dialectica

International Journal of Philosophy

Contents

EUROPEAN SOCIETY FOR ANALYTIC PHILOSOPHY

Vol 74 : 3
September 2020

dialectica

International Journal of Philosophy
Official Organ of the European Society of Analytic Philosophy

founded in 1947 by Gaston Bachelard, Paul Bernays and Ferdinand Gonseth

dialectica

vol 74
issue 3

September 2020

Contents

The Paradox of the Arche-fossil
An Analysis of Meillassoux's Challenge to Correlationism, Idealism included

F.A. Muller

In his influential *After Finitude. An Essay on the Necessity of Contingency* (2008), Quentin Meillassoux argues that *Correlationism* (an umbrella-term encompassing most varieties of Idealism) gives rise to an irresolvable paradox, called "the Paradox of the Arche-fossil," which is essentially a clash between philosophical principles and scientific findings. This irresolvable paradox of Correlationism then paves the way for the "Speculative Turn" and the ensuing rise of burgeoning "speculative realism" in Continental Philosophy: noumenal reality, as-it-is-in-and-of-itself, "the Great Outdoors," is back on the Continental stage, open for speculative thought and even metaphysical knowledge. We attempt to provide a thorough and charitable analysis of the Paradox of the Arche-fossil. Our analyses lead us to conclude that Meillassoux's argument fails, due to an ambiguity with regard to the concept of being that cannot be repaired. We end by directing attention to another ominous threat to Correlationism, ignored by Meillassoux and all "speculative realists" alike, which is still breathing.

1 Correlationism

In his influential monograph *After Finitude. An Essay on the Necessity of Contingency* (2008), Quentin Meillassoux argues that "Correlationism" gives rise to a paradox, "The Paradox of the Arche-fossil" (2008).[1] This criticism paves the way for the "Speculative Turn," and the ensuing rise of burgeoning

1 Correlationism was further considered by Meillassoux in his 2012 lecture "Iteration, Reiteration, Repetition: A Speculative Analysis of the Sign Devoid of Meaning," given at Freie Universität Berlin, Germany and published in Meillassoux (2016), and in his London lecture of 2008, published as *Time Without Becoming* (2014).

"speculative realism" in Continental Philosophy.[2] In this opening section, we explain what "Correlationism" is; in the subsequent Section 2, we make our acquaintance with the Paradox of the Arche-fossil, and near the end of that section we provide an overview of what is to come in this paper after these two sections.

Meillassoux (2008, 16):

> By *correlation* we mean the idea that we only ever have access to the correlation between thinking and being, and never to either term considered apart from the other. We shall henceforth call *Correlationism* any current of thought that maintains the unsurpassable character of the correlation so defined. Consequently, it becomes possible to say that every philosophy which disavows naive realism has become a variant of Correlationism.[3]

On Meillassoux's most recent terminology, only four (major types of) metaphysical views are possible. We extend and refine his list, characterise these views as generally as possible (while remaining informative) below, and make ten elucidatory remarks.[4]

2 Brassier (2007), Bryant, Srnicek and Harman (2011), Harman (2007, 2011, 2018, 2019), Bryant (2011), Ennis (2011), Roffe (2012), Shaviro (2014), Toadvine (2014), Gijsbers (2015), Wiltsche (2017).

3 Slightly corrected English translation of the French of Meillassoux (2006, 18):

> Par "corrélation" nous entendons l'idée suivant laquelle nous n'avons accès qu'à la corrélation de la pensée et de l'être, et jamais à l'un de ces termes pris isolément. Nous appellerons *corrélationisme* tout courant de pensée qui soutiendra le caractère indépassable de la corrélation ainsi entendue. Dès lors, il devient possible de dire que toute philosophie qui ne se veut pas un réalisme naïf est devenue une variante du corrélationisme.

Presumably Meillassoux means we have only access to *correlates* of being (thoughts, senses, words), rather than to the *correlation*; these correlates are "unsurpassable."

4 Meillassoux's most recent terminology is that of his 2012 lecture (2016). He introduced another neologism for what is usually called "Absolute Idealism": "Subjectalism," also called (by a terminologically wavering Meillassoux) "subjectivist metaphysics," "Subjective Idealism" and "metaphysical subjectivism"; previously, he had classified Absolute Idealism as a variety of Correlationism. The conception of Correlationism of 2008 made every metaphysical view "a variant of Correlationism" save "naive realism," and thereby turned every contemporary philosopher into a correlationist; this made the term too broad for interesting philosophical use; the new conception of 2012 makes the one of 2008 subdivide into "2012-Correlationism" and Absolute Idealism.

(α) Direct Realism (naive realism; dogmatic metaphysics in Kant's sense, as opposed to Kant's "critical" metaphysics). Every human being, every subject (S), can directly access "the absolute," Being, reality as-it-is-in-and-of-itself (henceforth: $\mathcal{R}$); by directly accessing $\mathcal{R}$, S can and does obtain knowledge about, and understanding of $\mathcal{R}$.[5] For the Direct Realist, objectivity resides in the direct access that we have to $\mathcal{R}$. Some correspondence theory of truth is part and parcel of Direct Realism: true propositions (or sentences, or assertions, or beliefs) are *made true* by features of $\mathcal{R}$; those features have come to be called "truth-makers."

The pejorative adjective "naive" indicates that it is naive to believe that we can compare our sense data, concepts, thoughts, propositions, principles, models, theories, etc., "directly" with $\mathcal{R}$, as we can compare a face with a portrait by holding the portrait next to the face and looking at both. To say that to know or to understand $\mathcal{R}$ directly and unmediatedly is to utter a "performative contradiction," as Meillassoux (2014, 10) puts it, "through which you refute what you say or think by your very act of saying it or thinking it."

(β) Correlationism Every human being, every subject (S), cannot access $\mathcal{R}$ directly, but only some correlate of it, some re-presentation of it, which we symbolically abbreviate as: $\Phi[\mathcal{R}]$. This $\Phi[\mathcal{R}]$ is the joint result of $\mathcal{R}$ and the means and modes of cognition of S (Φ).[6] Everything that S knows is knowledge about $\Phi[\mathcal{R}]$, everything S understands is understanding of $\Phi[\mathcal{R}]$, and everything S perceives is perception of $\Phi[\mathcal{R}]$. Thus we codify Correlationism schematically and symbolically as follows:[7]

$$\Phi : \mathcal{R} \rightsquigarrow \Phi[\mathcal{R}].$$

The world $\Phi[\mathcal{R}]$ depends on reality ($\mathcal{R}$) and on us (Φ), and "$\rightsquigarrow$" is this dependency relation. (One might be tempted to replace the dependency relation "$\rightsquigarrow$" with a mapping relation "$\mapsto$"; but then we would be saying that $\Phi[\mathcal{R}]$ *supervenes* on $\mathcal{R}$, and this would presuppose that we can speak meaning-

5 The phrase "to access" is an umbrella-term for: to observe, to experience, to talk and to think about, to become aware of, to understand, to know about, and perhaps more.

6 The word "cognition" is an umbrella-term for our capacities for thought, talk and perception.

7 Harman (2011, 4) characterises Correlation as the conjunction of two theses (in our terminology): (a) the "correlation" between $\mathcal{R}$ and $\Phi[\mathcal{R}]$ via Φ is central; and $\mathcal{R}$ cannot be accessed without $\Phi[\mathcal{R}]$; and (b) $\mathcal{R}$ cannot be known and cannot be understood. Close enough.

doi: 10.48106/dial.v74.i3

fully *about* $\mathcal{R}$, which would make us already occupy a philosophical position, whereas we are still in the business of describing the possible philosophical positions.[8])

(β.1) Weak Correlationism Subject S can mention and posit $\mathcal{R}$ meaningfully, and can think and talk meaningfully about $\mathcal{R}$, but cannot obtain any knowledge about $\mathcal{R}$ or acquire any understanding of $\mathcal{R}$. Reality ($\mathcal{R}$) is epistemically inaccessible for S.

(β.2) Strong Type Correlationism Subject S can only mention and posit $\mathcal{R}$ meaningfully, but cannot further think or talk meaningfully *about* $\mathcal{R}$, let alone obtain knowledge about $\mathcal{R}$ or acquire understanding of $\mathcal{R}$. Reality ($\mathcal{R}$) is epistemically and linguistically inaccessible for S. Features of Φ are sought that are characteristic of the type, or species, of S, i.e. *homo sapiens*. For living organisms different from us, a different world obtains, say $\Phi_X[\mathcal{R}]$ for species X.

For Correlationists (β.1) and (β.2), objectivity resides in the agreement among subjects ("intersubjective agreement") about $\Phi[\mathcal{R}]$; a correspondence theory of truth can only be an intra-world correspondence, internal to $\Phi[\mathcal{R}]$, between true propositions and other features of $\Phi[\mathcal{R}]$, rather than of $\mathcal{R}$ as Direct Realism (α) would have it.

(β.3) Strong Token Correlationism As (β.2), with the difference that no specific features of Φ characteristic of the type, or species, to which subjects belong are sought for; every token S, or some comparatively small group $\mathcal{G}$ of tokens, has "its own world": $\Phi_S[\mathcal{R}]$ and $\Phi_{\mathcal{G}}[\mathcal{R}]$, respectively.[9] Objectivity can be buried in the graveyard of philosophically useless concepts.

(γ) Absolute Idealism Human beings, subjects (S), have only access to what is presented to S by the means and modes of cognition of S. This is all there is, and we codify it symbolically and schematically by: $\Phi[\cdot]$. There is no reality ($\mathcal{R}$) separate and distinct from $\Phi[\cdot]$; we can neither mention nor posit $\mathcal{R}$ meaningfully let alone truthfully; the proper meaning of "reality" is

8 Roughly, $\Phi[\mathcal{R}]$ supervenes on $\mathcal{R}$ iff differences in $\Phi[\mathcal{R}]$ imply differences in $\mathcal{R}$; to speak of differences in $\mathcal{R}$ is to speak *about* $\mathcal{R}$.

9 Harman (2011, 14) calls strong type Correlationism (β.2) just "strong correlationism," and strong token correlationism (β.3) "very strong correlationism."

$\Phi[\cdot]$. The distinction between reality and the world collapses, which suggest "the absolute-idealist identity": $\mathcal{R} = \Phi[\cdot]$. As Meillassoux (2008, 28) puts it, with an allusion to Quine: "To be is to be a correlate."[10] Objectivity is as in Correlationism (β.1) and (β.2). Often $\Phi[\cdot]$ is identified with Descartes' *res cogitans* and is "purely mental," which makes Absolute Idealism a form of substance monism.

(δ) Metaphysical Realism Human beings, subjects (S), can mention and posit $\mathcal{R}$ meaningfully, can think and talk meaningfully about $\mathcal{R}$, can and do access $\mathcal{R}$, *by means of* Φ; they can obtain, and perhaps even do obtain knowledge about $\mathcal{R}$, and they can and perhaps even do acquire understanding of $\mathcal{R}$, *by means of* representation $\Phi[\mathcal{R}]$, never directly. The metaphysical realist rejects *ab ovo* the distinction between unknowable reality ($\mathcal{R}$) and knowable world ($\Phi[\mathcal{R}]$). Reality ($\mathcal{R}$) has a specific structure, a specific composition of specific types of entities, which is all independent of (the existence of) S. Reality ($\mathcal{R}$) is cognitively and linguistically accessible for S, notably by the means and methods of science.[11] Objectivity resides in some correspondence theory of truth, which makes true propositions expressed in $\Phi[\mathcal{R}]$ correspond to features of $\mathcal{R}$.

Some versions of Metaphysical Realism have idealist elements, and thus become realist-idealist hybrids (β–δ), such as Koch's (2006) view that the existence of embodied subjects necessitates that spacetime be a feature of $\mathcal{R}$.

(ε) Pragmatism Human beings, subjects (S), have only access to what is presented to S by the means and modes of cognition of S. This is all there is, and we codify it symbolically and schematically by: $\Phi[\cdot]$. The question whether there is a reality ($\mathcal{R}$) separate and distinct from $\Phi[\cdot]$ is not worth thinking about: a redundant and useless issue, devoid of any consequences for our lives. Objectivity is as in Correlationism (β.1) and (β.2); and truth is warranted assertibility or what experts agree on in the limit of inquiry.

Pragmatism resembles (γ) Absolute Idealism, but is never accompanied by the claim that $\Phi[\cdot]$ is identified with Descartes' *res cogitans* and is "purely men-

10 Quine (1948, 15): "To be is the value of a bound variable." French original in Meillassoux (2006, 39): *"être, c'est être un corrélat"*.

11 See Sebold (2014, 13–50), Chapter II "Metaphysical Realism and its Discontents," for a brief contemporary statement of Metaphysical Realism. For spatio-temporal reasons, we must gloss over the concept of "subject-independence," aka "mind-independence."

doi: 10.48106/dial.v74.i3

tal." Pragmatism is not a form substance monism; it rebuffs such pragmatically meaningless metaphysical classifications.

A number of terminological remarks follow next, mainly in order to prevent confusion.

First, the traditional term "(metaphysical) Idealism" encompasses both Correlationism (β) and Absolute Idealism (γ).[12]

Secondly, Meillassoux (2016) also uses "materialism" for Direct Realism (α), "objective Idealism" for strong type Correlationism ($\beta.2$), "subjective Idealism" for both weak Correlationism ($\beta.1$) and strong token Correlationism ($\beta.3$), and "subjectivist materialism" for Metaphysical Realism (δ). We shan't. Further, Meillassoux (2016) calls ($\beta.1$) and ($\beta.2$) transcendental versions of Correlationism, and ($\beta.3$) the post-modern version.

Thirdly, Meillassoux's "Speculative Realism" falls under Metaphysical Realism (δ), because Meillassoux claims to know things about $\mathcal{R}$, such as the necessity of contingency, the falsehood of Leibniz's venerable Principle of Sufficient Reason, mathematics as a means to access $\mathcal{R}$ epistemically, the existence of arche-fossils in $\mathcal{R}$, and the coming into being of subjects in $\mathcal{R}$ (see next section). Thus Speculative Realism *opposes* Correlationism (β).

Fourthly, the adjective "speculative" in Speculative Realism should not be understood as indicating that only epistemically void guesswork is on the philosophical agenda when entering the transcendental level (as is clear from what we mentioned in the previous remark), but arguably better understood in Hegelian fashion. Hegel puts speculation opposite to reflection; reason encompasses both. Reflection is what we do when we gather knowledge of the world: "objectify" it, carve it up, structure it, assign properties and relations, as if the understanding and knowing subject is not there, is at no place specifically. Compare an eye looking at the world and never encountering itself. Speculation is what happens when the subject turns reflection on itself, on its "subjectivity," and becomes conscious of itself as an understanding and knowing subject. Compare to the eye looking in the mirror. Then knowl-

12 Russell (1912, 16) defines Idealism as "the doctrine that whatever exists, or at any rate whatever can be known to exist, must be in some sense mental." This is only Absolute Idealism (γ) and therefore somewhat restrictive. The Lemma on Idealism of the Stanford Encyclopedia of Philosophy makes the same mistake (cf. Guyer and Horstmann 2021).

edge about this subjectivity is not out of reach, which makes thought at the transcendental level not epistemically void.[13]

Fifthly, the ubiquitous term "Transcendental Idealism" (as opposed to Subjective, or Absolute, Idealism) seems co-extensive with Correlationism: all philosophers who call themselves Transcendental Idealists, or are classified as such by others, notably by historians of philosophy, turn out to fit the description of Correlationism (β), and conversely. But it may be that some Correlationists would resist being classified as "Transcendental Idealists," if only because the term "transcendental" has different meanings.

Sixthly, the fashionable terminology of "mediation" relates as follows: saying that our access to $\mathcal{R}$ is *mediated* is the same as saying that we can only access the correlate $\Phi[\mathcal{R}]$, rather than $\mathcal{R}$ directly; we can access $\mathcal{R}$ only indirectly, mediated by Φ. In the 20th century, "two principal 'media' of the correlation were consciousness and language," elucidates Meillassoux (2008, 6), consciousness being prominent in varieties of Phenomenology in the Continental Tradition, language being prominent in the Analytic Tradition.[14]

Seventhly, and confusingly, according to every metaphysical view in the taxonomy, something is real, be it $\mathcal{R}$, or $\Phi[\mathcal{R}]$, or the means and modes of mediation (Φ). So when we call Correlationism (β) and Absolute Idealism (γ) varieties of *Anti-Realism*, this does not imply that nothing is real according to these views, let alone that everything is somehow "an illusion."

Eightly, Eddington's two tables (the solid, brown wooden thing we sit, work and eat at, and the material object mereologically composed of zillions of atomic nuclei of protons and neutrons, and electrons somehow zooming around them, obeying the laws of quantum mechanics), and Sellars' two images (manifest and scientific) amount to the same distinction *within* $\Phi[\mathcal{R}]$, rather than *between* $\Phi[\mathcal{R}]$ and $\mathcal{R}$.[15]

Ninethly, (α) Direct Realism may be a straw man, which makes the criticism of stating it is performing a contradiction an act of burning a straw man. Perhaps the only form of realism is what we have called here Metaphysical

13 See Verene (2007, 7–9, 11) for an elaboration. For other speculative realists, such as G. Harman, "speculative" means indeed guesswork when it comes to $\mathcal{R}$; Harman (2019) follows Whitehead in claiming that it is folly to claim that we know anything about $\mathcal{R}$.

14 See the Appendix for examples. The use of "mediation" is not the same as in the vernacular. As one referee points out: "If I can reach you only via the phone, it would be false to say that I cannot reach you." In this sense, only the metaphysical realist can use "mediation": we reach $\mathcal{R}$ mediated by Φ; the (β) Correlationist is stuck with $\Phi[\mathcal{R}]$.

15 Christias (2016) has however argued that Meillassoux's "correlationist circle" echoes W.F. Sellar's "myth of the given."

doi: 10.48106/dial.v74.i3

Realism (δ). For in all honesty, which philosopher would claim we have knowledge of $\mathcal{R}$ *without* being aware and acknowledging that this knowledge is expressed in language, employs concepts, and depends on our specific sensory organs? Nothing in our analysis of Meillassoux's paradox of the arche-fossil depends on whether one maintains or dismisses that there is a difference between (α) Direct Realism and (δ) Metaphysical Realism.

Tenthly, and related to the previous remark: having *direct* epistemic access to $\mathcal{R}$ *perceptually*, via our sensory organs, is less controversial than having direct epistemic access to $\mathcal{R}$ *conceptually*, as (α) Direct Realism claims. Direct perceptual access to $\mathcal{R}$ is part of (α) Direct Realism, but cannot be dismissed as stating it being a performative contradiction.[16]

For a number of illustrations of correlationists from the history of philosophy, we refer to the Appendix. Needless to emphasize that nothing about Meillassoux's arguments depends on exactly who falls in which category, or even whether it is undecidable whether some token philosopher belongs to which type. That several if not all reputed philosophers fit into one of the delineated categories (α–γ) is sufficient to convince us that Meillassoux is not talking to himself. Below we shall occasionally mention a correlationist philosopher, notably Kant, for the sake of illustration, or sometimes as a foil.

Now we are, at last, ready to turn to the central argument of Meillassoux.

2 The Paradox of the Arche-fossil

The "Paradox of the Arche-fossil" is the contradiction that Meillassoux infers from two propositions ("W" alludes to World, "A" to Arche-fossil):

(W) Without subjects, there is no world.

(A) Subjects have come into being in the world.

Skimpily, the argument for the contradiction goes as follows.

If there were no subjects, there would be no means and modes of cognition either (Φ), and consequently there would not be a world $\Phi[\mathcal{R}]$. Hence Correlationism (β) implies (W). Proposition (A) is a well-established piece of scientific knowledge about certain entities, subjects, in a presupposed world $\Phi[\mathcal{R}]$: they came into an existing world (A) whilst that world did not exist

16 Meillassoux (2008, 2014, 2016) never talks about perception when talking about direct/naive realism.

before they came into it (W). Faced with this conflict between (W) and (A), the choice for Meillassoux is easy: farewell to Correlationism (W), because rejecting such a well-established scientific truth as (A) is irrational. Reason commands us to reject Correlationism (β).

Notice that Metaphysical Realists (δ) are not in trouble, because no contradiction ensues from their understanding of propositions (W) and (A):

(W*) Without subjects, there is no $\Phi[\mathcal{R}]$—but there always is $\mathcal{R}$.

(A*) Subjects have come into being in $\mathcal{R}$, equipped with their means and modes of cognition (Φ).

This understanding is pretty standard in science; natural scientists generically hold that they find out things about reality ("A new species has been discovered"; "The existence of the Higgs boson has been established"; "This area is poisoned by radio-active radiation"; "The existence of gravitational waves has been confirmed"). Metaphysical Realism (δ) is the default philosophical view of the natural scientist.

The purpose of the current paper is to analyse Meillassoux's deceptively simple argument in detail and to find out: whether the argument is deductively valid; and if it is, whether it is as lethal for Correlationism as Meillassoux claims it to be.

The metaphysical dispute between Idealism and Realism has always been taken to be a quintessential *philosophical* debate, one on which *science* cannot have any bearing. But if Meillassoux is correct, then science does bear on this debate: Idealism would have been slain by a contemporary continental philosopher crucially using science. Surely this would be one of the greatest ironies in the history of Western Thought, in the light of both the detached relations between science and Continental Philosophy generally and the pervasive anti-realism in Continental Philosophy (Sebold 2014; Braver 2007 *passim*). Another irony is that the argument strikingly resembles the legendary dispute in 1951, in a Parisian café, between A.J. Ayer, G. Bataille, M. Merleau-Ponty and G. Abrosina about whether the sun existed before there were human beings, which is often seen as the historical event where the Analytic-Continental Divide was first noticed.[17] Is Meillassoux finally settling this

17 Sebold (2014, 1–3), Vrahimis (2012).

(A^0) Without subjects, there is no sun.

doi: 10.48106/dial.v74.i3

dispute, going back to the cradle of the Great Divide in philosophy, in favour of the analytic philosopher Ayer, and against his continental colleagues Bataille and Merleau-Ponty?

