central phenomena discussed in the literature on the theory and practice of secular mindfulness, decentering, and one of the most influential contemporary philosophical accounts of attention—namely, Wayne Wu’s theory of attention as selection for action. I begin by presenting and critically examining Victor Lange’s recent attempt to show that decentering constitutes a counterexample to Wu’s account. I then argue that Lange’s conception of decentering is inadequate, and propose an alternative understanding according to which decentering indeed serves as a counterexample to the view of attention as selection for action. Finally, I outline possible directions for further philosophical research on attention that accommodate the findings of secular mindfulness, particularly those concerning decentering.

KEYWORDS attention; decentering; defusion; Lange, Victor; mindfulness; Wu, Wayne.

INTRODUCTION

Across a growing body of literature concerned with the practice of mindfulness¹ and its potential benefits for mental health and well-being, one can encounter the widely shared idea that the central aspect of this practice is a change in our attitude towards our thoughts and other mental states. This phenomenon is called “decentering” (see Segal, Williams and Teasdale 2012), “reperceiving” (Shapiro et al. 2006), or “defusion” (Hayes, Strosahl and Wilson 2011).²

For Segal, Williams and Teasdale (2013, 91), to decenter “means to relate to thoughts, feelings, body sensations, and impulses to act as events passing in the mind and body, rather than identifying with them.” In the decentered perspective, “thoughts and feelings are seen as simply passing events in the mind that arise, become objects of awareness, and then pass away” (Segal, Williams and Teasdale 2013, 73). They draw on the idea of Jon Kabat-Zinn, who, inspired by Buddhist ideas and practices, has developed mindfulness exercises in which “we intentionally practice letting go of each thought that attracts our attention” and “just observe them as thoughts, as discrete events that appear in the field of our awareness . . . [and—P.S.] we intentionally decline getting caught up in the content [of them—P.S.]” (Kabat-Zinn 1990, 68). According to Shapiro and colleagues, who also draw on Kabat-Zinn’s idea, having attained the ability to reperceive, “rather than being immersed in the drama of our personal narrative or life story, we are able to stand back and simply witness it” (Shapiro et al. 2006, 377).

On the other hand, another crucial notion used in accounts of the mindfulness practice that leads to decentering is the notion of attention. The most

1. I refer here, and throughout the paper, to the secular theory and practice of mindfulness. While secular mindfulness is deeply inspired by Buddhist ideas and practices, there is considerable debate over the extent to which the former remains faithful to the latter—not to mention the internal Buddhist debates about how to understand and practice Buddhist meditation (Arbel 2016; Bodhi 2011; Dreyfus 2011; Dunne 2011; Fernell and Segal 2011; Gethin 2011; Grossman and Van Dam 2011; Nyanaponika 1968; Olendzki 2011; Polak 2024). Entering into these debates not only lies beyond the scope of my paper, but is also unnecessary for my purposes, which are to identify a particular mental phenomenon and highlight its relationship to a particular theory of attention.

2. Below, I will treat these terms, borrowed from the literature cited, as synonyms. Engaging with those who interpret them as referring to slightly different mental phenomena lies beyond the scope of the present paper. What is important for my purposes is, once again, to identify a particular mental phenomenon that serves as a counterexample to a specific theory of attention. I take this phenomenon to be the one referred to by all three terms—decentering, reperceiving, and defusion—yet my argument does not depend on the truth of this latter claim. It would remain valid even if only one of these terms referred to a phenomenon with the characteristics I outline in the final part of the paper.

influential definition of this practice, formulated by Kabat-Zinn, states that mindfulness is defined as “paying attention in a particular way: on purpose, in the present moment, and non-judgmentally” (Kabat-Zinn 1994, 4). This is why it is important to analyze how the phenomenon of decentering relates to philosophical accounts of attention. In this context, it is noteworthy that Victor Lange (2025) argues that decentering can be viewed as a counterexample to all those contemporary philosophical theories of attention that assume that “attending to a phenomenon implies selecting the phenomenon for further processing” (Lange 2025, 1552), and, in particular, to one of the most prominent among them—namely, Wayne Wu’s theory of attention as selection for action (hereafter SfA).

Lange is not alone in his attempt: Wu’s account is hotly disputed, and several authors have identified phenomena—such as attentional capture or automatic actions—that can be regarded as counterexamples to SfA (see, e.g., Buehler 2019, 2022; Watzl 2017, 2023). Nevertheless, Wu has responded to these critics, seeking to show that the phenomena pointed out by them do not, in fact, undermine SfA. His responses are not decisive but, I think, do possess a certain plausibility, particularly in light of the later development of his theory (Wu 2019, 2023b, 2025). Decentering, however, to which Lange points, is not only much more thoroughly developed in the literature than the rather sketchily described other cases, but has also not yet been considered in the ongoing philosophical debate as a counterexample to SfA. For this reason, it deserves particular attention.

I think that Lange highlights the right problem and comes to the right conclusion: decentering does, in fact, pose a serious threat to SfA. His reasoning is, however, flawed: if decentering were what he takes it to be, it would not constitute a counterexample to SfA. In this paper, I first sketch Lange’s argument and then show how Wu might respond to his reasoning. Finally, I present another account of decentering—one that is, in my opinion, much more faithful to what one encounters in the literature on mindfulness and experiences in its actual practice³—and show why it challenges SfA.

3. See the caveat offered in footnote 1 above.

I

Lange (2025, 1532–38) defines decentering as follows:

Undergoing a mental state, M, an agent, A, decenters from M if and only if A simultaneously

- (a) introspectively attends to M (sub-operation 1)
- (b) detaches from M (sub-operation 2). (Lange 2025, 1532)

According to him, decentering is to be deployed only in pathological mental states,⁴ and is constituted by two simultaneous sub-operations, the first of which consists in the intentional focusing of attention on the relevant mental state, thereby increasing the subject's awareness of that state, and the second of which results in an agent downregulating the influence the state has on her psychological processing. The second sub-operation, detachment, has two aspects. Lange, drawing here on Bernstein's model (Bernstein et al. 2015, 2019), describes them as disidentification and non-reactivity: "The process of disidentification involves an agent relating to the occurring mental state as external to herself and not as an expression of herself," while non-reactivity "means that a subject does not act bodily on the basis of the relevant state" and "avoids any mental interference with the state."

Lange understands disidentification from one's mental states as a subjective, practical stance toward those states. When a subject identifies with her mental state, "she views the state as a meaningful expression of herself," taking the content of the mental state seriously, "which makes it more prominent in governing her mind and associated actions." By contrast, when a subject disidentifies with a given mental state, "she (probably) does not view the state as meaningful or worth taking seriously," and "such subjective externalization (most likely) decreases a state's influence on one's mind and associated actions, compared to subjectively identifying with the state." Disidentification leads to non-reactivity.

4. Given the Buddhist roots of secular mindfulness, it is worth making clear that Lange uses the term "pathological" in the contemporary Western, discriminative sense, in which only certain ordinary mental states are so classified—not in the Buddhist sense, according to which all mental states fall within the realm of *samsara*, meaning existence permeated by suffering (*pathos*). Later in this paper I show that, on this point (as on several others), Lange misreads the idea of decentering as it has been developed in the mindfulness literature. Be that as it may, he states plainly that "disidentification and non-reactivity are ways of relating to the *pathological mental state*. They are not ways of relating to any other mental state" (2025, 1538; italics in the original).

Lange criticizes the view—which he attributes to Puc (2019)—that bare attention suffices for decentering,⁵ and argues for the claim that decentering should be understood as including the two conditions sketched above: i.e. as “something more than merely introspecting one’s mental states.” He holds that introspective attention and detachment may come apart; this is why he conceives decentering as “a complex operation where multiple, inter-related sub-operations work together,” and not as “exhaustively a matter of an epistemic change in how one introspects.”

Another crucial feature of decentering is, according to Lange, the directness of the detachment. In cases in which detachment is indirect, “an agent detaches from a mental state, *M*, by upregulating another mental state, *M**, such that her processing is determined by *M** with the indirect effect or by-product that she detaches from *M*.” But in decentering, Lange insists, detachment is direct: “in direct detachment, the downregulation of a mental state is not a side-effect of the upregulation of another mental state. The detachment is itself the instant operation performed by the agent.” Lange holds that “describing the detachment of decentering as direct is the best way to make sense of the instructions and the phenomenology of decentering.” Subjects who are taught how to decenter “receive no instructions in upregulating other . . . mental states. Neither does decentering involve the experience of upregulating another mental state. It involves the experience that the pathological state remains present but that there is an increased distance to it.”

II

Decentering, understood in the terms sketched above, constitutes, Lange holds, a counterexample to Wayne Wu’s theory of attention as selection for action (SfA).

Lange interprets Wu’s SfA theory as being based on two fundamental theses:

SfA – attention. An agent, *A*, attends to an input, *P*, if and only if *A* selects *P* for action.

SfA – attention control. An agent, *A*, executes attention control if and only if *A*’s relevant goals or intentions structure a selection of an input for action.
(Lange 2025, 1540)

5. Later in the paper, I argue that, contrary to Lange’s view, bare attention suffices for decentering, albeit in a slightly different sense from that proposed by Puc. Engaging with Puc’s position on this issue, however, lies beyond the scope of the present paper. I leave this question to future research, the direction of which I outline in the paper’s conclusion.

For the line of argumentation developed by Lange (2025, 1540–44), it is crucial that attention control be understood as “restricted to the structuring of the selection of an input for action, not the direct execution of the action.” Having interpreted SfA in this way, he points out that decentering involves attention but, at the same time, “this type of attention does not seem to be a matter of selecting the pathological state for an action.” By contrast, it is “a matter of attending to a state with the purpose of it not being selected for an action”: i.e. it amounts to selecting a mental state for non-action—which contradicts the *SfA-attention* thesis. A similar problem arises with respect to the second central thesis of SfA, namely *SfA-attention control*. In Lange’s view, “(1) the detachment of decentering is intentional and direct, and (2) the detachment is an execution of attention control.” This is why, “as an act of attention control, decentering is the operation of avoiding that a state produces or influences action”: that is, the subject’s intentions structure the selection of an input, but it is selected not for action—contrary to what *SfA-attention control* states.

Lange considers several possible replies on the part of the defender of SfA. The first one focuses on sub-operation 1, and consists in claiming that introspective attention is the selection of a given mental state (an input) for the action of increasing its conscious representation or salience. To this, he answers that “some philosophers have argued that it is hard to make sense of an increase in conscious representation or salience as a form of action.” His main argument tackles, however, sub-operation 2. He argues that defenders of SfA should show “that we can adequately describe sub-operation 2 as an agent’s intentions or goals structuring the selection of the pathological state for some action,” and claims that they may try to do this in three ways—but none of those ways are successful.

One way is to claim that “sub-operation 2 involves the selection of the pathological state for the action of turning it off.” Lange—silently assuming that turning off the mental state amounts to the inhibition of its neural correlate—answers that this proposal is hardly reconcilable with the theory that inhibition at the neural level is never direct but always involves giving priority to other neural mappings over that which is to be inhibited. Given this fact, if sub-operation 2 amounted to turning off the pathological mental state a subject wishes to detach from, the detachment in this case would be of an indirect character. But even if this difficulty could be resolved, the defender of SfA faces another problem. According to Wu—as interpreted by Lange—“an input is selected for action only if it implies that the input produces further effects in processing,” whereas selection for turning off “involves that a state is selected for not producing further effects.”

Another possible way of defending SfA considered by Lange is the claim that sub-operation 2 may be understood as “the structured selection of the pathological thought for the action of meta-representing it.” He answers that “such meta-representation would plausibly not be sufficient for the detachment involved in sub-operation 2.” According to him, it is possible for a subject to meta-represent a mental state as “external to me/not an expression of me” or “not to be reacted upon,” and still fail to execute the intended psychological control over that state—for example, to act upon a state even when it is represented as “a state not to be reacted upon.” From this perspective, decentering involves not only meta-representing but also some form of psychological control. He also rejects a further possible reply, according to which “the meta-representation involved in sub-operation 2 is of a non-propositional format such that if a mental state is encoded in this format, it ensures detachment.” He is skeptical about the non-propositional character of meta-representation and, further, claims that even such non-propositional meta-representation cannot help here, for the same reasons that undermine the third possible SfA defender’s reply—namely, that sub-operation 2 of decentering involves selecting a pathological mental state for a stance of disidentification and non-reactivity.

Lange (2025, 1545–46) regards stance as a form of activity, and agrees that before a subject establishes the intended stance toward the pathological state, she must select the relevant state for this stance. Nevertheless, he holds that after the maintenance of the stance is established, the subject’s “mental operation is not primarily a matter of selection of the state for this stance,” but is “primarily a matter of her executing, or we might say acting, the stance.” He concludes that “it is by actively executing the stances that the subjects manipulate the grip the pathological state has in their processing and downregulates it. This is how sub-operation 2 involves phases that are not adequately described as a process of selection of pathological states for stances, but direct executions of the stances.” Lange rejects a possible reply on the part of the defender of SfA based on the analogy between maintaining the stance toward a mental state and keeping the body frozen in a “freezing dance.” He claims that even if keeping the body “frozen” may be explained as the selection of proprioceptive inputs for some action—because preventing the body from moving may demand active engagement of the subject (e.g., actively rejecting impulses to move)—this is not an adequate analogue for decentering. This analogy suggests that decentering can be conceived as freezing one’s mental life. Lange remarks, however, that the analogy is flawed because “decentering is not about freezing or keeping your pathological state constant. It is about continuously downregulating

it. Upon successful decentering, the pathological state disappears over time and there is no need to introspect or detach from it.”

III

In the present section, I will argue that decentering, as Lange understands it, does not in fact constitute a counterexample to SfA—especially as the concept is developed in Wu’s more recent works. Lange’s argument amounts to the claim that decentering is a case of *attention without action*. Wu responds to such an argument by claiming that:

Once it is clear what acting comes to, namely input-output coupling in an action space, then it seems that guidance is present in every movement, even the subtlest movements of the mind. . . . In general, my response to putative counterexamples that attempt to show attention without agency is to ask whether the form of attention at issue can be part of intentional action, and then to uncover the structure of such expressions of intentional agency, revealing its component action-relevant capacities. Putative counterexamples will then be revealed to have the relevant structure, often unnoticed when we focus on the automatic forms but brought to light once we recognize that action has its own internal architecture. (Wu 2023b, 82)

When considering the (in)compatibility of SfA and decentering, it is crucial to understand Wu’s view on both action and selection. The basic structure of action is connected to the fact that action is possible in situations in which a subject faces multiple possible behavioral paths. Action differs from mere reflexes in that reflexes occur when there are “no additional behavioral paths beyond the one path taken (this includes the path of not acting). Thus, the behavior space consists of a simple one-one mapping from target to response.” This passage indicates that, according to Wu, such a plurality of options includes situations in which the only possible alternative is between acting in a specific way or not acting at all. When one has to choose between acting or not acting in response to X, one selects X for action: “for agency to be possible, there must be behavioral options, even if it is just the option of not acting” (Wu 2014, 89–90).

While explaining the notion of behavioral space, Wu uses the notions of input and output. The behavior space “is constituted (1) by inputs, which are the agent’s psychological states at that time, such as her seeing, feeling, remembering, entertaining, and so on, and (2) by outputs, which can be further psychological states.” In the behavior space, there are many possible paths from inputs to outputs, so when a subject acts, “one of the potential

paths is actualized,” and her action is “a specific input guiding a response, given what the agent intends” The course of action is set by the intention—“the agent responds in light of how she *takes things*” (Wu 2023a, 62–65).

In the case of mental action—what Wu calls “movement of the mind”—such traversal along a particular path in the behavior space, from an input mental state to an output mental state, “is best depicted in terms of intentional content.” During the course of mental action, such content is transformed over time (Wu 2023a, 70–71). If the change in content is small, a movement of the mind may be considered “short.” Wu regards covert perceptual attention as “the shortest movement,” in which “the input state is modified in the output state.” Wu leaves open the question of “how one precisely draws a distinction between states that change and states that induce distinct states”; what matters for him is “changes in intentional content as a way to track the progress of a mental action” (Wu 2023a, 71). In the case of movement of the mind, “there can be a small distance between input and output, e.g., while the input might be the flashing of a specific visual image, the output would be the maintaining of that very image. Input and output are nearly identical, the latter simply involving a response to the image” (Wu 2014, 93). In other words, even if there is no change in intentional content, but there is some subjective response to the input state—such as maintaining that mental state—attention is still present, even in “the limiting case of maintaining attention on a target.” It can be counted as mental action because the subject, by maintaining a particular mental state, intentionally remains on a particular path in the behavior space. According to Wu, it is this case—not decentering—that can be considered analogous to keeping one’s body “frozen” (Wu 2023a, 75).

The question arises whether decentering, as Lange describes it, should be regarded as a mental action in the sense explained by Wu. Decentering, Lange holds, involves a particular intention: a subject identifies a particular mental state as a pathological one and intends to downregulate its influence on her processing—in other words, she intends to respond in some way to that state. Some initial taking up of mental states sets the introspective attention: the subject focuses on a pathological mental state as a particular input chosen from among many possible inputs. This is sub-operation 1, which, according to Lange, increases the subject’s awareness of that mental state. The subject then intentionally—in the light of how she takes things—responds to that input by performing sub-operation 2. In this stage, she detaches from that state; to put it in Lange’s words, she performs an act of “not taking it seriously,” interpreting the relevant state as “not an expression of herself” or “not meaningful,” and then actively executes the stance

of non-reactivity toward the relevant state. All of Lange's descriptions of what occurs during both sub-operations constitutive of decentering indicate that it involves a change in intentional content sufficient, in Wu's view, for detecting a movement of the mind. Moreover, as Lange explicitly states, this movement is brought about intentionally by the subject, as an actualization of one among many possible paths in the subject's behavior space. This becomes even clearer when one considers what Lange insists on—namely, the directness of the downregulation of the relevant mental state.

Lange claims that detachment, being the direct execution of a stance toward a pathological mental state, excludes the possibility of understanding decentering as the selection of the state for this stance. The question, however, is what the role of sub-operation 1 is in decentering. In Lange's view it is a necessary component of decentering. If so, then decentering involves the selection of the pathological mental state—attention to it (sub-operation 1)—for the action, i.e., execution of the particular stance toward that state (sub-operation 2). Lange (2025, 1551–52) insists that SfA (like any other theory of attention) must take into account the two-dimensional, complex nature of decentering, which “involves agents intentionally manipulating two causal relations of the same mental state at the same point in time”: i.e. “the causal relation of increasing conscious representation of the state and the causal relation of the state's determination of further processing.” I think, however, that this conception of “decentering as a multi-dynamical unit of attention and attention control, in which agents relate to the same mental state in two different ways” poses no problem for Wu's theory.

To see this, it is sufficient to notice that Wu distinguishes attention as (mental) action (*attending*) and attention in action (*attention*) (Wu 2023a, 75; Wu 2023b, 63). When we are considering attending as mental action “it will have the complex structure that is revealed by reflection on the behavior space and the non-deliberative Many-Many Problem” (Wu 2023a, 75), meaning that “attending as action involves a coupling of input to output” within the action space (Wu 2023b, 62). Wu acknowledges, however, that we frequently “isolate attention in the context of action, as when we speak of doing things in a way that depends on attention,” and in that case, when we treat attention as a component of action, “we need only have in view the input that will inform action” (Wu 2023a, 75). With the above distinction in place, one can easily explain decentering (as Lange presents it) in SfA categories. The whole process of decentering has the structure not of a reflex, but of an action—it is a particular path, intentionally taken by the subject, within a behavioral space constituted by many possibilities. Even if it may be disputable whether the whole process can be counted as attending as action, it—as Lange himself

insists—includes as its necessary component sub-operation 1, attention to the pathological state, where this guides sub-operation 2, detachment from that state (explained by Lange as downregulating the influence that this state has on the subject's further mental processing).

The account of decentering (as Lange describes it) in SfA terms will be even clearer if one includes Wu's development of the crucial notion of selection. Wu acknowledges that his slogan "attention is selection for action" has come in for critique (Watzl 2017) on the basis of the observation that its grammar suggests accomplishment at a particular point in time ("I selected this mental state for action at noon"), whereas attention is instead a continuous process ("I've been attending to this mental state since noon") (Wu 2023a, 66). Wu responds to this critique by acknowledging the ambiguity of the slogan and explaining that the notion of selection should not be understood as implying such a non-processual nature of attention. He agrees that attention is a process; this is why he considers the very term "attention" to be unfortunate, since it suggests a state. Therefore, he prefers to use the term "attending" (Wu 2023b, 63–64). This is why, in his later writings, he uses the term "guidance" instead: "Attention is mental guidance in action, the agent's taking things informing response" (Wu 2023b, 65). Or, to put it in more elaborate terms:

To speak of attention as selection for action is to speak of a way that the subject is attuned during action to relevant information such that it is deployed to inform the subject's response . . . action is constituted by a response guided by the agent's attunement to certain features of the world, including features of the subject him- or herself. There are, then, two necessary "aspects" of attention so conceived: (1) the attunement ("selection"), and (2) the link between the response and that to which the subject is attuned ("for action"). (Wu 2023a, 67)

In light of the above explanation, one may account for Lange's decentering by saying that sub-operation 1 amounts to attunement to the relevant mental state, whereas sub-operation 2 constitutes the link between the response (detachment) and the mental state to which the subject is attuned.

If the line of reasoning offered here is sound, then Lange's rejoinders to the possible SfA defender's replies are questionable. His rejoinder to the idea that sub-operation 2 of decentering is the selection of a pathological mental state for a stance of disidentification and non-reactivity is based on an understanding of selection as taking place at a particular point in time. It is only such an understanding that enables one to say that a subject attends

to the pathological state towards which the stance of disidentification and non-reactivity is to be adopted just before establishing the intended stance toward that state, but not while maintaining the stance. If one understands selection as a continuous process of guidance, one should acknowledge that this process of selection (or, better, selecting) constantly guides the active execution of the stance. Similarly, with regard to the idea that decentering may be conceptualized as the selection of a pathological mental state for turning off, it can be said that Lange (2025, 1544) wrongly perceives a discrepancy between the idea that “an input is selected for action only if it implies that the input produces further effects in processing” and the claim that selection for turning off “involves that a state is selected for *not* producing further effects.” The further effects the selected input produces in processing are changes to the input itself: i.e. changes in the intentional content of the relevant mental state—changes brought about during action (movement of the mind) guided by the input selected for that action.

IV

Lange’s failure to show that decentering challenges SfA does not mean that there is no discrepancy between decentering and Wu’s account of attention. This is because Lange’s understanding of decentering is problematic: if decentering had the nature Lange claims it has, any attempts to decenter from one’s thoughts would have to fail. For, as the experience of many mindfulness practitioners suggests, and as empirical studies confirm (Wegner and Erber 1992; Wegner 1994), attempts to directly detach from any mental state result not in downregulating the influence of that mental state—much less in turning that state off—but, contrary to the subject’s intentions, in upregulating its role in further mental processing (and frequently in further outer behavior) on the part of the subject who tries to detach from it. This is why, in many mindfulness schools, the process of decentering is conceived and taught quite differently from how Lange describes it. And it is this process that, I will try to show, poses a serious challenge to the SfA theory of attention.

First of all, contrary to Lange’s claim, decentering does not apply only to pathological mental states. As Jon Kabat-Zinn writes, in mindfulness practice “we treat all our thoughts as if they are of equal value” (1990, 68). Moreover, as Segal, Williams and Teasdale (2013, 150) insist, “the issue is not learning how to switch thoughts off, but how best we can change the way we relate to them: seeing them as they are—simply—as streams of thinking, events in the mind, rather than getting lost in them.” The crucial point, however, is that the aim described above is not achieved by any

direct intervention on the part of the subject. On the contrary, the only way to decenter is to “leave the thoughts alone,” without creating even “a hidden agenda that will get rid of unwanted experience if we simply allow it” (Nairn, Choden and Regan-Addis 2019, 74). The process is aptly described through a metaphor used by Dahl and colleagues:

To illustrate the difference between meta-awareness and experiential fusion, let us consider an example. Imagine that you are watching an enthralling movie. In one moment, you might be experientially fused with the movie, to the point when you are no longer consciously aware that you are sitting in a movie theater. In the next moment, you might suddenly become aware of your surroundings and the fact that you are viewing images on a screen. In both moments, you may be attentive to the movie, but only in the second moment are you also aware of the process of watching the movie. (Dahl, Lutz, and Davidson 2015, 516)

The above image clearly illustrates what occurs during decentering/reperceiving/defusion. In the “fused” state, one is focused on a particular thought at the expense of other stimuli (especially present bodily and perceptual ones; Smallwood and Schooler 2006; Smallwood, Baracaia, et al., 2003; Smallwood et al., 2007; Schooler et al., 2011). A subject focused on that thought tends to experience its content as reality itself. For instance, if the thought is about an emotionally significant past event, the person reacts emotionally as if the event were occurring in the present. In extreme cases—such as flashbacks in post-traumatic stress disorder (PTSD)—the emotional reaction may be identical to that experienced during the original traumatic event. As described by Hayes and colleagues (Hayes, Strosahl, and Wilson, 2011), one looks at the world *from* the thought without looking *at* the thought itself; that is, while thinking about something, one loses awareness of the process of thinking—loses the awareness that these are merely one’s thoughts about something. This is why the core of the process of decentering/reperceiving/defusion is a broadening of the field of awareness. That is achieved by dispersing attention so that it is not focused on any particular stimulus (i.e. mental state), whether a thought, bodily sensation, or perception of some external stimulus, but rather embraces as wide a range of experiences as possible. Such dispersed attention, which results in simultaneous awareness of both thought and bodily experiences (e.g., the breath) and the actual external surroundings of the subject, brings about the experience of the thought as merely a mental event passing through the mind, which in turn fosters an attitude of non-reactivity.

As the experience of many mindfulness practitioners shows, attention is repeatedly captured by particular thoughts; that is, it becomes focused on a single thought at the expense of the rest of one's ongoing experience. Mindfulness practice consists in the subject's gently broadening the focus of her attention whenever she notices that she is "lost in thinking." A key strategy for avoiding "getting lost in thinking" or "getting caught up in thinking" is to use "mindfulness support": i.e., to retain some peripheral awareness of, for example, one's own breath (Nairn, Choden and Regan-Addis 2019, 27–29; Choden and Regan-Addis 2018, 60–61). It is important, however, that awareness of one's breath does not block out "awareness of other things going on in and around us", because "maintaining peripheral awareness is a key element" of this process (Nairn, Choden, and Regan-Addis 2019, 27). As a result of such repeated practice, one can develop the ability to stabilize dispersed attention. When such dispersed attention is stabilized, and one simply registers all incoming stimuli in one's awareness without any further bodily or mental reaction toward them, one can "let go of support" and "rest in the midst of all" (Nairn, Choden, and Regan-Addis 2019, 28–29 and 181–84). Contrary to Lange's account, it is the broadening of the focus of one's attention resulting in non-conceptual meta-awareness of the mental states being attended to (Dunne, Thompson and Schooler 2019) that, in turn, brings about decentering/reperceiving/defusion without any additional actions on the part of the subject. In the context of this perspective, I suggest understanding these three terms as expressions of three simultaneous and interrelated aspects of one process: the broadening of the focus of one's attention means that no mental state occupies the center of the attentional field (decentering); as a result, one becomes aware of one's mental states as just mental events passing through one's mind (reperceiving); this, in turn, results in not being fused with the content of those mental states (defusion).