Parenthetically, Merleau-Ponty is nowhere mentioned by Meillassoux (2008). Yet witness how close Merleau-Ponty was to the Paradox of the Arche-fossil, and how he hinted at a resolution of sorts (the sentence italicised by this author points to a rejection of (W)):

> For what precisely is meant by saying that the world existed before any human consciousness? An example of what is meant is that the earth originally issued from a primitive nebula from which the combination of conditions necessary to life was absent. But every one of these words, like every equation in physics, presupposes our pre-scientific experience of the world, and this reference to the world in which we live goes to make up the proposition's valid meaning. [...] Laplace's nebula is not behind us, at our remote beginnings, but in front of us in the cultural world. What in fact do we mean when we say that there is no world without a being in the world? *Not indeed that the world is constituted by consciousness*, but on the contrary that consciousness always finds itself already at work in the world. (Merleau-Ponty 2002, 502)

To repeat, the purpose of this paper is to analyse Meillassoux's argument that leads to this paradox for Correlationism (β), in order to find out whether the argument is valid, beginning in the next section (Section 3). For the sake of brevity, we shall call this argument for the Paradox of the Arche-fossil "the *archument*." By considering responses by Meillassoux to two criticisms of the archument, we provide more rigorous and precise presentations of the archument (Section 4, Section 5). We then argue, on the basis of our analyses, that the archument is invalid (Section 6). Then we present another analysis of the archument with a so-called tenseless concept of being, and reach the same conclusion (Section 7). Subsequently, we drill deeper by addressing the pivotal question that gave rise to the Paradox of the Arche-fossil in the first place (the paradox implies that the correlationist is unable to answer this

(W^0) The sun has come into the world before there were subjects in the world.

The Great Divide between Continental and Analytical Philosophy thus began, if only partly, as a divide between Correlationism (β) and Metaphysical Realism (δ).

question); we argue that Kant's Correlationism can provide an answer to this pivotal question (Section 8).

Before we continue, one caveat concerning Meillassoux's criticism of Correlationism (and thus of Idealism) consists essentially of pointing out a clash between Correlationism and science: the premises of the various rational reconstructions of his argument we shall provide in the course of this paper are either premises that Correlationism is committed to (W) or morsels of well-established scientific knowledge (A). Meillassoux wholeheartedly accepts science, and does not accept, endorse or defend the Correlationist premises; he only accepts, endorses and sometimes defends *that* they are premises that Correlationism is committed to. A Correlationist accepts, endorses and defends these premises, which means that pointing out that Meillassoux does not accept, endorse or defend these premises is irrelevant. Needless to say that whether these premises *are* Correlationist will be a topic of unremitting concern throughout this paper.

3 The Archument

Meillassoux (2008, 9) lists the following morsels of scientific knowledge, truths established by empirical inquiry beyond reasonable doubt.[18]

(A1) The universe is about 13.5 billion years old.

(A2) The accretion of planet Earth began about 4.6 billion years ago.

(A3) Life emerged on planet Earth about 3.5 billion years ago.

(A4) Human life, *homo habilis*, arose about 2 million years ago.

(A5) *Homo sapiens* came into being about 0.5 million years ago.

Of course (A3) and (A5) jointly imply premise (A) of the previous section, because human beings are subjects, and because Meillassoux tacitly assumes that there were no other subjects elsewhere in the universe at earlier times—a tacit assumption we shall, for the sake of argument, subscribe to throughout

18 We have added proposition (A5), because it is doubtful whether *homo habilis* already mastered reasoning with the Kantian epistemic categories, say, and understood the world in these terms; for that we need *homo sapiens*. French original in Meillassoux (2006, 24).

doi: 10.48106/dial.v74.i3

this paper. Meillassoux calls *ancestral* propositions like (A1)–(A5) and (A), which are about the universe, Earth notably included, at times when no life had emerged yet on Earth; and he calls currently existing objects that carry proof of this ancestry *arche-fossils* (2008, 10); we call them *present arche-fossils*, and objects that are or carry proof of this ancestry but no longer exists, *past arche-fossils*, such as dinosaurs and entirely degenerated skeletons of dinosaurs. Present arche-fossils are part of the empirical evidence that has turned the ancestral propositions into morsels of well-established scientific knowledge.

In Chapter 1, "Ancestrality," Meillassoux (2008, 10–11) addresses the following aporia:

(Q) How is Correlationism able to think meaningfully, and to understand and to know ancestral propositions?

Meillassoux argues that Correlationism (β) is unable to understand and to know ancestral propositions: it can understand and know them only on pain of contradiction, and this contradiction is the "Paradox of the Arche-fossil." Again, the archument for this paradox runs as follows.

We, subjects, human beings, *make* reality ($\mathcal{R}$) knowable and understandable, by our means and modes of cognition, enabled by our sensory organs and brains (Φ), which form a necessary condition for the possibility of human knowledge and understanding. The moulding and grinding of $\mathcal{R}$ results in the world: $\Phi[\mathcal{R}]$. The world is a *re-presentation* via Φ of what is *presented* to us ($\mathcal{R}$). Without human beings, our means and modes of cognition (Φ) are also absent, and there is not and cannot be world $\Phi[\mathcal{R}]$ either (W). Before there were human beings, there was no world $\Phi[\mathcal{R}]$. Yet subjects have come into being in the world (A). For a long time, there were no human beings in the world; they evolved from other organisms, which in turn somehow evolved from lifeless chemical substances. So once there was $\Phi[\mathcal{R}]$ while there could not be and therefore was not $\Phi[\mathcal{R}]$, which is a contradiction.

This was the archument once again. Meillassoux (2008, 17–18):[19]

19 French original in Meillassoux (2006, 35–36):

> Il n'y a pas de compromis possible entre le corrélat et l'archifossile: l'un des deux étant admis, l'autre est de ce fait disqualifié. [...] Face à l'archifossile, *tous les idéalismes convergent et deviennent également extraordinaires*—tous les corrélationismes se révèlent comme des idéalismes extrêmes, incapables de se résoudre à admettre que *ces évènements d'une matière sans homme* dont nous parle la science

Correlationists are essentially Creationists. *Take that!*

Should we conclude that the Paradox of the Arche-fossil is the silver bullet for Correlationism (β) and by implication for metaphysical Idealism? Or has something gone awry?[20] What, then, exactly has gone awry? First we consider two objections to the archument and Meillassoux's defence (Section 4, Section 5), for this will yield ingredients for a clarified and manifestly deductive valid version of the archument further on.

4 A Subterfuge of Lacunae

The first objection that Meillassoux (2008, 18) considers targets the alleged privilege of the temporal *ancient* over the spatial *distant*. If the Correlationist

> ont effectivement pu se produire tels que la science en parle. Et notre corrélationiste se trouve alors dangereusement proche de ces créationistes contemporains: de ces croyants pittoresques qui affirment aujourd'hui, selon une lecture "littérale" de la Bible, que la Terre n'aurait pas plus de 6000 ans, et qui, se voyant objecter les datations plus anciennes de la science, répondent, impavides, que Dieu a créé il y a 6000 ans, en même temps que la Terre, des composés radioactifs indiquant un âge de la Terre beaucoup plus anciens—cela pour éprouver la foi des physiciens. Le sens de l'archifossile serait-il pareillement d'éprouver la foi du philosophe dans les corrélats, même en présence de données qui indiquent un écart abyssal entre ce qui existe et ce qui apparaît?
>
> There is no possible compromise between the correlation and the arche-fossil: once one has acknowledged one, one has thereby disqualified the other. […] Confronted with the arche-fossil, *every variety of idealism converges and becomes equally extraordinary*—every variety of correlationism is exposed as an extreme idealism, one that is incapable of admitting that what science tells us about these occurrences of matter independent of humanity effectively occurred as described by science. And our correlationist then finds himself dangerously close to contemporary creationists: those quaint believers who assert today, in accordance with a "literal" reading of the Bible, that the earth is no more than 6,000 years old, and who, when confronted with the much older dates arrived at by science, reply unperturbed that God also created at the same time as the earth 6,000 years ago those radioactive compounds that seem to indicate that the earth is much older than it is—in order to test the physicist's faith. Similarly, might the arche-fossil not be meant be to test the philosopher's faith in correlation, even when confronted with data which seem to point to an abyssal divide between what exists and what appears?

20 In contrast, Toadvine (2014) suggests we swallow that the world did not exist before *homo sapiens* came into being, which means that Toadvine accepts the archument and rejects premise (A). Toadvine is in the company of the Creationists who believe that God created the world about 6,000 years ago, including planet Earth filled with *apparent* arche-fossils—to test us.

doi: 10.48106/dial.v74.i3

cannot understand and cannot know ancestral propositions like (A1)–(A5), then the Correlationist also cannot understand and cannot know "distant propositions," that is, propositions about locations in the world, in the universe, where there are no subjects, never have been subjects and never will be subjects. Think of space-time regions outside the light-cone of the history of all terrestrial subjects: there *cannot* be a causal connection between such regions and any region occupied by some actual terrestrial subject.[21] For this is impossible according to well-established scientific knowledge, specifically Einstein's Theory of Relativity.

One might very well think that invoking the spatial distant next to the temporal ancient makes things worse for the Correlationist, for now we also seem to have a "Paradox of the Distant Location" within arm's length. Meillassoux sees it otherwise, and judges that this invocation of the spatially distant is meant to transform the archument into a trivial one, by "identifying it with a familiar and inconsequential anti-Idealist argument" (2008, 18). For the same can be said about craters on the far side of the moon, and, we might add, about locations deep inside Earth, where no man has ever gone and presumably never will go. The problem is however not the actual absence or physical impossibility of the presence of *human witnesses to events* in the world, or of *observers of past ancestral objects*. Meillassoux (2008, 19) holds that a Correlationist can understand and accept subjunctive conditionals like: If some subject had been then and there, that subject would have witnessed events that occurred then and there. And if the physical *possibility* of the existence of witnesses at certain spacetime regions is sufficient for understanding and knowing propositions about events occurring in those regions, or about objects that exist in those regions, then we are done. Correlationism would perhaps stand tall. The ancestral propositions (A1)–(A5) could be understood and known after all.

But when subjunctive conditionals with perhaps conceptually impossible antecedents, and certainly physically impossible antecedents are false, then the ones mentioned above with ancestral antecedents are false, because there is no conceptually possible world in which subjects witness the coming into being of $\Phi[\mathcal{R}]$. In fact, it seems conceptually, and therefore physically, impossible for there to be a world in which humans supposedly witness how the species to which they belong comes into being. But note that Meillassoux's

21 We gloss over travelling through stable wormholes, which seems impossible; see Cuyubamba, Konoplya and Zhidenko (2018).

talk about *witnesses* and their *possible* presence is not the issue. The issue is whether ancestral propositions have truth-conditions in $\Phi[\mathcal{R}]$ that Correlationism can accept. As long as these truth-conditions in $\Phi[\mathcal{R}]$ do not involve human witnesses, the issue of witnesses is a red herring.[22]

The response of Meillassoux raises the following question: if, for the Correlationist, there is no such thing as the Paradox of the Spatial Distant, and the Correlationist can understand and know modal propositions, specifically subjunctive conditionals, about distant locations in the world, then why can the Correlationist not do so with the temporal ancient? Why does the Paradox of the Arche-fossil, the Paradox of the Temporal Ancient, arise at all?

The problem, Meillassoux contends, is the possibility of understanding there being a world at all, of understanding how $\Phi[\mathcal{R}]$ could have come into being. The relevant relation under investigation is one between subjects and the world, as in proposition (W), which is not *spatio-temporal* but *logical-conceptual*: without subjects, the coming into being of the world, of the givenness of being, of $\Phi[\mathcal{R}]$, cannot be understood, because when it happened, the necessary cognitive conditions and capacities were not realised to understand and to know that subjects had come into being. Meillassoux (2008, 22):[23]

> So the challenge is therefore the following: to understand how science can think a world wherein spatio-temporal givenness itself came into being at a time and in a space which preceded every variety of givenness.
>
> We now see that the sophisticated nature of this first rejoinder consists in trying to occlude one lacuna by another, in trying to mask the non-being of the given by a given of non-being, as though the former could be reduced to the latter. But this switching of absences, this subterfuge of lacunae, cannot disguise the fundamental difference between our two voids, and thereby the difference between the two arguments: the trivial argument from the unperceived and the valid argument from the ancestral.

22 Truth-conditions in $\Phi[\mathcal{R}]$, because truth-conditions in $\mathcal{R}$ are Correlationist impossibilities, let alone knowledge conditions. We remark that humans are of course needed to ascertain whether truth-conditions obtain.

23 The part on p. 18 starting with *** and ending again with *** on p. 26 of the translation, Meillassoux (2008), is not present in the original French (2006). This absence is nowhere mentioned by the translator, R. Bassier.

doi: 10.48106/dial.v74.i3

We gloss over unpacking this response in detail, although we do want to point out that the first sentence in this quotation is anathema to Correlationism, and even smacks of nonsense: givenness, re-presented reality $\Phi[\mathcal{R}]$, the spatio-temporal world as a whole with every actually existing concrete entity in it conceptualised by us, cannot possibly *come into being in that same world* $\Phi[\mathcal{R}]$, because if this were possible, then $\Phi[\mathcal{R}]$ had to be there already; and if $\Phi[\mathcal{R}]$ was already there, it need not come into being anymore. What we shall attempt to do next is to present "the valid argument from the ancestral."

Let us begin again with propositions (W) and (A), using our correlationist relation $\mathcal{R} \rightsquigarrow \Phi[\mathcal{R}]$:

(W) If there are no subjects in $\Phi[\mathcal{R}]$, then there is no $\Phi[\mathcal{R}]$ ("non-being of the given").

(A) Subjects have come into being in $\Phi[\mathcal{R}]$ ("givenness of being").

Can we rigorously deduce a contradiction from these propositions?

Proposition (W) *seems* to have the following consequence:

(W0) For every time t, if there are no subjects in $\Phi[\mathcal{R}]$ at time t, then there is no $\Phi[\mathcal{R}]$ at that time t ("non-being of the given" at time t).

Proposition (A) says that at some time in $\Phi[\mathcal{R}]$, about 2 million years ago (A3), subjects (human beings) *came into* $\Phi[\mathcal{R}]$. Then at some time, in fact at any time much earlier, there were no subjects in $\Phi[\mathcal{R}]$, there was $\Phi[\mathcal{R}]$ without subjects in it, viz. the ancestral propositions (A1)–(A5). Then $\Phi[\mathcal{R}]$ was ancestral. In other words:

(A6) For some moment in time t_0 in $\Phi[\mathcal{R}]$, there are no subjects in $\Phi[\mathcal{R}]$ at t_0.

Clearly (A6), which is a consequence of (A), presupposes that there is $\Phi[\mathcal{R}]$, in which we are considering a particular moment in time, t_0, when there are no subjects. Then by *modus ponendo ponens* via (W0)—seemingly implied by (W)—we have that at t_0 there is no $\Phi[\mathcal{R}]$, which flatly contradicts the presupposition of (A6). This means that (A) and (W) are inconsistent, which is the Paradox of the Arche-fossil again.

In this reconstruction of the archument, we have taken (W0) as a consequence, or proper replacement, of (W). Wrongly, of course, for time is part of what subjects bring to the table with Φ; time is a constitutive component

of $\Phi[\mathcal{R}]$, rather than a feature of $\mathcal{R}$. For example, for Kant, time is the inner form of *Anschauung*, knowable by introspection, which form is a cognitive capacity of subjects, constitutive of the Kantian phenomenal world $\Phi[\mathcal{R}]$, rather then a feature of $\mathcal{R}$ (1787, B50–B51). Speaking about there being, or not being, $\Phi[\mathcal{R}]$ is speaking at a level, call it the *transcendental level*, where there is no time. Whereas the concept of existence (or being) that is expressed in the consequent of (W0) must by conceptual necessity be *tenseless*, the concept of existence (or being) expressed in ancestral propositions (A2)–(A5) obviously is *tensed*, and applies to everything in $\Phi[\mathcal{R}]$, at the *phenomenal level*. Well, what holds for Kant, holds for every Correlationist (β): time and tense do not apply to $\mathcal{R}$; they only apply to $\Phi[\mathcal{R}]$. Nothing can come into being in $\mathcal{R}$, full stop.

We want to mention that the distinction between tensed being (and existence) and tenseless being (and existence) was first sharply drawn by the Idealist McTaggart (1908). In this famous paper on "The Unreality of Time," McTaggart distinguished the *A-Theory*, the *B-Theory* and the *C-Theory* of time and being. In both the B- and the C-Theory, the concept of being is timeless and tenseless, and therefore, according to McTaggart, incapable of capturing the essence of being, whereas existence in the A-Theory is temporal and tensed, capable of capturing the essence of time.[24] According to the A-Theory, the conjugation of verbs in tenses in language reflects the ontic categories of past, present and future; time is the change in ontic status of events from past via present to future. According to the B- and C-Theory, events *be* or *not be*; the English language does not have a tenseless conjugation of verbs to express this—metaphysics outruns language, whence putting the infinitive in italics as a means of expressing the tenseless mode. Mulder (2014, sec. 6.2) demonstrates that the A-Theory and the B-Theory belong to distinct clusters of concepts, which defy inter-translation; the A-Theory comes with tensed being, tensed predication and endurantism (objects have no temporal parts), whereas the B-Theory (and C-Theory) comes with tenseless being, tenseless predication and perdurantism (objects have temporal parts).

We return to the archument. To repeat, at the transcendental-level (shortly: τ-level), when we want to talk about $\mathcal{R}$, only the tenseless conception of existence is available. Time is entirely absent at the τ-level. As soon as one wants to apply time to $\mathcal{R}$, as soon as one wants "to interpret time realistically"

24 But leads to trouble, which makes McTaggart Idealistically, as well as controversially, conclude that time is not real, by which is meant: not a feature of $\mathcal{R}$.

doi: 10.48106/dial.v74.i3

(rather than "idealistically"), then one has left the Correlationist building (β) and entered the Metaphysical-Realist building (δ).

Proposition (W) we now express as follows:

(W1) If there *be* no subjects in $\mathcal{R}$, then there *be* no world $\Phi[\mathcal{R}]$ either ("non-being of the given").

To summarise, the deduction of a contradiction from (W) and (A) was achieved by the wrong phrasing of (W), as (W0) rather than as (W1); and when the different concepts of existence (and being) that figure in (W) and (A) are expressed (correctly and) differently, we obtain (W1); and from (W1) and (A) no contradiction ensues. Who claims it does follow, commits the fallacy of equivocation. The Paradox of the Arche-fossil no longer arises.

The first objection against the archument concerned the unequal treatment of space and time, and was addressed by its propounder Meillassoux. Although we did not expound Meillassoux's defence in its entirety (due to some poignant unclarities), we did clarify the crucial role of time in the argument, which has resulted in the analysis of the archument above. Correlationism is still standing because the archument against Correlationism fails on our analysis. We next move to the second objection addressed by Meillassoux.

5 An Amphiboly

Meillassoux (2008, 22–23) envisions a critic of his archument charging him with having confused the phenomenal level (φ-level) and the transcendental level (τ-level). At the φ-level, we talk and think about what's going on in the world, what there is in the world, how it all hangs together, etc.; this forms the subject-matter of science. At the τ-level, we talk and think about the knowing and understanding subject, about the necessary conditions for the possibility of knowledge and understanding, about what is beyond all possible experience, about $\mathcal{R}$ and its relation to $\Phi[\mathcal{R}]$, about the being of $\mathcal{R}$ and the being of $\Phi[\mathcal{R}]$, etc. The connection to the previous section is that the tenseless *be* is the concept of being (and existence) at the τ-level, whereas tensed being is the standard concept of being (and existence) at the φ-level. Meillassoux (2008, 22) compares:[25]

25 This cited part belongs to the added text in the translation, absent from the original French, Meillassoux (2006). See footnote 23.

> Now, these two levels of thought—the phenomenal and the transcendental—are like the two sides of a flat sheet of paper: they are absolutely inseparable but they never intersect. But your mistake [= Meillassoux's mistake, according to an imaginary critic of the archument, FAM] consists precisely in allowing them to intersect—you have turned a structure, which should have remained flat, into a Möbius strip.

The virtual critic of Meillassoux (2008, 23) continues by saying:

> Consequently, your conception of a "time of science", in which both bodies and the manifestation of bodies arose, is "amphibolous"—it conflates the objective being of bodies, which do in fact emerge and perish in time, with the conditions for the objective knowledge of the objective being of bodies, which have nothing to do with any sort of time.

Meillassoux goes on to explain—*contra* Kant it seems—that there is no such thing at the τ-level called a *transcendental subject*; there are only *objects* (material beings) and embodied *subjects* (a particular type of material beings) at the φ-level, in the world, $\Phi[\mathcal{R}]$: human beings. Let's adopt the following criterion (enter Meillassoux's "the temporality of the conditions of instantiation"):

> *Subject Criterion.* Concrete entity S is a *subject* at time t in the world, $\Phi[\mathcal{R}]$, iff at time t in the world, $\Phi[\mathcal{R}]$, S is a living embodied being that possesses the following familiar capacities: sensory (seeing, smelling, touching, hearing, etc.), cognitive (remembering, reasoning, comparing, understanding, knowing), cogitative (thinking, believing, accepting, imagining), affective (feeling), and connative (wanting, desiring).

The further defence against this charge of "amphiboly" that Meillassoux propounds is as follows (2008, 25, his emphasis):[26]

> that the transcendental subject remains indissociable from its incarnation in a body; in other words, *it is indissociable from a determinate object in the world*. [...] when we raise the question of the emergence of thinking bodies in time, we are also raising

26 Same as in footnote 25.

doi: 10.48106/dial.v74.i3

551 the question of *the temporality of the conditions of instantiation,*
552 *and hence of the taking place of the transcendental as such.*

553 The last emphasised phrase we take to mean that at a certain time (however
554 vaguely delineated this time is in our evolutionary chronology), a certain kind
555 of object comes into the world that meets the Subject Criterion displayed above.
556 Meillassoux (2008, 25) further rejects the existence of a transcendental subject
557 as an uninstantiated universal, akin to a Platonic form ("indissociable from a
558 determinate object" in the world): there are only instantiated "transcendental"
559 subjects, that is, things in $\Phi[\mathcal{R}]$ that meet the Subject Criterion displayed
560 above.

561 What makes subjects transcendental is that they can engage in transcen-
562 dental thought, at the τ-level (see above).[27] Little if anything in the archument
563 hinges on the transcendentality of the subject. Presumably therefore Meillas-
564 soux does not elaborate on the meaning of "transcendentality," apart from
565 what we just have reported—which seems quite sufficient for his archument.
566 We need not be detained further by it, and move to present a rigorous and
567 clarified expression of premise (W).

568 6 Correlationism Reclaimed

569 To recapitulate, we began in our analysis of the archument with (W) and (A),
570 and deduced a contradiction from incorrect (W0) and correct consequence
571 (A6) of (A). In order to avoid committing the fallacy of equivocation with
572 regard to being and existence, we invoked the distinction between a tensed and
573 a tenseless conception of being, which resulted in (W1), a τ-level proposition,
574 rather than (W0), from which no paradox arose for Correlationism (end of
575 Section 4). Keeping (W) and (A) *both* entirely at the φ-level can be done, but
576 that does not lead to a paradox either: this is the way of Absolute Idealism (γ),
577 as Meillassoux (2008, 14, 27) admits.[28] Absolute Idealism was never in the
578 target area of the archument. (Parenthetically it is also the way of Metaphysical
579 Realism (δ), as we pointed out in Section 2, with (W*) and (A*).) Since a
580 version entirely at the φ-level does not give rise to a paradox, and a version

27 For Kant, it means that the subject has the capacity to think at the τ-level, of having thoughts that *transcend* all possible sensory experience and imagination, of knowing *a priori* about necessary conditions of the possibility of having sensory experiences ($\Phi[\mathcal{R}]$). For Hegel, it means that the subject has the capacity for speculative thought (see Section 1 and the *fourth* terminological remark)

28 Harman (2011, 15–16) agrees.

entirely at the τ-level is not in the cards to begin with (subjects definitely do not evolve into being at the τ-level, which makes τ-version of A conceptually impossible), we next present a version of the archument of which one of the premises explicitly connects the τ- and the φ-level.

The first premise is the following φ-level ancestral proposition:

(A6) For some moment of time t_0, there are no subjects in $\Phi[\mathcal{R}]$ at t_0.

In fact, a stronger ancestral position is licensed by science, which also implies (A):

(A7) For every moment of time before 3.500 billion years ago, there were no subjects in $\Phi[\mathcal{R}]$.

Next, a version of (W), which connects φ-level (antecedent) to τ-level (consequent) explicitly:

(W2) If, for every time t, there is no subject in $\Phi[\mathcal{R}]$ at t, then $\Phi[\mathcal{R}]$ *not be*.

Logically equivalent to (W2) is the contraposed proposition:

> If $\Phi[\mathcal{R}]$ *be*, then for some moment in time t_0, there is some subject in $\Phi[\mathcal{R}]$ at t_0.

The converse of this last version of (W2) also seems to provide the weakest sufficient condition for there *be* $\Phi[\mathcal{R}]$; when we combine this with (W2), and move from "moments in time" to "spacetime regions," we obtain the following crisp and clear criterion, which explicitly connects the φ-level to the τ-level:

(W3) $\Phi[\mathcal{R}]$ *be* iff for some spacetime region, there is some subject in $\Phi[\mathcal{R}]$ in the region.

Obviously criterion (W3) is compatible with every ancestral proposition, (A1)–(A7). Since the ancestral propositions are true, and they presuppose there *be* $\Phi[\mathcal{R}]$, it is also true that there *be* $\Phi[\mathcal{R}]$. Then both conditions of the biconditional (W3) are also true, which makes (W3) and (W4) true. Then any of (A1)–(A7) and (W3) are consistent. Thus the Paradox of the Arche-fossil no longer obtains. Correlationism can be reclaimed.

By way of a closing remark of the current section, we point that there is a version of premise (W), connecting the φ- and the τ-level, that does lead to a contradiction with premise (A):

doi: 10.48106/dial.v74.i3

(W4) If $\Phi[\mathcal{R}]$ *be*, then at every moment of time t, there are subjects in $\Phi[\mathcal{R}]$.

But Correlationism is not committed to such a strong necessary condition for there *be* the world $\Phi[\mathcal{R}]$ as in (W4). Correlationism remains reclaimed.

7 Sub Specie Aeternitatis

So far we have argued that: (i) at the τ-level, only the concept of tenseless being (and existence) is applicable, whereas at the φ-level, the concept of tensed being (and existence) is applicable; (ii) the archument, which led to the Paradox of the Arche-fossil for Correlationism (β), committed the fallacy of equivocation by confusing these concepts; the archument turned out to be amphibolous; and (iii) if these concepts of being (and existence) are properly distinguished in more precise rephrasings of (A) and (W), no Paradox of the Arche-fossil arises. We have employed the tensed conception of being (and existence) for $\Phi[\mathcal{R}]$. But in metaphysics the A-Theory of tensed being is not the only view on time and being. In fact, a sizeable number of philosophers defend that a tenseless conception of being (and the B-Theory of time) is *better* than the rival tensed conception. This raises the question whether the archument is possible when it only employs the concept of tenseless existence, also for ancestral premise (A), at the φ-level. Let's see.

We now look at the world, including its spacetime, $\Phi[\mathcal{R}]$, *sub specie aeternitatis*, all of whose events *be*. The crucial ancestral proposition expressed in accordance with the tenseless conception (and the B-Theory of time) is the following, where time is an ordering relation between events that *be*:

(B) There *be* no subjects earlier than some time t_0.

In the language of Relativity Theory, the ancestral proposition reads:

(B*) There *be* no subjects below some 3-dimensional space-like hypersurface in 3+1-dimensional spacetime in $\Phi[\mathcal{R}]$.

The appropriate version of (W), which should be acceptable for every Correlationist, is:

(W5) $\Phi[\mathcal{R}]$ *be* iff in some space-time region in $\Phi[\mathcal{R}]$ there *be* at least one subject.

Obviously propositions (B) and (W5), and (B*) and (W5), are perfectly compatible. When we add that above the hypersurface mentioned in (B*) there are spacetime regions where subjects *be*, like the terrestrial regions that we, human beings, inhabit, then we deduce with (W5) that $\Phi[\mathcal{R}]$ *be*. Correlationism still remains reclaimed.

We have now exhausted all possible versions of the archument, with different concepts of being, either tensed, tenseless, or both but then explicitly connected by a premise acceptable by Correlationist lights. Every version of the archument does not give rise to a paradox, and the only valid archument we could muster needed a premise to which no Correlationist will subscribe to. We conclude that Meillassoux's paradox of the arche-fossil collapses.

8 An Aporia

Meillassoux has declared to present the Paradox of the Arche-fossil not as a silver bullet for Correlationism (β), but as raising an *aporia* for Correlationism, i.e. the following aporia (which we mentioned earlier):

(Q) *How is Correlationism able to think meaningfully, and to understand and to know ancestral propositions?*

Meillassoux argued that Correlationism can do so only on pain of being caught in a contradiction, the Paradox of the Arche-fossil. Careful analysis has however led us to conclude that no such paradox arises. But if our analyses are correct, then how *should* the Correlationist answer aporia (Q)?

Well, that is going to depend on which variety of Correlationism one considers, on how Φ is construed and understood. For example, Wiltsche (2017) has provided the answer on behalf of Husserlian Phenomenology (by curiously injecting it with a constructive-empiricist serum). Let us sketch, for the sake of concreteness, answer to Q on behalf of Kant's Transcendental Idealism, which we have classified as Strong Type Correlationism ($\beta.2$) in Section 1.

Kant has provided one of the most refined and elaborate views about the correlation $\mathcal{R} \rightsquigarrow \Phi[\mathcal{R}]$. Kant's two forms of sensible intuition (*Anschauung*), space (outer) and time (inner), do not come with *restrictions* on the values that spatial and temporal variables may assume, notably not with the temporal restriction such that events that happened earlier than about 2 million years ago are un-intuitive for us. The same holds for our synthetic knowledge *a priori* of these two forms of sensible intuition. (As a matter of fact, obviously

doi: 10.48106/dial.v74.i3

these events are not un-intuitive, as Meillassoux emphasizes.) Every subject can in principle understand every proposition about any space-time region in the world when expressed in terms of the Kantian categories—or different, non-Kantian categories for that matter. When space and time are "given to us," when these forms of sensible intuition are hard-wired in our minds during the nine months we float around in the amniotic fluid of our mother's womb, then every single spatial point and region in space, and every moment and interval in time, are in principle "given to us" in one fell swoop. There are no exempted spatio-temporal regions in the world such that propositions about *these* regions cannot be thought, understood or known by subjects, including propositions about times long before we ourselves appeared on the terrestrial scene. Recall that our synthetic knowledge *a priori* of these forms of sensible intuition are Euclidean geometry (for space) and pure chronometry: both include *all* spatial points and regions, and *all* moments and temporal intervals; they are sensible-intuition-wise all on a par. A few citations from Kant's *Kritik*:[29]

> We present space as an *infinite* given magnitude. (1787, B39) Geometry is a science that determines the properties of space synthetically and yet *a priori*. (1787, B40) To say that time is *infinite* means nothing more than that any determinate magnitude of time is possible only through limitations [put] on a single underlying time. Hence the original presentation time must be given as *unlimited*. (1787, B48) We present the time sequence by a line progressing *ad infinitum*, a line in which the manifold constitutes a series of only one dimension. (1787, B50)

This is, roughly, the Kantian answer to Meillassoux's aporia (Q). Meillassoux's claim that Kantian correlationists cannot answer (Q) is not credible. They can answer it. The transition *from* an ancestral world (i.e. a spatio-temporal "part"

29 References to Kant's *Kritik* in standard fashion. Translations from Kant (1996). Originals in Kant (1787, our italics): "Der Raum wird als eine *unendliche* gegebene Grösse vorgestellt" (B39–40); "Geometrie ist eine Wissenschaft, welche die Eigenschaften des Raumes synthetisch und doch a priori bestimmt" (B40); "Die *Unendlichkeit* der Zeit bedeutet nichts weiter, als dass alle bestimmte Grösse der Zeit nur durch Einschränkungen einer einigen zum Grunde liegenden Zeit möglich sei. Daher muss die ursprüngliche Vorstellung Zeit als *uneingeschränkt* gegeben sein" (B48); "Und, eben weil diese innere Anschauung keine Gestalt gibt, suchen wir auch diesen Mangel durch Analogien zu ersetzen, und stellen die Zeitfolge durch eine ins *Unendliche* fortgehende Linie vor, in welcher das Mannigfaltige eine Reihe ausmacht, die nur von einer Dimension ist" (B50).

of $\Phi[\mathcal{R}]$) without human beings, even without animal beings, *to* a world full of them (i.e. another spatio-temporal "part" of $\Phi[\mathcal{R}]$) can be understood and known by means of the theory of evolution, which theory does not scandalise the Kantian forms of sensible intuition and epistemic categories one scintilla. At the transcendental level, we can think "speculatively" about $\mathcal{R}$, that $\mathcal{R}$ exists tenselessly, but at that level there is neither time nor coming-into-being of transcendental subjects. These subjects came into being in $\Phi[\mathcal{R}]$, and their capacity of reason enables them to think about themselves as thinking, observing, knowing and understanding subjects, which is what it means to say that they are *transcendental* subjects.

So much on behalf of the Kantian Correlationist.

When the "Speculative Turn" is based on Meillassoux's allegedly successful criticism of Correlationism (β) whilst Correlationism is, in fact, still standing, must we then conclude that the "speculative" realists have taken a turn without a reason? No, we should not. Metaphysical Realism (δ) is, has been, and presumably will be, a respectable metaphysical view. The speculative realists have joined the ranks. After all, there are no knockdown arguments in philosophy, as D.K. Lewis (1983, x) once remarked, certainly not when it comes to such general metaphysical theses as the varieties of Correlationism (β) and Metaphysical Realism (δ). All we can conclude is that Meillassoux's attempt to knock down Correlationism by confronting it with an allegedly irresolvable paradox has crumbled. Notwithstanding the fact that no silver bullets have been fired, Continental Speculative Realism can steam ahead.

Unless Donald Davidson is right, who posed a threat for Correlationism (β) as well as for Metaphysical Realism (δ). The final section briefly considers this more ominous threat.

9 Collapse

In our schematic and symbolic expression of the correlation

$$\mathcal{R} \rightsquigarrow \Phi[\mathcal{R}]$$

we have attempted to say as little as possible about the correlation ($\rightsquigarrow$) between reality ($\mathcal{R}$) and the experienced and conceptualised world ($\Phi[\mathcal{R}]$). Deliberately so, for as soon as one begins to assert things about this relation, one lands in a specific version of Correlationism (β) or Metaphysical Realism (δ), which all rely on the correlation. In his sensational paper "On the Very Idea of

doi: 10.48106/dial.v74.i3

a Conceptual Scheme," Davidson (1973) essentially argued, in our current terminology, that to speak of "the conceptual part" of $\Phi[\mathcal{R}]$ (our "conceptual scheme") is *unintelligible*.

In a nutshell, besides arguing that all attempts to elucidate the correlation between $\mathcal{R}$ and $\Phi[\mathcal{R}]$ metaphorically fail miserably (as e.g. the Strong Correlationists (β.2) and (β.3) will applaud), Davidson's central argument runs as follows.

If the correlation $\mathcal{R} \rightsquigarrow \Phi[\mathcal{R}]$ is intelligible (Premise), then the possibility of there being a distinct correlation, say $\mathcal{R} \rightsquigarrow \Psi[\mathcal{R}]$, is also intelligible. For the conceptual schemes of $\Phi[\mathcal{R}]$ and $\Psi[\mathcal{R}]$ to be genuinely distinct, it is impossible to translate them into each other (or the two distinct languages that express these conceptual schemes). For if it were possible to translate them, they would not be genuinely distinct. But earlier Davidson had argued, on the basis of his truth-conditional theory of meaning and his idea of radical interpretation (inspired by Quine's idea of radical translation), that untranslatable languages are not possible. (Which implies that conceptual relativism is impossible too.) Davidson (1973, 20):

> For we have found no intelligible basis on which it can be said that schemes are different. It would be equally wrong to announce the glorious news that all mankind—all speakers of a language, at least—share a common scheme and ontology. For if we cannot intellibly say that schemes are different, neither can we say that they are one.

Thus Premise leads to an untenable impossibility claim, and therefore must be repudiated: the correlation is *not* intelligible.

The ramifications of repudiating the intelligibility of the correlation are too ominous not to be mentioned here. Remarkable is that no continental philosopher in the speculative realism movement pays attention to it.

Davidson suggested that the intelligibility of the conceptual part of the correlation is a "third dogma" of "empiricism," which, after Quine's identification of two other untenable dogmas of empiricism, has to be abandoned too. But if Davidson is right, and the correlation is unintelligible, then not only empiricism must go, but every version of Correlationism (β) and Metaphysical Realism (δ) must go too, speculative realism not excluded. Since Direct Realism (α) is incoherent because to state it is to perform a contradiction, the only two metaphysical views that remain standing are Absolute Idealism (γ) and Pragmatism (ε).

Against Absolute Idealism (γ), a Wittgensteinian line of argument can be employed. When Absolute Idealists use the words "matter," "material object," "mental" etc. in the same manner as everybody else (which they must, otherwise they could not communicate with anybody else), and when usage is constitutive for meaning, then to claim that matter is mental (or ideal) and that everything is mental (or ideal) is simply incoherent. If such a Wittgensteinian line of argument is successful, then Pragmatism (ε) is the last man standing. All rival philosophical views (α–δ) will then have collapsed. A knock down argument after all?

Presumably not. Whether the criticism that Davidson's argument has received amounts to a definitive refutation, I dare not say.[30] I dare say that a critical analysis of Meillassoux's archument has appeared that threatens to refute it.

10 Appendix: Correlationists

We provide a few examples to illustrate that Meillassoux's terminology, including our extension and refinements, fits philosophy (these illustrations are necessarily brief and sketchy, and are *stricto sensu* otiose for the audience of this journal).

Berkeley was an Absolute Idealist (Subjectalist, γ). In his *Treatise concerning the Principles of Human Knowledge* (1710, pt. I.VI), we read:

> VI. Some Truths there are so near and obvious to the Mind, that a Man need only open his Eyes to see them. Such I take this Important one to be, to wit, that all the Choir of Heaven and Furniture of the Earth, in a word all those Bodies which compose the mighty Frame of the World, have not any Subsistence without a Mind, that their Being is to be perceived or known; that consequently so long as they are not actually perceived by me, or do not exist in my Mind or that of any other created Spirit, they must either have no Existence at all, or else subsist in the Mind of some eternal Spirit.

30 Some critical attempts: Quine (1981), McGinn (1982), Larson (1987), Child (1994), Hacker (1996), Baghramian (1998), Forster (1998), Ayers (2004), Coll Mármol (2007), McDowell (2009), Wang (2009), Coleman (2010).

doi: 10.48106/dial.v74.i3

According to Berkeley, reality ($\mathcal{R}$) is identified to the whole of all and only human minds, $\Phi_S[\mathcal{R}]$ for every subject S, and God.

Meillassoux says that Absolute Idealism "absolutizes" $\Phi[\mathcal{R}]$, makes do with only $\Phi[\mathcal{R}]$, and renounces $\mathcal{R}$, hence better denoted as: $\Phi[\cdot]$. Solipsism is a variety of absolute Idealism. Besides Berkeley, Fichte's rejection of $\mathcal{R}$ makes him an absolute idealist too, Meillassoux (2016) claims, and lists more absolute idealists:

> Sensation was absolutized (Maupertius' and Diderot's hylozoism), as was reason (Hegelian idealism), freedom (the Schelling of 1809), perception (Bergson and the image in itself, in the first Chapter of *Matter and Memory*), will (Schopenhauer), wills in their mutual conflict (Nietzsche's will to power), the self in its initial germ state (Deleuze's "larval selves" in *Difference and Repetition*), etc.

Hume was a Correlationist: $\Phi_S[\mathcal{R}]$ is the passing show, the stream of consciousness (ideas and impressions, which include sensations, desires, passions, sentiments) of subject S. Recall Hume's view about causality: there are no necessary connections in nature; everything our sensory experience tells us is that there are regularities between impressions, one event of a certain type followed by an event of another type, a "constant conjunction in temporal order." Clearly Hume speculated about $\mathcal{R}$, and claimed that we do not and cannot know about (necessary connections in) $\mathcal{R}$. Concerning perception, the prime candidate of (α) Direct Realism for direct access to $\mathcal{R}$, Hume was sceptical; from *An Enquiry concerning Human Understanding* (1758, 153):

> It is a question of fact, whether the perceptions of the senses be produced by external objects, resembling them: how shall this question be determined? By experience surely; as all other questions of a like nature. But here experience is, and must be entirely silent. The mind has never anything present to it but the perceptions, and cannot possibly reach any experience of their connexion with objects. The supposition of such a connexion is, therefore, without any foundation in reasoning.

We have not and cannot have a clue whether $\Phi[\mathcal{R}]$ represents $\mathcal{R}$; $\mathcal{R}$ is epistemically inaccessible for us. This makes Hume a Correlationist (β), arguably of the weak variety ($\beta.1$)

In Kant's transcendental objective (i.e. intersubjective) Idealism, $\Phi[\mathcal{R}]$ is the perceivable, knowable and understandable phenomenal world, *die Welt für uns*, brought about by $\mathcal{R}$ and Φ; Φ is constituted by the twelve epistemic categories (causality, quantity, etc.) and the two forms of *Anschauung* (time, space); and $\mathcal{R}$ is the hardly knowable noumenal reality, reality as-it-is-in-and-of-itself, *die Welt an sich*. Meillassoux (2014, 11) emphasizes that according to Kant, subjects can know four things about noumenal reality ($\mathcal{R}$): (i) the thing-in-itself exists independently of us (there are not only phenomena); (ii) it affects us and produces representations of it in us; (iii) it is not contradictory; and (iv) it is not spatio-temporal, because space and time are forms of *Anschauung* and pertain to its phenomenal representation. These are indisputably traces of the Metaphysical Realism (δ). But since our knowledge of $\mathcal{R}$ seems to be exhausted by these rather trivial items, which pale in comparison to the amount of gathered scientific knowledge about $\Phi[\mathcal{R}]$, even in Kant's day, Kant seems more appropriately classified as a strong type Correlationist ($\beta.2$).