I think that even if, in the practice leading to the decentered state of mind, one can detect some selection for action (in Wu's sense), those periods in which one successfully rests in the decentered state of mind pose the real challenge for the SfA theory of attention. Wu (2014, 93) himself comments on the case in which "an embarrassing image of last evening's faux pas might involuntarily flash in one's head," stating that "if the thought amounted to nothing more than that, no attention emerges. There is just the fleeting image." From his perspective, "it is only when that image engages further activity—when one ponders it, laments it, or just sustains the image as one internally cringes—that attention emerges." Indeed:

[Even—P.S.] in attentional capture, attention enters the scene when the item that does the capturing not only alters the shape of one's mental states, but does so in a way that engages a response. Otherwise, there is only the mental registering of a change but no attention to it. The point, then, is that attentional capture is more than one's mental states changing in response to a sudden stimulus. Attention is not just a shift in consciousness or an alteration in one's mental states given a new stimulus. Rather, it is only when this change in one's mind engages with something further that attention comes on the scene. Without this further ingredient, attention is not present. The proposal is that this ingredient is selection for action: the change engages a response. (Wu 2014, 93)

Such a view, Wu admits, assumes that "one can be conscious of X without attending to X" (Wu 2014, 107)

I suggest that, in the above passage, Wu adequately describes what is going on in one's mind when one successfully decenters or defuses from one's thoughts. From his perspective, however, this is a case in which there is no attention involved in the process: one attends to nothing in particular—that is, one does not attend at all. Why not agree with Wu, and why align oneself with the view of many authors writing on mindfulness, who describe such a decentered state as involving the work of attention?

There are several reasons to invoke the notion of attention here. First, it is necessary to distinguish the state of mindful decentering—defined as a state in which a subject possesses a vivid and clear awareness of what is occurring in her mind, a state that it is tempting to describe as *an attentive one*—from the dull state of mind in which a subject does not attend to anything in particular and, as a consequence, is (almost) unaware of her mental states. This seems to be the difference between not attending at all (attending to nothing) and attending to everything in a balanced way. The use of the notion of attention in explaining the above difference can be justified by the empirical, neuroscientific work of Posner and Petersen (1990; 2012), work on which Wu himself draws. As Wu himself remarks, "Posner and Petersen identified three networks associated with functions commonly attributed to attention: '(a) orienting to sensory events; (b) detecting signals for focal (conscious) processing; and (c) maintaining a vigilant or alert state'" (Posner and Petersen 1990, 26; cited in Wu 2014, 27–28). It is the necessity of taking into account the last function—maintaining a vigilant or alert state—that makes invoking attention in decentering indispensable.

Secondly, "resting in the midst of all" and remaining in the decentered state of mind is—even if it sounds somewhat paradoxical—a form of activity or action, albeit not an activity involving the selection of something for

action.⁶ It is not a reflex (in Wu's sense), for it occurs within a behavioral space that includes other possible behavioral paths. A mindfulness practitioner who, in the decentered state, rests in the midst of all can, at any moment, cease to do so and instead select a particular stimulus for further action—for example, she may select an unpleasant thought that has just arisen in her mind in order to downregulate it. This process can naturally be described as involving the narrowing of one's previously dispersed attention and its focusing on a particular stimulus. From this perspective, both possible courses of action—continuing to rest in the midst of all and selecting something (for further action)—may be regarded as forms of the same activity: namely, the exercise of attention in different ways.

Wu might try to answer that, in the case of mindful decentering (resting in the midst of all), a subject exhibits vigilance but not attention (one does not pay attention but is vigilant), but I think he will not succeed in this way. He presents us with two slightly different accounts of vigilance, but neither of them is suited to the task. In Wu (2014) he states that in most accounts of vigilance the notion of attention is used. He further remarks that “a change in vigilance is typically measured by the vigilance decrement, and the latter is tied to certain properties of task performance, namely changes in detection rate and reaction time,” and suggests that “it looks like vigilance, even if it is different from selection for action, supervenes on it,” for “*changes* in vigilance are measured by changes in selection for action as measured by behavioral outcomes” (2014, 94). His conclusion there is that “vigilance is a property of selection for action over time: vigilance is a measure of how effective subjects are in selecting for action. One assesses vigilance by assessing how subjects sustain selection for action in an experimental setting.” On the other hand, in Wu (2023b) he presents a slightly different account of vigilance, claiming that it is “a propensity to attend to task-relevant targets” and holding that “vigilance's expression is attention . . . but to be vigilant regarding X is not yet to attend to X” (2023b, 110). As an example of a vigilant person, he describes a detective who watches for a thief at a moment when the latter has not yet appeared at the crime scene. A vigilant person has a propensity to attend, but is not attending, for there is no item to attend to yet.

6. It is important to distinguish between the claim that attention is selection for action and the claim that attention—or, more precisely, attending—is itself an activity or an action. Wu endorses both claims (as I have discussed above), whereas some of his critics—for example, Watzl—accept only the second. This is why, in the case of decentering, one can indeed identify a form of activity: decentering may be conceived as a specific mode of attending. This is not, however, to suggest that this activity or action constitutes selection for action.

It is evident that both of Wu's accounts of vigilance fail to explain the alert state of a mindful subject. In both accounts, vigilance is connected with attention in such a way that the more vigilant a person is, the more probable it is that she will attend to—i.e. select for action—the particular stimulus. In the case of mindful decentering, a person is alert (vigilant, attentive) in a way that makes the selection of a particular stimulus for action (i.e. attention in the SfA sense) less probable. That is why, if alertness/vigilance is connected to attention—as Posner and Petersen hold, and as teachers of mindfulness suggest—attention should be explained in a way other than the one Wu proposes.

CONCLUSION

Mindfulness practice, and the mental phenomena that mindfulness practitioners experience during such practice, may provide important data for philosophical theories concerning issues connected with consciousness, attention and the like. But philosophers must adequately understand what is experienced during mindfulness practice. Otherwise, they may find themselves in a situation comparable to that of Lange, who highlights the real problem—namely, the discrepancy between decentering and SfA—but has come to the right conclusion for the wrong reasons. Nevertheless, his initial observation stands: decentering (even if conceived differently than in his account) constitutes a threat to Wu's theory of attention as selection for action.

Given what has been written above about decentering—especially the fact that its key aspect consists in dispersing one's attention as widely as possible, as opposed to keeping it focused on a particular stimulus—I suggest that we look for a theory according to which attention is understood as an activity of structuring one's mental life, such as Sebastian Watzl's account (Watzl 2017) or certain phenomenological conceptions (see, e.g., Arvidson 2006; Gurwitsch 2010). It is, of course, possible that these theories, too, will need to be refined in order to accommodate decentering—or perhaps other forms of mental experience characteristic of mindfulness practitioners. In any case, a broad field of research lies ahead.

REFERENCES

- Arbel, Keren. 2016. *Early Buddhist Meditation: The Four Jhānas as the Actualization of Insight*. Abingdon: Routledge.
- Arvidson P. Sven. 2006. *The Sphere of Attention. Context and Margin*. Dordrecht: Springer
- Bernstein, Amit, Yuval Hadash, and David M. Fresco. 2019. "Metacognitive Processes Model of Decentering: Emerging Methods and Insights." *Current Opinion in Psychology* 28: 245–51. <https://doi.org/10.1016/j.copsyc.2019.01.019>.

- Bernstein, Amit, Yuval Hadash, Yael Lichtash, Galia Tanay, Kathrine Shepherd, and David M. Fresco. 2015. "Decentering and Related Constructs: A Critical Review and Metacognitive Processes Model." *Perspectives on Psychological Science* 10 (5): 599–617. <https://doi.org/10.1177/1745691615594577>.
- Bodhi, Bhikkhu. 2011. "What Does Mindfulness Really Mean? A Canonical Perspective." *Contemporary Buddhism* 12 (1): 19–39. <https://doi.org/10.1080/14639947.2011.564813>.
- Buehler, Denis. 2019. "Flexible occurrent control." *Philosophical Studies* 176, 2119–37 (2019). <https://doi.org/10.1007/s11098-018-1118-3>.
- . 2022. "Seing Circles: Inattentive Response-Coupling." *Ergo* 9 (62), 1671–91. <https://doi.org/10.3998/ergo.3587>
- Choden, and Heather Regan-Addis. 2018. *Mindfulness Based Living Course*. Winchester: O-Books.
- Dahl, Cortland J., Antoine Lutz, and Richard J. Davidson. 2015. "Reconstructing and Deconstructing the Self: Cognitive Mechanisms in Meditation Practice." *Trends in Cognitive Sciences* 19 (9): 515–23. <https://doi.org/10.1016/j.tics.2015.07.001>.
- Dreyfus, Georges. 2011. "Is Mindfulness Present-Centred and Non-Judgmental? A Discussion of the Cognitive Dimensions of Mindfulness." *Contemporary Buddhism* 12 (1): 41–54. <https://doi.org/10.1080/14639947.2011.564815>.
- Dunne, John D., Evan Thompson, and Jonathan Schooler. 2019. "Mindful Meta-Awareness: Sustained and Non-Propositional." *Current Opinion in Psychology* 28: 307–11. <https://doi.org/10.1016/j.copsyc.2019.07.003>
- Dunne, John. 2011. "Toward an Understanding of Non-Dual Mindfulness." *Contemporary Buddhism* 12 (1): 71–88. <https://doi.org/10.1080/14639947.2011.564820>.
- Fennell, Melanie, and Zindel Segal. 2011. "Mindfulness-Based Cognitive Therapy: Culture Clash or Creative Fusion?" *Contemporary Buddhism* 12 (1): 125–42. <https://doi.org/10.1080/14639947.2011.564828>.
- Gethin, Rupert. 2011. "On Some Definitions of Mindfulness." *Contemporary Buddhism* 12 (1): 263–79. <https://doi.org/10.1080/14639947.2011.564843>.
- Grossman, Paul, and Nicholas T. Van Dam. 2011. "Mindfulness, by Any Other Name...: Trials and Tribulations of Sati in Western Psychology and Science." *Contemporary Buddhism* 12 (1): 219–39. <https://doi.org/10.1080/14639947.2011.564841>.
- Gurwitsch Aron. 2010. *The Field of Consiousness*. In: *The Collected Works of Aron Gurwitsch (1901-1973)*, edited by Richard M. Zaner and Lester Embree, Dordrecht Heidelberg London New York: Springer
- Hayes, Steven C., Kirk D. Strosahl, and Kelly G. Wilson. 2011. *Acceptance and Commitment Therapy: The Process and Practice of Mindful Change*. 2nd ed. New York: Guilford Press.
- Kabat-Zinn, Jon. 1990. *Full Catastrophe Living*. New York: Delta.
- Lange, Victor. 2025. "Decentering and Attention." *Philosophical Psychology* 38 (4): 1530–57. <https://doi.org/10.1080/09515089.2023.2263034>.
- Nairn, Rob, Choden, and Heather Regan-Addis. 2019. *From Mindfulness to Insight*. Boulder, CO: Shambhala.
- Nyanaponika, Thera. 1968. *The Power of Mindfulness*. Kandy: Buddhist Publication Society.
- Olendzki, Andrew. 2011. "The Construction of Mindfulness." *Contemporary Buddhism* 12 (1): 55–70. <https://doi.org/10.1080/14639947.2011.564817>.
- Polak, Grzegorz. 2024. *Nikāya Buddhism and Early Chan: A Different Meditative Paradigm*. Sheffield: Equinox.

- Posner, Michael I., and Steven E. Petersen. 1990. "The Attention System of the Human Brain." *Annual Review of Neuroscience* 13: 25–42. <https://doi.org/10.1146/annurev.ne.13.030190.000325>.
- . 2012. "The Attention System of the Human Brain: 20 Years After." *Annual Review of Neuroscience* 35: 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>.
- Puc, Jan. 2019. "In Defence of Bare Attention: A Phenomenological Interpretation of Mindfulness." *Journal of Consciousness Studies* 26 (5–6): 170–90.
- Schooler, Jonathan W., Jonathan Smallwood, Kalina Christoff, Todd C. Handy, Erik D. Reichle, and Michael A. Sayette. 2011. "Meta-Awareness, Perceptual Decoupling and the Wandering Mind." *Trends in Cognitive Sciences* 15 (7): 319–26. <https://doi.org/10.1016/j.tics.2011.05.006>.
- Segal, Zindel V., J. Mark G. Williams, and John D. Teasdale. 2013. *Mindfulness-Based Cognitive Therapy for Depression*. 2nd ed. New York: Guilford Press.
- Shapiro, Shauna L., Linda E. Carlson, John A. Astin, and Benedict Freedman. 2006. "Mechanisms of Mindfulness." *Journal of Clinical Psychology* 62 (3): 373–86. <https://doi.org/10.1002/jclp.20237>.
- Smallwood, Jonathan, and Jonathan W. Schooler. 2006. "The Restless Mind." *Psychological Bulletin* 132 (6): 946–58. <https://doi.org/10.1037/0033-2909.132.6.946>.
- Smallwood, Jonathan, Merrill McSpadden, and Jonathan W. Schooler. 2007. "The Lights Are On but No One's Home: The Decoupling of Executive Resources When the Mind Wanders." *Psychonomic Bulletin & Review* 14 (3): 527–33. <https://doi.org/10.3758/bf03194102>.
- Smallwood, Jonathan, Simona F. Baracaia, Michelle Lowe, and Marc Obonsawin. 2003. "Task Unrelated Thought Whilst Encoding Information." *Consciousness and Cognition* 12 (3): 452–84. [https://doi.org/10.1016/S1053-8100\(03\)00018-7](https://doi.org/10.1016/S1053-8100(03)00018-7).
- Watzl, Sebastian. 2017. *Structuring Mind: The Nature of Attention and How It Shapes Consciousness*. Oxford: Oxford University Press.
- Wegner, Daniel M. 1994. "Ironic Processes of Mental Control." *Psychological Review* 101 (1): 34–52. <https://doi.org/10.1037/0033-295X.101.1.34>.
- Wegner, Daniel M., and Ralph Erber. 1992. "The Hyperaccessibility of Suppressed Thoughts." *Journal of Personality and Social Psychology* 63 (6): 903–12. <https://doi.org/10.1037/0022-3514.63.6.903>.
- Wu, Wayne. 2014. *Attention*. London: Routledge.
- . 2019. "Structuring Mind: The Nature of Attention and How It Shapes Consciousness, by Sebastian Watzl. Oxford and New York: Oxford University Press, 2017. Pp. xii + 322." *Mind*, Volume 128, Issue 511, July 2019, Pages 945–53, <https://doi.org/10.1093/mind/fzy058>
- . 2023a. "Attending as Mental Action: Mental Action as Attending." In *Mental Action and the Conscious Mind*, edited by Michael Brent and Lisa Miracchi Titus, 61–78. London: Taylor & Francis.
- . 2023b. *Movements of the Mind. A Theory of Attention, Intention and Action*, Oxford, England; New York, NY: Oxford University Press;
- . 2025. "Attention as selection for action defended." *Philosophy and Phenomenological Research*, 110, 421–41. <https://doi.org/10.1111/phpr.13101>

Doxastic Responsibility and the Challenge of Doxastic Voluntarism

Insights from Cases of Self-Deception

Ewa Odoj

ABSTRACT In the article, I present the debate on doxastic voluntarism and its relation to doxastic responsibility. I outline the discussion in the literature, focusing on Alston's argument against doxastic responsibility, and then present my own position in this debate. I defend a conception of doxastic freedom that remains consistent with the principle of alternative possibilities. To this end, I provide an epistemological analysis of the phenomenon I call "doxastic self-deception." I also introduce the notions of "doxastic strategy" and "alethic impurism"—a view concerning the possibility of pragmatic reasons for beliefs. I conclude that doxastic responsibility is possible because we have the ability to self-deceive, and at the same time possess metacognitive capacities that enable cognitive self-control.

KEYWORDS alethic impurism; doxastic freedom; doxastic responsibility; doxastic strategy; self-deception

The issues of doxastic voluntarism and responsibility emerged in contemporary analytic epistemology at the end of the twentieth century as part of the ethics of belief. Doxastic voluntarism is the view that we exercise freedom with respect to which beliefs we hold. Proponents of the opposite view—doxastic involuntarists—argue that we do not have such freedom. The debate between doxastic voluntarism and involuntarism is closely connected with the issue of doxastic responsibility since, according to a basic intuition about freedom in general, a subject can be held responsible only for something over which they had influence.

While the topic has not assumed the status of one of the central problems of epistemology, its importance for the discipline is often underestimated. If epistemology refers to the normative sphere, then a fundamental question is whether its normativity should be construed in deontological terms, as grounded in intellectual obligations. As William Alston has argued, one of the most basic concepts in epistemology—justification—already presupposes both doxastic responsibility and freedom. Any full account of the so-called third condition of knowledge must therefore engage with this issue.

This article is divided into two parts. The first has an expository character. It begins with a consideration of the philosophical significance of the issue of doxastic freedom and responsibility, before turning to the classic texts and debates that first drew epistemologists' attention to the issue. I then examine Alston's influential argument against doxastic responsibility, as well as the conceptual argument against doxastic voluntarism, which is often regarded as the strongest case against the possibility of doxastic freedom. The first part concludes with a discussion of the main strategies for rejecting Alston's argument that have been presented in the extensive debates concerning his position. Finally, I point to important issues that require further examination in order to adequately address the problem of doxastic freedom and responsibility.

The second part is a presentation of my own position on the issue. I develop this view by beginning with four very similar examples of processes of belief acquisition. Two of them are appropriate from the perspective of epistemic normativity, while the other two are inappropriate. I argue that the inappropriate cases of belief acquisition are instances of a phenomenon I call "doxastic self-deception." In presenting this concept, I make use of the notion of a "doxastic strategy," understood as the subject's balancing between two aspects of the epistemic aim. In this way, I arrive at the question of whether there can be non-epistemic reasons for belief. I introduce and defend a position on this matter that I call "alethic impurism." This is a position concerning the possibility of pragmatic, that is non-epistemic,

reasons for beliefs. It parallels, to some extent, impurism (also referred to as “pragmatic encroachment”) in relation to the so-called threshold problem for fallibilism, with the key distinction that the form of impurism I defend pertains solely to standards of assertion rather than standards of knowledge. This perspective enables me to formulate a view on doxastic freedom and responsibility in which two elements play a central role: our capacity for doxastic self-deception and our metacognitive abilities.

PART I

1. The philosophical significance of the topic of doxastic voluntarism

Our ways of thinking and speaking about both our own and others’ beliefs frequently betray an underlying intuition that the domain of mental representation is normative in character. We commonly employ expressions such as, “you ought not to think that,” or “how could he have believed such a thing?” Such normative judgments may carry different meanings, but two are of central importance from an epistemological perspective: first, normativity as proper functioning of cognitive faculties, and second, normativity in the deontological sense, tied to the fulfillment of intellectual duties. The former notion is relatively uncontroversial—for instance, when a subject’s cognitive faculties (such as vision) fail to operate properly, perhaps due to biological deterioration, the subject forms beliefs about the world that she would not have had if those faculties had been functioning properly.

The issue of whether there exist distinctly epistemic duties is considerably more complex. One may ask whether the normativity of the cognitive domain reduces *solely* to the proper functioning of cognitive mechanisms.¹ If such mechanisms are understood simply as biological functions of the organism, it becomes puzzling why we so often speak or think of someone’s mental representation as subject to blame—captured, for example, in the indignant exclamation, “how could he have believed that!?” On the ordinary, intuitive conception, blame presupposes that the agent had some measure of control over that for which she is blamed. By contrast, we do not, in general, treat the biological functions of organisms as something for which a subject is (at least directly) responsible or which remains under their control.

1. The foremost advocate of such position within analytic philosophy was Alvin Plantinga (see, e.g., Plantinga 1993). It is noteworthy that Plantinga was both intellectually influenced by, and personally close to, William Alston—whose argument against epistemic responsibility is regarded as the classic starting point of the debate on this issue.

The problem acquires particular significance once we turn to beliefs that are connected with morally relevant actions. In typical situations, individuals act in accordance with their own beliefs, and we intuitively regard this as appropriate. Taking this into account, and assuming that the domain of our beliefs is not something for which we are responsible, it becomes unclear how we can ascribe moral blame to anyone if that person does not themselves believe that they have done something wrong. Put differently, if agents act on what they take to be true, and if their beliefs are beyond the scope of responsibility, then objective moral blame collapses into merely subjective moral blame. Yet this contradicts our moral intuitions and the social practice of holding people responsible for their actions.

The problem that emerges here can be illustrated by means of the following trilemma:

1. People have no influence over their own beliefs.
2. People have the right to act according to their own beliefs.
3. Sometimes people are blameworthy for their actions that resulted from their beliefs, even though they do not assign blame to themselves.

All three of the above theses seem to align with our intuitions, and at least *prima facie*, there are reasons to consider them true. Nevertheless, the conjunction of any two of the above theses entails the negation of the third. If people have no influence over their own beliefs and have the right to act in accordance with them, then they are not subject to objective moral blame if they do not assign themselves subjective blame. If people are morally blameworthy for their actions and have no influence over their beliefs, then they cannot always act in accordance with what they themselves judge. Conversely, if people are sometimes subject to objective moral blame and can act in accordance with their beliefs, this means they do have influence over their own beliefs. For if they can act in accordance with their beliefs, and we assign them blame for that action, then the blame concerns the beliefs they hold. From this perspective, it becomes clear why the debate on doxastic responsibility and freedom emerged under the banner of the ethics of belief.

The problem of doxastic freedom and responsibility is not limited to matters that are explicitly moral in nature. Our actions are intimately connected to our representation of the world. The importance of this insight is not restricted to the moral domain. We are accountable for ourselves—both in our long-term and immediate decisions—and the ways in which we develop as persons depends on how we perceive reality. We identify with our beliefs, treating them as a core part of our identity. This is particularly true for

worldview beliefs, including religious,² moral, and political convictions. On the one hand, we regard these beliefs—especially our own—as the product of recognizing the truth about reality. On the other hand, we associate them with personal freedom and responsibility to a much greater extent than we do with other kinds of beliefs. Taken together, these observations raise the question: do we truly have grounds to ascribe responsibility and freedom to ourselves and others in the domain of our mental representation of the world?

2. *The origin of the debate*

Clifford and James

The classic texts that serve as reference points in contemporary discussions on the ethics of belief are William Clifford's essay *The Ethics of Belief* (Clifford 2002) and William James's critical response entitled *The Will to Believe* (James 2002).³ Clifford addresses the issue of intellectual duties, and in his essay he presents the famous maxim: *It is wrong, always, everywhere, and for anyone, to believe anything upon insufficient evidence*. This statement became the earliest formulation of a position known as evidentialism, which postulates the existence of an epistemic obligation to accept beliefs solely on the basis of sufficient evidence. The thesis has since become one of the most widely discussed topics in the ethics of belief.

Evidentialism is, above all, a position that presupposes only epistemic deontologism—namely, the view that in our intellectual life we have duties regarding what we believe (the content of these duties being specified by the thesis of evidentialism). Some epistemologists, however, extend this position to the concept of epistemic justification, arguing that the essence of justification—as the third condition of knowledge—lies in fulfilling the epistemic obligation to possess sufficient evidence for one's beliefs. Although Clifford's formulation of the evidentialist thesis remains a key point of reference for many evidentialists, its precise formulation continues to be a matter of debate (e.g., Conee & Feldman 2004; Chignell 2018).

Clifford's essay has sparked discussions concerning the relationship between epistemic and moral obligations. Clifford appears to defend the rather controversial thesis that all epistemic responsibility is, at the same

2. The significance of the problem of doxastic responsibility and freedom is particularly pronounced from the perspective of religious doctrines that foresee reward or punishment for faith. If faith consists at least in part of beliefs (e.g., regarding the existence of a Creator of the world), how could believers and non-believers be held accountable for it if they had no influence over whether they possess such beliefs?

3. For a thorough review of doxastic voluntarism in earlier philosophical periods, see: Boespflug and Jackson (2024).

time, moral responsibility. The question of the relationship between intellectual and moral duties is a complex and compelling issue at the intersection of epistemology and ethics.⁴ However, most proponents of epistemic deontologism regard intellectual duties (including the maxim of evidentialism) as *sui generis* epistemic obligations. Contemporary evidentialists tend to distance themselves from Clifford's radical claim that the maxim of evidentialism is always of a moral nature.

Clifford does not formulate his accusations explicitly, but readers of his essay have little doubt that his charge of moral irresponsibility is directed primarily at religious individuals—specifically, those who rely on divine providence when making important decisions and who, according to Clifford, thereby act in violation of the evidentialist maxim. Shortly after the publication of Clifford's essay, a defense of such religious individuals was formulated by William James in his essay *The Will to Believe*. According to James, there are situations in which it is not possible to fulfil the demands of the evidentialist thesis. In such cases, it is both intellectually and morally permissible to adopt a belief despite the absence of sufficiently strong evidence. James identifies three criteria that must characterize a belief in order for it to be considered a justified exception in respect of the evidentialist demand: the living option (the subject is psychologically predisposed to accept the belief), the forced option (there is no possibility of avoiding the decision, as suspending judgment is equivalent to rejection the belief), and the momentous option (the belief is of vital importance and has significant consequences). In such cases, if the subject lacks the opportunity to obtain strong evidence in support of a particular belief, they may still adopt it without violating their intellectual or moral obligations. According to James, a paradigmatic example of such a case is religious faith.

James's classic essay is frequently cited by those who defend a more limited interpretation of the evidentialist thesis, highlighting a significant set of exceptions to the rule (e.g., Pace 2011). A common contemporary view holds that the range of exceptions to the evidentialist thesis is even broader. These exceptions concern not only religious beliefs but also other worldview-related beliefs, including those regarding morality, values, and even political preferences. Defenders of evidentialism, in turn, maintain that the thesis remains valid in its categorical form—provided that the concepts it involves (e.g., belief, evidence, sufficiency of evidence) are adequately specified (e.g., Feldman 2006).

4. On the relationship between moral and epistemic duty, see, e.g., Haack (2001); Shaffer (2006); Bergeron (2006); Booth (2012).

Williams and Alston

The third classic text cited in discussions on doxastic voluntarism is Bernard Williams' essay *Deciding to Believe*, in which he explicitly addresses the problem of doxastic freedom. Williams argues that we lack freedom concerning the beliefs we hold. His position is aptly captured in the following passage from this essay:

If I could acquire a belief at will, I could acquire it whether it was true or not; moreover I would know that I could acquire it whether it was true or not. If in full consciousness I could will to acquire a "belief" irrespective of its truth, it is unclear that before the event I could seriously think of it as a belief, i.e. as something purporting to represent reality. (Williams 1973)

Williams' reasoning appears to capture a fundamental intuition about our cognitive life, and as a result, it has gained many adherents and is often taken for granted. The authors note that Williams' position can be interpreted in two ways: as a logical impossibility of adopting beliefs by an act of will, or as a psychological impossibility. The most important defender of the psychological interpretation was William Alston. He argued that it is simply psychologically impossible to genuinely assert something whose truth one does not believe:

My argument for this, if it can be called that, simply consists in asking you to consider whether you have any such powers. Can you, at this moment, start to believe that the U.S. is still a colony of Great Britain, just by deciding to do so. If you find it too incredible that you should be sufficiently motivated to try to believe this, suppose that someone offers you \$500,000,000 to believe it, and you are much more interested in the money than in believing the truth. Could you do what it takes to get that reward? (Alston 1988, 263)

The logical—or conceptual—interpretation refers to the nature of belief. It holds that truth, as the constitutive aim of beliefs, makes it impossible to adopt beliefs by an act of will. Given its significance, this topic will be addressed in a dedicated section of the present article.

The fourth classic text concerning doxastic voluntarism is William Alston's article *The Deontological Conception of Epistemic Justification* (Alston 1988). While addressing a topic closely connected to the ethics of belief, Alston goes beyond its scope. He observes that the standard epistemological understanding of justification typically assumes a deontological framework, identifying justification with the fulfilment of epistemic

duties. Alston challenges this tradition by arguing against the very existence of epistemic duties. His line of reasoning can be presented as follows:

1. If a subject is obliged to perform an action, then she must have the ability to perform it.
2. We do not have freedom over what beliefs we have.
3. So we cannot have duties in this respect.