In Schopenhauer's subjective Idealism, $\Phi_S[\mathcal{R}]$ is *my* active representation, *meine Vorstellung, die Welt für mich*, and $\mathcal{R}$ is *die Wille*, and is knowable, but only subjectively (from a "1st-person perspective") as wanting, desiring, longing, craving, hoping, intending, yearning. Schopenhauer (1958, vols. II, 197): "I call the Will the thing-in-itself." Schopenhauer (1958, vols. I, 3):[31]

> Therefore no truth is more certain, more independent of all others, and less in need of proof than this, namely that everything that exists to know, and hence the whole of this world, is only object in relation to the subject, perception of the perceiver, in a word, *representation*. [...] Everything that in any way belongs and can belong to the world is inevitably associated with this being-conditioned by the subject, and it exists only for the subject. The World is representation.

Since the Will, i.e. $\mathcal{R}$, can be known subjectively, Schopenhauer seems to fall somewhere between metaphysical Realism (δ) and strong token Correlationism ($\beta.3$).

31 Original German, in Schopenhauer (1844, Zweiter Band, Kapitel 18): "[...] nenne den Willen das Ding an sich"; and (1844, Erster Band, Erstes Buch, §1): "Keine Wahrheit is also gewisser, von allen anderen unabhängiger und eines Beweises weniger bedürftig, als diese, dass Alles, was für Erkentniss da ist, also die ganze Welt, nur Object in Beziehung auf das Subjekt ist, Anschauung des Anschauenden, mit Einem Wort, Vorstellung."

doi: 10.48106/dial.v74.i3

In Husserl's Phenomenology, $\Phi[\mathcal{R}]$ is consciousness and $\mathcal{R}$ is reality. Husserl (2003, 156): "Transcendental Idealism says: a nature without co-existing subjects of possible experience regarding it is unthinkable; possible subjects of experience are not sufficient."[32] For Heidegger's Existential Phenomenology, roughly, $\Phi[\mathcal{R}]$ is *Dasein* and $\mathcal{R}$ is *Sein*, and their "essential togetherness," their *Ereignis*, is the correlation, as Meillassoux (2008, 8) declares. Both Husserl and Heidegger are weak Correlationists (β.1).

In Russell's empiricist Phenomenalism, $\Phi_S[\mathcal{R}]$ are the sense data, the sensations, of subject S, with which S is intimately acquainted, and $\mathcal{R}$ comprises the entities that cause $\Phi_S[\mathcal{R}]$, or are the entities that S constructs out of his sensations. Similar posits hold for every other subject, and the whole of all events ($\mathcal{R}$) is justified abductively as the best explanation of every $\Phi_S[\mathcal{R}]$ and their similarities. Russell further assumes that the structure between the sense data mirrors relations between the causes in $\mathcal{R}$ of these sense data. This makes Russell move in the direction of Metaphysical Realism (δ).[33]

Wittgenstein's *Tractatus Logico-Philosophicus* (1922), which expounds a metaphysical theory of meaning, is based on a (δ) metaphysical realist theory of $\mathcal{R}$. Wittgenstein (1922) deemed *facts* to be truth-makers:

> 1.1 The world is the totality of facts, not of things.
>
> 4.01 The proposition is a picture of reality.
>
> 4.022 The proposition *shows* its meaning. The proposition *shows* how the facts are, *if* true.

This also smells of (α) Direct Realism. (Facts still are the most popular truth-makers, but may ultimately be redundant, as Betti 2015 *passim* has argued).

In the realism debate in philosophy of science, realists are Metaphysical Realists (δ), who take $\Phi[\mathcal{R}]$ to include prominently the theories and models that constitute our current scientific knowledge; they take $\mathcal{R}$ to be very knowable. Van Fraassen's (1980) famous constructive empiricism is an interesting combination: with respect to observable part of $\mathcal{R}$, the view is a metaphysical realist one (δ), and with respect to unobservable part of $\mathcal{R}$, the view is weak

32 Original German: "Der transzendentale Idealismus sagt: Eine Natur ist nicht denkbar ohne mit existierende Subjekte möglicher Erfahrung von ihr; es genügen nicht mögliche Erfahrungssubjekte."

33 See for example, Russell's Lecture III, "Our Knowledge of the External World," in the book bearing the same title (1914).

Correlationist (β.1). All scientific knowledge is about the observable part of $\mathcal{R}$; the unobservable part is epistemically inaccessible by us. (To find out where to draw the line in $\mathcal{R}$ between what is observable and what is unobservable is according to Van Fraassen a subject-matter of scientific inquiry rather than philosophical analysis.)

Anti-realists in Analytic Philosophy, like Putnam (1981) (after having denounced Metaphysical Realism), Dummett (1975), Rorty (1979), and Brandom (2008), fall under Pragmatism (ε) as characterised here.

So much for this hodge-podge of illustrations of Meillassoux's taxonomy. These should suffice to convince us that Meillassoux is not talking to himself but about many renowned philosophers, whose views indisputably fit in the category of Correlationism (β).*

F.A. Muller
0000-0002-7343-4649
Utrecht University
Erasmus University Rotterdam
f.a.muller@uu.nl

References

Ayers, Michael R. 2004. "Sense Experience, Concepts and Content: Objections to Davidson and McDowell." in *Perception and Reality: From Descartes to the Present*, edited by Ralph Schumacher, pp. 239–262. Paderborn: Mentis Verlag, https://philpapers.org/archive/AYESEC.pdf.

Baghramian, Maria. 1998. "Why Conceptual Schemes?" *Proceedings of the Aristotelian Society* 98: 287–306, doi:10.1111/1467-9264.00038.

Berkeley, George. 1710. *A Treatise Concerning the Principles of Human Knowledge*. Dublin: Aaron Rhames.

—. 1949. *The Works of George Berkeley Bishop of Cloyne. Volume Two: The Principles of Human Knowledge, First Draft of the Introduction to the Principles, Three Dialogues between Hylas and Philonous, Philosophical Correspondence with Johnson*. Bibliotheca Britannica Philosophica. London: Thomas Nelson and Sons Ltd. Edited by Thomas Edmund Jessop.

—. 1988. *A Treatise Concerning the Principles of Human Knowledge*. Oxford: Oxford University Press. Edited by Jonathan Dancy.

* I am grateful to Gijs Leegwater, Geurt Sengers and Stefan Wintein (Erasmus University Rotterdam), Menno Lievers, Jesse Mulder (Utrecht University), Arjen Kleinherenbrink (Radboud University Nijmegen), Victor Gijsbers (Leiden University), and two referees of this journal for suggestions, comments and corrections.

Betti, Arianna. 2015. *Against Facts*. Cambridge, Massachusetts: The MIT Press, doi:10.7551/mitpress/9780262029216.001.0001.

Brandom, Robert B. 2008. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199542871.001.0001.

Brassier, Ray. 2007. *Nihil Unbound: Enlightenment and Extinction*. Basingstoke, Hampshire: Palgrave Macmillan, doi:10.1057/9780230590823.

Braver, Lee. 2007. *A Thing of This World: A History of Continental Anti-Realism*. Evanston, Illinois: Northwestern University Press.

Bryant, Levi R. 2011. *The Democracy of Objects*. Ann Arbor, Michigan: Open Humanities Press, doi:10.3998/ohp.9750134.0001.001.

Bryant, Levi R., Srnicek, Nick and Harman, Graham, eds. 2011. *The Speculative Turn: Continental Materialism and Realism*. Melbourne: Re.Press.

Child, William. 1994. "On the Dualism of Scheme and Content." *Proceedings of the Aristotelian Society* 94: 53–71, doi:10.1093/aristotelian/94.1.53.

Christias, Dionysis. 2016. "Sellars, Meillassoux, and the Myth of the Categorial Given: A Sellarsian Critique of ''Correlationism" and Meillassoux's 'Speculative Realism'." *Journal of Philosophical Research* 41: 105–128, doi:10.5840/jpr201662970.

Coleman, Martin. 2010. "On the Very Good Idea of a Conceptual Scheme." *The Pluralist* 5(2): 69–86, doi:10.1353/plu.2010.0002.

Coll Mármol, Jesús Antonio. 2007. "Conceptual Schemes and Empiricism: What Davidson Saw and McDowell Missed." *Theoria (San Sebastian), Secunda época* 22(2): 153–165, doi:10.1387/theoria.465.

Cuyubamba, Marco Antonio, Konoplya, Roman A. and Zhidenko, Olexandr. 2018. "No Stable Wormholes in Einstein-dilaton-Gauss-Bonnet Theory." *Physical Review D* 98(4): 69–86, doi:10.1103/physrevd.98.044040.

Davidson, Donald. 1973. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47: 5–20. Reprinted in Davidson (1984, 183–198), doi:10.2307/3129898.

—. 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, doi:10.1093/0199246297.001.0001.

Dummett, Michael A. E. 1975. "The Philosophical Basis of Intuitionistic Logic." in *Logic Colloquium '73*, edited by H. E. Rose and John C. Shepherdson, pp. 5–40. Studies in Logic and the Foundations of Mathematics n. 80. Amsterdam: North-Holland Publishing Co. Reprinted in Dummett (1978, 215–247).

—. 1978. *Truth and Other Enigmas*. Cambridge, Massachusetts: Harvard University Press.

Ennis, Paul J. 2011. "The Transcendental Core of Correlationism." *Cosmos and History: The Journal of Natural and Social Philosophy* 7(1): 37–48.

Forster, Michael N. 1998. "On the Very Idea of Denying the Existence of Radically Different Conceptual Schemes." *Inquiry* 41(2): 133–185, doi:10.1080/00201749832185 0.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press, doi:10.1093/0198244274.001.0001.

Gijsbers, Victor. 2015. "The Boundless Correlation: Meillassoux, Ancestrality, and the Limits of Givenness." Unpublished manuscript.

Guyer, Paul and Horstmann, Rolf-Peter. 2021. "Idealism." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/spr2021/entries/idealism/.

Hacker, Peter M. S. 1996. "On Davidson's Idea of a Conceptual Scheme." *The Philosophical Quarterly* 46(184): 289–307, doi:10.2307/2956442.

Harman, Graham. 2007. "Quentin Meillassoux: A New French Philosopher." *Philosophy Today* 51(1): 104–117, doi:10.5840/philtoday200751163.

—. 2011. *Quentin Meillassoux: Philosophy in the Making*. Edinburgh: Edinburgh University Press. second edition: Harman (2015).

—. 2015. *Quentin Meillassoux: Philosophy in the Making*. 2nd ed. Edinburgh: Edinburgh University Press. First edition: Harman (2011), doi:10.1515/9780748693467.

—. 2018. *Speculative Realism: An Introduction*. Cambridge: Polity Press.

—. 2019. "Speculative Realism." in *Critical Terms in Future Studies*, edited by Heike Paul, pp. 287–290. London: Palgrave Macmillan, doi:10.1007/978-3-030-28987-4_44.

Hume, David. 1758. *An Enquiry Concerning Human Understanding*. 2nd ed. London: Andrew Millar of the Strand.

Husserl, Edmund. 2003. *Transzendentaler Idealismus. Texte aus dem Nachlass (1908–1921)*. Husserliana [HUSS] n. 36. Den Haag: Kluwer Academic Publishers. Edited by Robin D. Rollinger, with Rochus Sowa, doi:10.1007/978-94-007-1062-7.

Kant, Immanuel. 1787. *Kritik der reinen Vernunft*. 2nd ed. Riga: Johann Friedrich Hartknoch.

—. 1996. *Critique of Pure Reason*. Indianapolis, Indiana: Hackett Publishing Co. Translated by Werner S. Pluhar, introduced by Patricia Kitcher.

Koch, Anton Friedrich. 2006. "The Subjectivity Thesis and Its Corollaries." *Philosophical Inquiry* 28(3/4): 9–20, doi:10.5840/philinquiry2006283/42.

Larson, David T. 1987. "Correspondence and the Third Dogma." *Dialectica* 41(3): 231–238, doi:10.1111/j.1746-8361.1987.tb00890.x.

Lewis, David. 1983. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

McDowell, John Henry. 2009. *The Engaged Intellect: Philosophical Essays*. Cambridge, Massachusetts: Harvard University Press, doi:10.2307/j.ctv1p6hp01.

McGinn, Marie. 1982. "The Third Dogma of Empiricism." *Proceedings of the Aristotelian Society* 82: 89–101, doi:10.1093/aristotelian/82.1.89.

McTaggart, J. McT. Ellis. 1908. "The Unreality of Time." *Mind* 18(68): 457–484, doi:10.1093/mind/xvii.4.457.

Meillassoux, Quentin. 2006. *Après la finitude. Essai sur la nécessité de la contingence*. Paris: Éditions du Seuil. Préface d'Alain Badiou.

—. 2008. *After Finitude: An Essay on the Necessity of Contingency*. London: Continuum International Publishing Group. Translation of Meillassoux (2006) by Ray Brassier.

—. 2014. *Time Without Becoming*. Milano: Mimesis Edizioni. Lecture at Middlesex University, London, UK, 8 May 2008, edited by Anna Longo.

—. 2016. "Iteration, Reiteration, Repetition: A Speculative Analysis of the Sign Devoid of Meaning." in *Genealogies of Speculation: Materialism and Subjectivity since Structuralism*, edited by Armen Avanessian and Suhail Malik, pp. 117–198. London: Bloomsbury Academic, doi:10.5040/9781474220354.ch-005.

Merleau-Ponty, Maurice. 1945. *Phénoménologie de la perception*. Tel. Paris: Gallimard.

—. 2002. *Phenomenology of Perception*. 2nd ed. London: Routledge. Translation of Merleau-Ponty (1945) by Colin Smith, doi:10.4324/9780203994610.

Mulder, Jesse M. 2014. "Conceptual Realism: The Structure of Metaphysical Thought." PhD dissertation, Utrecht: Department of Philosophy, https://hdl.handle.net/1874/294803.

Putnam, Hilary. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511625398.

Quine, Willard van Orman. 1948. "On What There Is." *The Review of Metaphysics* 2(5): 21–38. Republished as Quine (1951), reprinted in Quine (1953, 1–19).

—. 1951. "On What There Is." *Proceedings of the Aristotelian Society, Supplementary Volume* 25: 217–234. Reprint of Quine (1948), doi:10.1093/aristoteliansupp/25.1.125.

—. 1953. *From a Logical Point of View: 9 Logico-Philosophical Essays*. Cambridge, Massachusetts: Harvard University Press. Cited after the revised edition: Quine (1961).

—. 1961. *From a Logical Point of View: 9 Logico-Philosophical Essays*. 2nd ed. Cambridge, Massachusetts: Harvard University Press. Revised edition of Quine (1953), reprinted 1980.

—. 1981. "On the Very Idea of a Third Dogma." in *Theories and Things*, pp. 38–42. Cambridge, Massachusetts: Harvard University Press.

Roffe, Jon. 2012. "Time and Ground: A Critique of Meillassoux's Speculative Realism." *Angelaki. Journal of the Theoretical Humanities* 17(1): 57–62, doi:10.1080/0969725x.2012.671661.

Rorty, Richard M. 1979. *Philosophy and the Mirror of Nature*. Princeton, New Jersey: Princeton University Press. Reissued as Rorty (2009).

—. 2009. *Philosophy and the Mirror of Nature*. Princeton, New Jersey: Princeton University Press. Thirtieth-Anniversary Edition, with an introduction by Michael Williams and an afterword by David Bromwich; first edition: Rorty (1979), doi:10.2307/j.ctvc77b6z.

Russell, Bertrand Arthur William. 1912. *The Problems of Philosophy*. Home University Library of Modern Knowledge n. 35. London: Williams and Norgate.

—. 1914. *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*. LaSalle, Illinois: Open Court Publishing Co. Revised edition: Russell (1926).

—. 1926. *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*. 2nd ed. London: George Allen & Unwin. Revised edition of Russell (1914).

Schopenhauer, Arthur. 1819. *Die Welt als Wille und Vorstellung*. Leipzig: Bibliographisches Institut, F.A. Brockhaus.

—. 1844. *Die Welt als Wille und Vorstellung*. 2nd ed. Leipzig: Bibliographisches Institut, F.A. Brockhaus. First edition: Schopenhauer (1819).

—. 1859. *Die Welt als Wille und Vorstellung*. 3rd ed. Leipzig: F.A. Brockhaus. First edition: Schopenhauer (1819).

—. 1958. *The World as Will and Representation*. Indian Hills, Colorado: The Falcon's Wing Press. Translation of Schopenhauer (1859) by E.F.J. Payne.

Sebold, Richard. 2014. *Continental Anti-Realism: A Critique*. Lanham, Maryland: Rowman & Littlefield.

Shaviro, Steven. 2014. *The Universe of Things: On Speculative Realism*. posthumanities n. 30. Minneapolis, Minnesota: University of Minnesota Press, doi:10.5749/minnesota/9780816689248.001.0001.

Toadvine, Ted. 2014. "The Elemental Past." *Research in Phenomenology* 44(2): 262–279, doi:10.1163/15691640-12341288.

Verene, Donald Phillip. 2007. *Hegel's Absolute: An Introduction to Reading the* Phenomenology of Spirit. Albany, New York: State University of New York Press.

Vrahimis, Andreas. 2012. "'Was There a Sun Before Men Existed?' A.J. Ayer and French Philosophy in the Fifties." *Journal for the History of Analytical Philosophy* 1(9), doi:10.4148/jhap.v1i9.1449.

Wang, Xinli. 2009. "On Davidson's Refutation of Conceptual Schemes and Conceptual Relativism." *Pacific Philosophical Quarterly* 90(1): 140–164, doi:10.1111/j.1468-0114.2009.01332.x.

Wiltsche, Harald A. 2017. "Science, Realism and Correlationism: A Phenomenological Critique of Meillassoux' Argument from Ancestrality." *European Journal of Philosophy* 25(3): 808–832, doi:10.1111/ejop.12159.

Wittgenstein, Ludwig. 1922. *Tractatus logico-philosophicus*. International Library of Psychology, Philosophy and Scientific Method. London: Kegan Paul, Trench, Trübner & Co.

Certainty and Assertion

Jacques-Henri Vollet

It is widely held that assertions are partially governed by an epistemic norm. But what is the epistemic condition set out in the norm? Is it knowledge, truth, belief, or something else? In this paper, I defend a view similar to that of Stanley (2008), according to which the relevant epistemic condition is epistemic certainty, where epistemic certainty (but not knowledge) is context-sensitive. I start by distinguishing epistemic certainty, subjective certainty, and knowledge. Then, I explain why it's much more plausible to think that "certain," rather than "know," is context-sensitive. After that, I respond to an important worry raised by Pritchard, according to which the proposed view is too strong to accommodate our current practice of assertion. I then show that the main linguistic and conversational data advanced in the recent literature in favour of the knowledge condition are best explained by the certainty view. Finally, I offer two principled considerations: the certainty view is the only one compatible with three independently plausible claims and it fits very well with the common thought that knowledge does not entail certainty.

According to many philosophers, assertions are partially governed by an epistemic norm, at least in the minimal sense that they must satisfy a relevant epistemic condition.[1] But what is this epistemic condition? The main proposals include truth (Weiner 2005; Whiting 2013), belief (Bach 2005), knowledge (Williamson 2000; DeRose 2009; Ichikawa 2017), reasonableness or justification to believe (Douven 2006; Lackey 2007; Kvanvig 2009), and warrant (Brown 2010, 2011; Gerken 2017). Many writers are monist, but some defend pluralism (Levin 2008). Some take the relevant epistemic condition to be in some way sensitive to the context (DeRose 2009; Brown 2010, 2011; Gerken 2017; Ichikawa 2017) while others contend that it is invariant or insensitive.

1 The proposal defended in this paper is meant to be neutral on whether the relevant epistemic condition states a norm constitutive and/or individuative of assertions. For discussion, see Engel (2008), Pagin (2015), Gerken and Petersen (2020). See also Fassio (2017) for the distinction between (constitutive) norms and regulation conditions.

In this paper, my aim is to introduce and defend a view similar to that of Stanley (2008), according to which the relevant epistemic condition is epistemic certainty, where "certainty" is understood in a context-sensitive way. In section 1, I introduce the certainty view by clarifying the relations between epistemic certainty, subjective certainty and knowledge. I explain why I think it is *epistemic* certainty, rather than *subjective* certainty, which primarily matters for assertion. In section 2, I show that, in contrast to "know," there are strong linguistic and conversational reasons to think that "certain" is context-sensitive. In section 3, I respond to Pritchard's worry that the proposed account is too strong to accommodate our current practice of assertion. Section 4 shows that the certainty view easily explains all the conversational and linguistic data recently put forth in favour of the knowledge view, while also explaining data recalcitrant to the knowledge view. Finally, I adduce two principled considerations in favour of the certainty view in section 5.[2, 3]

1 Epistemic Certainty, Subjective Certainty, and Knowledge

According to the proposal defended in this paper, the epistemic condition relevant for appropriate assertion is *epistemic* certainty:

2 As we will see, I often follow Stanley (2008), but I consider data and responses to objections that he does not consider, as well as new reasons to think that the certainty norm is superior to its competitors. I also make claims and arguments that Stanley might want to reject. For example, I suggest that being in a position to know is often sufficient for epistemic certainty, a claim which is inconsistent with the testimonial argument he proposes at p. 52. Some other differences will be noted in due course.

3 While the present paper was under review in this journal, two papers were published defending views congenial to the one proposed here (Petersen 2019; Beddor 2020). There is some overlap between these papers and the present one, but also important differences. Let me mention the most striking. Beddor (2020) mainly defends a certainty norm for practical reasoning and only briefly mentions some basic data in favour of a similar norm for assertion. Petersen (2019) defends a certainty norm for assertion, but the way in which this norm is understood carries assumptions that one might want to reject. For example, Petersen's arguments often rely on the assumption that the certainty norm is *additional* to the knowledge norm. Petersen's certainty norm includes a belief component which, on the view defended in the present paper, should be merely derivative. In contrast to Petersen's certainty norm, the certainty norm defended in this paper is overtly gradable and not binary. It must also be noted that the ways in which this paper and Petersen's deal with Pritchard, Williamson and Turri's objections do not coincide, but complement each other. Finally, while Petersen offers arguments which are not considered here (e.g. the isolated second-hand knowledge argument and the concessive knowledge attributions argument), the present paper discusses more data, establishes the bad prospects of competing views, and suggests two further motivations for the certainty view (section 5).

CN-E. S (epistemically) ought to assert that p only if p is *epistemically* certain for S.

Some basic clarifications are required in order to get the proposal right. First, we can think of epistemic certainty as a high degree of epistemic justification (many would say the highest degree) and of subjective certainty as a high degree of confidence (many would say the highest degree). Second, in ordinary and philosophical contexts, we often use the expression "S is certain/sure that p" to refer to subjective certainty. But while we mainly use "It is certain for S that p" to refer to epistemic certainty in philosophical contexts, this expression is not frequently used in ordinary contexts.[4] However, the notion of epistemic certainty is not technical. Suppose there is a televised poker tournament where the broadcasters and audience can see the hands but the players cannot. One player bets confidently at a point when it is not certain, given the information available to her, that her hand is the best, though the broadcasters and audience can see that it is. It seems perfectly natural to say something like "She can't be sure that her hand is the best" or "She should not be so sure that her hand is the best." These are statements about lack of epistemic warrant for subjective certainty, i.e about lack of epistemic certainty.

A third clarification concerns the relation between knowledge and epistemic certainty. It's natural to think that epistemic certainty entails knowledge-level justification (or being in a position to know). Perhaps more surprisingly, the view defended in this paper also takes it that knowledge or knowledge-level justification does *not* entail epistemic certainty.[5] At first sight, this seems

4 As an anonymous referee pointed out to me, we find slightly different expressions, like "It is certain that p for S" (e.g. "It is certain that there will be punishment for the prisoners"), typically used to express the speaker's certainty that a proposition p being about a subject S is true. Another example is "It is certain, for S, that p" (e.g. "It is certain, for the prisoners, that there will be punishment"), typically used to express that S is certain that p. The expression "It is certain for S that p" when used in ordinary language sometimes seems to be equivalent to the further expression "S takes for certain/granted that p." For example, "It is certain for John that it will rain" seems equivalent to "John takes for certain that it will rain," and this seems to express John's subjective certainty. Still, "It is certain for S that p" may also be taken as equivalent to "According to S, it is certain that p." For example, "According to John, it is certain that it will rain." Here, we seem to express John's belief that the proposition that it will rain is certain, which is to express epistemic certainty.

5 According to Stanley (2008, 35), to say that p is epistemically certain for S is to say that S "knows that p (or is in a position to know that p) on the basis of evidence that gives one the highest degree of justification for one's belief that p." This definition implies that certainty is an absolute notion. As we will see below, however, we can distinguish certainty (a contextually-influenced

doi: 10.48106/dial.v74.i3

to clash with infallibilist approaches to knowledge. On closer inspection, however, it should be clear that, on pain of scepticism, everyone should grant that knowledge does not require satisfying absolutely maximal epistemic standards (e.g. Cartesian certainty).[6] In addition, and following for example Williamson's influential non-sceptical infallibilist view, it's natural to think that there is some epistemic space between knowledge-level standards and absolutely maximal epistemic standards.[7] This fact is reflected in ordinary language. The expression "I know that *p* with certainty (/for sure)" does not appear redundant and expresses something stronger than "I know that *p*" (although not necessarily something as strong as the satisfaction of Cartesian standards). Thus, the ordinary notion of certainty seems to capture a degree of justification (or confidence) between knowledge-level justification (or confidence) and absolutely maximal certainty. This notion of certainty is the notion invoked by the view defended in this paper.

Let me add two further clarifications. In section 4 below, we will see that expressions of *subjective* certainty and uncertainty are also highly relevant for assertions. On this basis, one might think that what really matters is subjective certainty:[8]

> CN-S. S (epistemically) ought to assert that *p* only if S is *subjectively* certain that *p*.

degree of justification, potentially above knowledge-level justification) and absolute certainty. I do not take the certainty view of assertion to demand absolute certainty. It is worth noting that Stanley also distinguishes certainty from absolute certainty later in his paper (see 2008, 54).

6 See Brown (2011, 2018). Of course, if it is also true that certainty does not require the satisfaction of absolutely maximal epistemic standards, as I will argue below, this point is not enough to show that knowledge does not require certainty. But, at least, it is enough to show that knowledge does not *have to* require certainty (at least in some sense) and to suggest that there is an epistemic space between knowledge and absolute certainty. Thanks to a reviewer for raising this point.

7 Williamson's probability one infallibilism has it that you know that *p* only if *p* has probability one given your evidence (i.e. only if *p* has the highest degree of evidential probability). In this sense, knowledge requires epistemic certainty. But Williamson also grants that our epistemic position with respect to *p* can be improved further by knowing that we know (that we know…) that *p*. As he writes (2009, 339): "Thus some propositions with evidential probability 1 will have epistemic advantages over other propositions with evidential probability 1." In addition, although Williamson assumes that knowledge requires some kind of subjective certainty, namely outright belief, he also admits that there are weaker and stronger forms of outright beliefs (see 2000, 99). See also Wedgwood (2012) and Gao (2019).

8 Stanley favours the epistemic norm over the subjective norm only "for the sake of discussion" (2008, 52) and does not argue for its superiority, as I will do.

Further, I will take the data concerning subjective certainty as favouring CN-E and not CN-S. This may seem problematic.

Some reflection on the relation between epistemic and subjective certainty can dispel these worries, though. To begin with, a natural thought is that epistemic certainty is the epistemic norm of subjective certainty:[9]

> ECNSC. If *p* is epistemically certain for S (and S considers whether *p*) then S (epistemically) ought to be subjectively certain that *p*, and if *p* is not epistemically certain for S, then S (epistemically) ought not to be subjectively certain that *p*.

If so, if we accept either CN-E or CN-S, there will be something inappropriate if a subject asserts that *p* while lacking epistemic certainty or subjective certainty (see Stanley 2008, 51–52). Indeed, suppose we accept CN-E and suppose that S asserts that *p* while lacking epistemic certainty. It follows that S violates the supposed norm of assertion (CN-E). Assume that S asserts that *p* while lacking subjective certainty. S violates either the norm of subjective certainty (ECNSC) or the supposed norm of assertion (CN-E). Alternatively, suppose we accept CN-S. If S asserts that *p* while lacking epistemic certainty, then S violates either the norm of subjective certainty (ECNSC) or the supposed norm of assertion (CN-S). If S asserts that *p* while lacking subjective certainty, then S violates the supposed norm of assertion (CN-S). In brief, given ECNSC and the subject's adherence to either CN-E or CN-S, if it's appropriate for S to assert that *p*, then S does not lack epistemic and subjective certainty with respect to *p*.

Still, why should we prefer CN-E over CN-S? The main reason is the following. Suppose the norm of assertion is CN-S. It follows that if, after reading his favoured guru's book, John is certain or completely convinced that he is a cabbage, then John can appropriately assert "I am a cabbage." But it is far from clear, to say the least, that John's assertion is epistemically warranted. Further, in general, we should not allow normative reasons (and permissions) to be generated merely by the adoption of unjustified attitudes. This point has been largely and forcefully made in the literature with regard to normative requirements in general and there is no reason to think that similar considerations do not apply here.[10] Yet, according to CN-S, if John wants to

9 As Klein (1998) writes: "Presumably a person would want the degree of belief in a proposition to parallel the degree of epistemic warrant for it."

10 See e.g. Broome (2013). See also Williamson (2000, 260–261).

doi: 10.48106/dial.v74.i3

assert that he is a cabbage while he is not certain that he is a cabbage, he can warrant his assertion by adopting the attitude of certainty. CN-S has implausible consequences.[11] It's important to stress, though, that in this framework the apparent plausibility of CN-S can be explained: necessarily, if S satisfies ECNSC and CN-E, then either S is subjectively certain or she does not assert. But, again, that does not imply that there is a norm of assertion such as CN-S.

2 The Context-Sensitivity of Epistemic Certainty

An additional claim essential to the proposed view is that "certain" is context-sensitive. To illustrate, consider the following cases (inspired by Huemer 2007):

> AIRPORT. Mary is picking up Sam from the AIRPORT, but she is a little late, so she calls Sam on his cell phone.
> Mary: Where are you?
> Sam: I'm on the ground; we've just landed.
> Mary: Is it possible that you're still in the air?
> Sam: No, it's certain that I'm on the ground. I can see it through the window.

> EPISTEMOLOGY CLASS. John is teaching a class about philosophical scepticism. After reviewing Descartes' sceptical scenarios in the *First Meditation*, John gestures at the table at the front of the room and asks the class: "So, is it certain that there is a table here?" A student replies: "No, it isn't certain. We might be the victim of an evil demon."

A visual experience that *p* seems to be sufficient for an appropriate assertion of "It is certain that *p*" in AIRPORT, but insufficient for an appropriate assertion of

11 Here are some further reasons to think that the norm is CN-E and not CN-S. First, CN-S is incompatible with the plausible claim that "selfless assertions" (where speakers assert in accordance with their evidence but against their beliefs) can be epistemically appropriate (see Lackey 2007, 608). Second, when challenged, we defend our assertions by putting forth reasons to think that they are true, not by citing our strong convictions that they are true. Third, suppose the evidence is sufficient, but one irrationally does not believe that *p* and, as a result, one does not assert that *p*. The mere fact that one is not convinced that *p* (while it's clear that the evidence sufficiently indicates that *p*) does not seem to justify—although it can explain—the fact that one refrains from asserting that *p*.

"It is certain that *p*" in EPISTEMOLOGY CLASS. This suggests that the epistemic standards that must be satisfied for the truth of sentences such as "It is certain that *p*" and "It is uncertain that *p*" shift with the context (i.e. the aim of the discussion, the interests of the speakers, etc.).[12]

Similar pairs of cases were initially offered in favour of contextualism about "know."[13] However, a potential problem for contextualism about "know" is that, prima facie, there are no clear linguistic reasons to think that "know" is context-sensitive and some linguistic reasons to think otherwise.[14] In addition, we seem to be ignorant of the alleged context-relativity of "know" whereas we are not ignorant of the context-sensitivity of other uncontroversial context-sensitive terms, like indexicals, gradable adjectives, etc (see Schiffer 1996). Crucially, things are radically different with "certain." From a linguistic point of view, it is highly plausible that "certain" is context-sensitive. For it is uncontroversial that "certain" (contrary to "know," and like "tall") is gradable. It can take the comparative form (more… than…), it can be modified by degree modifiers such as adverbs (very, extremely, totally, absolutely…), and it can be combined with "how" to form questions. We can say that John is more certain than Bill that the bank is open, that John is absolutely certain that the bank is open, and we may ask how certain John is. (Similar examples can be provided involving an epistemic sense of "certain.") If so, there is a scale associated with this adjective. Like "tall," it's very plausible to think that the degree on the scale which picks out the appropriate threshold for "certain" is context-sensitive.

12 Since these cases are first-person cases, they are not sufficient to rule out that the variation is due to the subject's situation, rather than the speaker's (thanks to a reviewer for raising this issue). Still, the linguistic considerations below lend further independent support to the suggestion that "certain" is context-sensitive. We must also keep in mind that adopting a contextualist semantics for "certain" is not incompatible with holding a view like pragmatic encroachment about certainty, that is, a view in which truth-irrelevant factors of the subject's situation (such as the rationality for the subject of acting on the proposition) are part of the truth-conditions of "certainty" attributions (see e.g. Stanley 2005 for a defence of pragmatic encroachment about knowledge and evidence).

13 See DeRose (2009), Cohen (1999), Blome-Tillmann (2014) and Ichikawa (2017). For an alternative explanation of these pairs of cases appealing to a norm of assertion which has features compatible with those of the certainty norm defended here, see Vollet (2020).

14 See Stanley (2005) and Blome-Tillmann (2014) for possible replies. My point here is just that it is much easier to accept that "certain" is context-sensitive.

doi: 10.48106/dial.v74.i3

Someone may object that "certain" is an *absolute maximum-standard* gradable adjective, which always picks out the highest degree on the scale.[15] Even so, recent linguistic theories have suggested that absolute adjectives may also be context-sensitive (what counts as "the highest degree" on the scale may be context-sensitive).[16] Further, it's common (and useful) to distinguish a relative (or non absolutely maximal) sense of "certain" from an absolute sense of "certain," so that we can distinguish "certainty" from "absolute certainty," and various degrees of certainty in between.[17] For example, it does not seem weird at all to say "I'm sure that I have hands, but I am even more certain that I exist." This makes sense of the idea that non-sceptical infallibilists about knowledge can grant that we can know that p with "more or less certainty," where "less certainty" does not imply that the proposition is uncertain or that we are not certain of it in some important sense.[18] Thus, even if "absolute certainty" were not context-sensitive, it remains that the threshold for "certain" could

15 On the distinction between the different kinds of gradable adjectives (relative *vs* absolute and minimum-standard *vs* maximum-standard) see Kennedy and McNally (2005). The claim that "certain" is an absolute maximum-standard adjective, so that certainty implies absolute certainty, is defended by Unger (1975).

16 See Cruse (1980); McNally (2011). One way of developing a context-sensitive semantics for "certain," understood as a maximum-standard absolute adjective, is suggested by Lewis (1979, 353–354), according to whom there are context-sensitive standards of precision for the correct use of absolute terms (like "flat"). Some differences (say, some bumps) may be irrelevant in some contexts or given the object under consideration (e.g. a pavement, a table). Beddor (2016, ch. 3) develops this suggestion by assuming that certainty is the highest grade on the scale and that the granularity of the scale is context-dependent. On this model, propositions which do not count as equally certain on a fine-grained scale can count as equally certain on a less fine-grained scale. This view aims to reconcile the idea that "certainty" is a maximum-standard adjective with the intuition that we can truly ascribe "certainty" to many things in many contexts.

17 Like Stanley, I'm inclined to think that "certain" is not a maximum-standard adjective and that "[t]he semantic function of 'absolutely' is to raise the degree on the scale above that for 'certain' " (2008, 54). See also Popper: "it is not impossible to improve even on the most certain of certainties" (1972, 79). Many other authors reject the absolutely maximalist view of "certain" by relativizing the standards of certainty (see, among others, Firth 1967; Ayer 1973, 232; Miller 1978; Williamson 2000, 213, 254). Klein (1981, 181–189) distinguishes two concepts of absolute certainty: absolute certainty in the actual world and absolute certainty in all the possible worlds.

18 To illustrate, consider subjective certainty. Assume Williamson's view that outright belief comes in degrees. Then, even if "outright believing" at the first degree may count as being certain (for Williamson, it involves being willing to rely on p, at least in some situations), it does not count as the highest grade of certainty (for we may not be willing to act on p in any situation). Or assume Wedgwood's view that to outright believe p is to be disposed to act on p in normal situations. Then you can outright believe p, and in this sense be certain that p, even if you do not have the highest degree of certainty, for you could have a stronger disposition to act on p in normal and in (some further) abnormal situations.

be context-sensitive. In contrast to "know," there is no linguistic objection to the idea that "certain" (or "absolutely certain") is context-sensitive and some good reasons to think that it is.

It is also important to emphasize that we are not ignorant of the context-sensitivity of "certain." Consider the following dialogue:

TALL. John: Robert is tall.
Paul: But is Robert taller than six feet?
John: No.
Paul: So Robert is not tall, right?
John: I did not mean he is *that* tall.

Clearly, John's last answer is perfectly understandable, for "tall" is context-sensitive.

Consider now the following dialogue, with "know":

TRAFFIC JAM 1. John: I know that Robert will be here at 10 a.m.
Paul: But can you rule out that he will be late due to an exceptional traffic jam?
John: No.
Paul: So you don't know that Robert will come at 10 a.m, right?
John: I didn't mean *that* knowledge (/I did not mean "I know" or "knowledge" in that sense).

John's last utterance is rather puzzling and this is easily explained if "know" is not context-sensitive (see Stanley 2005, 52–53). John should either grant that he does not know, or else challenge the relevance of the traffic jam possibility.

Consider now a similar dialogue with "certain":

TRAFFIC JAM 2. John: It's certain that Robert will be here at 10 a.m.
Paul: But can you rule out the possibility that he will be late due to an exceptional traffic jam?
John: No.
Paul: So, it is not certain that Robert will come at 10 a.m., right?
John: I did not mean it is *that* certain (/I was not considering such a level of certainty/I didn't mean it is absolutely certain).

doi: 10.48106/dial.v74.i3

John's last answer is much less puzzling than John's last answer in TRAFFIC JAM 1, and as acceptable as that of John in TALL.[19] The claim that "certain" is context-sensitive, like "tall," is immensely plausible—certainly much more plausible than the claim that "know" is context-sensitive.

In sum, linguistic and conversational considerations give us good reasons to adopt a context-sensitive view of "certain," and if we adopt a shifty view at all, it's much more natural to adopt a shifty view about "certain" than about "know."

If "certain" is context-sensitive, we must determine the context which is relevant to assess the epistemic appropriateness of assertions. For this purpose, it seems quite natural to invoke the speaker's context and relativise the certainty norm of assertion as follows:

19 An anonymous referee pointed out to me that TRAFFIC JAM 2 seems no more natural than the following dialogue involving "know":

> TRAFFIC JAM 3. John: I know that Robert will be here at 10 a.m.
> Paul: But can you rule out that he will be late due to an exceptional traffic jam?
> John: No.
> Paul: So, you don't know that Robert will come at 10 a.m., right?
> John: I did not mean it is *that* certain (/I was not considering such a level of certainty/I didn't mean it is absolutely certain).

I agree that John's last utterance seems equally acceptable in TRAFFIC JAM 3. But this is what we should expect. If "certain" is context-sensitive, "know with certainty" is context-sensitive, and to assert "I know that p" does not necessarily mean that one knows that p with the level of certainty suggested by Paul. Therefore, John's last answer is easily understood as meaning "I did not mean that I knew that p with this level of certainty, but only that I knew that p."

We can make sense of Paul's challenge if we assume that Paul is mistaken about the level of certainty associated with John's knowledge. Compare:

> BASKETBALL. John: Robert is a basketball player.
> Paul: But Robert is not taller than five feet!
> John: Yes.
> Paul: So Robert can't be a basketball player, right?
> John: I did not mean he is not a very short basketball player.

Even if we expect a basketball player to be taller than five feet, that's not a necessary condition for being a basketball player. Similarly, even if Paul expects John's knowledge to be associated with a fairly high degree of certainty (presumably due to the context of John's assertion), it remains the case that knowledge does not entail this degree of certainty, as John rightly notes.

CN.R. S (epistemically) ought to assert that *p* in C only if S satisfies the epistemic standards of epistemic certainty which are operative in C.[20]

With this relativised formulation in mind, we are now in a position to assess the main objection to the proposed view, namely, that it is too demanding to accommodate our current practice of assertion.

3 Pritchard's Worry

The certainty view of assertion does not have many advocates nowadays. It is often thought that this view is too demanding.[21] In particular, following Pritchard (2008), we might worry that the epistemic standards of "certain" cannot be sufficiently low to accommodate our current practice of assertion. More precisely, according to Pritchard (2008, 60), "*prima facie* at any rate, one would expect that the threshold for 'certain' should be fairly high in all contexts." If so, we have some reason to expect that these standards will not be met often enough to warrant most of our assertions. However, while we can agree that the (contextually-influenced) standards of *absolute* certainty should be fairly high in all contexts, we must distinguish certainty and absolute

20 See DeRose (2009, 99) for a similarly relativised knowledge norm. As it is assumed that the degrees of certainty depend on the degrees of justification, we can see CN.R as belonging to the family of gradable justification or warrant norms for assertion (and action), such as those defended by Brown (2010), McKinnon (2015), Locke (2015) or Gerken (2017). However, there are substantial differences between these views and the one proposed here. According to these views, truth is not required and there are contexts in which a warrant or justification insufficient for knowledge can be sufficient for epistemic assertability (or actionability). An apparent advantage of such views is that they can easily handle cases in which the asserted belief is false, or not justified, but in which, intuitively, it would not be reasonable to criticise the assertion (see e.g. Gerken 2017). Yet, these potentially problematic cases can also be handled by proponents of the certainty view if they interpret them as cases where the assertion is epistemically inappropriate but excusable (see, among many others, Kelp and Simion 2017 for a similar move in defence of the knowledge view). In addition, a disadvantage of these alternative views is that they have to explain why the linguistic and conversational data involving "knowledge" are invariant across contexts (see section 5). Finally, I should add that although in this paper I focus on the certainty norm for assertion, I think good arguments can be marshalled in favour of a similar certainty norm for action (see Vollet 2017; Beddor 2020). One such argument might use a commonality assumption between the epistemic norms of assertion and action (see Gerken and Petersen 2020, sect. 3 for a very good overview of this "commonality" issue). Thanks to a anonymous reviewer for raising these issues.