Alston maintains that epistemic deontologism, understood as a concept of justification, cannot be sustained in view of the two intuitively compelling premises (1) and (2). Premise (2) appears to be true primarily in light of the argument against doxastic voluntarism as presented by Williams (in two variants—psychological and conceptual). Premise (1), on the other hand, pertains to the standard understanding of duty and seems to correspond to our basic intuitions. It is known as the Kantian principle “ought implies can.” Intuitively, it seems that we cannot ascribe to someone the obligation to do something over which she had no control.

Alston’s article has provoked increased interest in the problem of epistemic freedom and its relation to deontological concept/conception of justification and epistemic normativity. The article received numerous responses from epistemologists. Some of them (mainly those sympathetic to externalism) endorsed his conclusion regarding the necessity of rejecting deontologism, at least as a condition of justification and a source of normativity in epistemic domain. In Section 4, I will present the different strategies that have been adopted to reject Alston’s argument. First, however, it is worth examining in more detail the conceptual argument against doxastic voluntarism, which offers the strongest support for premise (2) of Alston’s argument.

3. Truth as the constitutive aim of belief and the problem of doxastic voluntarism

To hold a belief that *p* is, by definition, to regard *p* as true—that is, to assert its truth. A mental state of a different character is not a belief at all, but rather a desire or an imagining. This observation is largely uncontroversial. Within the epistemological debate on the aim of belief, it is widely accepted that belief is constitutively connected to truth, which functions as its goal or norm. Proponents of the conceptual argument against doxastic voluntarism rely on precisely this point. Their claim is that a subject cannot simply choose which beliefs to hold, since for a mental state to qualify as a belief, the subject must remain responsive to how reality presents itself to her, rather than to how she wishes it to be.

The constitutive link between belief and truth is widely regarded as the strongest basis for rejecting doxastic voluntarism, and Williams’s statement

quoted in the previous paragraph is considered a classic argument of this kind. Yet Williams's wording itself has faced serious criticism. Particularly controversial is his claim that "*moreover I would know that I could acquire it whether it was true or not*" [emphasis mine]. Advocates of the conceptual strategy against doxastic voluntarism have accordingly proposed refinements of Alston's argument and advanced further considerations in support of the necessary tie between belief and truth—a tie that, by its very nature, precludes voluntary control over one's beliefs. The core of this conceptual approach to doxastic involuntarism is well expressed in a formulation defended by Barbara Winters: "it is impossible to believe that one believes *p* and that one's belief of *p* originated and is sustained in a way that has no connection with *p*'s truth" (Winters 1979, 243).

Particularly noteworthy is the extended argumentation for the claim that we cannot believe at will by Pamela Hieronymi. These analyses are broadly in line with Williams' thought, but go well beyond it (see Hieronymi 2006; see also the discussion on Hieronymi's argumentation in Setiya 2008 and Hieronymi 2009). Hieronymi distinguishes two types of reasons for adopting a belief: constitutive reasons, otherwise called "content-related reasons," and extrinsic reasons, which can also be described as "attitude-related reasons":

Of course, if you take certain reasons to show that *p*, you therein believe *p*. Thus the reasons taken to bear positively on whether *p*—those taken to be content-related reasons for the belief that *p*—are also what I will call "constitutive reasons" for the belief that *p*. They support the commitment constitutive of the belief. By finding such reasons convincing, you therein believe. (Hieronymi 2006, 51)

Extrinsic reasons for adopting the belief that *p* are, for instance, considerations such as its perceived usefulness, importance, or desirability, independently of whether *p* is in fact true. Yet the presence of such reasons does not suffice to make the subject affirm *p* as true. On this account, beliefs are not voluntary: their formation depends solely on the evidence to which the subject has access, and therefore, ultimately, on how the world actually is—rather than on how one might wish it to be, or on any other factors unrelated to the truth of *p*.

At this point, the close connection between conceptual doxastic involuntarism and the principle of evidentialism becomes apparent. The latter holds that one is obligated to endorse only those beliefs for which one possesses adequate evidential support. The claim that only evidence can constitute

a genuine reason for belief received a robust theoretical grounding in Nishi Shah's well-known thesis of transparency:

TrT 1 *the deliberative question* whether to believe that *p* is transparent to the question whether *p* (Shah & Velleman 2005, 497).

According to Shah: (1) belief differs from other cognitive propositional attitudes in that it is regulated for truth (the descriptive part of the concept of belief); (2) part of the concept of belief is also a standard of being correct if and only if it is true (the normative part of the concept) (see Shah 2003, 2006; Shah & Velleman 2005).

This position provides a clearer grasp of the basic intuition underlying the conceptual critique of doxastic voluntarism. Shah's analyses are also meant to support evidentialism and, at the same time, to undermine pragmatism, understood here as the view that reasons for belief may be of a non-epistemic kind—in other words, that they may consist in something other than evidence. Shah's account has generated extensive discussion (see, e.g., Sullivan-Bissett 2018; McHugh 2013). In the light of Shah's attempt to exclude pragmatic reasons for belief, McHugh reformulated the transparency thesis in the following way, which will play a role in the second part of this article:

TrT 2 Pragmatic considerations cannot occur to a thinker, within doxastic deliberation, as relevant to what to believe (McHugh 2013, 448).

The discussions reviewed above indicate that the problem of doxastic voluntarism is closely connected with the role of truth as the goal of belief, the validity and applicability of the evidentialist thesis, and whether there can be pragmatic (non-epistemic) reasons for belief.

4. Strategies for dealing with the problem of doxastic responsibility

Alston's argument against the existence of epistemic duties rests on two key premises: (1) If a subject is obliged to perform an action, then she must have the ability to perform it; and (2) We do not have freedom over what beliefs we have. The possible responses to Alston's argument can be divided into two groups: those that challenge the first premise, and those that challenge the second. Each of these lines of response can be developed in several distinct variants.

Undermining the first premise

A widely discussed way of defending the thesis of the existence of doxastic responsibility is to either reject premise (1) or propose a different interpretation of it. The Kantian phrase “ought implies can” refers primarily to the fact that responsibility presupposes the subject’s ability to perform what is the object of responsibility. However, the question arises whether one can sensibly speak of responsibility with respect to something that the subject must do (rather than merely can). Alston assumes in his reasoning that one cannot. As he writes: “one can be obliged to do A only if one has an effective choice as to whether to do A” (Alston 1988, 259). Some authors criticize Alston’s reasoning by rejecting this assumption, at least on epistemological grounds.

One way of rejecting Alston’s argument by undermining premise (1) is to recognize that not all areas in which we deal with responsibilities are governed by the “ought implies can” principle invoked in premise (1). For example, Richard Feldman argues that doxastic responsibilities are a type of role responsibility. Just as a teacher should clearly explain the material to their pupils, and a parent should take good care of their child, so people as cognitive agents should adopt their beliefs in an appropriate manner. Feldman emphasizes that, in the case of role obligations, it is not required that the subject be able to perform them or that they have voluntarily undertaken the role. A poor teacher may not be able to teach well, and an incompetent parent may not be able to care adequately for a child, yet they are still burdened with the obligation to fulfill their role properly. According to Feldman, deontology on epistemic grounds does not connect with such a duty, for which it is necessary that the subject possess freedom. Therefore, in his view, Alston’s argument is flawed (see Feldman 2001, 2008).⁵

A strategy that gained considerable popularity was first proposed by Matthias Steup, who applied the compatibilist position known from the free will debate to the problem of doxastic responsibility. First, Steup points out that the libertarian notion of freedom as being uncaused is inapplicable on epistemological grounds. Beliefs are always directed toward the truth and cannot be a kind of mental coin toss. In relation to our actions, however, we are also dealing with decision-making based on an examination of reasons for and against, which we regard as a kind of freedom. Steup argues that it is precisely this sort of freedom that we exercise in the process

5. For a similar strategy to counter Alston’s argument, and for a discussion of Feldman’s position, see, e.g., Chuard and Southwood (2009); Altschul (2014); Chrisman (2008); McHugh (2012).

of adopting beliefs. Beliefs are completely determined by the evidence, but this does not prevent them, according to compatibilist proponents, from being considered free. According to Steup, as an effect of doxastic deliberation we make doxastic decisions and execute them by adopting beliefs, and this constitutes a kind of doxastic freedom that we possess (see Steup 2000).⁶ Steup's position has met with considerable criticism. Critics, among others, point out that relatively few beliefs arise as a result of doxastic deliberation, while the vast majority arise spontaneously and unreflectively, for example, as a result of simple visual perception. This latter type of beliefs does not seem to fit the picture of doxastic freedom sketched by Steup.

The concept of doxastic compatibilism has received considerable attention, although different authors emphasize different aspects of our epistemic situation as sources of doxastic responsibility, despite the determination of beliefs by evidence. For instance, Sharon Ryan highlights the intentionality of beliefs (see Ryan 2003; also Steup 2012, 2017). According to her, beliefs are not analogous to actions that are clearly involuntary, such as pathological compulsive behavior or blushing in an embarrassing situation. In such cases, it is more accurate to say that something happens to the subject rather than that the subject performs these actions. Ryan points out that, as with the acquisition of beliefs, our free actions are sometimes performed unconsciously and automatically—for example, pressing particular keyboard keys while typing on a computer.⁷

Williams, and with him many other authors, is guided by the intuition that truth as the aim of belief is incompatible with acquiring beliefs at will. Authors such as Shah and Montmarquet argue that Williams' intuition wrongly presupposes that an action with a goal cannot be performed freely because it cannot be carried out independently of that goal. Shah reduces Williams' argument to absurdity by formulating an analogous argument for the thesis that one cannot lie freely, since in order to lie it is necessary to pursue the goal of deceiving someone (see Shah 2002). As Montmarquet points out, an action is not rendered involuntary by the controlling influence of reason. In the same way, in his view, the adoption of beliefs should not be considered deprived of freedom merely on the basis that, having its

6. In the following years, Steup developed his concept of doxastic compatibilism; see Steup (2008, 2012, 2017).

7. For criticism of Ryan's position, see Buckareff (2006a). See also other criticisms of doxastic compatibilism: Booth (2009, 2014); Buckareff (2006b); Schmitt (2015); Tebben (2014); Bayer (2015); Peels (2014); Wagner (2017).

purpose—truth—it is directed by evidence (see Montmarquet 1986).⁸ Many authors, especially those advocating doxastic compatibilism, present such reasons for the lack of conflict between freedom of belief and the evidence in favor of those beliefs.

Undermining the second premise

The second line of defense of epistemic deontology is to reject premise (2), which holds that we lack freedom over what beliefs we hold. One option here is to identify those instances of belief for which we do in fact have freedom. Authors try to point to examples where the voluntary adoption of a belief is tied to a change in the world that makes the belief true. These are cases in which, on the one hand, the belief is adopted voluntarily in the strongest sense of the term, and on the other, the belief still stands in the right relation to truth as its aim. For instance, Feldman (2001) argues that I can bring about my belief that the light is on simply by switching on the light. Such examples, however, merely weaken Alston's argument or show the need for certain qualifications. Since they are rare and highly specific, a conception of doxastic voluntarism resting solely on them would be of little interest for the issue of doxastic responsibility. Ginet, by contrast, draws attention to situations in which a subject has some, though not conclusive, evidence for a belief (e.g., that he locked the door before leaving) but, due to the undesirable consequences of doubt (e.g., having to return home to check), chooses to adopt the belief (see Ginet 2001). However, it is doubtful whether in this case one should really speak of belief, and not merely of making a decision to act as if a certain state of affairs were true (see also Cohen 1992).

Another way to reject premise (2) is to point out that, although we cannot adopt beliefs through a direct act of will, we have several means available to indirectly influence our beliefs (see, e.g., Nottelmann 2007; Peels 2017b, 2017a). This position is called "indirect belief voluntarism." For example, we can acquire additional evidence on a given topic, exercise our cognitive competence, or become aware of and avoid common reasoning errors. According to some authors, the possibility of engaging in such activities is sufficient to make us responsible for our beliefs, providing an adequate basis for epistemic deontology. This approach is particularly popular among proponents of virtue epistemology (see, e.g., Audi 2008;

8. This issue has developed into extensive discussions about the analogy—or lack of it—between actions and reasons for actions, and beliefs and reasons for them. On this issue see, e.g., Buckareff (2006a, 2008); Cohen (2016); Roeber (2016).

Kruse 2017; Montmarquet 1993, 2008a, 2008b). According to this approach in epistemology, doxastic responsibility is tied to the subject's proper epistemic virtue—a stance or attitude that ensures she appropriately engages in the pursuit of truth.

Importantly, advocates of indirect doxastic voluntarism share with involuntarists the view that a subject must ultimately remain committed to the evidence available to them when forming beliefs. Indirect doxastic voluntarism is thus a position that, on the one hand, defends doxastic responsibility, but on the other hand acknowledges the central insight of the argument against the possibility of doxastic freedom. On the third hand, proponents of this position do not reject the principle of alternative possibilities, as doxastic compatibilists do. Their strategy for addressing the doxastic responsibility problem consists in relocating the locus of freedom—the subject's capacity to influence—from the moment of assertion itself to the preceding activities aimed at it: namely, the gathering and evaluation of evidence.

Another strategy for rejecting Alston's argument can be understood as a reformulation of premise (2). According to several authors, epistemic deontologism should not be defined by the very vague notion of belief. In their view, we have duties with respect to more subtle epistemic attitudes, which they define as commitment, or—especially in recent years—credence. These attitudes fall within the scope of our influence, and therefore, when deontologism is understood in this way, the problem identified by Alston does not arise (see, e.g., Tebben 2018). Credence is the subjective assessment of probability that a subject assigns to a given proposition. It could be argued that although we have no control over our outright beliefs, we do have some influence over our degrees of confidence (see Jackson 2019a; Gaultier 2020).⁹

5. Important issues to be addressed

Debates over the compatibilist approach highlight a deeper issue in the epistemology of belief that needs to be clearly articulated and thoroughly examined in order to adequately address the problem of doxastic freedom and responsibility. Our notion of freedom refers mainly to action, while beliefs are treated by epistemologists (e.g., when analyzing the notion of knowledge) as states. Consequently, some authors believe that only indirect doxastic voluntarism is possible, because only with respect to such types of actions as seeking new evidence can we speak of freedom (see, e.g., Audi

9. On the relationship between belief and credence, see Jackson (2020).

2001; Buckareff 2006a, 2006b). Others distinguish the moment of acceptance of a belief as appropriate for the attribution (or denial) of freedom. For example, Shah and Velleman distinguish belief as a doxastic attitude and judgment as a cognitive mental act of affirming a proposition (see Shah & Velleman 2005; also McHugh 2011). On the other hand, Sosa, in his extended epistemological approach, formulates a position on epistemic normativity in which the comparison of beliefs to actions (such as the archer's arrow) plays a key role, and he treats normativity in epistemology as a kind of performance normativity (see Sosa 2009, 2015).¹⁰

In recent years, there have been voices in the literature arguing that responsibility for beliefs should be modeled not on responsibility for actions, but on the subject's responsibility for her states, such as emotional ones (see Schmidt 2020). There are also voices suggesting that the problem of doxastic voluntarism should be considered in terms of the subject's freedom of intention (see, e.g., McHugh 2014, 2017; Flowerree 2017; for critical discussion see also Shepherd 2018).

Another key issue within the epistemology of belief, and one that plays a decisive role in addressing the problem of doxastic responsibility, concerns the question of what type of belief should be regarded as exemplary. Some defenders of doxastic voluntarism treat doxastic freedom as if the typical case of belief acquisition were the situation of weighing evidence for and against, with other cases being merely more automatic and less conscious variants of this process (e.g., Steup 2000). Other authors strongly object, arguing that conscious evaluation of evidence applies only to an extremely narrow range of beliefs and cannot be treated as the paradigmatic case—or even as a significant type of belief acquisition at all. According to them, what is most typical is the spontaneous, unreflective emergence of new beliefs in our minds (e.g., Plantinga 1993). It is likely that doxastic freedom and responsibility must be theoretically conceived differently in the case of reflective beliefs and in the case of spontaneous ones, and that any attempt to develop a single theory encompassing all types of beliefs we hold is destined to fail due to its incompleteness.

PART II

1. Towards a conception of doxastic voluntarism

As noted in the first part of the article, some authors argue that, just as the presence of reasons for action does not deprive those actions of freedom, beliefs should not be considered unfree merely because they are supported

10. On the criticism of Sosa's position, see, e.g., Chrisman (2020).

by evidence. This line of thought is developed in particular by advocates of a compatibilist approach to doxastic responsibility. On this view, the principle “ought implies can” requires only that the subject be capable of performing the action in question. The necessity of performing that action, they contend, does not undermine responsibility. Note, however, that even if there are extremely strong reasons for a certain action, the subject is still responsible for taking it because she may act otherwise. Even if it would be an extremely irrational choice, its realization remains possible for the subject. Meanwhile, the core of the intuition about the lack of doxastic freedom lies in the fact that the subject is incapable of consciously adopting beliefs against the reasons she perceives. Thus, there is a highly significant disanalogy between actions and the reasons for them, on the one hand, and beliefs and the evidence supporting them, on the other.

In defending the compatibilist conception of doxastic freedom, the aforementioned Steup contrasts compatibilism with libertarianism, according to which free choice means a completely arbitrary choice, unguided by anything. He then rightly rejects freedom so understood as the basis for formulating a conception of doxastic voluntarism (see Steup 2000). However, this is not the only available alternative. Incompatibilism can also mean that the subject has the ability to go against even strong reasons available to her—that is, she has the ability to act irrationally. Doxastic incompatibilism, understood in this way, may imply that it is possible for a subject to accept a belief in spite of, or independently from, the available evidence. However, the core problem with doxastic freedom arises from the fact that, due to the constitutive character of truth as the aim of belief, this is not possible.¹¹ This suggests, I think, that doxastic freedom should be sought in the phenomenon I call “doxastic self-deception.”

The idea of truth as the aim of belief, which also grounds the transparency thesis, suggests that, *prima facie*, if *S* states that a proposition *p* is true,

11. Freedom of action arises in situations of dilemma: i.e. when a subject has very strong reasons both for and against taking a given action and must decide which reasons to follow. It seems that in such cases rationality permits different choices on the part of the subject. The question, then, is whether we encounter analogous situations in epistemology. Defenders of the position known as permissivism argue that there are circumstances in which rationality allows for different doxastic attitudes, such as accepting the belief that *p* or suspending judgment about *p*. The problem of doxastic voluntarism appears to have significant implications for the debate on the plausibility of the permissivist thesis. One possible form of doxastic freedom could consist in the subject's capacity to choose among doxastic attitudes when faced with a similar body of evidence—for example, between accepting a belief and suspending judgment. On the relation between permissivism and the problem of doxastic voluntarism, see, e.g., Nickel (2010); Roeber (2019, 2020).

then *S* is not in a position not to believe it, and if *S* does not state that a proposition *p* is true, then *S* is not in a position to believe it. When the subject acts epistemically appropriately, it appears that she ultimately has no control over whether she adopts a particular belief, since her assertions must faithfully reflect the evidence she possesses. From this perspective, if doxastic freedom exists, it might manifest as a form of self-deception: the subject has evidence in favor of *p*, yet in accepting $\sim p$ or suspending judgment regarding *p*, she fails to recognize—or refuses to acknowledge—that she possesses such evidence. Following this reasoning, I begin my inquiry with an analysis of doxastic self-deception, a phenomenon that provides a crucial point of reference for the conception of doxastic voluntarism I defend.

2. Examples of doxastic self-deception

I will now present four examples of belief formation. Two of them illustrate epistemically proper belief acquisition, while the other two are analogous cases of doxastic self-deception, in which—as I will argue—doxastic freedom becomes apparent. Comparing these examples will enable us to draw conclusions, in particular concerning doxastic voluntarism.

Christopher and the Unhealthy Chickens (Low-Stakes Example)

Several leading media outlets have reported that chicken meat available in stores poses health risks—particularly for young children—due to a new strain of virus affecting poultry in the country. Christopher, the father of a small child who regularly eats chicken, responds to this news by saying to his wife: “The media have reported that chicken meat is harmful to children, so we should not give it to our child.” Christopher generally accepts media reports as reliable sources of information, although he is aware that sometimes the media are focused on sensationalism rather than providing accurate knowledge.

Andrew and the Unhealthy Chickens (High-Stakes Example)

Several leading media outlets have reported that chicken meat available in stores poses health risks—particularly for young children—due to a new strain of virus affecting poultry in the country. Andrew is the father of a small child with severe autism, which manifests, among other things, in extreme food selectivity. Andrew’s child eats only pasta with chicken. Andrew is aware that if he removes chicken from his child’s diet, the child will not receive adequate nutrition. Andrew says to his wife: “The media have reported that chicken meat is harmful to children. We need to check

this information in serious sources.” In other matters, Andrew generally accepts media reports as reliable sources of information.

Thomas and the Unhealthy Chickens (Example of Self-Deception Similar to Low Stakes)

Several leading media outlets have reported that chicken meat available in stores poses health risks—particularly for young children—due to a new strain of virus affecting poultry in the country. Thomas and his wife are vegetarians, but they must prepare chicken for dinner relatively often because their child insists on eating it. The parents are not happy with this situation. Thomas says to his wife: “The media have reported that chicken meat is harmful to children, so we should not give it to our child.” Thomas generally accepts media reports as reliable sources of information, although he is aware that sometimes the media are focused on sensationalism rather than providing accurate knowledge.

Gregory and the Unhealthy Chickens (Example of Self-Deception Similar to High Stakes)

Several leading media outlets have reported that chicken meat available in stores poses health risks—particularly for young children—due to a new strain of virus affecting poultry in the country. Gregory and his wife dislike cooking and do not want to devote time to it. They have learned to prepare a quick chicken dish that their child enjoys and which they usually serve. Gregory is unwilling to remove chicken from his child’s diet, since changing meals would require him to spend time on an activity he dislikes. Gregory says to his wife: “The media have reported that chicken meat is harmful to children. That is probably nonsense. Let’s wait until we see serious scientific evidence.” In other matters, Gregory generally accepts media reports as reliable sources of information.

3. Doxastic strategy

The cases outlined above are worth analyzing in terms of two dimensions of the epistemic goal. It is widely acknowledged that the epistemic goal—truth—can be realized in two ways: by acquiring true beliefs and by avoiding false ones. An exclusive emphasis on either dimension inevitably distorts our doxastic practices. One who aims only at maximizing the acquisition of true beliefs risks accepting an excess of falsehoods, whereas one who concentrates solely on avoiding error is led into skepticism. The proper pursuit of the epistemic goal therefore requires that the subject strike a balance between these two aspects in her cognitive life.

Christopher and Andrew possess the same evidence. Christopher is in a low-stakes situation. The evidence he has—namely, the media reports—he regards as sufficient to adopt the belief that chicken meat should be withdrawn, although in some situations his attitude toward the reliability of media reports is more cautious. Christopher is guided by the first aspect of the epistemic goal, namely the acquisition of true beliefs. Andrew, by contrast, is in a high-stakes situation. For him, it is very important not to accept a falsehood concerning the presence of chicken meat in his child's diet; therefore, he requires stronger evidence and, in the meantime, suspends judgment.

There are no strict epistemic rules determining how strong one's evidence must be in order to be justified in making an assertion. How many instances must I observe before I am entitled to generalize? How many hypotheses, and with what degree of scrutiny, must I consider before I can conclude which is the most plausible? How vivid must a memory be before it can ground a belief? How much, and what kind, of testimonial evidence is required for me to adopt a belief on its basis? The thesis of evidentialism contains the key characterization of evidence as "sufficient," but it does not specify what this precisely means. In each case, it is the subject who must determine whether the available evidence meets this standard. This gives rise to the threshold problem for beliefs. While analogous to the threshold problem for knowledge, the focus here is not on the standards for knowledge, but rather on the standards governing assertion.¹²

In determining how strong the evidence must be in a given case, the subject must be guided by the two aspects of the epistemic goal mentioned above. The balance between acquiring true beliefs and avoiding false ones is achieved by weighing satisfaction with the available evidence against exercising greater caution and demanding stronger evidence. In this way, in each instance of doxastic deliberation, the subject determines what counts as sufficient evidence. I refer to this as a "doxastic strategy."¹³

A doxastic strategy can be understood as an attitude of the subject that, under particular circumstances, guides them either to accept the available evidence as sufficient or to exercise greater caution and require stronger

12. Assuming the Knowledge Norm of Assertion (KNA), which states that one should only assert a proposition *p* if one knows that *p* is true (Williamson 2000), the standards governing assertion are subordinate to the standards governing knowledge. However, in my position I focus exclusively on the doxastic responsibility of the subject: that is, I am only concerned with how the situation appears from the subject's perspective. I do not address the question of whether the external conditions for knowledge are satisfied.

13. A similar idea can be found in Helm (1994).

evidence. While this process is most apparent during doxastic deliberation, I contend that it often operates in a more automatic and unconscious manner, even when the subject is not consciously attending to the strength of the evidence. An doxastic strategy is primarily a dispositional stance or an attitude of the subject that may, but does not necessarily, involve deliberate reflection on the evidence.

It can be said that our epistemic circumstances often force us to adopt a doxastic strategy that takes into account the subject's non-epistemic context, because, as I noted earlier: 1) the epistemic goal has two aspects that can pull the subject in opposite directions; 2) the subject must address both aspects, since focusing on only one leads to a distortion of cognition; 3) the evidentialist thesis does not specify precisely how strong the evidence needs to be to count as sufficient. A fourth reason can also be added: as skeptics have observed and, rightly, fallibilists take into account, the vast majority of our beliefs can never be completely infallible—that is, they may turn out to be false despite our best efforts.

4. Doxastic self-deception

Christopher finds himself in unremarkable circumstances, where little depends on whether he adopts a particular belief. He is likewise not in any particularly epistemically sensitive situation. He does not perceive either his own subject-related conditions or the surrounding environment as especially error-prone. Accordingly, he is content to rely on the available, generally reliable evidence. His epistemic stance prioritizes acquiring beliefs over avoiding falsehood. By contrast, Andrew faces high-stakes practical circumstances. Given that much depends on the adoption of his belief, he takes a very cautious epistemic stance, suspending judgment until stronger evidence becomes available. Due to his practical situation, his epistemic orientation leans more toward avoiding error than toward broadening his set of beliefs. In both cases, the doxastic strategies adopted by the men are oriented toward achieving the epistemic goal of truth, while also taking practical considerations into account. Neither Christopher's nor Andrew's epistemic attitudes raise concerns. Comparing their situations illustrates that epistemic normativity permits the influence of non-epistemic factors on a subject's evidential expectations, and thus on their doxastic strategy.

Gregory's case bears some resemblance to Andrew's high-stakes situation, as both are oriented toward the second aspect of the epistemic aim. Yet, unlike Andrew, Gregory's stance stems from an unwillingness to accept a particular belief rather than from a heightened concern for truth. Thomas's case, in turn, appears analogous to Christopher's. However, his reliance

on the first aspect of the epistemic aim is motivated by a desire to embrace the belief that chicken meat is unhealthy for non-epistemic reasons. These instances illustrate cases in which the influence of non-epistemic factors is inappropriate from the standpoint of epistemic normativity.

What distinguishes the cases of Thomas and Gregory from those of Christopher and Andrew? Intuitively, we evaluate the epistemic stance of Thomas and Gregory as improper. As in the other two cases, non-epistemic factors shape the doxastic strategies they adopt. The crucial difference, however, is that Thomas and Gregory select a strategy motivated by the desire to accept—or to avoid accepting—a *particular belief*, rather than by a genuine concern for truth in the matter at hand. Their stance, therefore, can aptly be characterized as a form of doxastic self-deception.

Building on the foregoing analyses, I propose to characterize epistemic self-deception as follows:¹⁴

- SD Low: S desires to hold the belief that *p*, and thus settles for relatively weak evidence in its favor, adopting an doxastic strategy oriented toward the first aspect of the epistemic aim (acquiring true beliefs); or
- SD High: S does not wish to hold the belief that *p*, and thus requires stronger evidence for it, adopting an doxastic strategy oriented toward the second aspect of the epistemic aim (avoiding errors).