21 See Kvanvig (2009, 143), Turri (2010), Benton (2020), Gerken (2017, 138).

doi: 10.48106/dial.v74.i3

certainty. Further, even if we grant that the threshold for "certain" should be at least as strong as knowledge-level justification in all contexts and stronger at least in some contexts, we need not assume that the epistemic standards for knowledge are particularly high. On some accounts, they are even as weak as a true belief.[22] If the standards for knowledge are that weak, standards stronger than knowledge-level standards need not be that high. Of course, the weaker the knowledge-level standards are supposed to be, the less plausible is the thought that knowledge is (in many or some contexts) sufficient for certainty. But then, the more plausible is the thought that certainty does not demand very high standards. Conversely, the higher the knowledge-level standards are supposed to be, the more plausible is the claim that in many contexts they are sufficient for certainty, so that in many contexts certainty can be reached.

In addition, it's natural to assume that knowledge is the "floor" of certainty, so that in low-standards contexts "S has knowledge-level justification for p if and only if p is certain for S" is true, and that these low-standards contexts are fairly frequent. This assumption explains why we might be naturally inclined to think that knowledge entails certainty. If this assumption is correct, indeed, in many contexts it would be false to say "S knows that p but S is not certain that p ($/p$ is not certain)." We can combine this natural assumption with the consideration that, with regard to many propositions, our epistemic position is stronger than knowledge-level. For many propositions, we have testimonies coming from various sources, repeated visual and tactile experiences, knowledge that we know the relevant proposition, etc.[23] It follows that even if we grant that certainty requires fairly high standards in all contexts, this is not a compelling reason to think that certainty is not often reached.

Pritchard (2008, 60–61) also proposes the following case:

> JENNY. Jenny steps off the train in an unfamiliar town and asks the first person that she meets for directions. This person is indeed knowledgeable about the area and communicates this knowledge to Jenny, who promptly heads off to where she needs to go.

22 See Sartwell (1991, 1992) and DeRose's discussion about the "floor" of "knows" (2009, 13–18).

23 The claim that one's epistemic position is often stronger than knowledge-level justification can be accepted even if one thinks that knowledge-level justification requires evidential probability one (see footnote 7). Importantly, we may think that safety or reliability also matters for the strength of one's epistemic position. Non-sceptical views of knowledge grant that knowledge does not require maximal safety or reliability. For more on this issue, see Brown (2011, 161–162) and Schulz (2017).

According to Pritchard, "it is hard to see why Jenny (or anyone else for that matter) would regard her as being certain of what she believes, whether the certainty in question is of the subjective or epistemic variety." Thus, champions of the certainty view would encounter the following dilemma: either they must say that Jenny herself cannot appropriately flat-out assert the relevant proposition about the direction which, according to Pritchard, would contradict our intuitions and amount to conceding that "far more of our assertions are improper than we typically suppose" (2008, 61); or else they must grant that Jenny's assertion would be appropriate, and the standards of certainty met, which is implausible, for "no-one would surely describe Jenny as certain of what she knows" (2008, 61).

Let me start with the second horn of the dilemma. Why can't we see Jenny or the relevant proposition as certain in this context? Jenny is described as promptly acting on what she has been told, which suggests that she acts unhesitatingly, and therefore that she is in some sense certain in this context. Further, the person who communicates the information to Jenny is supposed to be knowledgeable about the area. On this basis, we may think that what she says is certain. For example, were Jenny to raise a doubt by asking, "But is it certain?", it would seem perfectly appropriate for this person to reply, "Yes it is! I've been living here for 10 years!" All that Pritchard says in favour of the fact that Jenny is uncertain is that she has a low degree of confidence. But "low" is context-sensitive. We may think that, in Jenny's context, this degree of belief counts as sufficiently high for certainty.

Consider the first horn of the dilemma. Suppose we accept that Jenny is uncertain and conclude that she herself cannot appropriately flat-out assert the proposition regarding the direction. First, contrary to what Pritchard suggests, this would not lead us to concede that much, for most of the propositions we assert in ordinary life are more warranted for us than this proposition is for Jenny. Second, it seems to me that if we suppose that Jenny is uncertain, the idea that her flat-out assertion would be inappropriate is rather intuitive, in particular if that supposition is fleshed out in more detail. Suppose, for example, that we say that Jenny is uncertain because she feels doubtful about the truth of the proposition or because she hesitates to act on it. Now, imagine that someone asks Jenny for the direction, and she unhesitatingly flat-out asserts the proposition in question. It seems that we should regard her flat-out assertion as inappropriate. We would expect her to hedge the assertion by saying something like, "This is the right direction, I believe." Alternatively, it would be very natural for her to say, "I was told that this is the right direction."

doi: 10.48106/dial.v74.i3

Lastly, it pays to note that the possible intuition that unconfident Jenny can warrantedly assert the target proposition can be explained by invoking the notion of conditional assertion. When we make assertions, we often speak loosely. We leave aspects of the asserted content implicit. In particular, as Bach (2010, 122–125) notes, many assertions seem to involve an implicit assumption of normality or an implicit *ceteris paribus* clause. For example, if unconfident Jenny asserts "This is the right direction," we may argue that she asserts loosely and merely commits herself to the conditional "Provided things are normal (i.e. if what I've been told is true), this is the right direction." Understood in this way, Jenny's flat-out assertion does not constitute a counterexample to the certainty view of assertion.

Finally, Pritchard (2008, 63) notes that, "it is in fact very easy to get people to concede that they are not certain of something that they believe, even when no additional practical considerations are being raised." If that is true, that is problematic for the certainty view, for the most plausible explanation why people easily concede that they are not certain is precisely that, in fact, they are not certain. Pritchard proposes the following case to support his claim:

> CAR PARK. My wife and I are heading out of the shopping arcade and I stride purposively towards the part of the car park where I believe our car to be. Nothing in my behaviour indicates any doubt on my part on this score and, indeed, I do know that my car is parked at the relevant location. My wife asks me whether I'm certain that it is parked there, whether there is any possibility that I could be wrong.

According to Pritchard (2008, 63–64), "I would be unlikely to say 'yes'." I must say that I do not share Pritchard's intuition here. At least, we should note that it is also far from clear that an ordinary subject (in a normal situation with low stakes) would be likely to say "no." In contrast, it is clear that "Yes, I'm fairly/pretty/reasonably certain that the car is parked there" would be a very natural reply.[24] By using "fairly/pretty/reasonably certain" in this way, it does not seem that the subject retracts, qualifies or hedges his or her assertion (compare: "No, but I think/Probably it's parked there"). Rather, in doing so the subject seems to maintain the assertion, which suggests that he or she thinks he or she satisfies the epistemic norm. If CN.R is true, that is what we

24 Thanks to an anonymous reviewer for this suggestion.

should expect, for mentioning that we are reasonably certain (that is, certain although less than absolutely certain) indicates that we satisfy the certainty norm of assertion.

Perhaps the case is underdescribed. For example, it is unclear with what degree of clarity the subject is supposed to remember where he has parked the car or to what extent he has paid attention when he parked. Undoubtedly, there are cases where we are unsure, for example when we do not remember very well. But in many ordinary cases with low stakes, it appears to me that the most natural reply would be that we have no doubt or that we are certain enough. On this point, I can just encourage the reader to check whether people are willing to insist that they are (fairly) certain of where they have parked their car, what they have eaten at lunch, what job they have, where they live, etc.

4 Linguistic and Conversational Data

Let us now turn to the consideration of conversational and linguistic data. We can show that appealing to a certainty condition allows us to explain all the data put forth in favour of the knowledge account of assertion while also explaining data recalcitrant to this account.[25, 26]

4.1 *Moore's Paradoxical Sentences*

It is common to appeal to Moore's paradoxical sentences to defend one or another view about the epistemic condition required for appropriate assertion. Moore (1942, 543) notes that it sounds incoherent to assert "*p* but I do not believe that *p*." Yet, it is clear that this sentence does not express a semantic

25 If we maintain that knowledge entails certainty and that "know" is context-sensitive, the certainty account will collapse into the knowledge account. The defender of the knowledge norm will then be in a position to explain the data involving "certain" (see Ichikawa 2017, 185–186). However, the claims that knowledge entails certainty and that "know" is context-sensitive are far from trivial and rather controversial.

26 As anticipated in section 1, these data involve indifferently epistemic and subjective certainty, but that is not problematic. As explained, if we assume ECNSC these data can be accounted for either by CN-S or CN-E, and I have offered independent reasons to think that CN-E is superior to CN-S. Also, I assume below that "it is certain that p" expresses epistemic certainty, which might be doubted given the considerations of footnote 4. However, I think that in the present context this expression is naturally understood as meaning, "In my opinion, it is certain that *p*," which I think expresses the assertor's belief that *p* is epistemically certain.

doi: 10.48106/dial.v74.i3

contradiction, for it may well be true that *p* and that the subject who utters this sentence does not believe *p*. So, how are we to explain the fact that asserting this sentence sounds incoherent?

A popular explanation appeals to the norm of assertion. Suppose, for example, that you should not assert what you do not believe. Suppose that you are seen as following this norm. Then, if you assert that *p*, you are seen as believing that *p* (or at least as taking yourself to believe that *p*). In other words, in virtue of the supposed belief norm of assertion, by asserting that *p* you represent yourself as believing that *p*. In the second half of your assertion, though, you say that you do not believe that *p*. Thus, such an assertion sounds incoherent because it represents the assertor as believing and not believing that *p*.

Consider the following sentences:

(1) *p* but I do not believe that *p*
(2) *p* but I do not know that *p*
(3) *p* but I am not certain that *p*
(4) *p* but it is not certain that *p*

Assertions of these sentences sound incoherent.[27]

As Stanley notes, if we embrace the certainty view of assertion, we can easily explain in a unified way why assertions of (1)–(4) sound incoherent. Suppose you follow the certainty norm of assertion. When you assert that *p*, you take *p* to be epistemically certain. Given the plausible bridge principle (ECNSC)

27 See also Williamson (2009), Kvanvig (2009), Turri (2010) and Gerken (2017). Some philosophers might think that (3) does not always sound incoherent. Consider Radford's case (1966): In a quiz, Albert is asked when Queen Elisabeth died. Albert thinks he does not know the answer. Yet, he reliably answers "Elisabeth died in 1603." Albert does not trust his answer, and hence, according to Radford, Albert does not believe that Elisabeth died in 1603. Still, according to Radford, he knows the answer. We might think that this case illustrates a situation where a subject can assert without infelicity "*p* but I am not certain that *p*." If so, assertions of (3) are not always infelicitous. But first, suppose we accept that Albert can felicitously assert (3). Then, it must also be granted that Albert can without incoherence assert (1) and (2), for Albert does not believe that Elisabeth died in 1603, and yet he says "Elisabeth died in 1603"; also, Albert believes that he does not know that Elisabeth died in 1603, and yet he says "Elisabeth died in 1603." If Radford's case shows that asserting (3) is not always incoherent, then it shows that the strategy consisting in appealing to Moore's paradoxical sentences is misguided in the first place. Second, note that it is far from clear that if Albert utters "Elisabeth died in 1603 but I am not certain," he is really asserting the first half of this sentence. Albert is participating in a quiz. As Radford stresses, Albert takes his answer to be a mere guess. So, arguably, Albert's utterance of "Elisabeth died in 1603" is not an assertion, but a guess.

according to which if p is epistemically certain for you, then you should be subjectively certain that p (at least if you consider whether p), if you assert that p. Subjective certainty rightly based on epistemic certainty entails knowledge. So, when you assert that p, you represent p as epistemically certain for you, and you represent yourself as believing with certainty that p, and hence as knowing that p. But when you assert the second half of sentences (1)–(4), you deny a necessary condition for epistemic and/or subjective certainty (namely, you deny that you believe that p, that you know that p, that p is certain, and that you are certain that p). You represent yourself as having incoherent attitudes.

4.2 *Rejection of Two Explanations Consistent with the Knowledge Norm*

Advocates of epistemic conditions weaker than certainty—such as knowledge—must propose a specific explanation for the infelicity of (3) and (4). A first explanation proposed by knowledge normers appeals to a supposed "contextual" connection between the epistemic standards of knowledge and those of certainty. When considering sentences similar to (3) and (4), Williamson (2000, 254) writes:

> What seems to be at work here is a reluctance to allow the contextually set standards for knowledge and certainty to diverge. Many people are not very happy to say things like "She knew that A, but she could not be certain that A."

According to this proposal, we can explain the infelicities of asserting (3) and (4) in the following way. By saying that p, you represent yourself as knowing that p (assumption of the knowledge norm). In virtue of a general reluctance to dissociate the standards of knowledge from those of certainty, we expect you to be reluctant to say "I cannot be certain that p," for this would amount to representing yourself as endorsing

(5) I know that p but I am not certain that p (/it is not certain that p).

Now, if we assume that knowledge is the norm of assertion and that we are reluctant to make assertions like (5), this may explain why we are reluctant to assert (3) or (4).

However, while it can be granted that an assertion of (5) sounds incoherent, Williamson's explanation ultimately relies on the claim that people are reluc-

doi: 10.48106/dial.v74.i3

tant to assert sentences like, "She knew that A, but she could not be certain that A." But Williamson does not explain why people are so reluctant.[28]

The main problem for Williamson's approach, however, is that it is far from clear that we are reluctant to assert such sentences. As Stanley (2008) notes, asserting the following sentence does not sound incoherent:

(6) S knows that p, but being a cautious fellow, she is not certain that p.[29]

We may reinforce this line of thought by noting that assertions of the following sentence do not sound incoherent:

(7) S knows that p but she does not know that she knows that p. That's why she is not certain that p.[30]

28 Williamson's proposal seems to be that in many (if not most) contexts, the context-sensitive epistemic standards of "certain" and the invariant epistemic standards of knowledge are identical. In these contexts, we can explain the infelicity of (3) and (4) with the knowledge norm, for in such contexts if you do not satisfy the epistemic standards of certainty (in the relevant sense) you do not satisfy the epistemic standards of knowledge. Williamson adds that in contexts in which "certain" is used with higher standards, like when we use "absolutely certain"—so that the epistemic standards of knowledge and certainty diverge—"assertability goes with knowledge, not with the highest possible standards of certainty" (2000, 254). I agree, but as Turri (2010, 458–459) notes, when we use "certain" we generally invoke "ordinary standards of certainty and assuredness, not the highest possible standards." Therefore, since these standards can be stronger than those required by knowledge, although weaker than standards of absolute certainty, Williamson's solution "does not speak to this problem" (Turri 2010, 459).

29 Of course, we are reluctant to assert the epistemic version of (6):

(6*) S knows that p, but since she is a cautious fellow, it is not certain that p for S.

This can be explained by the fact that the epistemic certainty of a proposition relative to a subject does not depend on whether this subject is cautious but on his evidence.

30 An anonymous referee pointed out to me that (7) sounds unnatural as an utterance of ordinary language. I agree. The point is just that this sentence does not sound incoherent, not that we commonly say that kind of things. This referee also notes that the following sentences, involving epistemic certainty, do not sound particularly good, with (7**) sounding particularly bad:

(7*) S knows that p but she does not know that she knows that p. That's why it is not certain (for her) that p.

(7**) We knew that p but we did not know that we knew that p. That's why it was not certain for us that p.

Here too, I agree that we do not often say this, but I feel no contradiction in asserting these sentences. Perhaps modifying these sentences with "certain enough" would make them more acceptable:

If we are not always reluctant to make third-person "knowledge" ascriptions while denying third-person certainty, it is hard to see how Williamson's approach can explain the infelicity of (3), (4) and (5).

A second possible explanation of the infelicity of (3), (4) and (5) consistent with the knowledge norm for assertion could appeal to the distinction between warranted assertion and knowledge that the assertion is warranted, and to the claim that certainty is necessary for knowledge of knowledge.

The distinction between warranted assertion and knowledge that the assertion is warranted is sometimes used by proponents of the knowledge norm to explain why assertions of the following sentence sound infelicitous:

(8) *p* but I do not know whether I know that *p*.

According to proponents of the knowledge norm, indeed, knowledge that one knows is not required for epistemically appropriate assertion. So, as Sosa (2009) emphasizes, if they are right, why is it that asserting (8) sounds infelicitous?

In reply, advocates of the knowledge norm sometimes appeal to the thought that there is something bad in doing something appropriate without knowing that it is appropriate; and they say that an assertion of (8) sounds infelicitous because the subject represents herself as not knowing that the assertion is warranted (i.e. known), which is somehow wrong (see Benton 2013).

Now, suppose we understand certainty in terms of knowledge that one knows, as some writers suggest (e.g. Turri 2010, 459). The kind of explanation offered for the infelicity of (8) can be used by the knowledge normer to explain the infelicity of (3), (4) and (5). Someone asserting "*p* but I am not certain that *p* (/it is not certain that *p*)" would represent herself as knowing that *p* (in virtue of the supposed knowledge norm of assertion), but as not knowing that she knows that *p*. Hence, she would represent herself as not knowing

(7***) S knows that *p* but she does not know that she knows that *p*. That's why it is not certain enough (for her) that *p*.

(7****) We knew that *p* but we did not know that we knew that *p*. That's why it was not certain enough (for us) that *p*.

Consider also:

(7*****) We began to learn that *p*, but it wasn't certain enough yet.

The fact that it's not incoherent to use the expression "it is not certain enough that *p*" alongside "know that *p*" in this way shows that we allow the epistemic standards for knowledge and those for certainty (in the relevant sense of "certainty") to diverge.

doi: 10.48106/dial.v74.i3

that the assertion is warranted. And someone asserting "I know that *p* but I cannot be certain that *p*" would represent herself as being certain that she knows that *p* (in virtue of the knowledge norm of assertion), but as not being certain that *p*. Hence the infelicity.

While interesting, there are several reasons to think that this explanation is ultimately unsuccessful. A first point to note is that it is far from clear that the problem with someone who asserts "*p*" while she cannot be certain that *p* has to do with the fact that she cannot know that she knows that *p*. Indeed, asserting (9) seems equally infelicitous:

(9) *p*, but it is not certain that I know that *p*.

Yet it may well be true that the subject knows that she knows that *p*, without knowing that she knows that she knows that *p*. But according to the proposal under examination, not knowing that one knows that one knows that *p* (or, in other words, not knowing that *p* is certain, or not being certain that one knows that *p*) should make no difference for an appropriate assertion that p, since what is required is merely warrant (i.e. knowledge) and knowledge that one has warrant (i.e. knowledge of knowledge that *p* or certainty that *p*).

Secondly, it is unclear that certainty is always necessary for knowing that one knows. Indeed, it seems that you can know that you know that *p* even if *p* is still uncertain. Suppose it is still uncertain that you know that you know that *p*. Your epistemic position with respect to *p* could be better and, if so, we may think that *p* is not certain. More generally, if we think that knowledge does not entail certainty, it's clear that you can know that you know that *p* even if *p* is not certain.

Thirdly, the proposal under examination crucially relies on the assumption that, in some sense, it is always bad to do something for which one has warrant if one does not know that one has warrant for doing it. Suppose that this assumption is true. It is then very plausible to think that the strength of the (alleged) requirement to know that one has warrant varies with the importance of being warranted.[31] But if so, in contexts in which it is not at all important to make a warranted assertion, it should be possible to assert without obvious infelicity (3), (4), (5) and (8). Yet, assertions of these sentences *always* sound infelicitous.

In addition, it is hard to see why such a requirement should always be in force, in particular when it comes to *epistemic* warrant. Suppose that our

31 Williamson (2005) suggests something along these lines.

epistemic position with respect to the target proposition is good enough for assertion, given the norm of assertion. Suppose that, for whatever general reason, a further relevant second-level epistemic position with respect to the first-level epistemic warrant is always required. Now, presumably, the same kind of reason should lead us to think that a relevant third-level epistemic position with respect to the second-level epistemic position is also always required. And so on. This obviously leads to an infinite regress. Since we have to stop somewhere, it is natural to stop at the first-level.

Fourthly, if the approach under examination is correct, it undermines a fundamental argument for the knowledge account. This argument is based on the fact that it is appropriate to challenge an assertion by using "know" (I consider this argument below). But if the present approach is correct, an advocate of a norm of truth, for example, could argue that when we require knowledge of the asserted proposition, we require knowledge that the assertion is warranted, i.e. we require that it is known that the assertion is true.[32]

If these considerations are correct, there is little hope for the proponent of the knowledge norm in appealing to a distinction between warrant and knowledge of warrant. We cannot exclude that another proposal could be offered by the advocate of the knowledge norm or some other weaker condition. However, let me point out that, in contrast to rival views, the certainty view provides a unified and very straightforward explanation of the infelicity of asserting Moorean sentences.[33]

4.3 *Appropriate Challenges*

In support of the knowledge account, Williamson (2000, 252–253) emphasizes that it is natural to challenge an assertion by asking "How do you know that *p*?" or "Do you know that *p*?" (see also Unger 1975, 263–264). Of course, these data suggest that knowledge is necessary, but they do not show that knowledge is sufficient. Therefore, they do not speak against the certainty

32 Similarly, Weiner (2005, 235–236) argues that truth is the epistemic condition for proper assertions but that secondary property requires a reasonable belief that one satisfies this condition, i.e. a reasonable belief that the proposition is true. Pagin (2015, 19) considers the possibility of reinforcing the notion of secondary property by appealing to knowledge instead of a reasonable belief.