In doxastic self-deception, then, the choice of doxastic strategy is guided by the subject's preferences concerning a particular belief, rather than by the pursuit of the epistemic aim of standing in the appropriate relation to the truth of that belief.

14. The literature contains a considerable number of philosophical treatments of self-deception, most of which aim to explicate the structure of the phenomenon and to resolve the paradoxes it generates. These discussions typically revolve around questions such as: Is self-deception an intentional act on the part of the subject, or rather a delusion driven by her desires? Does self-deception yield a genuine belief? Does the self-deceiver simultaneously entertain two contradictory beliefs? The most influential accounts have been developed by A. Mele and E. Funkhouser (see, e.g., Funkhouser 2019; Mele 2001; see also Baghrmian & Nicholson 2013). I do not aim to enter these debates in detail, for my research has a different focus. I intentionally employ the term *doxastic self-deception* rather than simply *self-deception*, as I am not concerned with the phenomenon in its everyday, colloquial sense. Instead, the notion I advance is broader, encompassing all situations in which a subject, in a biased manner, influences how she evaluates the evidence available to her in support of a given belief.

I term such situations “doxastic self-deception” because the subject, on the one hand, has preferences regarding the assertion (or suspension of judgment) of a particular belief, but in order for the belief to qualify as a mental state of an actual belief—rather than, for example, a desire—the subject must, in a sense, pretend to themselves that they are faithful to the evidence. Since the constitutive aim of beliefs is truth, the subject should adopt only those beliefs for which she has sufficient evidence. Yet our epistemic circumstances not only permit, but often require, the subject to adopt a doxastic strategy that takes into account both epistemic and practical factors. By manipulating the strategy according to personal bias, the subject can shape their beliefs in an intellectually dishonest manner. Given the nature of belief, this influence must occur through a form of self-deception.

5. Alethic impurism regarding beliefs

The cases of Christopher and Andrew demonstrate that assessing the required strength of evidence can legitimately involve considering the subject’s practical circumstances. Here, the issue of pragmatic reasons for belief comes into play: can non-epistemic factors influence whether a belief is adopted, or do only epistemic factors matter?¹⁵ Comparing the four examples discussed above leads to the following conclusion:

AI^I Pragmatic reasons may influence the doxastic strategy adopted by the subject—whether the subject expects stronger evidence or is satisfied with the available evidence—as long as the subject remains in an alethic stance, that is, cares about knowing the truth regarding the matter.

This thesis constitutes the central claim of the position I call “alethic impurism.”

The alethic impurism I aim to defend corresponds, to some extent, to impurism regarding the conditions of knowledge. A notable account of such impurism has been developed by Fantl and McGrath, who operate within an evidentialist framework—a framework that resonates with the present line of inquiry. Drawing on their analyses, I propose the following definition of evidentialist impurism:

15. Pragmatists about beliefs argue that there can be genuine practical reasons for beliefs, while opponents of this position argue that reasons for beliefs should always be of an exclusively evidential character (see, e.g., Reisner 2018; Rinard 2019; Sharadin 2018; Bondy 2019; Schmidt 2022).

EvI^K How high the probability of p given S 's evidence e must be in order for S to know p may vary with S 's practical circumstances.¹⁶

Alethic impurism, however, is a position solely regarding the standards of assertion, not of knowledge.¹⁷ Therefore, the thesis should properly be reformulated in the following form:

EvI^B How high the probability of p given S 's evidence e must be in order for S to *believe* p may vary with S 's practical circumstances.

Comparing the cases of Christopher and Andrew with those of Thomas and Gregory reveals the need to add to the above formulation a qualification concerning the subject's proper epistemic stance. With this modification, we can articulate a second formulation of the alethic impurism thesis:

All² How high the probability of p given S 's evidence e must be in order for S to believe p may vary with S 's practical circumstances, insofar as the subject remains in an alethic stance.

According to the alethic impurist view, non-epistemic factors may influence the formation of beliefs, but only under specific conditions: (1) exclusively by shaping the doxastic strategy the subject adopts—namely, the required strength of evidence she demands; and (2) only so long as the subject remain in an alethic stance. Thus, according to alethic impurism, Thomas and Gregory fail to act in an epistemically proper manner, for in choosing their doxastic strategies they do not remain oriented toward truth. Ultimately, only evidence can serve as a reason for belief, though non-epistemic factors may affect how strong the evidence must be, provided the subject's aim remains the pursuit of truth in the matter at hand.¹⁸

The two conditions that must be met for non-epistemic circumstances to legitimately influence the subject's assertion make alethic impurism compatible with the transparency thesis—for, within this framework, *the deliberative question whether to believe that p is transparent to the question*

16. See Fantl and McGrath (2009).

17. I agree with Jennifer Nagel that the arguments advanced by impurists pertain to the standards of assertion rather than to those of knowledge. See Nagel (2008).

18. The analyses of alethic impurism presented here can be used to defend the position of permissivism: i.e. the claim that the same set of evidence can justify more than one epistemic attitude.

whether p. As cited above, McHugh reformulates the transparency thesis in the following way:

TrT² Pragmatic considerations cannot occur to a thinker, within doxastic deliberation, as relevant to *what* to believe (McHugh 2013, 448; my emphasis).

Alethic impurism aligns with this thesis, yet it is also compatible with:

All³ Pragmatic considerations can occur to a thinker, within doxastic deliberation, as relevant to *whether* to believe.

Pragmatic factors can affect whether a subject will adopt a belief. For example, in a high-risk situation a subject might withhold belief because their doxastic strategy prioritizes avoiding errors. Pragmatic considerations, however, cannot serve as a reason to adopt a strategy that inherently favors a specific belief. In other words, the subject must remain unbiased. The alethic attitude I describe¹⁹ can also be understood simply as intellectual honesty.²⁰

6. Doxastic freedom as doxastic self-control

Alethic impurism yields significant insights regarding doxastic responsibility. First, our epistemic conditions make us not function purely as

19. As one of the reviewers has noted, an interesting comparison can be made between my position and the distinction proposed by Williamson in “Justifications, Excuses, and Sceptical Scenarios” (forthcoming). Williamson introduces three interdependent norms: “Let J be a truth-related norm of belief. Then DJ is the norm of being the sort of person who complies with J, and ODJ is the norm of doing in the given situation what the sort of person who complies with J would do.” (p. 12) There are two important truth-related norms for belief that I consider in this article. The first is the evidentialist norm of adjusting one’s beliefs to one’s evidence. As I argue, when analyzing the cases of doxastic self-deception, compliance with this norm is not sufficient for an agent to be a doxastically responsible. Alethic impurism also requires that the agent be in the alethic stance. The requirement concerns a disposition of the subject and, following Williamson, is thus a secondary norm, DJ. This suggests that there is some truth-related norm J, of which DJ is derivative, that is more fundamental than the evidentialist norm, and my examples of epistemic self-deception provide some support for this claim. One plausible candidate for that norm is to believe only what one knows. It should be emphasized, however, that my considerations concern doxastic responsibility rather than the epistemic justification, so they are related to Williamson’s analysis, but Williamson goes beyond the scope of my reflection when he examines the epistemic status of beliefs in the brain-in-a-vat scenario.

20. For interesting reflections on intellectual humility related to the issues addressed in the article, see Carter and Gordon (2020); Tanesini (2020).

mechanisms in the epistemic domain. Rather, we act like agents—at least to a certain extent, autonomously setting our goals (whether to adopt beliefs or avoid error) and determining how to achieve them (by specifying the required strength of evidence). Situations in which this is clearly apparent also demonstrate that, in the epistemic realm, we are subject—at least in part—to normative evaluation appropriate for rational, free agents. Second, we are capable of doxastic self-deception, which consists in the subject acting much as she would in proper belief formation, and yet—so to speak—deceiving herself by concealing from herself the true motives driving her actions. This suggests that in the epistemic sphere, under certain circumstances, one can speak of the subject’s “doxastic blame.” Third, properly fulfilling one’s epistemic duties crucially depends on the subject’s alethic attitude, which I have also referred to as “intellectual honesty.” This is a subjective factor in the sense that it cannot be pinned down by strict, intersubjectively measurable conditions, such as the strength of evidence. Furthermore, it is theoretically possible that the assessment of doxastic responsibility for subjects in two situations could differ radically, even if in both cases externally observable conditions were identical, provided that the subjects exhibited different epistemic attitudes.²¹

The cases of Thomas and Gregory show that, depending on one’s preferences, a subject can influence which beliefs she adopts. If, for some reason, a particular belief does not suit her, she can continually disregard the evidence and expect an ever stronger evidence for the relevant belief, whereas if the belief particularly satisfies her, she can quickly settle for the available evidence. In this way, the subject can, to some extent, manipulate her beliefs while still grounding them in evidence, which is a condition for belief as such. But can the ability to self-deceive be considered an instance of doxastic freedom?

The notion of freedom involves two fundamental intuitions: conscious control and the ability to act otherwise. In the case of doxastic self-deception,

21. The position presented here provides the tools for an interesting interpretation of a classic text by William James, entitled “The Will to Believe.” This article is often interpreted as contradicting the evidentialist thesis and doxastic responsibility, or denying that truth is the aim of belief. Alethic impurism provides a way of interpreting James’ position in such a way as to resist these objections. The believer and non-believer could adopt different strategies concerning religious beliefs. One could be more focused on increasing the chance of having true beliefs, and the other could find it more important to protect himself from errors. Following our “passions,” which James writes about, can be understood as taking up one of these two strategies, depending on the non-epistemic factors the subject is placed in. For this way of interpreting James’ position, see Odoj (2014), and for a similar interpretation of James’ position, see also Pace (2011).

the condition of consciously undertaking an action and controlling it cannot be fulfilled. The subject, in a sense, must conceal from herself that she is influencing her own belief-formation process. She is at once both the deceiver and the deceived. From this perspective, understanding doxastic voluntarism through the lens of doxastic self-deception appears irreconcilable with this intuition. Second, in line with the principle “ought implies can,” a free action is one that the subject could have performed differently, though she was not determined to do so. Put differently, a free action is one that could have been otherwise. As repeatedly noted above, this intuition appears at odds with the notion of doxastic freedom itself. I think, however, that a closer analysis reveals how both intuitions can be reconciled with the conception of doxastic freedom I defend.

The cases of Thomas and Gregory could have turned out differently if they had reflected on their own motives and the circumstances they found themselves in, leading them, contrary to their initial impulse, to adjust their stance. In doing so, they would have employed their capacity for metacognition—simplified, the ability to think about one’s own thinking. Metacognition is a higher-order cognitive ability that develops in humans (and, to some extent, in certain animals). Joëlle Proust defines it as follows: “Metacognition is the set of capacities through which an operating cognitive subsystem is evaluated or represented by another subsystem in a context-sensitive way” (Proust 2013, 4).

Doxastic freedom is possible because we are capable of doxastic self-deception: that is, of acting in ways that are epistemically improper, even culpable.²² Metacognition equips us with a capacity for self-control that, to some extent and within the limits of its development, allows the subject to act otherwise—to correct her stance and proceed in an epistemically appropriate way. Doxastic freedom, then, does not lie in choosing whichever beliefs one prefers, but in the ability to regulate the processes responsible for belief formation so that they accord with the norms of assertion. It arises because, while we are prone to doxastic self-deception, we are also able to monitor and control ourselves through metacognition. To that extent, which beliefs we hold does depend on us, and the conception of doxastic freedom defended here remains consistent with the principle that “ought implies can.” While the subject’s influence on her beliefs through doxastic self-deception cannot occur consciously, the capacity for regulating her own

22. A similar intuition to the effect that what I call ‘doxastic self-deception’ is an example of doxastic freedom can be found in Booth (2007) and in Funkhouser (2003); see also Adler (2002) and McCormick (2015).

cognitive processes can indeed be exercised in a deliberate and controlled way. Metacognition to a certain extent enables the subject to recognize her own biases. The conception of doxastic freedom I have defended is thus consistent with the two general intuitions about freedom discussed above. This, in turn, provides grounds for speaking of doxastic responsibility in the sense that Alston criticizes.

If responsibility requires the possibility of influence, then, on the present account, the scope of doxastic responsibility is defined by the extent of a subject's metacognitive development. Thus, the doxastic responsibility of a young child or of a person with significant cognitive impairments is diminished—or perhaps absent altogether—in comparison with that of a mature adult with well-developed metacognitive abilities. Furthermore, doxastic responsibility, so understood, varies in accordance with individual cognitive differences among persons.²³

7. *Objections*

Serious objections can be raised against the conception advanced here. One might contend, for example, that there is no substantive difference between unmotivated delusion or cognitive bias and motivated irrationality, i.e. self-deception.²⁴ Likewise, it could be argued that invoking the phenomenon of self-deception (or, more broadly, motivated irrationality) allows, at most, for the isolation of a limited class of beliefs that remain resistant to Alston's objection. After all, not all of our beliefs can plausibly be construed as potential objects of self-deception. Perceptual beliefs, for instance, seem to arise in a way that precludes such influence on the part of the subject.

In the position I defend, however, the crucial element is the component of doxastic self-control through metacognitive abilities. Even if, in a given case, it is difficult to unambiguously identify the subject's bias, for doxastic responsibility it suffices to point to the subject's capacity for exercising self-control over her own cognitive processes. If such capacities are present, they provide the subject with the possibility of influence, whether the source of error lies in an improper epistemic attitude—on which the preceding analyses have focused—or in an innocent mistake, such as an error made

23. The phenomenon of doxastic self-deception might also be analyzed in terms of subjective probability and the propositional attitude called 'credence.' For interesting remarks on how the notion of degrees of belief affects the arguments against the evidentialism thesis formulated by proponents of pragmatic encroachment, see Ganson (2008); Jackson (2019b).

24. On the relationship between delusions, biased beliefs and self-deception, see Bayne and Fernandez (2010).

in overly hasty calculation. Cases in which a subject's bias is clearly visible serve as useful illustrations of a phenomenon that, within the complexities of life for beings such as ourselves, may manifest with varying degrees of clarity. For that reason, such cases are valuable in offering epistemological insight into important questions of doxastic freedom and responsibility. This does not mean, however, that doxastic responsibility applies only to cases as clear-cut as those exemplifying doxastic self-deception.²⁵

The conception of doxastic responsibility developed in this article seems to presuppose a relatively robust understanding of doxastic blame. One potential objection is that we frequently engage in self-deception when it serves to achieve a positive outcome, such as personal well-being or growth. A typical example is self-deception regarding one's state of health: a person might believe they are healthier than they actually are, which can help them cope with the challenges of daily life. In response, it should be emphasized that the analyses presented here concern solely epistemic duties—also referred to as “intellectual duties”—which represent only one set of the many responsibilities associated with human life. Duties often conflict with one another. Even if self-deception about one's health could be construed as blameworthy in relation to a person's epistemic duties, it is evident that, all things considered, the individual should not be blamed, as she is fulfilling other, arguably more pressing personal duties, such as maintaining her well-being.

One might raise a doubt as to whether, in the examples of doxastic self-deception described, the men in fact hold an outright belief—assert something—rather than being in some other attitude such as, for instance, acceptance or merely a disposition to act (as in a situation in which I do not remember whether I locked the door, but in order not to waste time going back I assume that I did and simply drive on).²⁶ For my argument, the cases of Thomas and Gregory are particularly important, although this issue may concern all four men. I think that, from a psychological point of view, every option is possible in each of the four types of case and to varying degrees: a person's attitude may range from full assertion, through partial

25. On the subject's responsibility for self-deception and more broadly ill-formed beliefs, see Holroyd, Scaife, and Stafford (2017); Ellis (2022); McHugh and Davison (2020); Dominguez (2020); Levy (2014, 2017); Sie and Voorst Vader-Bours (2016); Madva (2016); Washington and Kelly (2016); Frankish (2016). For critical discussion, see Bortolotti (2020). On the impact of biases on our belief-forming processes, see Siegel (2020). On the implications of positions on the structure of self-deception present in the literature for the problem of responsibility for self-deception, see Nelkin (2012).

26. I am grateful to an anonymous reviewer for drawing my attention to this point.

credence combined with a certain disposition to act, all the way to a mere disposition to act *as if* a given proposition were true, without asserting that it is in fact true. For my purposes, it is sufficient that cases of outright belief in two types of self-deception situation are psychologically plausible. From the perspective of the alethic impurism I defend, it is in fact desirable that the subject adopts only a certain disposition to act, rather than an outright belief. If this is so, then either they spontaneously display the appropriate alethic attitude, or they have managed to adopt this attitude by mastering their competing tendencies, thanks to their metacognitive capacities.

One might also object that my position does not differ from indirect doxastic voluntarism.²⁷ I do not dispute that exercising indirect control over our beliefs constitutes an important aspect of doxastic responsibility. Nevertheless, I contend that if epistemic responsibility were reduced solely to indirect doxastic voluntarism, a significant problem would emerge. As fallibilists rightly emphasize, almost none of our beliefs can be regarded as absolutely indubitable. There is always—even if only to the slightest degree—the possibility that new evidence will arise, or that the evidence we currently possess is flawed. Thus, if we were to assume that there is a duty to examine the available evidence and seek new evidence until absolute certainty is achieved, we would inevitably lapse into skepticism. Proponents of indirect doxastic voluntarism are thus obliged to clarify under what conditions a subject bears a duty of deeper inquiry. This raises the further question of what it means for evidence to be sufficient. When is it appropriate to trust evidence, and when must it be verified? This question marks the starting point of my investigation: I aim to show that the problem of doxastic responsibility ultimately depends on a more fundamental factor—the subject's alethic stance. This stance cannot be reduced to additional acts, such as the verification of evidence or seeking for new evidence. Rather, the appropriate execution of such supplementary acts depends precisely on the underlying alethic attitude of the subject.

SUMMARY

This paper has proposed an approach to understanding doxastic voluntarism. The position can be characterized as strong, in the sense that the concept of doxastic freedom I defend is compatible with Kant's principle of "ought implies can," understood in an incompatibilist manner. To the extent that the nature of the cognitive domain allows, I have sought to preserve the intuition that the concept of freedom entails the principle

27. I am grateful to a second anonymous reviewer for drawing my attention to this point.

of alternative possibilities. In this way, my account offers a defense of epistemic deontologism against the objection raised by Alston. Furthermore, it is sufficiently robust to support the attribution of doxastic responsibility, and even doxastic blame.

In developing this conception, I seek to reconcile both the intuition underlying the conceptual argument against doxastic voluntarism—with its strong emphasis on the constitutive role of truth as the aim of belief—and the intuitions supporting the existence of pragmatic reasons for belief. This is possible because I use the phenomenon of doxastic self-deception, with its paradoxical structure, as a foundation for articulating the concept of doxastic freedom. At the same time, I indicate that the sources of doxastic responsibility should be sought in the subject's metacognitive abilities, which enable self-control of one's own cognitive processes.

REFERENCES

- Adler, Jonathan. 2002. *Belief's Own Ethics*. Cambridge, MA: MIT Press.
- Alston, William P. 1988. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2: 257–99. <https://doi.org/10.2307/2214077>.
- Altschul, Jon. 2014. "Epistemic Deontologism and Role-Oughts." *Logos & Episteme* 5 (3): 245–63. <https://doi.org/10.5840/logos-episteme2014531>.
- Audi, Robert. 2001. "Doxastic Voluntarism and the Ethics of Belief." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 93–111. Oxford: Oxford University Press.
- . 2008. "The Ethics of Belief: Doxastic Self-Control and Intellectual Virtue." *Synthese* 161 (3): 403–18. <https://doi.org/10.1007/s11229-006-9092-2>.
- Baghramian, Maria, and Anna Nicholson. 2013. "The Puzzle of Self-Deception." *Philosophy Compass* 8 (11): 1018–29. <https://doi.org/10.1111/phc3.12083>.
- Bayer, Benjamin. 2015. "The Elusiveness of Doxastic Compatibilism." *American Philosophical Quarterly* 52 (3): 233–51.
- Bayne, Tim, and Jordi Fernandez, eds. 2010. *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*. New York: Psychology Press.
- Bergeron, Melissa. 2006. "The Ethics of Belief: Conservative Belief Management." *Social Epistemology* 20 (1): 67–78. <https://doi.org/10.1080/02691720500512291>.
- Boespflug, Mark, and Elizabeth Jackson. 2024. "Doxastic Voluntarism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2024 ed. Stanford, CA: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2024/entries/doxastic-voluntarism/>. Accessed November 20, 2025.
- Bondy, Patrick. 2019. "The Epistemic Norm of Inference and Non-Epistemic Reasons for Belief." *Synthese* 198 (2): 1761–81. <https://doi.org/10.1007/s11229-019-02163-3>.
- Booth, Anthony R. 2007. "Doxastic Voluntarism and Self-Deception." *Disputatio* 2 (22): 115–29. <https://doi.org/10.2478/disp-2007-0003>.
- . 2009. "Compatibilism and Free Belief." *Philosophical Papers* 38 (1): 1–12. <https://doi.org/10.1080/05568640902933379>.
- . 2012. "All Things Considered Duties to Believe." *Synthese* 187: 509–17. <https://doi.org/10.1007/s11229-010-9857-5>.

- . 2014. "On Some Recent Moves in Defence of Doxastic Compatibilism." *Synthese* 191 (8): 1867–80. <https://doi.org/10.1007/s11229-013-0378-x>.
- Bortolotti, Lisa. 2020. *The Epistemic Innocence of Irrational Beliefs*. Oxford: Oxford University Press.
- Buckareff, Andrei A. 2006a. "Compatibilism and Doxastic Control." *Philosophia* 34 (2): 143–52. <https://doi.org/10.1007/s11406-006-9013-0>.
- . 2006b. "Doxastic Decisions and Controlling Belief." *Acta Analytica* 21 (1): 102–14. <https://doi.org/10.1007/s12136-006-1017-7>.
- . 2008. "Action and Doxastic Control: The Asymmetry Thesis Revisited." *European Journal of Analytic Philosophy* 4 (1): 5–12.
- Carter, Adam, and Emma Gordon. 2020. "Intellectual Humility and Assertion." In *The Routledge Handbook of Philosophy of Humility*, edited by Mark Alfano, Michael P. Lynch, and Alessandra Tanesini, 335–45. London: Routledge.
- Chignell, Andrew. 2018. "The Ethics of Belief." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018 ed. Stanford, CA: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/entries/ethics-belief/>. Accessed November 20, 2025.
- Chrisman, Matthew. 2008. "Ought to Believe." *Journal of Philosophy* 105 (7): 346–70.
- . 2020. "Performance Normativity and Here-And-Now Doxastic Agency." *Synthese* 197 (12): 5137–45. <https://doi.org/10.1007/s11229-017-1641-3>.
- Chuard, Philippe, and Nicholas Southwood. 2009. "Epistemic Norms without Voluntary Control." *Noûs* 43 (4): 599–32. <https://doi.org/10.1111/j.1468-0068.2009.00721.x>.
- Clifford, William K. 2002. "The Ethics of Belief." In *The Theory of Knowledge: Classic and Contemporary Readings*, 3rd ed., edited by Louis P. Pojman, 515–18. Belmont, CA: Wadsworth.
- Cohen, L. Jonathan. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.
- Cohen, Stewart. 2016. "Reasons to Believe and Reasons to Act." *Episteme* 13 (4): 427–38. <https://doi.org/10.1017/epi.2016.22>.
- Conee, Earl, and Richard Feldman. 2004. *Evidentialism: Essays in Epistemology*. Oxford: Clarendon Press.
- Dominguez, Noel. 2020. "Moral Responsibility for Implicit Bias." In *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*, edited by Erin Beeghly and Alex Madva, 153–73. New York: Routledge.
- Ellis, Jon. 2022. "Motivated Reasoning and the Ethics of Belief." *Philosophy Compass* 17 (6): e12828. <https://doi.org/10.1111/phc3.12828>.
- Fantl, Jeremy, and Matthew McGrath. 2009. *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
- Feldman, Richard. 2001. "Voluntary Belief and Epistemic Evaluation." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 77–92. Oxford: Oxford University Press.
- . 2006. "Clifford's Principle and James's Options." *Social Epistemology* 20 (1): 19–33. <https://doi.org/10.1080/02691720600631645>.
- . 2008. "Modest Deontologism in Epistemology." *Synthese* 161: 339–55. <https://doi.org/10.1007/s11229-006-9088-y>.
- Flowerree, A.K. 2017. "Agency of Belief and Intention." *Synthese* 194 (8): 2763–84. <https://doi.org/10.1007/s11229-016-1138-5>.
- Frankish, Keith. 2016. "Playing Double: Implicit Bias, Dual Levels, and Self-Control." In *Implicit Bias and Philosophy*, vol. 1, edited by Michael Brownstein and Jennifer Saul, 23–46. Oxford: Oxford University Press.

- Funkhouser, Eric. 2003. "Willing Belief and the Norm of Truth." *Philosophical Studies* 115 (2): 179–95. <https://doi.org/10.1023/A:1025094823262>.
- . 2019. *Self-Deception*. London: Routledge.
- Ganson, Dorit. 2008. "Evidentialism and Pragmatic Constraints on Outright Belief." *Philosophical Studies* 139 (3): 441–58. <https://doi.org/10.1007/s11098-007-9133-9>.
- Gaultier, Benoit. 2020. "Responsibility for Doxastic Strength Grounds Responsibility for Belief." In *The Ethics of Belief and Beyond: Understanding Mental Normativity*, edited by Sebastian Schmidt and Gerhard Ernst, 149–75. London: Routledge.
- Ginet, Carl. 2001. "Deciding to Believe." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 63–76. Oxford: Oxford University Press.
- Haack, Susan. 2001. "'The Ethics of Belief' Reconsidered." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, 21–33. Oxford: Oxford University Press.
- Helm, Paul. 1994. *Belief Policies*. Cambridge: Cambridge University Press.
- Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87 (1): 45–74. <https://doi.org/10.1111/j.1468-0114.2006.00247.x>.
- . 2009. "Believing at Will." *Canadian Journal of Philosophy* 35: 149–87.
- Holroyd, Jules, Robin Scaife, and Tom Stafford. 2017. "Responsibility for Implicit Bias." *Philosophy Compass* 12 (11): e12410. <https://doi.org/10.1111/phc3.12410>.
- Jackson, Elizabeth. 2019a. "Belief and Credence: Why the Attitude-Type Matters." *Philosophical Studies* 176 (9): 2477–96. <https://doi.org/10.1007/s11098-018-1136-1>.
- . 2019b. "How Belief-Credence Dualism Explains Away Pragmatic Encroachment." *Philosophical Quarterly* 69 (276): 511–33. <https://doi.org/10.1093/pq/pqz006>.
- . 2020. "The Relationship between Belief and Credence." *Philosophy Compass* 15 (6): e12668. <https://doi.org/10.1111/phc3.12668>.
- James, William. 2002. "The Will to Believe." In *The Theory of Knowledge: Classic and Contemporary Readings*, 3rd ed., edited by Louis P. Pojman, 519–26. Belmont, CA: Wadsworth.
- Kruse, Andrea. 2017. "Why Doxastic Responsibility Is Not Based on Direct Doxastic Control." *Synthese* 194 (8): 2811–42. <https://doi.org/10.1007/s11229-015-0951-6>.
- Levy, Neil. 2014. "Consciousness, Implicit Attitudes, and Moral Responsibility." *Noûs* 48 (1): 21–40. <https://doi.org/10.1111/j.1468-0068.2011.00853.x>.
- . 2017. "Implicit Bias and Moral Responsibility: Probing the Data." *Philosophy and Phenomenological Research* 94 (1): 3–26. <https://doi.org/10.1111/phpr.12352>.
- Madva, Alex. 2016. "Virtue, Social Knowledge, and Implicit Bias." In *Implicit Bias and Philosophy*, vol. 1, edited by Michael Brownstein and Jennifer Saul, 191–215. Oxford: Oxford University Press.
- McCormick, Miriam Schleifer. 2015. *Believing Against the Evidence: Agency and the Ethics of Belief*. New York: Routledge.
- McHugh, Conor. 2011. "Judging as a Non-Voluntary Action." *Philosophical Studies* 152 (2): 245–69. <https://doi.org/10.1007/s11098-009-9478-3>.
- . 2012. "Epistemic Deontology and Voluntariness." *Erkenntnis* 77 (1): 65–94. <https://doi.org/10.1007/s10670-011-9299-6>.
- . 2013. "Normativism and Doxastic Deliberation." *Analytic Philosophy* 54 (4): 447–65. <https://doi.org/10.1111/phib.12030>.
- . 2014. "Exercising Doxastic Freedom." *Synthese* 191 (11): 2745–62. <https://doi.org/10.1111/j.1933-1592.2011.00531.x>.

- . 2017. "Attitudinal Control." *Synthese* 194 (8): 2745–62. <https://doi.org/10.1007/s11229-014-0643-7>.
- McHugh, Nancy A., and Lacey J. Davidson. 2020. "Epistemic Responsibility and Implicit Bias." In *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*, edited by Erin Beeghly and Alex Madva, 174–90. New York: Routledge.
- Mele, Alfred R. 2001. *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.
- Montmarquet, James. 1986. "The Voluntariness of Belief." *Analysis* 46 (1): 49–53. <https://doi.org/10.1093/analysis/46.1.49>.
- . 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield.
- . 2008a. "Virtue and Voluntarism." *Synthese* 161 (3): 393–402. <https://doi.org/10.1007/s11229-006-9091-3>.
- . 2008b. "The Voluntariness of Virtue—and Belief." *Philosophy* 83 (3): 373–90. <https://doi.org/10.1017/S0031819108000739>.
- Nagel, Jennifer. 2008. "Knowledge Ascriptions and the Psychological Consequences of Changing Stakes." *Australasian Journal of Philosophy* 86 (2): 279–94. <https://doi.org/10.1080/00048400801886397>.
- Nelkin, Dana Kay. 2012. "Responsibility and Self-Deception: A Framework." *Humana.Mente Journal of Philosophical Studies* 20: 117–39.
- Nickel, Philip. 2010. "Voluntary Belief on a Reasonable Basis." *Philosophy and Phenomenological Research* 81 (2): 312–34. <https://doi.org/10.1111/j.1933-1592.2010.00380.x>.
- Nottelmann, Nikolaj. 2007. *Blameworthy Belief: A Study in Epistemic Deontology*. Dordrecht: Springer.
- Odoj, Ewa. 2014. "Is 'The Presumption of Atheism' in Fact a Neutral Procedure? A Critical Examination of Antony Flew's Position." *Roczniki Filozoficzne* 62 (2): 115–32.
- Pace, Michael. 2011. "The Epistemic Value of Moral Considerations: Justification, Moral Encroachment, and James' 'Will to Believe.'" *Noûs* 45 (2): 239–68. <https://doi.org/10.1111/j.1468-0068.2010.00768.x>.
- Peels, Rik. 2014. "Against Doxastic Compatibilism." *Philosophy and Phenomenological Research* 89 (3): 679–702. <https://doi.org/10.1111/phpr.12040>.
- . 2017a. *Responsible Belief: A Theory in Ethics and Epistemology*. Oxford: Oxford University Press.
- . 2017b. "Responsible Belief and Epistemic Justification." *Synthese* 194 (8): 2895–2915. <https://doi.org/10.1007/s11229-016-1038-8>.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.
- Proust, Joëlle. 2013. *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford: Oxford University Press.
- Reisner, Andrew. 2018. "Pragmatic Reasons for Belief." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 705–29. Oxford: Oxford University Press.
- Rinard, Susanna. 2019. "Believing for Practical Reasons." *Noûs* 53 (4): 763–84. <https://doi.org/10.1111/nous.12253>.
- Roeber, Blake. 2016. "Reasons to Not Believe (and Reasons to Act)." *Episteme* 13 (4): 439–48. <https://doi.org/10.1017/epi.2016.23>.
- . 2019. "Evidence, Judgment, and Belief at Will." *Mind* 128 (511): 837–59.
- . 2020. "Permissive Situations and Direct Doxastic Control." *Philosophy and Phenomenological Research* 101 (2): 415–31. <https://doi.org/10.1111/phpr.12594>.
- Ryan, Sharon. 2003. "Doxastic Compatibilism and the Ethics of Belief." *Philosophical Studies* 114 (1–2): 47–79.

- Schmidt, Sebastian. 2020. "Responsibility for Attitudes, Object-Given Reasons, and Blame." In *The Ethics of Belief and Beyond: Understanding Mental Normativity*, edited by Sebastian Schmidt and Gerhard Ernst, 149–75. London: Routledge.
- . 2022. "On Believing Indirectly for Practical Reasons." *Philosophical Studies* 179 (6): 1795–1819. <https://doi.org/10.1007/s11098-021-01730-0>.
- Schmitt, Margaret. 2015. "Freedom and (Theoretical) Reason." *Synthese* 192 (1): 25–41. <https://doi.org/10.1007/s11229-014-0547-6>.
- Setiya, Kieran. 2008. "Believing at Will." *Midwest Studies in Philosophy* 32 (1): 36–52. <https://doi.org/10.1111/j.1475-4975.2008.00164.x>.
- Shaffer, Michael J. 2006. "The Publicity of Belief, Epistemic Wrongs and Moral Wrongs." *Social Epistemology* 20 (1): 41–54. <https://doi.org/10.1080/02691720500512440>.
- Shah, Nishi, and J. David Velleman. 2005. "Doxastic Deliberation." *Philosophical Review* 114 (4): 497–34. <https://doi.org/10.1215/00318108-114-4-497>.
- Shah, Nishi. 2002. "Clearing Space for Doxastic Voluntarism." *The Monist* 85 (3): 436–45.
- . 2003. "How Truth Governs Belief." *Philosophical Review* 112 (4): 447–82.
- . 2006. "A New Argument for Evidentialism." *Philosophical Quarterly* 56 (225): 481–98.
- Sharadin, Nathaniel. 2018. "Epistemic Instrumentalism and the Reason to Believe in Accord with the Evidence." *Synthese* 195 (9): 3791–809. <https://doi.org/10.1007/s11229-016-1245-3>.
- Shepherd, Joshua. 2018. "Intending, Believing, and Supposing at Will." *Ratio* 31 (3): 321–30. <https://doi.org/10.1111/rati.12198>.
- Sie, Maureen, and Nicole Voort Vader-Bours. 2016. "Stereotypes and Prejudices: Whose Responsibility? Indirect Personal Responsibility for Implicit Biases." In *Implicit Bias and Philosophy*, vol. 2, edited by Michael Brownstein and Jennifer Saul, 90–114. Oxford: Oxford University Press.
- Siegel, Susanna. 2020. "Bias and Perception." In *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*, edited by Erin Beeghly and Alex Madva, 99–115. London: Routledge.
- Sosa, Ernest. 2009. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, vol. 1. Oxford: Oxford University Press.
- . 2015. *Judgment and Agency*. Oxford: Oxford University Press.
- Steup, Matthias. 2000. "Doxastic Voluntarism and Epistemic Deontology." *Acta Analytica* 15 (1): 25–56.
- . 2008. "Doxastic Freedom." *Synthese* 161 (3): 375–92. <https://doi.org/10.1007/s11229-006-9090-4>.
- . 2012. "Belief Control and Intentionality." *Synthese* 188 (2): 145–63. <https://doi.org/10.1007/s11229-011-9919-3>.
- . 2017. "Believing Intentionally." *Synthese* 194 (8): 2673–94. <https://doi.org/10.1007/s11229-015-0780-7>.
- Sullivan-Bissett, Ema. 2018. "Explaining Doxastic Transparency: Aim, Norm, or Function?" *Synthese* 195 (8): 3453–76. <https://doi.org/10.1007/s11229-017-1377-0>.
- Tanesini, Alessandra. 2020. "Humility and Self-Knowledge." In *The Routledge Handbook of Philosophy of Humility*, edited by Mark Alfano, Michael P. Lynch, and Alessandra Tanesini, 292–302. London: Routledge.
- Tebben, Nicholas. 2014. "Deontology and Doxastic Control." *Synthese* 191 (12): 2835–47. <https://doi.org/10.1007/s11229-014-0423-4>.
- . 2018. "Belief Isn't Voluntary, but Commitment Is." *Synthese* 195 (3): 1163–79. <https://doi.org/10.1007/s11229-016-1258-y>.

- Wagner, Verena. 2017. "On the Analogy of Free Will and Free Belief." *Synthese* 194 (8): 2785–810. <https://doi.org/10.1007/s11229-015-0851-9>.
- Washington, Natalia, and Daniel Kelly. 2016. "Who's Responsible for This? Moral Responsibility, Externalism and Knowledge about Implicit Bias." In *Implicit Bias and Philosophy*, vol. 2, edited by Michael Brownstein and Jennifer Saul, 11–36. Oxford: Oxford University Press.
- Williams, Bernard. 1973. "Deciding to Believe." In *Problems of the Self: Philosophical Papers 1956–1972*, 136–51. Cambridge: Cambridge University Press.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
- . Forthcoming. "Justifications, Excuses, and Sceptical Scenarios." In *The New Evil Demon: New Essays on Knowledge, Justification and Rationality*, edited by Julien Dutant and Fabian Dorsch. Oxford: Oxford University Press.
- Winters, Barbara. 1979. "Believing at Will." *Journal of Philosophy* 76 (5): 243–56.

Reading as an Affective and Discursive Event

Its Contribution to Reshaping Human Identity


Robert Grzywacz


ABSTRACT This paper deals with a twofold understanding of the notion of *event*. The first construal of the latter draws upon the philosophical framework of Marc Richir, in which that concept *event* corresponds to a process of phenomenalization occurring within a schematism that serves as a transcendental matrix for individual phenomena. It enables access to a sphere of fluctuating phenomena correlated with the non-intentional activity of *phantasia*, which precedes their symbolic institution. According to Richir, the experience of reading literature can exemplify this kind of phenomenalization, one that activates interaction with human affective experience.

The second concept of *event* referred to in the study is derived from the thought of Paul Ricoeur, and may be characterized as a discursive one, since the latter thinker emphasizes the transcendence of the merely event-referring dimension of discourse in favour of the meaning it conveys. In his elaborated theory of reading, Ricoeur describes the process as both active and passive: a wandering point of view on the world opened up by the text, a dynamic synthesis of sentential retentions and protentions, a bidirectional modification of the reader's expectations and memories, a search for meaning and a struggle with its absence, and a breakdown and reconstitution of narrative coherence. Yet Ricoeur's category of the *world* of the text appears to suggest a certain kind of symbolic and ontological institution. At the stage of existential appropriation of textual proposals, these proposals are directed toward the imagination (operating intentionally), not toward fantasy (non-intentional).

The paper examines some consequences of both views of the act of reading through the lens of two selected narratives from *Difficult Loves* by Italo Calvino. The aim of this final investigation consists in assessing, from the perspective of reading literature, the joint contribution of both thinkers to an event-oriented reshaping of human identity.

KEYWORDS affective event; Calvino, Italo; discursive event; identity transformation; Richir, Marc; Ricoeur, Paul; reading

✉ Robert Grzywacz, Ignatianum University in Cracow, Poland  robert.grzywacz@ignatianum.edu.pl
0000-0001-8353-6238

©  *FORUM PHILOSOPHICUM* 30 (2025) no. 2, 271–93
ISSN 1426-1898 E-ISSN 2353-7043

SUBM. 9 October 2025 Acc. 11 November 2025
DOI: 10.35765/forphil.2025.3002.13

INTRODUCTION

The aim of this text is to demonstrate, by means of selected examples, the dual operativity of the category of *event* within contemporary phenomenological and hermeneutical thought, albeit at the cost of its concretization. First, it will be necessary to briefly sketch the genesis and context of the use of this category within the aforementioned philosophical current, where one speaks of events in a manner distinct from that of occurrences within the analytic tradition. Second, the category of *event* will be situated against the broader backdrop of two contrasting philosophical positions: Marc Richir's transcendental phenomenology and Paul Ricoeur's hermeneutical phenomenology. The choice of precisely these authors and their views is not accidental. Both elaborate important anthropological projects in which reference to human faculties related to the domain of aesthetics played a decisive role. Richir, in this regard, draws upon phenomenological analyses of *phantasia* (Richir 2000, 61–182; 2004), whereas for Ricoeur the imagination is of paramount significance (Ricoeur 1991, 168–87; Amalric 2013). Moreover, both thinkers decisively engage in their reflections with Immanuel Kant's third Critique (the *Critique of Judgment*) (Richir and Carlson 2015, 79–92; Ricoeur 1998, 180), thereby creatively developing the insights of the Königsberg-based philosopher. As will become evident, each of them also proposes compelling theoretical accounts of the act of reading literature (more precisely, the novel) and of its significance for human identity.

Accordingly, the category of *event* will first be applied to the act of reading itself in both perspectives, to its unfolding and to the description of what transpires therein according to each of the selected thinkers. Subsequently, in terms of the *event*, we shall describe the effects that the accomplished act of reading exerts upon the reader in the light of both theories. This final moment—the consequences of reading upon the reader and his or her world—will be exemplified by the reading of a specific text: namely, two selected short stories by Italo Calvino from the collection *Difficult Loves*. All the stories in this volume include in their titles a reference to someone's "adventure": that of the traveler, the reader, the poet, and so forth. Yet more than to these various figures, the narrated "adventures" pertain to the play of gazes or of touch and their inherent polysemy (Wasilewska 2022). As such, they appear particularly susceptible to a twofold reading, inspired respectively by the perspectives of Richir and Ricoeur.

THE CONCEPT OF *EVENT*: A GLIMPSE

As is well known, the category of *event*, in its contemporary continental elaborations, has a Heideggerian provenance. It was, however, the author of *Beiträge zur Philosophie (Vom Ereignis)* (*Contributions to Philosophy [Of the Event]*) who, already in the very title of this work, employed the term *Ereignis*, which he further specifies in the following manner: “beings are brought into their *constancy* through the *downgoing* of those who ground the truth of *beyng*. *Beyng* itself requires this. It needs those who go down and has already *appropriated* them, assigned them to itself, wherever beings appear. That is the essential occurrence of *beyng* itself; we call this essential occurrence the *event*” (Heidegger 2012, 8). And further: “*Beyng* essentially occurs as the *event*” (ibid., 25). The *event* here signifies the very destiny of Being, its coming into unconcealment, its phenomenological concretization as though “of its own accord.” As such, it is originary, novel, and each time unique, unrepeatable. It comes forth or enters into manifestation not as conditioned by anything external, but rather as its own source or cause. It constitutes a gratuitous self-donation which, in its arrival or appearance, ruptures temporal continuity, insofar as it is unexpected, unheard of, and resistant to logical or systematic interpretation. Instead, the *event* itself opens new fields of possibility, and thus, for its recipient, provokes shock, wonder, and astonishment, since it disturbs the familiar order of things and exceeds every expectation as well as all that is recognizable (Gilbert 2020, 285–313). In terms of temporalization, it could be compared to the ever-renewed effusion of primal impression that establishes the originary point of the living present in Husserl’s analyses of inner time-consciousness (Husserl 1989, 44–47), with the paradigmatic example being the fact of birth.

Among the thinkers who have developed the concept of the *event*—setting aside the manifold differences and significant nuances—one must certainly mention, in addition to Heidegger himself, Emmanuel Levinas, Jean-Luc Marion, and Claude Romano. The latter, in particular, distinguishes between the *event* in its proper sense, briefly characterized here (*événementiel*—for instance, mourning, encounter, illness), and so-called “intra-worldly” facts (*événementiel*—for example, rainfall, lightning, day-break). Both are impersonal; yet while the former personally engages its addressee or participant, the latter do not, exhibiting instead the impossibility of unambiguously assigning them a subject (in that they happen to an open plurality of beings—Romano 2009, 23–31). The foregoing provisional list should undoubtedly be supplemented with at least the names of Henri Maldiney, Renaud Barbaras, Alain Badiou and Marc Richir (Prášek 2021, 59–60).

Even the foregoing, admittedly rather economical observations, allow us to surmise that within the traditions under examination the *event* is conceived in a manner quite different from its understanding in analytic philosophy, where the corresponding term designates that which either takes place or does not, occurs or fails to occur. At times it is equated with a kind of repeatable, or potentially repeatable, state of affairs; at other times it is construed as a particular—i.e., a contingent, datable, and localizable exemplification of a property. Such a momentary exemplification, according to some (J. Kim), exhibits a certain internal structure, while for others (D. Davidson) it is devoid of such structure and thus susceptible to multiple descriptions (Loux and Crisp 2017, 143–46).

What distinguishes the category of the *event* in the approaches that will be developed below is its capacity to open up a new horizon in its concreteness and singularity—an aspect that requires, on the one hand, a phenomenological perspective and, on the other, a hermeneutical one. In Richir's case, the event is defined primarily as the creation or formation of meaning (*Sinnbildung*—Richir 2000, 21–22; Schnell 2011, 69), whereas for Ricoeur it is above all discourse (Ricoeur 2016, 94–98, 107–08), which entitles us in this latter case to speak of a “discursive event.” Richir emphasizes the evental character of meaning-formation by pointing to its instability (“flickering”) in the register that precedes symbolic institution, while Ricoeur underscores the dialectical structure of discourse, in which the evental aspect is bound to the semantic one (stabilizing the momentariness of the event within the symbolic order). Such constitution, in turn, generates the necessity of interpretative pluralism. It therefore appears that the two positions under further consideration are not radically opposed, but rather that their complementary treatment may prove fruitful. Moreover, by introducing a distinction between contextual-evental understanding and an understanding that transcends given horizons of meaning, Romano suggests that the truly original work of art provides a paradigmatic example of such an event:

A work of art cannot be understood within the artistic context in which it is born, which it necessarily transcends if it is an original work. In this respect, every interpretation of an event must draw on interpretative possibilities in the event itself: an event alone provides the key for its own deciphering. (Romano 2009, 62)

Let us now turn more closely to both approaches in their fundamental presuppositions, and in relation to the act of reading literature.

THE TRANSCENDENTAL PHENOMENOLOGY OF MARC RICHIR, AND THE ACT OF READING

Marc Richir's transcendental project of phenomenology assigns a pivotal role to *phantasia*¹ (*Phantasie*), carefully distinguished from imagination (*Einbildung*). At this point, the Belgian philosopher explicitly draws upon Husserl's distinctions regarding image-consciousness (Husserl 2005, 1–200), as well as upon Kant's transcendental schematism as presented in the first part of the third Critique (Richir and Carlson 2015, 80–91; Kant 2002, 103–04, 149–58). The significance of this differentiation becomes apparent once one considers Richir's radical methodological move, namely the hyperbolic phenomenological reduction, which extends beyond Husserl's own versions of the reduction at all its levels. For, in Richir's account, the reduction excludes not only the moment of ontological positing of the investigated structures, but also encompasses the very intentional and eidetic structures themselves, thereby radicalizing Husserl's determinations concerning "the retaining-in-grasp of the entire multiplicity of variations as the foundation of essential seeing" (Husserl 1973, §87c, 342). In this way, Richir seeks to reach a domain of inquiry more primordial than the symbolic and ontological orders.

In doing so, the Belgian phenomenologist makes extensive use of Kant. Beginning from the basic conception of transcendental schematism, Richir criticizes its conceptual and determining version from the *Critique of Pure Reason* as disappointing and unconvincing (Kant 1998, A137/B176–A147/B187, 271–77). For him, this version constitutes an unpresentable mediating presentation: an a priori mechanism—a "third term"—which orders the sensible manifold through the categories and thereby makes possible the constitution of scientific objects as such. As such, this determining version of transcendental schematism does not bear a phenomenological character. In the third Critique, by contrast, Kant speaks of another kind of schemata—reflective schemata—that the intellect draws from the imagination (sic!) in order to apprehend the aesthetic object, creatively gathering impressions. Here, thus, emerges the concept of reflection without concept, which Richir regards as "genial." On this basis, he develops his own account of schematism, now phenomenological, as a transcendental rhythm, a "music without sound" (Richir and Carlson 2015, 82–83).

It is precisely at this point that the particular role of *phantasia*, as distinguished from imagination, comes into play. For according to Richir,

1. The author deliberately preserves the Greek term, deeming its modern renderings inadequate on account of the semantic shifts introduced within the tradition.

phenomenological schematism is unrepresentable, yet it articulates phenomena in their dispersion and overlapping (in their primordial phenomenalization)—which, from the perspective of reduced experience, does not occur in time or space. Rather, this schematism has a proto-temporal character, as if temporalizing and spatializing what phenomenalizes within the primordial and likewise unrepresentable *chôra* (a concept Richir borrows from Plato's *Timaeus*: 48e–52e, which may be rendered as an originary “receptacle”; Strózewski 2021, 305). In other words, it enables “the shadows of nothing” (phenomena as shadows deprived of internal consistency, or, as Richir also puts it, as mere phenomena) to relate to one another in dispersions, condensations, and explosions. The non-intentional register of primitive (*sauvages*) essences does not constitute pure chaos; rather, it organizes itself according to a specific rhythm—the phenomenological schematism—which manifests itself as a flickering: a dynamism of oscillation between appearing and disappearing. What sets these primitive essences into motion, however, is affectivity, which originally constitutes a dimension of absolute inwardness, yet displays an activity akin to a blind, anonymous motor (Richir and Carlson 2015, 164–65, 171–76).

For Richir, this primordial register of phenomenalization—that is, phenomenological schematism—constitutes precisely the domain of non-intentional *phantasia*, whose significance, in his view, surpasses that of intentional imagination. To better grasp this position, it is worth bearing in mind such examples as the process of humanization in the sense given by Donald W. Winnicott,² the sudden dazzling thought that comes to mind, or the emotional absorption induced by the reading of a fascinating novel (Richir and Carlson 2015, 238–42; Richir 2007). *Phantasia* represents for Richir the most archaic form of phenomenalization, whose transposition into the order of symbolic institution (language), and simultaneously into the domain of ontology and its simulacra (sign-substitutes of reality), only occurs through acts of intentional imagination. He conceives of “pure *phantasia*” as strictly intertwined with affectivity: we encounter it when the phenomenological concreteness of primitive essences is stirred by affectivity within the realm of lived corporeality or intercorporeality (*Leiblichkeit*),

2. The matter finds its clearest articulation in the author's own words: “I am proposing that there is a stage in the development of human beings that comes before objectivity and perceptibility. At the theoretical beginning a baby can be said to live in a subjective or conceptual world. The change from the primary state to one in which objective perception is possible is not only a matter of inherent or inherited growth process; it needs in addition an environmental minimum. It belongs to the whole vast theme of the individual travelling from dependence towards independence” (Winnicott 1984, 151).

but without this experience being counterbalanced by its object-side (as *Leibkörper*). And yet, already here a preliminary distinction between interiority and exteriority becomes possible, for “perceptive *phantasia*” in turn allows the phenomenological field to transcend itself toward alterity.

This primordial phenomenological field denotes, in fact, a multiplicity of living “absolute Heres” that, not yet situated in objective space nor mediated by it, unfold a primordial “receptor” (*chôra*), opening proto-space. At the same time, this originary spatial dissemination is accompanied by an equally originary temporalization, which consists, as it were, in a kind of perceiving:

“Perceptive” *phantasia* is in fact a concreteness of language, as if the *phantasiai* were “perceiving” one another. And this is indeed what occurs: when one thinks, one “perceives” in *phantasia* what is still to come (a protention of language) in the temporalization of sense, just as one “perceives,” in the same manner, what has already come (a retention of language) in the same temporalization—but this without any assignable present: what has been thought has meaning only in relation to what is still to be thought. There is thus already a gaze, but a gaze of the invisible within and upon the invisible. In other words, for me, the “perceptive” *phantasia par excellence* is the gaze that “sees” the unfigurable living “behind” the figured: and it is the extraordinary power of language to hold itself together in pro- and retro-“perception,” this power whereby it is capable of reflecting itself and thus of gazing upon itself, but from within, with a kind of distance vis-à-vis sense in its search for itself. (Richir and Carlson 2015, 175)

In this manner, the originary multiplicity of “absolute Heres,” simultaneously “perceiving” and being “perceived,” realizes a reciprocal “empathy” (*Einfühlung*) within the proto-spatial sphere of transcendental inter-facticity (*Zwischenleiblichkeit*). This means that each living body (*Leib*) is at once an actual and irreducible bearer and recipient of gazes—not in the sensory sense, but precisely in the sense of phantasiatic “perceiving”—gazes both actual and virtual. In this sense, it is not only a living tissue of corporeality, but also phantasized corporeality (*Phantasieleib*—Richir and Carlson 2015, 240–41; Richir 2006, 36–38). Empathy, moreover, functions as follows:

It is not that I somehow “leave behind” my lived body (*Leib*) as my absolute Here, but rather that, in glimpsing the other lived body (*Leib*) and its absolute Here which is over there, it is as if I were there myself. I am therefore not “really” (*reell*) there with my lived experiences, for I cannot, so to speak,

leave my lived body (and my absolute Here) “behind me.” And yet I am there nonetheless as if, that is to say, not *ipso facto* through intuitive representation in imagination (for in that case the other would be nothing but a “projection” of my lived experiences into its interiority, a mere imaginary duplication of myself). What is at issue is an intuitive representation that seems to be induced by the as if which, in Husserl, generally signals the modification through imagination—but here, for us, by the mobilization of *phantasia*, which must be carefully distinguished from imagination. This, first of all, insofar as it is not from the outset a figuration of an intentional object. Or, insofar as . . . the intentionality of *phantasia* is not, originally, an object-intentionality, but rather, as we have implied, a *spatializing* “intentionality.” (Richir 2006, 37)

This new phenomenological approach of Richir—outlined here only schematically and in its essential contours—has given rise to an intriguing conception of the act of reading literature (Richir 2003, 24–26; 2011, 15–24). At this point, as well, the thinker emphasizes a significant phenomenological distinction between imagination (*Einbildung*) and *phantasia* (*Phantasie*). Imagination, referring to the concept of “image,” has historically generated much ambiguity, intensified by the contemporary dominance of audiovisual technologies. Husserl posits that imagination is an intentional act that aims at an object. This object is endowed with an intentional sense, but it is not present “in flesh and blood” like a perceived object. This object (called *Bildsujet*) is nevertheless present in the act of aiming itself, but as unreal. It is quasi-positing as present by and in the act of imagining, rather than positing as actually present. Two cases may occur: either the object is quasi-positing through an “image” (*Bildobjekt*) that has a physical support, such as, for example, a photograph, or the object is quasi-positing through an “image” that lacks a kind of physical support. In the latter example, without any physical support, when one imagines something, one aims at it and “sees it in the mind,” not its “image.” The proof is, for example, that some details are not countable in imagination, which would be possible if one perceived the object directly. Then, the “image” cannot be considered to be an imperfect reproduction of reality; rather, it is an unreality that fundamentally eludes us, yet paradoxically serves as a figuration (*Darstellung*) of the intentional object. Richir adds that the fidelity of the figuration to the perceptual original is of little importance; the essence of the act of imagining lies in its intentionality towards the object, and the intuitive figuration of this object is not what is properly aimed at. Figuration is never positing for its own sake, but mediates a positing or a quasi-positing. It is a “perceptive semblance,” a perception distinct from intentional aiming or passive (sensual)

reception. Richir explains that without the imaginative intentionality of the object, the physical support would be reduced—for example, in the case of a picture—to a collection of spots. The physical support is thus merely a support, and the intuitive figuration of the object acts as a simulacrum between the subject of the imaginative act and its object. A character is not present in their portrait, nor a landscape in its photograph; they are present in the intentional act of imagining (as noematic correlates), but they are not there, in presence, in the world.