33 A possible objection here is that the certainty view does not seem to be able to explain the infelicity of asserting "*p* but it is not certain that *p* is certain." However, this infelicity can be explained by the assumption that in mentioning certainty of certainty in the second half of the assertion, the assertor raises the epistemic standards for appropriate assertion to higher-order certainty.

doi: 10.48106/dial.v74.i3

view, for according to this view being in a position to know is necessary for certainty, and if you are in a position to know that *p* and you consider whether *p*, you should know that *p*. In addition, it's crucial to note that we can challenge an assertion by invoking certainty (e.g. "Are your sure?") (see Stanley 2008, 51). If knowledge is the norm of assertion, this is a surprising fact.

An explanation consistent with the knowledge account, relying on the claims that we ought to know that we have warrant and that certainty is necessary for knowledge of knowledge, is proposed by Turri (2010). However, to repeat, it's far from clear that certainty that *p* is necessary for knowledge that one knows that *p*. Second, it is hard to see why a requirement that one knows that one has epistemic warrant should always be in force. Third, this strategy undermines the argument for the knowledge norm based on the appropriateness of knowledge-based challenges.

Another suggestion would be that we must distinguish assertions and guarantees. This would explain why "certain" is used in some challenges. However, it seems that any kind of assertion can be properly challenged by using "certain" or "sure." Further, asserting that *p* clearly seems to be a way of guaranteeing that *p*. It would be very odd to say, "*p* but I do not guarantee that *p* is true." Finally, if we still want to distinguish assertions and guarantees, given the distinction between certainty and absolute certainty, it seems more natural to associate guarantees with absolute certainty.

4.4 *Unified Explanation of Moorean Sentences and Appropriate Challenges*

Benton (2011) points out that a satisfying account of assertion should be able to explain in a unified way the Moorean data and the appropriateness of challenges. Indeed, appropriate challenges "can elicit a *de facto* Moorean paradox within a conversational context" (2011, 686). Consider:

A: It is snowing.
B: How do you know?
A: Oh, I don't.
B: Huh?
A: Still, it's snowing.

Benton (2011, 686) argues that this favours the knowledge account because this account explains the relation between Moorean data and challenges in terms of knowledge.

However, the certainty view of assertion fares equally well in explaining why challenges in terms of knowledge may elicit de facto Moorean paradoxes. According to this view, the same thing explains why utterances of Moorean paradoxical sentences sound paradoxical and why we can challenge an assertion by using "know": a necessary condition for an appropriate assertion that you know that *p* is that you are in a position to know that you are in a position to know that *p*, and hence, by factivity of knowledge, that you are in a position to know that *p*. If you are in a position to know that you know that *p* and you consider whether you know that *p*, you should know that you know that *p*, and by factivity of knowledge you should know that *p*.

Further, it is not difficult to imagine a conversation with "certain" eliciting a de facto Moorean paradox in terms of certainty:

A: It is snowing.
B: Is it certain/Are you sure?
A: Oh, no.
B: Huh?
A: Still, it's snowing.

The certainty view has a simple and unified explanation for all these data, by appealing to the epistemic certainty norm for assertion (CN-E) and the epistemic certainty norm for subjective certainty (ECNSC), whereas the knowledge account does not.

4.5 *Parenthetical Uses*

Other linguistic data which have been produced have to do with parenthetical uses. First, consider parenthetical uses expressing the assertor's mental state. When we want to express a mere belief in what we assert, we can use "believe" in a parenthetical position:

(10) It is, I believe, raining.
(11) It is raining, I believe.

As Benton (2011) notes, it is striking that we cannot (or do not) use "know" parenthetically in the same way. Consider:

doi: 10.48106/dial.v74.i3

(12) It is, I know, raining
(13) It is raining, I know.

This use of "know" in parenthetical position seems redundant. Benton argues that this constitutes a further argument in favour of the knowledge account. Indeed, on this account, in asserting that *p* you already represent yourself as knowing that *p*. But if so, it is redundant to parenthetically use "know" to express your knowledge that *p* when you are already asserting that *p*.

However, Benton fails to note that, like "know," "certain" cannot be (or is not) used in this way without redundancy. Consider:

(14) It is, it's certain, raining
(15) It is raining, it's certain.

The certainty view of assertion can explain why these parenthetical uses of "certain" *and* "know" are redundant. The knowledge account of assertion merely explains why this use of "know" is redundant.

A possible worry is that one might think that if certainty is the norm of assertion, then, as knowledge is typically weaker than certainty, the parenthetical use of "know" should have the same effect as the parenthetical use of "believe" in hedging the assertion. However, the use of "know" in parenthetical position does not hedge the assertion. As Blaauw (2012) notes, it can even have a reinforcing effect. Consider the following case (see 2012, 106):

> LAZY. John is having a fight with his wife Jill. Apparently, as Jill brings up repeatedly during their heated conversation, John is very lazy; a point that Jill supports with ample evidence. At one point, exasperated, John asserts,

(16) I am very lazy, I know!

What John says sounds natural, and he could also have said:

(17) I am, I know, very lazy!

Of course, it is striking that the use of "believe" in parenthetical position does not have this reinforcing effect, but, rather, typically hedges the assertion. Blaauw takes these considerations to favour the knowledge account. If you already express your knowledge that *p* in asserting that *p*, then by saying that you know that *p*, you can reinforce what is already expressed. We might think

that if certainty, rather than knowledge, is the epistemic norm of assertion, it is unclear why the parenthetical use of "know" can have this reinforcing effect whereas the parenthetical use of "believe" has a diminishing effect.

There is a reply, though. By asserting that you believe that *p*, you represent as certain the proposition that you believe that *p*, but you do not represent as certain the proposition that *p*. Thus, you do not represent yourself as satisfying the epistemic norm for asserting *p*, which is why the parenthetical use of "believe" can hedge the assertion. In contrast, by asserting that you know that *p*, you represent as certain the proposition that you know that *p*, and by factivity of knowledge, you represent as certain the proposition that *p*. As a result, you represent yourself as satisfying the epistemic norm with respect to *p*. This is why the parenthetical use of "know" does not hedge the assertion.

In sum, the set of data having to do with parenthetical uses are more easily explained by the certainty view than by the knowledge account.

4.6 *Responses to Prompts to Assert*

Benton (2020) notes that "a standard response when one feels not well-positioned to assert, in reply to a prompt like 'Is it the case that *p*?', is to answer 'I don't know'." According to him, this speaks in favour of the knowledge account. Indeed, "the query was about *p*, not about whether one knows that *p*," and thus the fact that it is appropriate to answer "I do not know that *p*" has to be explained. The knowledge account explains why this answer is appropriate: it is appropriate because it is appropriate to decline a prompt to assert by saying that one does not meet the epistemic norm for assertion.

Still, this line of argument also favours the certainty view of assertion. Suppose that *p* is not certain. A perfectly correct response to "*p*?" is "I'm not sure." Further, note that the certainty view explains why "I do not know" is always an appropriate answer, because being in a position to know is a necessary condition for epistemic certainty. In contrast, the knowledge account cannot explain why "I'm not certain" is always an appropriate answer, since, according to this view, certainty is not required for assertion. As a consequence, considerations concerning typical ways of declining prompts to assert favour the certainty view over the knowledge account.

doi: 10.48106/dial.v74.i3

5 Two Principled Considerations

The certainty view of assertion appears to provide a straightforward and unified explanation of all the main linguistic and conversational data. To reinforce the case for this account, I shall now present two arguments based on principled considerations.

The first principled consideration is the following. It is hard to deny that the warrant required for appropriate assertion varies with the context (e.g. with the audience). Few would deny, for example, that in normal circumstances you can assert that *p* on the basis of a mere testimony that *p*, whereas when it comes to testify that *p* before a court you should have first-hand knowledge. But assuming this variability, it is difficult to maintain the knowledge norm for assertion, while explaining the data involving "know," without endorsing a shifty semantics about "know."[34] Indeed, these data are invariant in the sense that, in all contexts, it's infelicitous to say "*p* and I do not know that *p*"; in all contexts we can challenge an assertion by asking "How do you know that *p*?"; etc. Yet, as explained above, from a linguistic perspective it's not prima facie plausible to think that "know" is context-sensitive. If we can propose an account compatible with an invariantist semantics for "know," capable of explaining the invariant character of the data involving "know" and compatible with the claim that the epistemic warrant for assertion is variable, this account will certainly have an advantage. Therefore, it is an advantage of the certainty view that it respects the three following assumptions:

A. The epistemic standards that we must satisfy for appropriate assertions shift with the context.

B. The linguistic and conversational data (infelicity of asserting Moorean sentences, appropriate challenges, etc.) about "certain," "know" and "believe" are invariant across contexts.

C. The epistemic standards of "know" are invariant across contexts.

To illustrate, consider the knowledge account and the warrant account to see how they fail to respect these three assumptions. Assuming the knowledge

34 See Benton (2020). DeRose (2009) proposes an argument in favour of contextualism about "know" from the claims that the warrant for assertion varies across contexts and that knowledge is the norm of assertion.

norm, we can (partially) explain B. But if we embrace C we must reject A, for we assume that the epistemic standard of appropriate assertion is knowledge and that the epistemic standard of knowledge is invariant across contexts. If we accept A instead, we must then reject C. Suppose now that we adopt Gerken's warrant account, according to which the warrant for epistemically appropriate assertion shifts with the context, in the sense that the warrant is sometimes weaker and sometimes stronger than knowledge-level warrant (see Gerken 2017). We can accept A and C, but it is unclear how to account for B. For example, in situations in which the warrant is supposed to be weaker than knowledge, we cannot appeal to the supposed norm to explain why it sounds infelicitous to assert "*p* but I do not know that *p*."[35]

In contrast, consider the certainty view. We can adopt A, for the epistemic standard of certainty is taken to shift with the context. What degree of justification counts as good enough for certainty is partially influenced by the context. We can also respect B: the norm of assertion always requires at least knowledge-level justification and, given ECNSC, if you assert that *p*, you consider whether *p*, and you should be subjectively certain that *p*. Therefore, you should know that *p*. Finally, the certainty view respects the claim that the epistemic standards of knowledge are invariant across contexts.

The fact that claims A, B and C, which are independently plausible, are fully compatible with the certainty view, whereas together they raise problems for rival weaker accounts, provides a further argument for the certainty view. Of course, this argument is limited, for the claim that "certain" is context-sensitive could be disputed. However, given the considerations developed in section 2, it appears that it is highly plausible to think that "certain" is context-sensitive. In this regard, "certain" strongly differs from "know."

The second principled consideration is that many philosophers find it plausible that knowledge does not entail (epistemic) certainty or the highest grade of justification. But if that is correct, first, we have to explain why asserting (5) sounds infelicitous:

(5) I know that *p* but I am not certain that *p* (/it is not certain that *p*).

35 To explain this infelicity, Gerken might want to appeal to considerations similar to those he advances to explain why, although knowledge is (according to him) not the norm of action, "know" is prominently used in epistemic assessments of action (see 2017). For general criticisms of this strategy, see Vollet (2018).

doi: 10.48106/dial.v74.i3

As Stanley (2008) emphasizes, the certainty view provides a pragmatic explanation of the infelicity. Again, given the certainty norm of assertion, by asserting that you know that *p*, you represent yourself as being certain that you know that *p*, and by the factivity of knowledge, as being certain that *p*. But this contradicts the second half of your assertion.[36]

Second, if knowledge does not entail epistemic certainty or the highest grade of justification, it is obscure why knowledge is bound to always be sufficient for appropriate assertion. If knowledge does not require certainty or the highest grade of justification then, in principle, a situation can arise where the difference between knowledge and certainty could matter for appropriate assertion. In the absence of reasons to think that knowledge is bound to always be sufficient, it is more natural to think that certainty, rather than knowledge, is always sufficient for assertion.[37]

In sum, it is plausible that the degree of justification required for warranted assertions shifts across contexts and that knowledge does not entail certainty. These two claims fit nicely with the certainty view of assertion whereas they are in tension with (many) rival weaker accounts.

6 Conclusion

The claim that certainty is the norm of assertion is often dismissed as implausible. In this paper, I've responded to the main objection that this view is too strong. I've also considered the main conversational and linguistic data advanced in the recent literature, and I have shown that the certainty view provides us with a straightforward and unified explanation of them. I have also argued that this account fits nicely with the plausible claims that the degree of justification required for appropriate assertion is variable and that knowledge does not entail certainty. I thereby hope to have shown that, on more careful reflection, the certainty account of assertion is a viable and respectable account.*

36 That this explanation cannot be used for the third-person cases is not particularly problematic given that these third-person cases do not seem similarly infelicitous. See the discussion of sentences (6) and (7) above.

37 See Brown (2011) for a more developed argument that invariantist and non-sceptical views of knowledge, even infallibilist ones, give us no reason to expect knowledge to be always sufficient for action/assertion.

* I would like to thank P. Engel, M. Gerken, D. Fassio, A. Logins, F. Teroni, J. Brown, C. Kelp, M. Simion, M. Smith, M. Schulz, P. Rich, J. Koscholke, R. Heil, S. Spatan, audiences at the University of Edinburgh, the University of St Andrews, the University of Besançon, the EENM 2016, the

Jacques-Henri Vollet
0000-0003-2005-0091
University of Geneva
jacquesvollet@yahoo.fr

References

AYER, Alfred Jules. 1973. "Wittgenstein on Certainty." in *Understanding Wittgenstein*, edited by Godfrey N. A. VESEY, pp. 226–245. Royal Institute of Philosophy Lectures n. 7. London: MacMillan Publishing Co. Reprinted in Ayer (1984, 120–140), doi:10.1007/978-1-349-15546-0_14.

—. 1984. *Freedom and Morality and Other Essays*. Oxford: Oxford University Press.

BACH, Kent. 2005. "The Emperor's New 'Knows'." in *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, edited by Gerhard PREYER and Georg PETER, pp. 51–90. New York: Oxford University Press, doi:10.1093/oso/9780199267408.003.0003.

—. 2010. "Knowledge In and Out of Context." in *Knowledge and Skepticism*, edited by Joseph Keim CAMPBELL, Michael O'ROURKE, and Harry S. SILVERSTEIN, pp. 105–136. Topics in Contemporary Philosophy n. 5. Cambridge, Massachusetts: The MIT Press, doi:10.7551/mitpress/9780262014083.003.0006.

BEDDOR, Robert [Bob]. 2016. "Reduction in Epistemology." PhD dissertation, New Brunswick, New Haven: Philosophy Department, Rutgers University, doi:10.7282/T3ZW1P1V.

—. 2020. "Certainty in Action." *The Philosophical Quarterly* 70(281): 711–737, doi:10.1093/pq/pqaa006.

BENTON, Matthew A. 2011. "Two More for the Knowledge Account of Assertion." *Analysis* 71(4): 684–687, doi:10.1093/analys/anr085.

—. 2013. "Dubious Objections from Iterated Conjunctions." *Philosophical Studies* 162(2): 355–358, doi:10.1007/s11098-011-9769-3.

—. 2020. "Epistemological Aspects of Hope." in *The Moral Psychology of Hope*, edited by Claudia BLÖSER and Titus STAHL, pp. 135–151. Lanham, Maryland: Rowman & Littlefield, doi:10.5771/9781786609731.

BLAAUW, Martijn. 2012. "Reinforcing the Knowledge Account of Assertion." *Analysis* 72(1): 105–108, doi:10.1093/analys/anr124.

BLOME-TILLMANN, Michael. 2014. *Knowledge and Presuppositions*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199686087.001.0001.

Collège de France, the University of Trieste, and several anonymous referees for this journal for very helpful discussions or comments on earlier versions of this paper. The work on this paper was supported by the Swiss National Science Foundation grants number 178039 and 169293.

BROOME, John A. 2013. *Rationality Through Reasoning*. Hoboken, New Jersey: John Wiley and Sons, Inc., doi:10.1002/9781118609088.

BROWN, Jessica A. 2010. "Knowledge and Assertion." *Philosophy and Phenomenological Research* 81(3): 397–405, doi:10.1111/j.1933-1592.2010.00369.x.

—. 2011. "Fallibilism and the Knowledge Norm for Assertion and Practical Reasoning." in *Assertion: New Philosophical Essays*, edited by Jessica A. BROWN and Herman CAPPELEN, pp. 153–174. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199573004.003.0007.

—. 2018. *Fallibilism: Evidence and Knowledge*. Oxford: Oxford University Press, doi:10.1093/oso/9780198801771.001.0001.

COHEN, Stewart. 1999. "Contextualism, Skepticism, and the Structure of Reasons." in *Philosophical Perspectives 13: Epistemology*, edited by James E. TOMBERLIN, pp. 57–89. Oxford: Blackwell Publishers, doi:10.1111/0029-4624.33.s13.3.

CRUSE, David Alan. 1980. "Antonyms and Gradable Complementaries." in *Perspektiven der lexikalischen Semantik. Beiträge zum Wuppertaler Semantikkolloquium vom 2.-3. Dezember 1977*, edited by Dieter KASTOVSKY, pp. 14–25. Bonn: Bouvier Verlag Herbert Grundmann.

DEROSE, Keith. 2009. *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. I*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199564460.001.0001.

DOUVEN, Igor. 2006. "Assertion, Knowledge, and Rational Credibility." *The Philosophical Review* 115(4): 449–485, doi:10.1215/00318108-2006-010.

ENGEL, Pascal. 2008. "In What Sense is Knowledge the Norm of Assertion?" *Grazer Philosophische Studien* 77: 45–59, doi:10.1163/18756735-90000843.

FASSIO, Davide. 2017. "Is there an Epistemic Norm of Practical Reasoning?" *Philosophical Studies* 174(9): 2137–2166, doi:10.1007/s11098-016-0792-2.

FIRTH, Roderick. 1967. "The Anatomy of Certainty." *The Philosophical Review* 76(1): 3–27, doi:10.2307/2182963.

GAO, Jie. 2019. "Against the Iterated Knowledge Account of High-Stakes Cases." *Episteme* 16(1): 92–107, doi:10.1017/epi.2017.30.

GERKEN, Mikkel. 2017. *On Folk Epistemology: How We Think and Talk about Knowledge*. Oxford: Oxford University Press, doi:10.1093/oso/9780198803454.001.0001.

GERKEN, Mikkel and PETERSEN, Esben Nedenskov. 2020. "Epistemic Norms of Assertion and Action." in *The Oxford Handbook of Assertion*, edited by Sanford C. GOLDBERG, pp. 683–706. Oxford Handbooks. Oxford: Oxford University Press, doi:10.1093/oxfordhb/9780190675233.013.49.

HUEMER, Michael. 2007. "Epistemic Possibility." *Synthese* 156(1): 119–142, doi:10.1007/s11229-005-4782-8.

ICHIKAWA, Jonathan Jenkins. 2017. *Contextualising Knowledge. Epistemology and Semantics*. New York: Oxford University Press, doi:10.1093/oso/9780199682706.001.0001.

1913 KELP, Christoph and SIMION, Mona. 2017. "Criticism and Blame in Action and Assertion." *The Journal of Philosophy* 114(2): 76–93, doi:10.5840/jphil201711426.
1915 KENNEDY, Christopher and MCNALLY, Louise. 2005. "Scale Structure, Degree Modification, and the Semantics of Gradable Predicates." *Language* 81(2): 345–381, doi:10.1353/lan.2005.0071.
1918 KLEIN, Peter D. 1981. *Certainty: A Refutation of Skepticism*. Minneapolis, Minnesota: University of Minnesota Press.
1920 —. 1998. "Certainty." in *The Routledge Encyclopedia of Philosophy*, edited by Edward J. CRAIG. London: Routledge. The Routledge Encyclopedia was made available online in 2002 and is now regularly updated, doi:10.4324/9780415249126-P005-1.
1923 KVANVIG, Jonathan L. 2009. "Assertion, Knowledge, and Lotteries." in *Williamson on Knowledge*, edited by Patrick GREENOUGH and Duncan PRITCHARD, pp. 140–160. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199287512.003.0010.
1926 LACKEY, Jennifer. 2007. "Norms of Assertion." *Noûs* 41(4): 594–626, doi:10.1111/j.1468-0068.2007.00664.x.
1928 LEVIN, Janet. 2008. "Assertion, Practical Reason, and Pragmatic Theories of Knowledge." *Philosophy and Phenomenological Research* 76(2): 359–384, doi:10.1111/j.1933-1592.2007.00136.x.
1931 LEWIS, David. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8(3): 339–359. Reprinted in Lewis (1983, 233–249), doi:10.1007/BF00258436.
1933 —. 1983. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.
1935 LOCKE, Dustin. 2015. "Practical Certainty." *Philosophy and Phenomenological Research* 90(1): 72–95, doi:10.1111/phpr.12036.
1937 MCKINNON, Rachel. 2015. *The Norms of Assertion: Truth, Lies, and Warrant*. Innovations in Philosophy. London: Palgrave Macmillan, doi:10.1057/9781137521729_1.
1939 MCNALLY, Louise. 2011. "The Relative Role of Property Type and Scale Structure in Explaining the Behavior of Gradable Adjectives." in *Vagueness in Communication. International Workshop, ViC 2009, held as part of ESSLLI 2009, Bordeaux, France, July 2009. Revised Selected Papers*, edited by Rick W. F. NOUWEN, Robert van ROOIJ, Uli SAUERLAND, and Hans-Christian SCHMITZ, pp. 151–168. Lecture Notes in Artificial Intelligence n. 6517. Berlin: Springer, doi:10.1007/978-3-642-18446-8_9.
1945 MILLER, Richard W. 1978. "Absolute Certainty." *Mind* 87(345): 46–65, doi:10.1093/mind/lxxxvii.1.46.
1947 MOORE, George Edward. 1942. "Reply to My Critics." in *The Philosophy of G.E. Moore*, edited by Paul Arthur SCHILPP, pp. 533–688. The Library of Living Philosophers n. 4. LaSalle, Illinois: Open Court Publishing Co.
1950 PAGIN, Peter. 2015. "Problems with Norms of Assertion." *Philosophy and Phenomenological Research* 93(1): 178–207, doi:10.1111/phpr.12209.
1952 PETERSEN, Esben Nedenskov. 2019. "A Case for a Certainty Norm of Assertion." *Synthese* 196(11): 4691–4710, doi:10.1007/s11229-018-1682-2.