Conversely, *phantasia* emerges and vanishes in flashes (*blitzhaft*), intermittently and discontinuously; it is protean and, crucially, non-present. Richir draws consequences from this, stating that *phantasia* does not fall under a classical (Husserlian) temporalization into presents equipped with their protentions and retentions, but rather under a different type of temporalization in presence, without an assignable present. Furthermore, its protean character renders it non-figurative of (intentional) objects, making it nebulous, internally mutable, more or less intense, and thus simultaneously non-positional (in that it is not posited by an Ego and does not posit an object) and non-intentional (in the classical sense), it being unable by itself to mediate the positing of any object.

The fact that some of these characteristics also appear in intuitive figuration within imagination suggests, according to Richir, that the act of imagination is instituted on the phenomenological basis of *phantasia*, through a kind of immobilization of temporalization in presence. *Phantasia* is transposed in the (intentional) act of imagining in the form of a “perceptive semblance,” which is the semblance of the object intentionally aimed at by the act of imagining. This explains, for instance, why a dream is always a mixture of nebulous *phantasiai* that have remained in their original phenomenological status, and (intentional) imaginations in which one recognizes a particular place, character, or action. The intentionality “projected” onto *phantasia* transmutes it into a “perceptive semblance” that functions as the simulacrum of the quasi-presentation of an object (e.g. an animal seen in a cloud). This intimate, though often hidden, connection between *phantasia* and imagination also explains the fundamental instability of intuitive figuration in imagination. By contrast, the attempts to fix this instability can even destroy the vivacity of *phantasia* and impoverish it to the level of stereotypes.

Richir’s theory of reading is elaborated against the background formed by the paradox of theater (Richir 2003, 27–29; 2011, 25–32). As Husserl observed, on stage, the character, like Richard III, for instance, “is” present, though not (sensually) perceived “in flesh and blood.” What is perceived

is the actor, the stage, the scenery. A good actor does not provide an intuitive figuration of an intentionally aimed object; if the actor is good, the character they embody is alive. The magic of theater lies in unfolding an intrigue among living characters. If the actor is poor, the spectator is forced to imagine the character, leading to boredom or the projection of their own phantasms. The paradox of theater is that the actor “lends” their entire living body (*Leib*) to the living body of the character they embody, without any intuitive figuration of that character in imagination. The character is in presence, within the temporalizing flow of the intrigue, but never truly present in an intentional act of imagination. If well “embodied” by the actor, the character is not present in an act of perception or imagination; it is in presence, intuitively non-figurable, in *phantasia*. And if it is there, in presence but not present, it is nevertheless the “object” of a “perception” in a different phenomenological sense, through *phantasia* itself. The “perceptive” *phantasia* in question differs from any sensual perception. In this case, it is a “perception” not of unreality or reality, but of a “consistency” or “concreteness” (*Sachlichkeit*) belonging to the character, which is indeed there, in presence, but outside of any assignable present. Thus, theater is effective if it engages *phantasia*, but degenerates if it appeals to acts of imagination. While imagination only ever aims at imaginary objects, “perceptive” *phantasia* opens up to a “concreteness” whose internal horizon is reality, even if that reality never existed. This “consistency” is in infinite transition (it being a kind of transitional object in a sense echoing that of Winnicott) between “pure” *phantasia* and reality, in a free play of *phantasia*, which explains the many possible interpretations of theatrical characters. A good actor achieves empathy (*Einfühlung*) for the character through non-positional and non-figurative *phantasia*. It is through the actor’s “perceptive” *phantasiai* that they pass into the spectators’ *phantasiai*, revealing “perceptive” *phantasia* as the most archaic basis of intersubjective encounter.

In literature, the case is analogous (Richir 2003, 29–33). The novel, like theater, is a temporal unfolding of an intrigue among living characters, but without the necessity of being embodied by actors. Any intuitive figuration of these characters (e.g., in film adaptations) is immediately disappointing, because no actor, however talented, can convey the complexity of the character as it imposed itself on the novelist and as it imposes itself on the attentive reader. Like in theater, the key in literature is the experience of “living” characters. Both art forms combat the threat posed by imagination, understood as a drive for detailed, fixed figuration that invalidates vivacity. When literary experience is effective, it engages “perceptive”

phantasia rather than acts of imagination. Temporality occurs in presence without an assignable present. Thus, characters attain a “concreteness” that is in presence but not in the present, and has reality as its horizon. Both the writer and the reader (analogous to the actor and spectator) undergo a process of empathetical involvement (*Einfühlung*) with the character. The intrinsic life and affectivity of characters are intuitively non-figurable, but “perceptively” apprehended through *phantasia*. In literature, there is no physical embodiment of the character; any visual adaptations are inherently disappointing, because an actor can never capture the complexity of a novel’s character. The novel’s freedom from theatrical conventions grants “perceptive” *phantasia* a significantly greater freedom in the novel than in the theater. The novelist “interprets” their characters in the course of writing; they do not invent them from scratch. Characters become alive as they begin to live their own lives, which the writer must follow and develop amidst the infinite, indeterminate possibilities of *phantasia*. Unlike in theater, where the actor’s “perceptive” *phantasiai* are transmitted to the spectators, in literature it is the novelist’s art, contained in the text, that awakens “perceptive” *phantasiai* in the reader.

When we read a novel (well), we imagine only in ephemeral and fleeting moments. By contrast, the essence of our activity as a reader lies paradoxically in both the concern for understanding the text and in the activity of *phantasia*. Crucial are the intrinsic lives of the characters, meaning the “movements of the soul” or “labyrinths of affectivity” (Richir 2003, 31), where affections are intimately linked to *phantasiai* and “perceptive” *phantasiai*, and these movements are grasped “live.” This is why the role of the reader turns out to be quite particular. The reader must experience empathy for the concrete life of the characters—meaning achieving their “perception” in *phantasia*. This requires time—the time of temporalization in presence, without an assignable present. If the reader maintains integrity towards the reading and the writer, they will be incapable of creating a physical portrait of characters, and any proposed portrait will seem false, or even horrifying in its poverty. Many indeterminacies are needed for *phantasia* to be able to live, somewhat like in a dream. Hence, the reader should avoid the “traps of imagination” (Richir 2003, 32)—especially overly meticulous descriptions, which destroy phenomenological vivacity and lead to boredom. A sensitive and intelligent reader can feel and understand these plays in a non-positional manner: i.e., without positing them in reality. For the above reasons, the text should not contain overly detailed descriptions of places or characters, as these appeal to imagination rather than *phantasia*, disallowing “phenomenological vivacity.” Rather, through its indeterminacies, the text creates space

for the reader's *phantasia* to freely operate and animate the characters and the depicted world. The text becomes a vehicle for "perceptive" *phantasiai* that awaken "life" in the reader. Thus, Richir corrects the definition of the novel, stating that beyond the romantic intrigue, what matters is the intrinsic life of the characters, which makes the story aesthetically interesting. The text, rather than being a document or testimony, allows us to "feel" the depicted world in *phantasia* better than a history handbook. Such an interest in the extreme lability and fluidity of affections, and how they can shift from one modulation to another, as well as in the reader's experience of the "life" of characters by means of empathy and "perceptive" *phantasiai*, qualifies the *mimesis* that is at work here as non-reflective, active and internal. This is why, in the case of reading, as the act is viewed by Richir, we have to do with an affective event:

One grasps an intimacy not truly at a distance, but in the *chôra*, in another seat, anterior to extension, to space. This can, moreover, be verified when one reads literature: in reading Stendhal, for example, one indeed seizes upon something of Stendhal himself, in his singularity. And yet one can never gain access to the man as he was: that belongs to the domain of intersubjectivity, but it is lost forever—even if one may look upon portraits. (Richir and Carlson 2015, 242)

PAUL RICOEUR'S HERMENEUTIC PHENOMENOLOGY AND ITS ACCOUNT OF READING

Quite differently, in Ricoeur we have to do with a stance that could be characterized, in Richirian terms, as transposition into the order of symbolic institution. Actually, in his philosophy, Ricoeur makes a crucial distinction between the epistemological and ontological functions of the life-world (*Lebenswelt*—Ricoeur 2004a, 371–77). This distinction highlights the difference between the validation and obligatory character of scientific ideas on the one hand, and the ultimate referent of all scientific and cultural idealizations on the other. In the epistemological order, the idea of science takes precedence, yet the ontological priority of the *Lebenswelt*, as the ultimate referent of all utterances, undermines consciousness's claim to absolute authority over the universe of meanings. The separation of these two aspects of the *Lebenswelt* helps to recognize the twofold belonging of humans: to the *Lebenswelt* (given prior to the question of obligation) and to the world of symbols and rules, which forms an interpretative framework for the former, giving meaning to the practical, situational human condition. This twofold belonging entails, in Ricoeur's view, valuing interpretation

over descriptive accounts of immediate data. Since interpretation becomes unavoidable due to the insurmountable distance that language establishes from what it expresses. The reflective use of language implies an irreversible departure from immediacy, and the universe of signs and logic becomes the domain of validity.

Ricoeur's philosophical project, deeply rooted in the philosophy of reflection, existentialism, phenomenology and hermeneutics, posits that self-understanding is progressively mediated by texts. This ongoing process, framed by a "hermeneutic circle," begins with a non-critical self-understanding, moves through a critique involving objectifications, and culminates in a more critical and enriched self-understanding, often summarized as "more explanation for better understanding" (Ricoeur 2016, 94–126, 159–83). One of the pivotal concepts permeating this thought is that of discourse. Ricoeur defines discourse as a meaningful event belonging to the sphere of social communication, inherently involving the dynamic: "someone says something to someone about something" (2016, 101). This comprehensive definition integrates all constitutive elements of linguistic communication, including sender, receiver, medium, code, message and context. His work navigates the complexities of meaning and addresses interpretive conflicts through hermeneutical reflection.

Discourse unfolds through an internal dialectic of event and meaning, where the ephemeral utterance is reidentified as stable meaning, possessing both sense (its objective aspect) and reference (its connection to an extra-linguistic world). Thus, "if all discourse is realised as an event, all discourse is understood as meaning" (Ricoeur 2016, 96). Furthermore, writing is considered the full manifestation of discourse, introducing distanciation. This includes the meaning surpassing the event of saying, emancipation from the author's intention and original audience, and emancipation from ostensive reference, opening up the "world" of the text (Ricoeur 2016, 94–105).

These forms of distanciation are central to Ricoeur's understanding of semantic innovation, particularly in poetic discourse, where metaphor, by suspending literal meaning, projects new possibilities of seeing and being in the world, serving as a heuristic fiction that "re-describes reality" (Ricoeur 2016, 54, 104, 236, 256). This framework extends to narrative discourse, encompassing both historical and fictional narratives, which Ricoeur views as dependent on emplotment for organizing events into meaningful wholes. These observations lead us directly to his conception of the threefold *mimesis*.

As it has already been stated, Ricoeur's philosophy fundamentally posits that self-understanding is progressively mediated by texts, a process framed

by a “hermeneutic circle” moving from non-critical self-understanding through textual objectifications to a richer, critical self-understanding. This profound engagement with language, time and human experience is articulated through his theory of threefold *mimesis*, which is central to his understanding of the act of reading literature. Ricoeur articulates the mediation between time and narrative through three interconnected moments of *mimesis*: prefiguration (*mimesis* 1), configuration (*mimesis* 2), and refiguration (*mimesis* 3), with the moment of emplotment (configuration) constituting the pivot of this analysis (Ricoeur 1984, 52–87).

The initial stage of prefiguration refers to the pre-understanding of the world of action. This means that the composition of a plot is grounded in a prior familiarity with human acting, its meaningful structures, symbolic resources, and temporal character. These features are described rather than deduced, emphasizing that “to imitate or represent action is first to pre-understand what human acting is” (Ricoeur 1984, 64). This pre-understanding, common to both poets and their readers, is crucial, for although literature “institutes a break,” it would be “incomprehensible if it did not give a configuration to what was already a figure in human action” (ibid.).

The moment of configuration or emplotment is the “kingdom of the as if” (ibid.), the realm of narrative composition where emplotment (*muthos*) takes center stage. As the pivot of the analysis, configuration is the rule-governed composition of a tale. The plot’s primary function is to “grasp together” and integrate into “one whole and complete story multiple and scattered events, thereby schematizing the intelligible signification attached to the narrative taken as a whole” (Ricoeur 1984, x). This process creates a “synthesis of the heterogeneous,” achieving a “discordant concordance” (Ricoeur 1984, 66, 69–70) by bringing together diverse elements such as circumstances, goals, means, interactions and results. Ricoeur explicitly compares this configuring act to Kant’s productive imagination and schematism, understanding it not as a psychological faculty but as a transcendental one that engenders a “mixed intelligibility” (1984, 68) between a story’s themes and its intuitive presentation of events. Ricoeur extends the notion of an imitated action to novels oriented toward character or toward an idea, and applies the model of emplotment analogously to historical narrative through concepts like quasi-plot, quasi-character and quasi-event (1984, 181–225). Thus emplotment serves a crucial mediating function, operating within its own textual field to integrate and mediate between the pre-understanding characterizing the moment of prefiguration and the subsequent refiguration. It is here that the “literariness” of the work of literature is instituted, as the creative imitation produces “quasi-things” and invents the “as-if.”

The third moment of *mimesis*, termed refiguration or application, represents the return from the world of the text to the reader's world, where the narrative configuration is applied to real-life experience. The act of reading is the "operator" that links refiguration to configuration, and it is in this act that the work acquires a meaning in the full sense of the term, at the intersection of the world projected by the text and the life-world of the reader. The text, by suspending the reference of ordinary language and deploying "poetic reference," effects a "redescription of reality" and "augments it with meanings." It thereby unfolds a "world in front of itself," which the reader can inhabit and project their "ownmost powers" into (Ricoeur 1984, 80–81). This textual world is "the whole set of references opened by every sort of descriptive or poetic text . . . read, interpreted, and loved. To understand these texts is to interpolate among the predicates of our situation all those meanings that, from a simple environment (*Umwelt*), make a world (*Welt*)" (Ricoeur 1984, 80). Reading involves appropriation, a terminal act of self-understanding where the reader makes one's own what was initially alien. This is not a passive reception but a creative act that expands the reader's horizon of existence. Ricoeur emphasizes the dynamic interplay between the reader's historico-cultural context and the text, leading to a "fusion of horizons" (Gadamer) and the generation of new questions. Reading, especially fictional narratives, offers "imaginative variations" (Husserl) on time, enabling the exploration of nonlinear features of phenomenological time that historical time might obscure. This refiguration can have a cathartic effect, transforming negative emotions into pleasure and enabling new evaluations of reality, thereby providing an impetus to action. Ultimately, this process leads to the formation of narrative identity, where the self is dynamically reshaped by cultural narratives and its own life-story. The moment of application, therefore, makes explicit the "interweaving reference" between history and fiction in the refiguration of human time, ultimately revealing and transforming human action (Ricoeur 1984, 76–87).

The transformation in question becomes possible by means of certain writing and textual strategies, as well as through the reader's response to them. These three moments might be referred to as, respectively, the rhetoric of fiction, poetics, and the phenomenological and hermeneutical aesthetic of reading (Ricoeur 1988, 160–79).

The first moment focuses on the author's techniques aimed at persuading the reader and creating an "intensity of the illusion" (Ricoeur 1988, 161). Rhetoric of fiction distinguishes between the real author (the biographical figure) and the implied author (the "second self" projected within the

work). The implied author takes the initiative in the “show of strength” that governs the relationship between writing and reading, employing persuasive strategies to impose a fictional world upon the reader. This includes techniques like “showing” rather than “telling” and establishing a reliable narrator. Even when history is read as a novel, the fiction-effect lowers the reader’s guard, suspending mistrust and establishing confidence, thereby facilitating the “reenact[ment]” or “rethink[ing]” of situations and trains of thought by giving them “vividness” (Ricoeur 1988, 186). The “dramatization” of the narrator, in turn, introduces “credibility,” which functions analogously to documentary evidence in historiography. An unreliable narrator can serve an ethical function, calling the reader to freedom and responsibility by challenging them to decipher the unreliability.

The stage of poetics or textual configuration concerns the inscription of the author’s strategy within the literary configuration of the text itself. For Ricoeur, drawing on Roman Ingarden, a text is inherently incomplete, offering “schematic views” and “places of indeterminacy” that require the reader’s active “concretization.” Wolfgang Iser further developed this idea with the concept of the “wandering viewpoint,” which acknowledges that the text’s totality is never perceived at once and that reading is a journey through the text. This involves a continuous interplay between modified expectations (protentions) and transformed memories (retentions) (Ricoeur 1988, 167–69). Literary texts “depragmatize” objects, transforming them rather than merely denoting them, which allows them to offer new perspectives on reality. Reading, in this sense, is not passive reception but a “creative act replying to the poetic act that founded the work” (Ricoeur 1988, 320, n. 58). Moreover, temporalization procedures employed in narrative involve the relationship between the act of enunciating and the uttered statement, and their connection to the non-fictional time of life and action. Meaning is created through the divergence and mutual shifts of the time needed for narration and the time of the narrated things. The actual refiguration of human temporal experience is the work of the combined intentionalities of historical and fictional narratives. Historical intentionality, aiming to grasp the past “as it really was” (Ricoeur 1988, 207), cannot do so without an imaginative component. This imagination simulates seeing, making the past appear as “what I would have seen, what I would have witnessed if I had been there” (Ricoeur 1988, 185). Fictional intentionality becomes “historicized” by adhering to rules of “probability or necessity” in plot construction, making its unreal events “past facts for the narrative voice” (Ricoeur 1988, 190). This crossing of intentionalities happens in reading, creating an imaginative figuration of a certain state of affairs.

The act of reading involves a moment of *stasis*, where the reader immerses himself or herself in the fictional world, “un-realizing” themselves to the extent of the unreality of the fictional world they enter. This immersion facilitates the subsequent refiguration of their own experience. Ricoeur suggests that “the more readers become unreal in their reading, the more profound and far-reaching will be the work’s influence on social reality” (1988, 179). This refiguration involves a profound transformation of the reader. The cathartic moment of application frees the reader from everyday experience, enabling a clarifying understanding that empowers them to make new evaluations of their own reality. This emotional-cognitive transposition of meaning allows the reader to re-evaluate their own world in light of the poetic vision. The dialectic of refiguration manifests itself in the tension between the freedom of imaginative variations and the author’s persuasive strategies, between identity and difference, and between referentiality and communicability. While the implied author projects a world and guides interpretation through rhetorical strategies, the reader actively actualizes the text’s meaning. This involves a “dispossession” of the reader’s immediate ego, leading to a self-understanding reshaped by the text’s revelatory power. There is an asymmetry where the real author is effaced in the implied author, but the real reader embodies the textual suggestions and concretizes the implied reader (Ricoeur 1988, 170–71). The act of reading is not solitary: it is deeply embedded in a community of readers. Indeed, the community establishes norms and canons, framing the textual world beyond purely subjective interpretation. The ideal outcome is a “fusion of horizons” that results in an analogous relationship between the textual world and the reader’s life, providing a new *impetus* for action (Ricoeur 1988, 179, 247).

The above findings clearly establish that, for Ricoeur, unlike in Richir, imagination is essential for “concretizing” characters and events and for imaginative variations on time. While fiction offers freedom of imaginative variations, this freedom is internally constrained by the author’s implied vision, creating a force of conviction. Consequently, the effect of *catharsis* is not just emotional cleansing but “clarification,” leading to new assessments of reality. Thus, in Ricoeur’s conception of the act of reading and of the mimetic and cathartic processes intertwined with it—processes more cognitive than merely affective—we are, in essence, confronted with an event that merits the designation of “discursive.”

TWO READINGS OF TWO “ADVENTURES” BY ITALO CALVINO

In the final part of the present study, we shall briefly relate both theories of reading to two short stories by Italo Calvino, contained in the collection

Difficult Loves (Calvino 2017), in order to examine the complementarity of these philosophical approaches when applied. The collection *Difficult Loves* is one of particular interest since it presents a notable articulation of the author's "fascination with the empire of the senses" and the "materiality of expression" (Cavallaro 2010, 14). The stories in this collection often highlight the "ineradicable absurdity haunting humanity's feeble attempts at communication" and "people's perverse proclivity to erect barriers, both consciously and unconsciously, against the possibility of embarking on forms of emotional exchange that could alter their lives radically and thus alleviate their daily exposure to loneliness and tedium" (Cavallaro 2010, 191). The "adventure" in the titles frequently points to internal processes of recognition, the shattering of illusions, and confrontations with one's own reality, rather than external heroic deeds. Let us turn our gaze more attentively to two narratives drawn from the aforementioned collection.

In "The Adventure of a Traveler" (Calvino 2017, 77–96), Federico V., a diligent man from northern Italy, frequently travels by night train to Rome to meet Cinzia U., his beloved. During his journey, he meticulously plans their future conversations and relishes the anticipation of their intimacy, considering the pillow he buys as a "daily letter to Cinzia" (Calvino 2017, 81). He deliberately ignores other passengers, like a salesman, to immerse himself in his romantic fantasies, humming French love songs. Later, his compartment is invaded by boisterous soldiers speaking an unintelligible dialect, with whom he surprisingly feels a sense of identification, perhaps to enhance his coming reunion with Cinzia. Upon arriving in Rome, he calls Cinzia, whose sleepy voice makes him realize the profound, incommunicable significance of his night's journey, which now fades with the "cruel explosion of day" (Calvino 2017, 96).

It seems that with this story we have to do with a deeply internal, almost entirely mental, "adventure," where the physical journey serves as a canvas for Federico's rich inner life and fervent anticipations. It is an "inner concentration" (Calvino 2017, 79) and a "travel in love," a "release of euphoria" that he feels "harmonized with the race of the train" (Calvino 2017, 84). The "adventure" is in the intense, subjective world Federico constructs, which ultimately proves untranslatable into external reality. The story highlights the gap between internal experience and outer reality, and the inherent difficulty in communicating profoundly personal feelings. Federico's rich inner world, filled with "loftier heavens" (Calvino 2017, 85), cannot be fully conveyed to Cinzia. This aligns with Calvino's broader concern about the "plague afflicting language" that "tends to level out all expression" and deaden "the spark that shoots out from the collision

of words and new circumstances" (Cavallaro 2010, 10). Federico's inability to recount "the significance of that night" (Calvino 2017, 96) mirrors this linguistic limitation, where the richness of his subjective experience is lost in the ordinariness of spoken words.

One might say that, from the perspective of affective life, of lived sensitivity, where we encounter the emergent sense of a nocturnal journey in the ceaseless interplay between Federico's inner world and external circumstances, Richir's theory of reading makes it possible to grasp the dynamism of empathy and the efficacy of immersion into the world of his experiences, so as to undergo something of them oneself. On the other hand, Calvino, through Federico's journey, questions the nature of "presence" and "absence." Federico's journey is a state of "vague anticipation of tomorrow" (Calvino 2017, 87), where the imagined presence of Cinzia is more vivid than any immediate physical reality. This also touches upon Calvino's "ongoing concern with the coexistence of absence and presence, invisibility and visibility, by intimating that the observable is at all time traversed by the imperceptible" (Cavallaro 2010, 130), suggesting that the true essence of Federico's love exists in this invisible, internal dimension. From the perspective of the games with time, which take into account the interplay between narrated time, the time of narrating, and the time of reading, as well as the effect of the historicization of fiction and the cognitive import of iconic augmentation, one can appreciate the complementary contribution of Ricoeur's theory of reading. Indeed, what we encounter here is a certain cognitive gain with regard to the character of the relation between Federico and Cinzia, insofar as it is lived by him.

In another story, "The Adventure of a Motorist" (Calvino 2017, 171–80), the anonymous narrator drives at high speed at night on the highway, heading towards "Y" (a woman), in city "B," from his home in "A." He questions his true desire—whether he wants to find her or for her to rush towards him. He also suspects his rival, "Z," might be on the same road, and perceives passing cars as potential signals from either of them. He fears that a direct meeting would lead to "communication that is already difficult on the telephone" becoming "even more obstructed, suffocated, buried as if under an avalanche of sand" (Calvino 2017, 176). He wishes to transform himself and his message into a "cone of light launched at a hundred and forty kilometers an hour" (*ibid.*), believing such a signal could be understood. After an unsuccessful phone call to Y, he joyfully assumes she is driving towards him, so he turns back, searching for her among the anonymous flashes of cars. He concludes that the "price to pay is high . . . not to be able to distinguish ourselves from the many signals that pass along this road, each with a meaning of its own

that remains hidden and indecipherable, because there is no longer anyone outside of here who is able to receive us and understand us" (Calvino 2017, 179–80). This "adventure" is a metaphorical quest for connection and love in a modern, technologically mediated, and fast-paced world. It is a quasi-philosophical reflection on the nature of identity and communication in an era where interactions are often reduced to anonymous "flashes." The "adventure" lies in the pursuit of an elusive, reciprocal signal, driven by hope and uncertainty, in a landscape of constant, impersonal movement. The story profoundly comments on the challenges of authentic connection and communication in contemporary society, where human interactions are often mediated and depersonalized by speed and technology. Individuals become like "beams of light," their identities blurred in a "digital noise."

Richir's account of the act of reading provides a conceptual framework within which it becomes possible to apprehend, in terms of "perceptive" *phantasia*, the motorist's desire to become a "cone of light"—perfectly encapsulating the transformation "in which individual subjectivities are transformed into the anonymous glimmers and luminous flashes," and where "the symbiotic relationship between body and automobile is conducive to a metamorphosis of the human into almost incorporeal, beaming and flickering effects" (Cavallaro 2010, 60–61). This highlights Calvino's interest in the "unstoppable metamorphosis" (Cavallaro 2010, 154) of narratives and identities. The driver's uncertainty about "Y" and "Z" makes them anonymous "flashes," reducing them to fragments within a system of fleeting signs. By contrast, according to the Ricoeurian conceptualization of reading, the story deals with the progressive narrative apprehension of the whole, an unceasing play with time enacted through anticipations and retrospections in the narrator's consciousness. Moreover, this interpretation resonates with Calvino's concern for the limitations of human discourse, where communication is "obstructed, suffocated, buried" (Calvino 2017, 176). This suggests a profound loneliness at the heart of modern connection, where even if a message is sent, there might be no one truly capable of receiving and understanding it, unless our solitude is transgressed by the wonder of communication:

The incommunicable is the psychic as such, that is to say, that non-intentional dimension of life, that manner in which lived experience chains itself together with itself, that sequence of events transversally bound by time, that belonging of events to the same series, the same sphere, the same closure. The psychic, in a word, is the solitude of life, which at intervals is rescued by the miracle of discourse. (Ricoeur 2004b: 67)

CONCLUSION

By way of conclusion, we may say that both philosophical conceptions of reading have proven their fertility and complementarity in concrete application. The non-figurative and non-reflective, yet active and internal *mimesis through empathy* appears to be more operative on the plane of affective life, of shifting moods and flashes of sensation at the threshold between inner and outer worlds, where the psyche enters into ceaseless interactions with its changing environment and circumstances. By contrast, the re-figurative *threefold mimesis*—narratively restructuring experience and enriching it cognitively—demonstrates its functionality on the level of discursive reconstruction of identity. Both are indispensable, for, as Prášek observes, on the one hand:

The proper subjective dimension of the self—its identity—is constituted by the living body feeling itself from inside through “synaesthesia,” a sort of unifying archaic kinesthesia. It is only thanks to this identity that the process of phenomenalization (events of sense) can sediment and form a personal or internal history—his or her ipseity in the sense of personal uniqueness. (2021, 76).

On the other hand, he states that “the concrete subjectivity, a person, is nothing but the result of this sedimentation finally modified through the symbolic institution into a ‘personal story’ one can narrate” (Prášek 2021, 76). In other words, a comprehensive account of identity, as well as of the experience of our freedom (Romano 2020, 50–58, 66–70), requires the cooperation of these two approaches, these two conceptions of mimetic processes. For not only is reading itself a play of belonging and distancing—where the former seems to be more adequately grasped by Richir, and the latter by Ricoeur—but a similar struggle, an oscillation between proximity and distance, also characterizes our identity-transformations and our lived experience of freedom. The problematic moment in Richir concerning distance and the critical verification of experience—its conformity with the textual schema—is aptly articulated by Ricoeur, while the Belgian phenomenologist, in turn, complements him with conceptualizations of the moment of affective belonging and of the motivational sources of subjective transformation. In this context, Calvino himself declared that his efforts were not “merely aimed at making a book but also at changing myself, the goal of all human endeavor” (Cavallaro 2010, 174–75). This conviction profoundly echoes the inner metamorphoses undergone by his characters in *Difficult Loves*.

BIBLIOGRAPHY

- Amalric, Jean-Luc. 2013. *Paul Ricœur et l'imagination vive: Une genèse de la philosophie ricœurienne de l'imagination*. Paris: Hermann.
- Calvino, Italo. 2017. *Difficult Loves*. Translated by William Weaver. Boston: Mariner Books.
- Cavallaro, Dani. 2010. *The Mind of Italo Calvino: A Critical Exploration of His Thought and Writings*. Jefferson, NC: McFarland.
- Gilbert, Paul. 2020. *Tournants et tourments en métaphysique*. Paris: Hermann.
- Heidegger, Martin. 2012. *Contributions to Philosophy (Of the Event)*. Translated by Richard Rojcewicz and Daniela Vallega-Neu. Bloomington: Indiana University Press.
- Husserl, Edmund. 1973. *Experience and Judgment*. Edited by Ludwig Landgrebe. Translated by James S. Churchill and Karl Ameriks. London: Routledge & Kegan Paul.
- . 1989. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, Second Book*. Translated by Richard Rojcewicz and André Schuwer. Dordrecht: Kluwer Academic Publishers.
- . 2005. *Phantasy, Image Consciousness, and Memory (1898–1925)*. Translated by John B. Brough. Husserliana 23. Dordrecht: Springer.
- Kant, Immanuel. 1998. *Critique of Pure Reason*. Edited and translated by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.
- . 2002. *Critique of the Power of Judgment*. Translated by Paul Guyer. Cambridge: Cambridge University Press.
- Loux, Michael J., and Thomas M. Crisp, eds. 2017. *Metaphysics: A Contemporary Introduction*. 4th ed. London: Routledge.
- Prášek, Petr. 2021. "Is Happening Subjectivity a Phenomenological Category? Notes on the Problem of Subjectivity in Ricoeur." *Research in Hermeneutics, Phenomenology, and Practical Philosophy* 13 (1): 58–82.
- Richir, Marc, and Sacha Carlson. 2015. *L'écart et le rien*. Grenoble: Jérôme Millon.
- Richir, Marc. 2000. *Phénoménologie en esquisses*. Grenoble: Jérôme Millon.
- . 2003. "Du rôle de la phantasia au théâtre et dans le roman." *Littérature* 132: 24–33.
- . 2004. *Phantasia, imagination, affectivité*. Grenoble: Jérôme Millon.
- . 2006. "Leiblichkeit et phantasia." In *Psychothérapie phénoménologique*, edited by Paulette Wolf-Fédida, 35–45. Paris: Anthropos.
- . 2007. "Le tiers indiscret: Ébauche de phénoménologie génétique." *Archivio di Filosofia* 36: 169–173.
- . 2011. "Imagination et Phantasia chez Husserl." *Eikasía: Revista de Filosofía* 40: 13–20.
- Ricoeur, Paul. 1984. *Time and Narrative*. Vol. 1. Translated by Kathleen McLaughlin and David Pellauer. Chicago: University of Chicago Press.
- . 1988. *Time and Narrative*. Vol. 3. Translated by Kathleen Blamey and David Pellauer. Chicago: University of Chicago Press.
- . 1991. *From Text to Action: Essays in Hermeneutics, II*. Translated by Kathleen Blamey and John B. Thompson. Evanston, IL: Northwestern University Press.
- . 1998. *Critique and Conviction: Conversations with François Azouvi and Marc de Launay*. Translated by Kathleen Blamey. New York: Columbia University Press.
- . 2004a. *À l'école de la phénoménologie*. Paris: Vrin.
- . 2004b. "Discours et communication." In *Paul Ricœur*, edited by Myriam Revault d'Allonnes and François Azouvi, 51–67. *Cahiers de L'Herne* 81. Paris: Éditions de L'Herne.
- . 2016. *Hermeneutics and the Human Sciences*. Translated by John B. Thompson. Cambridge: Cambridge University Press.

- Romano, Claude. 2009. *Event and World*. Translated by Shane Mackinlay. New York: Fordham University Press.
- . 2020. *La liberté intérieure: Une esquisse*. Paris: Hermann.
- Schnell, Alexander. 2011. *Le sens se faisant: Marc Richir et la refondation de la phénoménologie transcendante*. Bruxelles: Editions Ousia.
- Stróżewski, Władysław. 2021. *Wykłady o Platonie*. Kraków: Universitas.
- Wasilewska, Anna. 2022. "Przygody dotyku i spojrzenia." In *Trudne miłości*, by Italo Calvino, 219–27. Warszawa: Państwowy Instytut Wydawniczy.
- Winnicott, Donald Woods. 1984. *Playing and Reality*. London: Tavistock.

Second Reply to Fr. Chaberek: On Why Merely Biological Humans Can Survive, and on When Merely Traditional Doctrine Can Be Abandoned

Kenneth Kemp

I do not want to repeat what I have already said in response to Fr Chaberek's critique of the articles in which I argued that recent scientific work, even if sound, does not constitute a challenge to theological monogenism. I think that he continues to misread the relevant texts and that his arguments generally do not get him where he needs them to go, but whether this is so can be evaluated without much further comment from me. Why, for example, he does not see that showing that a term has two slightly different meanings is making a distinction, not committing an equivocation (Chaberek 2025, 281), or why he thinks that two different beings with different powers due to one having only a corporeal and the other a spiritual substantial form show only an accidental difference (ibid., 283–84), can be left as exercises for the reader.

I do think that there might be some value in a response to two points new to his latest article—first, the alleged impossibility of what it is (*pace* Chaberek), for reasons I have already explained, perfectly reasonable to call merely biologically human beings and, second, his overly restrictive conditions for when merely traditional theological beliefs¹ may be modified or abandoned.

1. In a sense, all revealed truths, in virtue of having been passed on to us, are “traditional.” By “merely traditional” I mean ideas not revealed, but nevertheless of long standing: for example, ideas originating in a too literal interpretation of “simple and metaphorical language adapted to the mentality of a people but little cultured” (Pius XII 1950, 38), or of passages in which Scripture “described and dealt with things in more or less figurative language, or in terms which were commonly used at the time” (Leo XIII 1893, 18).

Polygenism did not begin with the theory of evolution (see La Peyrère 1655; Agassiz 1850), but Darwin's version of that theory gave it more scientific respectability than it had previously had, an acceptability only reinforced by later discoveries in genetics (Ayala 1998). The scientific thesis, as usually presented, contradicts Catholic doctrine. What should a Catholic scholar do? St. Augustine's advice is this: "When [critics] are able, from reliable evidence, to prove [*veracibus documentis demonstrare*] some fact of physical science, we shall show that it is not contrary to our Scripture" (Augustine [415] 1982, 1.21.41). So, following Augustine's advice, I wrote several articles (Kemp 2011, 2020, 2023) in which I offered, as a consistency proof, a four-point scenario demonstrating that what science actually shows (thought not necessarily all that some scientists think it shows) is not contrary to our Scripture (or to doctrine). Central to the proof were four theses:

- T1. Adam and Eve were the first rational beings.
- T2. All other rational beings were descendants of Adam and Eve.
- T3. The body of Adam was a product of evolution.
- T4. Some of the descendants of Adam and Eve interbred with the not-fully-human beings in the population from which fully human beings emerged.

The first three, I think, are true. The fourth is possible, which is all that my consistency proof requires.

T1 and T2 are a statement (a complete statement) of theological monogenism.² Pope Pius XII rightly said that "it is in no way apparent how such an opinion can be reconciled with [Church teaching on] original sin" (Pius XII 1950, 37). I made the same point (e.g., Kemp 2024, 393). I did not anywhere say, despite what Chaberek (2024, 274) implies, that the truth of these is an open question. Indeed, I have published my reasons for thinking that various proposed versions of theological polygenism fail to resolve the problem articulated by Pope Pius (see Kemp 2011, 229).

T3, although contrary to a traditional belief, was recognized as theologically acceptable both by Pius XII (1950, 36) and by St. John Paul II ([1986] 1996a, 1996b).³ It was included in the scenario because it is presupposed by the genetics-based argument for scientific polygenism that I had considered (as well as by others). It is logically independent of the doctrine

2. Pius (1950, 37) formulated this as the rejection of two ideas—"[that] after Adam there existed on this earth true men who did not take their origin through natural generation from him as from the first parent of all" and "that Adam represents a certain number of first parents."

3. Statements, one might add, that have at least as much theological weight as Chaberek attributes to Pope Pelagius I's sixth-century letter to King Childebert.

of original sin. Indeed Pius said that accepting it was not forbidden immediately before his rejection of polygenism.

T4 is what is most important here. My scenario includes T4 as its explanation of the genetic variability on the basis of which scientific polygenists make their case. I say about it that “[T4] being scientifically possible and theologically orthodox, any scientific arguments over polygenism are theologically irrelevant” (Kemp 2024, 393).

T4 presupposes a distinction between rational beings (Pius’ “true humans,” which I called “philosophically, and theologically, human beings”) and what I called “merely biologically human beings,” beings like us with respect to reproductive and sensory powers” but lacking rational souls. The existence of something like such beings is at least suggested by T3, even if reproductive compatibility across the transition to rationality (at infusion of a rational soul), crucial to T4, is not guaranteed by it.

It is this distinction which requires greater clarity about Chaberek’s assertion that Adam and Eve are “the exclusive origin of humanity” (2024, 159 and 164) than his article provides. He says that that means “there were no ‘pre-Adamites,’ ‘co-Adamites,’ and ‘post-Adamites,’ i.e. that all people that are alive and ever lived, or will ever live, descended from a single pair of Adam and Eve” (Chaberek 2025, 278). I presume that his “post-Adamites” and “people” are to be understood to be rational animals. The ambiguity centers on “descended from a single pair.” Does it mean only (T2) that Adam and Eve were among the ancestors of all other rational beings, or does it mean, in addition, that (T2*) all the biological ancestors of rational beings going back to Adam and Eve (other than Adam and Eve themselves) were descendants of Adam and Eve (i.e., *contra* T4, every biological ancestor of rational beings in every post-Adamic generation was a descendant of Adam and Eve)? The sources he cited do not get him beyond T2, but Chaberek’s rejection of T4 requires T2*.

Chaberek seems to think that T4 constitutes polygenism. It could be called a merely biological polygenism, but it does not challenge theological monogenism, T1 and T2, which are sufficient to ensure consistency with the doctrine of original sin, Pius’ ground for rejecting polygenism. So in a theological context, it should be recognized as a form of monogenism.

Chaberek raises three objections to T4.

First, he argues that the beings that T4 presupposes could not have existed. He seems to think that human beings have no adaptations conducive to survival and therefore depend entirely on their rationality (2025, 284). It is not, however, clear why the merely biological human beings would have had any more trouble surviving in their ecological niche than do, say, vervet monkeys or baboons in theirs. Chaberek has made two mistakes here.

His first is his paying exclusive attention to “adaptations in . . . bodily structure,” as though nothing in an animal’s behavioral repertoire could contribute to survivability. In fact, some of the very features which zoologists would list as adaptations in the primates I just mentioned are features that we human beings also have—sociability, resilience in the face of environmental variation, and complex communication patterns. Since those traits (at the animal level) are not dependent on rationality, there is no reason to doubt that merely biological human beings had them as well.

In addition, it is important to note that non-rational animals vary in intelligence (e.g., in learning and in problem-solving). Merely biologically human beings were surely very intelligent in the sense just specified. Although knowledge of God, and even language, would have been beyond their reach, applying these skills in ways conducive to survival would not have been.

Chaberek’s second mistake is his claim that the human body has no physiological adaptations that contribute to survivability. The idea that human beings have “unadapted, unspecialized” bodies is a product of Chaberek’s imagination, not of the close look at reality that he rightly says is important. If it is something like “natural weapons” that Chaberek wants, here are two—those that make it possible for us to make fists (Morgan: 2013) and to throw things (Darlington: 1975): “fists increase the peak stresses that are imposed on the target and, therefore, the potential for injury [to the target]” (Morgan 2013, 241); “the human arm [is] an efficient unique sling. No other animal can throw as man does. . . . A skillful man has a good chance to break the skull of another with one stone at 30 meters” (Darlington 1975, 3750). Neither fist-making nor stone-throwing require reason. Virgil may never have said “Beware of merely biological human beings bearing stones,” but if he had thought of it, he probably would have.

Second, Chaberek objects that the difference between merely biologically human beings and fully human beings is scientifically undetectable (Chaberek 2025, 286). Of course there are important differences that are scientifically undetectable, such as between the baptized and the unbaptized. This, for the most part, is not one. Rational beings can generally be distinguished from non-rational beings by their behavior (e.g., use of language). The fact that this is not yet true in infancy (or perhaps not at all in other cases of severe cognitive disability) just shows that not everything of importance can be established by science. Since the merely biological human beings are long gone, it is not entirely clear what moral problems he thinks come with my ideas; he does not say. The only thing I can think of is the level of respect due to fossil bones.

Third, he claims their existence is contrary to Church teaching. It is not. The documents Chaberek cites affirm T1 and T2, points included in the scenario. Earlier ones denied T3, but nothing he cites shows that T4 is contrary to Church teaching. The most he can say is that, T4 being motivated by T3 plus something like the very recent genetic evidence of trans-specific polymorphisms, no one who did not already accept those two points would have any reason to think T4 was true. But even the idea that T4 might have been an implicit merely traditional belief will not help Chaberek here.

Chaberek concedes that sometimes “a given interpretation of Scripture or Catholic belief . . . should be abandoned,” but wrongly says that this may be done only when “a certain claim (i.e., proven beyond any doubt) about nature would oppose [it]” (2024, 274), or “when things that contradict our belief are ‘certain from reason and experience’” (2025, 275). The criterion “proven beyond any doubt” is too strong, as will be shown from Church practice below. The phrase “certain from reason and experience” he found in St. Augustine ([415] 1982, 1.19.39⁴), but in Augustine’s text it does not play the role in the justification of revision that Chaberek gives it.⁵

Chaberek’s argument seems to depend on first distinguishing observable, and therefore allegedly certain, facts from allegedly uncertain theories, and then emphasizing (by italicization) the word “theory” in a quotation from St. Augustine (Chaberek 2025, 277), despite the fact that neither that word nor anything like it occurs at all in the original text.⁶ St. Augustine distinguishes what is asserted to be contrary to faith but can be shown not to be so from what really is contrary to faith. In any case, the “reason and experience” (emphasis mine) of Chaberek’s criterion includes more than the “direct observation” to which Chaberek reduces it.

Chaberek cites as an example of a justifiable abandonment of a merely traditional belief the acceptance of heliocentric astronomy, which, he says, is “directly observed in real-time” (275). This he contrasts with claims about the origin of the human race, which, he says, are not. What these direct astronomical observations were or who made them he never says. He cannot, because there never were any such observations. The certainty of heliocentrism is based not on direct observation of the motion of the

4. Chaberek’s citation, “1.19.3,” is a typographical error.

5. The context is a characterization of certain matters as ideas which a non-Christian would “hold to as being certain from reason and experience.”

6. “*Quidquid autem de quibuslibet suis voluminibus . . .*” (Augustine, [415] 1982, 1.21.41). Taylor did add the word “theory” in his translation of the passage in question, but its absence in the original is sufficient to make Chaberek’s italicization, and reliance on it in connection with his fact-theory contrast, misleading.

Earth, but on inference. It is the best explanation of the retrograde motion of the superior planets (Copernicus [before 1514] 1959, Ass. 7), of stellar aberration (Bradley 1728, 646–49), and of stellar parallax (Bessel, 1838). It is, that is to say, inferred, not directly observed.

Merely traditional beliefs can be called into question, indeed sometimes be modified or even abandoned, in the face of sufficiently convincing inference, even if only under the guidance of the Church's Magisterium (*Catechism* ¶183). One can see this in the history of T3, which, unlike T4, does challenge a merely traditional belief. Pius XII allowed discussion of "the origin of the human body as coming from pre-existent and living matter" even while asserting that there were reasons both for that opinion and for its contrary (Pius XII 1950, 36)—so, in his judgment, before it met Chaberek's condition of having been "proven beyond any doubt." Thirty six years later, John Paul said that "there is no apparent difficulty in explaining the origin of man, as far as concerns the body, by the hypothesis of evolutionism," even though "that hypothesis puts forward only a probability, not a scientific certainty" (John Paul II [1986] 1996a).

In conclusion: magisterial documents show that scientific arguments in favor of animal origins of the human body (T3) have at least opened up the question of, if not forced a revision of, a traditional theological belief connected with anthropogenesis. My consistency proof shows that scientific arguments for biological polygenism would not, no matter how strong they might be, require any revision of the Church doctrine of theological monogenism (T1 and T2), which (though Chaberek does not always seem willing to acknowledge this point) I have consistently affirmed as true. The distinctive thesis of the consistency proof (T4) is inconsistent neither with the doctrine of original sin as traditionally understood, nor with theological monogenism, nor with anything that can be found in magisterial documents.

BIBLIOGRAPHY

- Agassiz, Louis. 1850. "The Diversity of Origin of the Human Races." *Christian Examiner and Religious Miscellany* 49 (1): 110–45.
- Augustine. (415) 1982. *De Genesi ad litteram libri duodecim*. Trans. John Hammond Taylor, SJ. *The Literal Meaning of Genesis. Volume I*. New York: Newman.
- Ayala, Francisco J. 1998. "Evolution and the Uniqueness of Humankind." *Origins: CNS Documentary Service* 27 (34): 565–74.
- Bessel, Friedrich Wilhelm. 1838. "Bestimmung der Entfernung des 61sten Sterns des Schwans." *Astronomische Nachrichten* 16 (5): 65–96. Ed. and trans. John Herschel, "A letter from Professor Bessel to Sir J. Herschel, Bart., dated Königsberg, Oct. 23, 1838." *Monthly Notices of the Royal Astronomical Society* 4 (17): 152–61. <https://doi.org/10.1093/mnras/4.17.152>.

- Bradley, James. "A Letter from the Reverend Mr. James Bradley Savilian Professor of Astronomy at Oxford, and F.R.S. to Dr. Edmond Halley Astronom. Reg. &c. Giving an Account of a New Discovered Motion of the Fix'd Stars." *Philosophical Transactions of the Royal Society of London* 35 (406): 637–61. <https://doi.org/10.1098/rstl.1727.0064>.
- Catechism of the Catholic Church*. 1994. Vatican City: Libreria Editrice Vaticana.
- Chaberek, Michał. 2024. "Original Sin, Monogenesis and Human Origins A Response to Kenneth W. Kemp." *Forum Philosophicum* 29 (1): 153–65. <https://doi.org/10.35765/forphil.2024.2901.08>.
- . 2025. "The Arches and the Spandrels: A Response to Kenneth W. Kemp (2)." *Forum Philosophicum* 30 (1): 273–87. <https://doi.org/10.35765/forphil.2025.3001.13>.
- Copernicus, Nicolaus. [Before 1514] 1959. *Commentariolus*. Trans. Edward Rosen in *Three Copernican Treatises*. New York: Dover.
- Darlington, P.J. 1975. "Group Selection, Altruism, Reinforcement, and Throwing in Human Evolution." *Proceedings of the National Academy of Sciences* 72 (9): 3748–52. <https://doi.org/10.1073/pnas.72.9.3748>.
- John Paul II. 1996a. "Created Things Have a Legitimate Autonomy." General Audience, 16 April 1986. In *God, Father and Creator: A Catechesis on the Creed*, 216–20. Boston: Pauline Books & Media.
- . 1996b. Message to the Pontifical Academy of Sciences. In *Origins: CNS [Catholic News Service] Documentary Service* 26 (25): 414–16.
- Kemp, Kenneth W. 2011. "Science, Theology, and Monogenesis." *American Catholic Philosophical Quarterly* 85 (2): 217–36. <https://doi.org/10.5840/acpq201185213>.
- . 2020. "God, Evolution and the Body of Adam." *Scientia et Fides* 8 (2): 139–72. <https://doi.org/10.12775/SetF.2020.017>.
- . 2023. "Evolution, Adam, and the Catholic Church." *Logos* 26 (1): 22–43. <https://dx.doi.org/10.1353/log.2023.0006>.
- . 2024. "Monogenism: A Reply to Fr. Chaberek." *Forum Philosophicum* 29 (2): 391–99. <https://doi.org/10.35765/forphil.2024.2902.09>.
- La Peyrère, Isaac. 1655. *Præ-adamitæ sive Exercitatio super versibus duodecimio, decimotertio, & decimoquarto, capitis quinti Epistolæ D. Pauli ad romanos*. Translated as *Men before Adam: A Discourse upon the Twelfth, Thirteenth, and Fourteenth Verses of the Fifth Chapter of the Epistle of the Apostle Paul to the Romans, By which are Prov'd, that the First Men were Created before Adam*.
- Leo XIII. 1893. *Providentissimus Deus*. Accessed October 12, 2025. https://www.vatican.va/content/leo-xiii/en/encyclicals/documents/hf_l-xiii_enc_18111893_providentissimus-deus.html.
- Morgan, Michael H. 2013. "Protective Buttressing of the Human Fist and the Evolution of Hominin Hands." *Journal of Experimental Biology* 216 (2): 236–44. <https://doi.org/10.1242/jeb.075713>.
- Pius XII. 1950. *Humani Generis*. Accessed October 12, 2025. https://www.vatican.va/content/pius-xii/en/encyclicals/documents/hf_p-xii_enc_12081950_humani-generis.html.

Is Saruman a “Peacemaker,” and Abortion “Murder”?

Report from the Debate

Michał Zalewski

On November 8th, 2024, another of the online debates organized periodically by Ignatianum University in Cracow took place. The subject of the discussion was the book *Argumenty Semantyczne: Pojęcie, Podział, Kryteria Oceny* [*Semantic Arguments: The Concept, Classification, and Criteria for Assessing*] by Jakub Pruś, published by Ignatianum University Press in 2023. The main idea of the book is that semantic arguments should be distinguished as a special type of argument, with a separate scheme and critical questions for assessing them. The book was debated by three invited guests: Prof. Agnieszka Lekka-Kowalik (The John Paul II Catholic University of Lublin), Prof. Adam Jonkisz (Ignatianum University in Cracow), and Dr. Szymon Makula (University of Silesia in Katowice).

In the book, a “semantic argument” is defined as “an argument in which the meaning of a given term is established in order to support the persuasive goal of the speaker.” Pruś explains that he has borrowed this term from the Polish logician Teresa Hołówka, but has elaborated it in terms very different to those she originally had in mind. To illustrate this type of arguing, he provides several examples supporting the conclusion that “Julius Caesar was a criminal.” One appeals to historians’ opinion (argument from authority), another compares Caesar to Stalin (argument from analogy), and the final one is constructed by establishing the criteria for correct employment of the term “criminal” (e.g., “if someone kills or orders the killing of many thousands of people, then he is a criminal”). This last argument would

count as semantic, since it fixes the meaning of a particular term in order to support its claim.¹

The debate started from a short presentation of the book, falling essentially into two parts that were concerned, respectively, with argumentation schemes for semantic arguments and criteria for assessing semantic arguments. Once the presentation had finished, each of the invited guests presented a short review of Pruś's work and made some critical comments.

1. SEMANTIC ARGUMENTS: ARGUMENTATION SCHEMES AND CRITICAL QUESTIONS

Pruś distinguished two basic schemes for semantic arguments: arguments using definition (including Peter of Spain's distinction between arguments *to* and *from* definition) and arguments from classification.

Arguments to definition aim to include certain objects within the extension of a given definition, with this sometimes being the very end served by the reasoning in question ("this is racism"). Obviously, the semantic modification will be introduced through persuasive definition,² via the definition's premise:

- *Individual Premise*: *A* possesses some property *F*.
- *Definition Premise*: For all *x*, if *x* possesses some property *F*, then *x* fits definition *D* (where such a definition is, in the given context, controversial).
- *Conclusion*: *A* fits definition *D*.

It may happen that the argument can start out from a definition, and include a certain property in it, in order to attribute this property to a certain object ("All forms of racism should be prosecuted, therefore this given action also should be prosecuted"). Thus, the arguments to definition and from definition can be used together.

- *Definition Premise*: *A* fits definition *D*.
- *Classificatory Premise*: For all *x*, if *x* fits definition *D*, then *x* is classified as having the property *G* (where such a classification is, in the given context, controversial).
- *Conclusion*: *A* has the property *G*.

However, persuasive definitions are not the only way of modifying meaning: one may also modify meaning simply by linking two properties

1. The first formulation of the concept of semantic argument, and its definition, were presented in (Pruś 2019, 2020). The first typology of semantic argument was presented in (Pruś 2021), and the criteria for assessing semantic arguments in (Pruś and Macagno 2024).

2. "Definition is persuasive if it is put forward to support one's claim in the discussion and there is an alternative definition of a given term" (Pruś and Aberdein 2022, 33).

together,³ and then attribute a new property to a given object (“if something is prosecuted, then it is morally wrong, therefore racism is morally wrong”).

- *Individual Premise*: A possesses some property *F*.
- *Classificatory Premise*: For all *x*, if *x* possesses property *F*, then *x* can be classified as possessing property *G* (where such a classification is, in the given context, controversial).
- *Conclusion*: A possesses property *G*.

These are the three most common types of semantic argument. Prúš provided three examples of such arguments. The first one refers to Tolkien’s novel *Lord of the Rings*:

P1. (Hobbes’s definition): Peace is the absence of war.

P2. (by definition): A warmonger is one who seeks to bring about acts of war.

P3. (by definition): A defender of peace is one who seeks to avoid acts of war.

P4. Théoden seeks to bring about acts of war.

P5. Saruman seeks to avoid acts of war.

C. Therefore, Théoden is a warmonger, and Saruman is a defender of peace.

The fixed meaning of “peace” in P1 (which implies P2 and P3) serves here to support the main conclusion, which seems rather controversial. This argument parallels a similar debate over whether Chamberlain or Churchill should be classified as a peacemaker (see: Harold 2013). In this case, one can clearly see that there is something wrong with that definition.

Another example is provided by a debate between three medical doctors in the *Journal of Obstetrics and Gynecology*. In the article “A Consideration of Therapeutic Abortion,” Samuel A. Cosgrove and Patricia A. Carter state:

The fetus is a human individual with all the potentialities of every human being, and that its destruction is murder, only justifiable in the most extreme circumstances involving direct and imminent threat to the mother’s life. (1944, 299)

In the next issue of the journal, T.W. Jones responded with the following:

In my belief, murder is an unpleasant and ugly word. His [Cosgrove’s] definition of murder, premeditated destruction of human life, is excellent as far

3. Douglas Walton calls this an *Argument from Verbal Classification* (Walton 2008, 2005; Walton and Macagno 2009).

as it goes, but it does not go far enough. One word is needed for completion, namely, malice. (1944, 895)

He then went on to offer the dictionary definition of murder to support his view:

I substantiate my viewpoint with definitions from Webster's International Dictionary: "Murder—n—The offense of killing a human being with malice, pretense, or aforethought, express or implied; intentional and unlawful homicide." (1944, 895)

In his reply, Cosgrove arrives at the opposite conclusion, albeit starting from the same definition of murder. He argues that while murder does indeed require malice, abortion itself entails an implied form of malice. To substantiate this claim, he appeals to the same dictionary definition of "malice" used by Jones. On this basis, Cosgrove concludes that abortion does, in fact, fall under the definition of murder.

So what do these two arguments, one on peace (and war) and the other on murder and abortion, have in common? In both of them we encounter a fixing of meaning (via persuasive definition), which latter is used to support the conclusion. How to assess such arguments? Prus described the six main criteria for evaluating semantic arguments, presenting these in the form of critical questions:

1. The transparency criterion: "Is the person establishing the meaning doing so in a transparent way?"
2. Consistency with usage: "Does the established meaning correspond to some common linguistic practice? Are we in some way violating established usage?"
3. Consequences: "Does the established meaning support acceptable values and interests? What will happen if we adopt such a meaning?" (This refers to social, political, and moral consequences.)
4. Authority: "Does the person establishing the meaning have the authority to do so? Is this required by some relevant office or document?" (This is especially important in legal definitions.)
5. Alternatives: "Are there any alternative ways of understanding the meaning in question, or could the given meaning be understood differently?"
6. Individual premise criterion: "Does the given object really have the property that qualifies or disqualifies it from falling

within the scope of the definition?" (e.g., in the case of a ban on vehicles in a park, is the given object actually a vehicle?).

In the wake of the presentation (summarized above) of the general notion of semantic arguments, along with their types and critical questions, the debate commenced.

2. KEY THEMES AND DISCUSSIONS

Instead of presenting its contents in their original chronological sequence we will group the comments made by the invited guests and the author himself into a few key themes.

2.1. *Rationality in Argumentation and the Pragmatic Assumption*

Agnieszka Lekka-Kowalik asked whether considerations pertaining to argumentation presuppose the rationality of the subject. Pruś responded that he did not believe argumentation contains rationality in its definition, although reasoning (a part of argumentation) may of course be flawed. He stressed that the pragmatic definition of a semantic argument, with the added note about the controversiality of the definition, is essential. Without it, even the Socratic syllogism could be wrongly classified as a semantic argument.

2.2. *"Getting at the Essence of Things" vs. Language Manipulation*

Lekka-Kowalik raised the problem of distinguishing "uncovering the essence of things" from manipulating language. Referring to Aldous Huxley's *Eyeless in Giza* (about "true freedom" and "true abstinence"), Pruś noted that adding adjectives such as "true," "real," or "authentic" can serve manipulative purposes. He argued that those who manipulate concepts (by subtly altering meanings under the guise of citing dictionary definitions) are essentially reshaping arguments in a way that makes them tautological. For Pruś, definitions should be treated as arguments subject to critical evaluation, not as unquestionable "revelations."

2.3. *Rational vs. Irrational Justifications for Changing Definitions*

In the debate about Pluto's planetary status, Pruś pointed out that the "irrationality" of the justification lay in weighing pragmatic consequences. The International Astronomical Union's decision to redefine "planet" stemmed from the discovery of many celestial bodies larger than Pluto. The choice boiled down to whether to recognize 9, 12, or perhaps 20+ planets. The decision, though not "rational" in the sense of capturing the "essence of a planet," was pragmatic, focusing as it did on the largest and most important bodies in the Solar System. He called this approach "soft

pragmatism,” encompassing both essentialists and constructivists, since ultimately both appeal to consequences. Even appeals to the “essence of things” can be reframed as consequences (e.g., the consequence of being inconsistent with the cognitive value of truth).

2.4. *The Problem of Defining Eristic and Terminological Consistency*

Adam Jonkisz identified a terminological inconsistency in the book: namely, that certain concepts were defined in terms of activities—such as *argumentation* and *persuasion*—whereas *eristic* was characterized as an art or skill. This discrepancy, he argued, undermines the internal coherence of the conceptual framework. In response, Pruś acknowledged the inconsistency and conceded that defining *eristic* as an activity would have been more appropriate. Such a redefinition ensures uniformity across the set of terms, thereby enabling the substituting of these concepts in more complex definitions. This, in turn, makes it possible to assess whether the terms mutually correspond, and whether the resulting system of definitions yields a comprehensive and coherent theoretical framework.

2.5. *The Strength of Semantic Arguments and Critical Questions*

Szymon Makuła raised the issue of the lack of a theory of “argument strength” in argumentation studies. He proposed using critical questions as a tool for assessing argument strength and asked why these particular questions, and not others, were chosen. He suggested that critical questions should be designed so that their answers also address other questions, thereby maximizing informational value. Specifically, he proposed replacing the first critical question (“Was the change in meaning made transparently?”) with “Does the argument provide justification for the change in meaning?” Justification implies transparency, but not vice versa. Pruś agreed that this proposal was “elegant,” stressing that changes in meaning (e.g., of the word *Negro*) require justificatory reasons, and that without justification the argument will be weak. He also added that in cases of unsettled meanings (e.g., abortion, gender), unjustified semantic arguments may still be considered, though his framework focuses on usefulness.

2.6. *Pragmatism vs. Essentialism: Reality and Definitions*

Agnieszka Lekka-Kowalik repeatedly raised the following issue: if definitions do not reach as far as reality, are we condemned to a “battle of definitions,” with disputes settled by force rather than by reality itself?

The “wetlands” example: Pruś noted that in the case of “wetlands,” the decision to narrow the definition was political, it being aimed

at enabling development. Even if scientists provided factual information, the final decision was political.

The “woman” example: Lekka-Kowalik countered with the example of defining “woman” as “a sexual object serving to provide men with pleasure.” She questioned whether such a definition can be accepted if it lacks any connection to the “essence of things.”

Pruś maintained that agreement or disagreement with such a definition stems from rejection of exploitation of one half of humanity by the other half, not from appeals to “essence.” He stressed that reality “speaks to us” through science and through others’ voices, and that his framework aims to furnish evaluative criteria that will prove useful across a range of philosophical positions (be they essentialist or constructivist).

Makula added that definitions are logical tools for resolving disputes about meaning or regulating ambiguity. Language changes for various reasons, including the discovery of new phenomena (e.g., depression, gender issues). This does not imply abandoning reality, but acknowledging its presence in usage, scientific data, or empirical testing of operational definitions. He noted that in everyday life people often act without strict definitions (e.g., of “trash”).

In conclusion, Pruś emphasized that his proposal was not a theory of truth, but a practical tool for evaluating arguments—analogueous to evaluating arguments from authority, which also do not require adopting a specific theory of truth or metaphysics. The focus should be on “applicability” and “usefulness.”

The debate revealed the complexity of semantic arguments, which underlie many disputes—scientific, political and social. The key insight that emerged was that definitions have consequences, and their acceptance often stems from pragmatic considerations, interests and values, rather than from uncovering the “essence of things.” Pruś’s proposal offers tools for the systematic evaluation of these arguments, taking into account their context and consequences, and linguistic usage. The challenge is to find a way to balance pragmatism with appeals to reality, in order to avoid purely power-based resolutions in debates.

REFERENCES

- Cosgrove, Samuel A., and Patricia A. Carter. 1944. “A consideration of therapeutic abortion.” *American Journal of Obstetrics and Gynecology* 48 (3): 299–314.
- Harold, James A. 2013. “Distinguishing the Lover of Peace from the Pacifist, the Appeaser, and the Warmonger.” *Forum Philosophicum* 18 (1): 5–17.
- Jones, T.W. 1944. “Is Abortion Murder?” *American Journal of Obstetrics and Gynecology* 48 (6): 895–96.

- Pruś, Jakub. 2019. „Argumentacja semantyczna – podstawowe pojęcia i problemy definicyjne.” *Res Rhetorica* 6 (4): 56–76. <https://doi.org/10.29107/rr2019.4.3>.
- . 2020. „Znaczenie znaczenia w argumentacji. Zarys argumentów semantycznych.” In *Słowo: struktura – znaczenie – kontekst*, edited by Ewa Szkudlarek-Śmiechowicz, Agnieszka Wierzbicka and Elwira Olejniczak. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- . 2021. “How Can Modifications of Meaning Influence Argumentation? The Concept and Typology of Semantic Arguments.” *Argumentation* 35 (3): 483–508. <https://doi.org/10.1007/s10503-020-09542-y>.
- Pruś, Jakub, and Andrew Aberdein. 2022. “Is Every Definition Persuasive? Douglas Walton on Persuasiveness of Definition.” *Informal Logic* 42 (1): 25–47. <https://doi.org/10.22329/il.v42i1.7211>.
- Pruś, Jakub, and Fabrizio Macagno. 2024. “When Meaning Becomes Controversial. Critical Questions for Assessing Semantic Arguments.” *Informal Logic* 44 (2): 208–48. <https://doi.org/10.22329/il.v44i2.8435>.
- Walton, Douglas. 2005. “Deceptive Arguments Containing Persuasive Language and Persuasive Definitions.” *Argumentation* 19: 159–86. <https://doi.org/10.1007/s10503-005-2312-y>.
- . 2008. “The case of redefining “Planet” to Exclude Pluto.” *Informal Logic* 28 (2): 129–54.
- Walton, Douglas, and Fabrizio Macagno. 2009. “Reasoning from Classifications and Definitions.” *Argumentation* 23: 81–107. <https://doi.org/10.1007/s10503-008-9110-2>.

Report from the 3rd International Christian Philosophy Conference, “Christian Philosophy Facing Naturalism”

Oskar Lange

The international conference “Christian Philosophy Facing Naturalism” took place at Ignatianum University in Cracow from September 24th to 25th, 2024. This is the third event in the Christian Philosophy conference series organized by the Institute of Philosophy at UIK. Conference participants included philosophers from all over the world, including Poland, the United States of America, India, Great Britain, Hungary, Italy, the Czech Republic, Germany, and France. Among the guests were keynote speakers Robert C. Koons, Charles Taliaferro, Jacek Wojtysiak, Włodzisław Duch and Georg Gasser. Over two days, dozens of scholars discussed Christian philosophy, naturalism, and the complicated relationship between these two perspectives. Appropriately construed, a range of topic areas associated with science, artificial intelligence, logic, history, theology and politics were also discussed.

The project was co-financed with state budget funding from the Minister of Education and Science, within the framework of the “Scientific Excellence II” Programme (for Poland) and the City of Cracow.

The conference featured a debate entitled “The controversy over the naturalistic image of the world and man” between Georg Gasser and Włodzisław Duch. At the outset, Gasser differentiated between reductive and non-reductive, metaphysical and methodological, and “hard” and “soft” naturalisms. In the context of the humanities, hard naturalism, in which we reduce all phenomena to natural phenomena, is most often spoken of. According to Gasser, this type of approach is limiting when we talk about human beings (for example, in a therapeutic context). Duch, meanwhile, described himself as agreeing with Popper about the existence of three

worlds: more precisely, on his view, mental phenomena result from physical phenomena. Referring, for example, to the operation of GPT-4o networks, he held that science can produce systems that exhibit characteristics analogous to the human mind.

In the subsequent discussion, the speakers revealed the more nuanced aspects of their positions. Gasser pointed to the permeating, holistic nature of the mind, which is in constant relation to reality. Duch, on the other hand, agreed that the physical processes corresponding to thinking are subject to interpretation, and even a full understanding of the physical structure of the brain is not enough to fully understand the human being.

In addition to this official debate, we may mention at least one more, which arose between Peter van Inwagen and Charles Taliaferro. In his presentation, Inwagen analyzed, among other things, the problems of dualism, theism and the simulation hypothesis, starting from an ontological definition of naturalism. On this definition, naturalism is a view according to which everything consists only of fundamental entities lacking mental or teleological properties.

Charles Taliaferro, in his lecture, referred to Peter van Inwagen's critique of the argument from reason. Taliaferro placed himself in the tradition of the Cambridge Platonists, and in this spirit reinterpreted the argument from reason by going all the way back to its Platonic roots. He also referred to contemporary critics of the argument, including Inwagen. For Inwagen, as interpreted by Taliaferro, the argument from reason refers to the principle of reason, which is an abstract object, and abstract objects have no effect on things. The author of the paper expressed a different position towards the nature of form and abstract objects by giving a classic example of solving a mathematical problem that requires both abstract and material conditions to be taken account of.

In his lecture, Robert C. Koons presented the Aristotelian-inspired argument from the First Mover of St. Thomas, supplemented with a temporal perspective describing God as an out-of-time, unchanging mover, being outside of the chain of infinite causes and effects. To illustrate this concept and fend off possible objections, he drew an analogy between the author of a book who, in creating a certain narrative, is outside of its time, and God. In his lecture he showed how such a perspective avoids the most popular arguments levelled against Aquinas' First Mover argument. In his view, the approach of St. Thomas serves to prove the existence of an extra-temporal cause, and extra-temporality is a quality possessed only by God.

In his lecture, Jacek Wojtysiak analyzed the concepts of naturalism, religion and explanation with a view to properly elaborating the question

“Does naturalism explain religion?”—which was the title of his talk. In his argument for the existence of God, he proposed two possible alternatives: the N-world, in which God does not exist, and the T-world, in which He does. He then asked the question “In which world is religion more likely to occur?” and argued that this is more likely in the T-World. He referred in his argument to such aspects as motivation (in the T-world, there is a God who wants religion to exist), realization (in the T-world, religion is necessary), and harmony (there is a correspondence between the epistemic and ontological aspects of religion). Toward the end of the presentation, Wojtysiak also presented an additional argument for belief in a deity by showing that, assuming a naturalistic definition of religion, there are rational reasons for religious belief (e.g., utility, coherence, non-contradiction).

Among the presentations by other participants were critical studies and presentations of many other arguments and positions in the debate between naturalism and Christianity. These included the following: A. Plantinga’s argument against naturalism (Christopher Oldfield, Piotr Biłgorajski, Piotr Bylica, Norbert Heger and Andrzej Zabołotny) Wittgenstein’s philosophy (Carl Humphries, Ines Skelac and Christian Kanzian), non-religious anti-naturalism (Maciej Jemioł, Oskar Lange and Bartosz Wesół), cognitive science and psychology (Stanisław Ruczaj, Szilvia Finta and Evelina Deyneka) and many others (Walter Menezes, Adam Świeżynski, Mirosław Rucki, Alexander Barrientos, Miles Kenneth Donahue, Luca Gasparinetti; Margherita Moro, Piotr Mazur, Jiří Baroš, Antonios Kaldas, Krzysztof Pięta, Piotr Duchliński; Jarosław Kucharski, Guido J.M. Verstraeten, Marcin Podbielski, Robert B. Tierney, Tymoteusz Mietelski, Jacek Surzyn, Finley I. Lawson, Andrzej Karpinski, Przemysław A. Lewicki). All of the presentations are available on the YouTube channel of Ignatianum University in Cracow.

Call for Papers

Christian Philosophy: Between Christian and Post-Christian Worldviews, 22–23 September 2026, Ignatianum University in Cracow

The term “post-Christian” is increasingly appearing in philosophical and cultural discourse, employed to describe various phenomena that supposedly follow on after Christianity. Most often, the term is used to describe a contemporary world in which Christianity either is no longer the dominant religion or is not recognised as such in the way that it was until recently. At the same time, although there is a post-Christian world, the Christian world has not ended. The problem of the “post-Christian picture of reality” therefore provokes discussion amongst both supporters and opponents of Christianity – especially because what is “post-Christian” cannot be understood in isolation from Christianity itself.

In a globalised world, we are witnessing a clash between Christian and post-Christian images of the world. While some recognise the permanence and validity of the picture of reality founded on the Christian religion, others are convinced that this has, for various reasons, been deformed or destroyed and belongs to an irreversible past, both in terms of cognition and at the level of social practice.

While within Western civilisation broadly construed a post-Christian worldview founded on ecological, gender-based or technological naturalism would seem to be dominating, in other parts of the globe the Christian worldview is only just gaining ground.

The situation in which Christian and post-Christian worldviews clash within culture and social life poses a serious challenge for philosophy. Christianity-inspired philosophy must define its place in relation to not only worldviews, but also phenomena, trends and concepts with anti-Christian overtones. At the same time, the post-Christian worldview raises many questions that need to be addressed.

Proposals

We invite proposals that address the problems of Christian and post-Christian worldviews. Our interests lie especially in the following topics and questions, but are not limited to them:

Selected Problems

- What are the main historical and systematic problems of the Christian worldview?
- Is an evolution of the Christian worldview possible, or even necessary?
- What is the difference between post-Christian worldviews and non-Christian or post-religious worldviews?
- What are the main aspects and characteristics of the relationship between Christian and post-Christian worldviews?
- Is the transition between Christianity and post-Christianity itself an irreversible phenomenon?
- In what way is post-Christianity influencing debates in ethics and/or politics?
- Does the post-Christian worldview lead to a dissolution of our deep need for religious truths or values?
- Why is the post-Christian worldview mostly dominated by materialistic and relativistic perspectives that reject God as a person and the spiritual values of Christianity?
- What kind of personal identity and individual existence is being presented within the post-Christian worldview?
- Why is it that, in the post-Christian world, religion is becoming a tool of political mobilisation and/or manipulation?
- What is the function of religion within the Christian and post-Christian worldviews?
- Is the very meaning of Christianity dissolved in the post-Christian worldview into a set of broad ideals about human behaviour and society?
- What is the position of the Christian and post-Christian worldviews on the truth-falsehood opposition?

Submissions

Proposal Submission: Please submit a 500-word abstract of your paper (in PDF format) by March 31, 2026, via EasyChair, using the following link: <https://easychair.org/conferences?conf=chp26>

Language: only proposals in English will be accepted for consideration.

We will be delighted to encounter all participants in person here at Ignatianum University in Cracow. However, the organisers plan to conduct this conference in hybrid mode, combining both online and on-site elements. Each conference participant will receive a certificate indicating also the mode of participation.

Keynote speakers

We are pleased to announce that the following individuals have agreed to give a lecture or participate in a panel discussion during the conference:

Jeffrey Bloechl – Boston College, USA

Chantal Delsol – University of Paris-Est Marne-la-Vallée, France

Piotr Gutowski – John Paul II Catholic University of Lublin, Poland

John Milbank – University of Nottingham, UK

Linda Trinkaus Zagzebski – University of Oklahoma, USA

Fees

The conference is open to the public (also via social media). Presenting participants will be charged a fee to help cover costs (materials, dinner, coffee breaks, etc.). For the exact amount of the conference fee, see below. Early submission (up to December 31, 2025) will attract a reduced fee (so-called ‘Early Bird registration’).

Regular participants

60/80/100 EUR (Early Birds/PhD Students/Regular Participants)

Online participants

30/40/50 EUR (Early Birds/PhD Students/Regular Participants)

Publication

We plan to record all presentations and then publish them on conference YouTube channel and on the conference Facebook fanpage. After the conference we plan to publish a special issue in a philosophical journal, containing articles based on the conference presentations. With this in mind, speakers are encouraged to prepare a paper (up to 10,000 words) and submit it by December 31, 2026. Each article will be subject to a process of double-blind peer review. *Forum Philosophicum*, an international journal for philosophy (listed in SCOPUS), has already agreed to publish a special issue in 2026 including materials from the conference. However, we are also open to collaboration with other journals.

Deadlines

- Submission of Proposals (Early Birds): December 31, 2025
- Submission of Proposals: March 31, 2026
- Notification of Acceptance: April 30, 2026
- Registration Deadline and Payment: June 30, 2026
- Conference Dates: September 22–23, 2026
- Paper Submission Deadline: December 31, 2026

For video materials documenting previous editions of our conference, go to:
 “Christian Philosophy: Its Past, Present, and Future”, September 22–23, 2020:
https://www.youtube.com/playlist?list=PL2RdbxAZiEAmJWEvkKKt60Kj_ET8sE03s

“Christian Philosophy and its Challenges”, September 20–22, 2022: <https://www.youtube.com/playlist?list=PL2RdbxAZiEAn5Dj9XIZBE7LYYGivqmL96>

“Christian Philosophy facing Naturalism”, September 24–25, 2024:
https://www.youtube.com/playlist?list=PL2RdbxAZiEAlYPFTTeKfUAtN_7Sb5Jgwy

More information can be found on our website:

<https://christianphilosophy.ignatianum.edu.pl>

...or on our Facebook profile:

<https://www.facebook.com/christianphilosophyconference/>

If you have questions, please contact the Conference Secretary at:

christianphilosophy2026@ignatianum.edu.pl

Note about *Forum Philosophicum*

Forum Philosophicum is a scholarly journal that seeks new insights into philosophical discussions concerning the relationship between philosophy and faith. The tradition we are rooted in is that of Christian philosophy, understood as the independent exercise of reason in the context of the faith. However, we also welcome proposals that flow from the concerns important to philosophers shaped by other religious traditions, in particular those of Jewish or Muslim heritage. We seek articles that are able to engage in critical debate with evolving views on rationality and its relationship to the irrational, the indeterminate, and the nonbinary, and look forward to receiving analyses that discuss the place of faith and the rationality of faith in such contexts. We encourage discussions of human nature and spirituality that seek to address the proposals of neuroscience and confront the declarations of posthumanists, as well as inquiries into that mode of philosophizing which, while resting on the hermeneutics of culture, gives special emphasis to the hermeneutics of sacred texts. We would like to distinguish our journal by means of its unique approach, rather than through the specificity of the subjects we are prepared to see raised in it: an approach that allows faith to shape, enlighten, and strengthen rationality itself. In this way, we want to offer a true Forum for the community of philosophers who view their faith as a source of inspiration, especially when seeking to face up to the new and challenging forms being given to the ageold problems of philosophy.

Forum Philosophicum is published in English by Ignatianum University Press, part of the Ignatianum University in Cracow. Its two annual issues, published in June and December, together furnish around 30 papers. Until recently, these were published in a range of languages recognized as appropriate for international conferences, but as of 2011 the journal has accepted papers exclusively in English. It is, in its basic version, a traditional

printbased journal, but also has a parallel online version, available via the EBSCO Academic Search Complete electronic database (since vol. 6, 2001) and the Philosophy Documentation Center online subscription service (since vol. 1, 1996). Since vol. 23, 2018, all articles have been published in full openaccess model based on Creative Commons Attribution (CC-BY) license. Selected earlier articles are published in open access on the journal's website. The journal is also indexed in, among others, The Philosopher's Index, Atla Religion Database, and the PhilPapers database.

SUBMISSIONS

Forum Philosophicum only accepts submissions in English. We publish papers, cycles of papers, discussions and book reviews. Papers submitted for consideration in *Forum Philosophicum* cannot exceed 15,000 words, including footnotes and bibliography. They should be preceded by an abstract not exceeding 200 words. Manuscripts should be submitted via our website at:

<https://czasopisma.ignatianum.edu.pl/fp>

All papers accepted for publication will go through a linguistic review process to ensure the utmost clarity and accessibility of submissions.

EDITORIAL BOARD

Jakub Pruś, Editor-in-Chief,

Ignatianum University in Cracow, Poland

Szczepan Urbaniak SJ, Deputy Editor,

Ignatianum University in Cracow, Poland

Maciej Jemioł, Editorial Secretary,

Ignatianum University in Cracow, Poland

Férdia StoneDavis, International Assistant Editor,

Margaret Beaufort Institute of Theology, Cambridge, UK

Daniel Spencer, International Assistant Editor,

University of St Andrews, UK

Rev. Mark Sultana, International Assistant Editor,

University of Malta, Faculty of Theology, Malta

Magdalena Jankosz, Language Editor,

Pontifical University of John Paul II, Poland

Michał Zalewski SJ, Associate Editor for Reviews,

Ignatianum University in Cracow, Poland

ADVISORY BOARD

Andreas Wilmes,

West University of Timisoara, Romania

Christopher Wojtulewicz,
Katholieke Universiteit Leuven, Belgium
Marcin Podbielski,
Ignatianum University in Cracow, Poland
Petr Dvořák,
Academy of Sciences of the Czech Republic
Dariusz Łukasiewicz,
Kazimierz Wielki University in Bydgoszcz, Poland
David Pratt,
Georgetown University, WA, USA
Andrey Tikhonov,
Southern Federal University at RostovonDon, Russia
Elizabeth Burns,
Department of Theology and Religious Studies, King's College
London, UK
Alex R Gillham,
St. Bonaventure University, NY, USA
Joeri Schrijvers,
NorthWest University of Potchefstroom, South Africa
Till Kinzel,
Technische Universität Braunschweig, Germany

CONTACT INFORMATION

Forum Philosophicum

Ignatianum University in Cracow
ul. Kopernika 26, 31501 Kraków, Poland
(main building, room 310)
E-MAIL forum.philosophicum@ignatianum.edu.pl
PHONE +48 12 39 99 661
FAX +48 12 39 99 501
WEBSITE <https://czasopisma.ignatianum.edu.pl/fp>

PUBLISHER INFORMATION

Ignatianum University Press
ul. Kopernika 26, 31501 Kraków (Poland)
Wydawnictwo WAM (main building, room 358)
PHONE +48 12 39 99 620
FAX +48 12 39 99 501
EMAIL: wydawnictwo@ignatianum.edu.pl

REVIEW PROCESS

The review process for papers submitted to *Forum Philosophicum* adheres to the following guidelines:

1. A paper submitted to the Editor may be returned to the author if it does not meet the criteria described in the section Submissions, or if it does not discuss a topic within a subjectarea of interest to *Forum Philosophicum*.
2. Each paper will be sent to two reviewers for blind review. At least one of the reviewers must be affiliated to an institution in a different country than the author.
3. Members of the Board, including the Editor, are not allowed to write reviews. No specialist who is known to be related to the author or who may reasonably be thought to have collaborated with the author or supervised his / her work is eligible as a reviewer.
4. Reviewers are explicitly asked to focus on the research and philosophical merits of a paper; they are told that any linguistic and/or technical deficiencies pertaining to an otherwise good paper may be dealt with separately by the editors.
5. The reviewers are asked to conclude their review with a clear opinion, stating that a paper is (a) publishable without revisions, or (b) may be published after minor revisions have been introduced, or (c) may be resubmitted after major revision, or (d) should not be considered for publication at all.
6. The final decision concerning which of the papers that have obtained at least one positive review are to be published is taken by the EditorinChief, in consultation with the Editorial Board.
7. The names and affiliations of all reviewers are published annually on the website of the journal.

LEGAL ISSUES

COPYRIGHT *Forum Philosophicum* is published by Ignatianum University Press under a Creative Commons Attribution 4.0 International License. Authors of papers that have been accepted will be asked to grant the publisher the royaltyfree right to first publication, as well as a nonexclusive right to distribute and archive the paper in electronic format, including through organizations specializing in the distribution and archiving of scholarly journals. It should be understood that the fact of publication constitutes the sole form of remuneration to be received by authors.

FINANCING

All authors whose research is supported by special sources of financing, such as grants, research programs, etc., are asked to prepare an information note mentioning the support they have received. *Originality Forum Philosophicum* will only accept manuscripts of papers not previously published in, or submitted to and still currently under review by, another journal or collection.

The Polish Ministry of Science and Higher Education requires the editors of all Polish journals which they evaluate to follow certain basic rules of scientific honesty, as defined by Committee on Publication Ethics. While we strongly believe that the authors submitting papers to *Forum* have no intention whatsoever of breaking those rules, and while the rules pertaining to joint authorship are rarely applicable to papers submitted to a philosophical journal, we consider it important to reiterate some of those rules to our authors. Thus, we remind authors that it is not acceptable to submit papers that reflect anything other than original research or reflections conducted by the author himself / herself. Concealing the true authorship of a part of a paper, even if the paper is in itself an original contribution, also constitutes a breach of such rules. Since all papers are reviewed by specialists, those seeking to make a submission should work on the assumption that any and all instances of plagiarism or ghostwriting will be detected. For jointly authored papers, a clear indication must be given in the paper itself of the manner and extent of each and every author's contribution. Papers with so-called "guest authors," whose contribution to the thesis of the paper is actually minimal, cannot be accepted. All cases of academic dishonesty, if detected, will be reported to the organizations with which the authors are affiliated.