1954 POPPER, Karl Raimund. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford:
1955 Oxford University Press.
1956 PRITCHARD, Duncan. 2008. "Certainty and Skepticism." in *Philosophical Issues 18:*
1957 *Interdisciplinary Core Philosophy*, edited by Ernest SOSA and Enrique VILLANUEVA,
1958 pp. 58–67. Malden, Massachusetts: Wiley-Blackwell, doi:10.1111/j.1533-6077.2008.
1959 00137.x.
1960 RADFORD, Colin. 1966. "Knowledge – by Examples." *Analysis* 27(1): 1–11, doi:10.230
1961 7/3326979.
1962 SARTWELL, Crispin. 1991. "Knowledge is Merely True Belief." *American Philosophical*
1963 *Quarterly* 28(2): 157–165.
1964 —. 1992. "Why Knowledge is Merely True Belief." *The Journal of Philosophy* 89(4):
1965 167–180, doi:10.2307/2026639.
1966 SCHIFFER, Stephen. 1996. "Contextualist Solutions to Skepticism." *Proceedings of the*
1967 *Aristotelian Society* 96: 317–333, doi:10.1093/aristotelian/96.1.317.
1968 SCHULZ, Moritz. 2017. "Decisions and Higher-Order Knowledge." *Noûs* 51(3): 463–483,
1969 doi:10.1111/nous.12097.
1970 SOSA, David. 2009. "Dubious Assertions." *Philosophical Studies* 146(2): 269–272, doi:10
1971 .1007/s11098-008-9255-8.
1972 STANLEY, Jason. 2005. *Knowledge and Practical Interests*. Lines of Thought. Oxford:
1973 Oxford University Press, doi:10.1093/0199288038.001.0001.
1974 —. 2008. "Knowledge and Certainty." in *Philosophical Issues 18: Interdisciplinary Core*
1975 *Philosophy*, edited by Ernest SOSA and Enrique VILLANUEVA, pp. 35–57. Malden,
1976 Massachusetts: Wiley-Blackwell, doi:10.1111/j.1533-6077.2008.00136.x.
1977 TURRI, John. 2010. "Prompting Challenges." *Analysis* 70(3): 456–462, doi:10.1093/an
1978 alys/anq027.
1979 UNGER, Peter. 1975. *Ignorance: A Case for Scepticism*. Oxford: Oxford University Press,
1980 doi:10.1093/0198244177.001.0001.
1981 VOLLET, Jacques-Henri. 2017. "Knowledge, Certainty and Practical Factors." Thèse de
1982 doctorat, Genève: Département de philosophie, Université de Genève, doi:10.130
1983 97/archive-ouverte/unige:95005.
1984 —. 2018. "The Warrant Account and the Prominence of 'Know'." *Logos & Episteme*
1985 9(4): 467–483, doi:10.5840/logos-episteme20189436.
1986 —. 2020. "Certainty and Assertion." *Dialectica* 74(3): 475–507, doi:10.48106/dial.v74.
1987 i3.02.
1988 WEDGWOOD, Ralph. 2012. "Outright Belief." *Dialectica* 66(3): 309–329, doi:10.1111/j.
1989 1746-8361.2012.01305.x.
1990 WEINER, Matthew. 2005. "Must We Know What We Say?" *The Philosophical Review*
1991 114(2): 227–251, doi:10.1215/00318108-114-2-227.
1992 WHITING, Daniel. 2013. "Stick to the Facts: On the Norms of Assertion." *Erkenntnis*
1993 78(4): 847–867, doi:10.1007/s10670-012-9383-6.

1994 WILLIAMSON, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University
1995 Press, doi:10.1093/019925656X.001.0001.
1996 —. 2005. "Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowl-
1997 edge." *The Philosophical Quarterly* 55(219): 213–235, doi:10.1111/j.0031-8094.2005.
1998 00396.x.
1999 —. 2009. "Reply to Critics." in *Williamson on Knowledge*, edited by Patrick GREENOUGH
2000 and Duncan PRITCHARD, pp. 279–384. Oxford: Oxford University Press, doi:10.1
2001 093/acprof:oso/9780199287512.003.0017.

Reliable Knowledge

A Reply to Turri

Jonathan Dixon

Recently John Turri (**?**) has argued, contra the orthodoxy amongst epistemologists, that reliability is not a necessary condition for knowledge. From this result, Turri (**?**, **?**, **?**, **?**) defends a new account of knowledge—called abilism—that allows for unreliable knowledge. I argue that Turri's arguments fail to establish that unreliable knowledge is possible and argue that Turri's account of knowledge is false because reliability must be a necessary condition for knowledge.

Many epistemologists agree that knowledge must be reliably produced. For example, Goldman holds that justification is necessary for knowledge and that justification "is a function of the reliability of the process or processes that cause it" (1979, 345); Sosa holds that knowledge is produced by a disposition "that would in appropriately normal circumstances ensure (or make very likely) the success of any relevant performance issued by it" (2007, 29); and Williamson claims that "no reason has emerged to doubt the intuitive claim that reliability is necessary for knowledge" (2000, 100).[1] Recently John Turri (2015a) argued against this orthodoxy by providing two theoretical arguments for the possibility of unreliably produced knowledge. If either of Turri's arguments is sound then all accounts of knowledge that require reliability are false and most epistemologists have been on the wrong track in understanding the nature of knowledge. Realizing this, Turri (2015b, 2017, 2016a, 2019) defends a new account of knowledge, called *abilism,* which allows for knowledge to be unreliably produced.

1 See Turri (2015a, footnote 1) and Goldman and Beddor (2021) for a more complete list of epistemologists who think that knowledge must be reliably produced. However, there are a few contemporary philosophers who indirectly deny this claim. Sartwell (1991, 1992) argues that true belief alone is sufficient for knowledge; and because Hetherington (1998, 1999, 2016) argues that victims of Gettier-cases do possess the relevant knowledge, *a fortiori*, he holds that lucky (and so unreliable) processes can produce knowledge.

After providing some background and clarifying terms in § 1, in § 2 and § 3 I explain why each of Turri's (2015a) theoretical arguments for unreliable knowledge fail. And I conclude in § 4 with reasons why abilism is false and why reliability must be a necessary condition for knowledge.

1 Background and Clarifying Terms

Turri's (2015a) theoretical arguments for unreliable knowledge rely on what is called an *achievement account of knowledge*. This is roughly the family of views which hold that an agent S has knowledge of P just in case S's true belief in P manifests S's cognitive achievement.[2] While there are many ways of spelling-out the details of this account of knowledge and there are many challenges to this family of views,[3] I will set these issues aside and grant for the sake of argument that knowledge is a kind of cognitive achievement. My arguments below show that *even if* we grant this, both of Turri's (2015a) arguments for the possibility of unreliable knowledge fail.

The next thing I should explain is what Turri means by "reliability" and "achievement." Turri's definition of "reliability" is in line with how it is standardly understood: a process, disposition, or ability is (epistemically) reliable when and only when (significantly) more than half of its produced beliefs are true; and a process, disposition, or ability is (epistemically) *un*reliable when and only when less than half of its produced beliefs are true (2015a, 530).[4] While Turri (2015a) does not provide a definition of "achievement," the important thing for Turri is that *achievements need not be reliably produced* because "achievement can issue from even highly *un*reliable ability" (2015a, 531). An agent has an unreliable ability to Φ iff in using this ability to Φ the agent fails to Φ most of the time. For example, a novice musician who plays a chord for the first time, a child who takes his first step or speaks his first sentence, and a rookie golfer who makes par for the first time are all examples of achievements for Turri even though they fail to achieve their desired ends most of the time (2015a, 531–532). In sum, for Turri, achievements involve

2 This general account of knowledge is part of the ongoing research program in epistemology called *performance-based epistemology* which is exemplified by virtue epistemology. See Sosa (2007), Zagzebski (2009), and Greco (2010) for more details on this research program, virtue epistemology, and achievement accounts of knowledge. See Bradford (2015) for a general account of achievements.

3 For example, see Pritchard (2008, 2009) and Lackey (2007, 2009).

4 See Alston (1995) for a detailed characterization of epistemic reliability.

simply attaining one's intended outcome through one's (un)reliable process, disposition, or ability. I will also assume this understanding of "reliability" and "achievement" in what follows.

Lastly, it is worth pointing out that Turri's account of achievement is unique among those who hold an achievement account of knowledge because it does not require that achievements manifest one's *competence* which involves the reliability of processes, dispositions, or abilities (see Sosa 2007, 2015; Zagzebski 2009; and Greco 2010). Turri (2016b) explicitly points out this omission and Turri (2015b, 2017, 2016b; and 2019) endorses this as a beneficial feature of his achievement account of knowledge because it avoids problems Turri sees for these authors' accounts of knowledge.

2 Against Turri's First Argument

Turri's first argument for the possibility of unreliable knowledge is

1. Achievements don't require reliable abilities. (Premise)
2. If achievements don't require reliable abilities, then unreliable knowledge is possible. (Premise)
3. So unreliable knowledge is possible. (From 1 and 2) (2015a, 531) [5]

Turri supports the first premise by referencing the examples he provides of achievements issuing from unreliable abilities mentioned above. Turri supports the second premise by saying that if knowledge is a kind of intellectual achievement and achievements generally do not necessarily issue from reliable processes, abilities, or dispositions, then "absent a special reason to think otherwise, we should expect [knowledge] to share the profile of achievements generally" (2015a, 532). In short, Turri's argument attempts to shift the burden of proof on those who believe reliability is a necessary condition of knowledge to show why knowledge, as an intellectual achievement, cannot issue from unreliable abilities.

Turri's first argument fails to convincingly shift the burden of proof because it faces a dilemma: Either the first premise is false or the argument as a whole begs the question. The first premise is false if it is interpreted to mean "*all* achievements don't require reliable abilities." There are many achievements

5 Turri (2015a, 2015b, 2017, 2016a, 2016b; and 2019) never specifies how "possibility" should be understood. I likewise will not assume any particular account of "possibility."

doi: 10.48106/dial.v74.i3

that require reliable abilities. More specifically, achieving some goal often requires reliably performing some action. For example, winning a competitive darts or archery tournament often requires one to reliably hit their intended mark.[6] Indeed, achieving the goal of performing some action with 90%+ accuracy (e.g. hitting a bullseye in archery, hitting a baseball, playing a piece of music, or walking) requires performing this action with 90%+ accuracy. So, the proper interpretation of the first premise must be something like "*some* achievements don't require reliable abilities." However, if this interpretation is placed back into the argument above then it begs the question. The second premise would now read "if *some* achievements don't require reliabilities, then unreliable knowledge is possible." But since Turri has said nothing against the possibility that knowledge is the kind of intellectual achievement that requires reliability (like the ones listed above), Turri has not provided adequate reasons to think that *knowledge* is the kind of achievement that can be unreliably produced—which is the purpose of the argument. So, in order for this argument to conclude "unreliable knowledge is possible," it must beg the question and consequently fails to shift the burden of proof.

Turri anticipates and responds to this dilemma[7] by claiming that it can be avoided if we interpret the first premise as a proposition "about dominant tendencies, or what is typical, or what is natural and normal for a kind" (2015a, 534). For example, the propositions that "humans don't have eleven fingers" or "cats don't have two faces" express tendencies about how humans and cats' anatomy are typically constituted (Turri 2015a, 534). Although there are exceptions to these claims, these exceptions do not render these claims false when these claims express such tendencies. So, if premise one is understood as a tendency proposition, Turri claims his argument "would still be plausible because, as already mentioned, we would expect knowledge to fit the profile of achievements generally, unless we're given a special reason to think otherwise" (2015a, 534).

This response still fails for the reasons mentioned above. Even if we grant that premise one is a tendency proposition, Turri has not established that achievements have a general tendency to be unreliable. As argued above, there are a large number of achievements that require reliability. Turri's few examples of unreliable achievements are insufficient to establish that premise

6 However, in order to achieve some goal, one need not reliably achieve that goal (e.g. to win the archery competition one need not reliably win the archery competition). Thanks to an anonymous reviewer for helping to clarify this.

7 Turri (2015a, fn. 7) attributes this dilemma to Bruce Russell.

one is a tendency proposition. Furthermore, Turri has provided no positive reason to think that *knowledge* is kind of achievement that can be unreliably produced—which (again) is the purpose of the argument. So, Turri's first argument fails to shift the burden of proof because it either has a false premise or begs the question.

A better strategy for Turri to establish that unreliable knowledge is possible is to take a more direct route by providing an example where one intuitively knows some proposition P even though one's true belief that P was formed by an unreliable cognitive process, i.e. one that produces more false than true beliefs. This is what Turri's second argument for unreliable knowledge attempts to do. In § 4 I will take on the burden of proof and argue that reliability is a necessary condition for knowledge.

3 Against Turri's Second Argument

Turri's second and more direct argument for the possibility of unreliable knowledge involves explanatory inference (aka, inference to the best explanation or IBE). As Turri notes, IBE is used in scientific reasoning and in everyday life to provide probable explanations for a set of data or certain phenomena. What best explains the fact that humans and chimpanzees have so many anatomical similarities? We have a common ancestor. What best explains the appearance of a new jug of milk in the fridge? My spouse bought it at the store. Turri claims that this kind of reasoning supports the possibility of unreliable knowledge:

> The epistemic efficacy of explanatory inference supports the view that unreliable knowledge is possible. Inference to the best explanation yields knowledge if the explanation that we arrive at is true. But even when it is true, the best explanation might not be very likely. So our disposition to infer to the best explanation might not be reliable. So unreliable knowledge is possible. (2015a, 536)

That is, even though IBE is often unreliable, the explanations it provides (when true) can yield knowledge. More specifically, some hypothesis "H" can

doi: 10.48106/dial.v74.i3

best explain a set of data "D" in our world even if there is a greater number of (nearby) possible worlds were D obtains and H is false (Turri 2015a, 536–537).[8]

To illustrate this argument, Turri provides a case study involving the television show *House M.D.* Gregory House (the protagonist) is a world-renowned medical doctor who has an incredible ability to diagnose patients where other doctors have failed. Simply put, he is the best of the best. However, despite being the best, House misdiagnoses patients a lot. Indeed, nearly every episode follows the same structure where House misdiagnoses the patient several times before coming to the right diagnosis just in the nick of time to save the patient's life. Turri contends that House's method for diagnosing patients is IBE—House infers a hypothesis/diagnosis that best explains the data/symptoms. And with each failed diagnosis House gains new insights to symptoms that inform his subsequent diagnoses. Given this description of House's track record, Turri argues that House's reliability is considerably less than .5. But despite House's unreliability, when he ends up correctly diagnosing his patient "House knows what disease that patient has" (Turri 2015a, 538). In short, this case study shows that IBE "can yield knowledge, even though it doesn't yield the correct verdict most of the time" (Turri 2015a, 539). Turri summarizes his second argument as

1. If House knows, then unreliable knowledge is possible. (Premise)
2. House knows. (Premise)
3. So unreliable knowledge is possible. (From 1 and 2)

> The argument is valid. Line 1 is supported by the fact that House's method usually produces false beliefs. Line 2 is supported by intuition, and by the fact that millions of viewers, including trained epistemologists, detect no incoherence in the story line, week after week, over many seasons. (2015a, 539)

I believe that both premises of Turri's second argument are false because Turri misrepresents House's medical abilities and knowledge. While Turri is right that House's diagnostic track record is well below .5, Turri takes the lesson here to be that, despite his track-record, "House knows" the correct diagnosis when he gets it right via IBE because House has a special ability to figure

8 There is controversy about whether IBE can provide explanations and/or produce knowledge (e.g. van Fraassen 1989). So, Turri's second argument has the important caveat that one must first accept that *IBE can produce explanations / knowledge* before this argument can be persuasive.

out the right diagnosis more often than any other doctor. This misrepresents House's abilities because, contra Turri, House is remarkable at getting the right diagnosis not because he knows the correct diagnosis more often than any other doctor, but because he has a remarkable ability to propose novel diagnostic hypotheses worthy of consideration and testing. But this ability to come up with possible explanations of patient's symptoms does not itself allow House to know that his diagnoses are correct *until* the treatment actually works (or when the reliable test results confirm his diagnosis).[9]

To illustrate these points, consider the following case that parallels Turri's House example:

> Jessica has very poor eyesight and is legally blind without her glasses. However, despite her eyesight, Jessica has a special ability to correctly identify pictures without her glasses. While others who are similarly handicapped can only identify pictures 5% of the time on average, Jessica is able to correctly identity such images 25% of the time on average. Now imagine that Jessica is presented with an image of a basketball that she, and others with her eyesight, phenomenologically describes as a blurry spot of reddish orange. Without her glasses Jessica infers incorrectly three times in a row that the picture is of an orange fruit, the Sun, and then a Lego piece. After each incorrect answer or hypothesis Jessica is told new information about the image that reveals why her answers were incorrect, e.g. it is not a fruit for her orange fruit hypothesis, it is an object you can touch for her Sun hypothesis, and it is an object that is bigger than a Lego piece. After all of this Jessica then answers correctly, but is not yet told that she is correct.

The crucial question to now ask is: At this point, does Jessica *know* what the picture is of? Intuitively, the answer is no. While Jessica, like House, has a special ability to get it right more often than her peers, this is not because she knows the correct answer more often, but because she is better at coming up with worthy hypotheses.[10] And, like House, Jessica does not know

9 I am indebted to Hilary Kornblith for a discussion on these points.

10 This misrepresentation of House's abilities is related to another debate concerning the nature of IBE. There is a tradition, going back to Pierce, of distinguishing abduction from IBE. Traditionally, abduction is concerned with hypothesis construction while IBE is concerned with selecting the hypothesis that is most likely to be true from a set of hypotheses. And traditionally, it is a feature

doi: 10.48106/dial.v74.i3

her hypothesis is correct until it's confirmed. Thus, premise two of Turri's argument is false because before the proposed treatment works (or when a reliable test result confirms a diagnosis) House does not know whether his hypothesized diagnosis is correct. Premise one is also false because if we plug this understanding of what House knows back into the antecedent of this premise, it renders the consequent false. That is, if "House knows" is understood to be true only after his hypothesized diagnosis has been tested and confirmed, then House's knowledge is not an instance of unreliable knowledge.[11]

4 Why Reliability is a Necessary Condition for Knowledge

So far, I have argued that Turri (2015a) has not provided adequate reasons to reject the orthodox view that knowledge requires reliability. In this final section I will directly argue against Turri's (2015b, 2017, 2016a, 2019) abilist account of knowledge[12] and argue that reliability must be a necessary condition for knowledge.

of abduction to be unreliable in order to produce a variety of hypothesis to be tested. Furthermore, there has been a recent trend to conflate these two (see McAuliffe 2015 for a defense of these points). On this understanding, House does not use IBE but uses abduction and Turri conflates these two when he writes "House and his team explicitly reason abductively" (2015a, 537) and "House's method for trying to solve the case *just is* to employ inference to the best explanation" (2015a, 540, his emphasis).

11 Turri's argument that IBE can produce unreliable knowledge also concerns two other recent debates about the nature of IBE. First, it is related to the issue of whether van Fraassen's Bad Lot objection shows that IBE is an unreliable inference form. The Bad Lot Objection argues that IBE is an inadequate inference form because it has no way of discerning whether a set of hypotheses are all false and so would lead to a false conclusion in these cases. Schupbach (2014) recently argues that van Fraassen's Bad Lot Objection does not establish that IBE is an unreliable inference form any more than we can show that modus ponens is unreliable by plugging in false premises (see Dellsén 2017 for a response). If this were true, then it would be inappropriate to characterize either IBE or modus ponens as unreliable. Second, many have argued that IBE is a heuristic that approximates objective Bayesian reasoning. On this understanding, IBE's primary function is to locate the "most probable available explanatory hypothesis to serve as a working hypothesis in an agent's further investigations" (Dellsén 2018). This understanding dovetails with my explanation of House's use of IBE above. In short, these issues concerning the nature and reliability of IBE impact and potentially undermine the soundness of Turri's argument that IBE can unreliably produce knowledge.

12 Turri (2015a) also calls his account of knowledge *ecumenical reliabilism*. While there are small differences between these accounts, they can safely be considered together since they both hold that unreliable knowledge can occur when agents successfully achieve their desired ends *through* their abilities. This claim is the focus of the rest of the paper.

Turri defines *abilism* in the following ways:

> Abilism defines knowledge as true belief manifesting the agent's cognitive ability or powers (2016a, 225);
>
> Knowledge is approximately true thin belief manifesting *cognitive ability* (2015b, 321; and 2017, 164);
>
> Knowledge is an accurate representation produced by cognitive ability (2019).[13]

Turri's terminology of cognitive abilities "producing" or "manifesting" true beliefs serves to explain why certain unreliable processes can produce knowledge. Turri (2016b) takes the following example from Sosa (2007) to elucidate these concepts: An archer hitting a bullseye manifests her athletic ability only when her hitting the bullseye is *based on* or *the result of* or *because of* her abilities. If a gust of unexpected wind interferes with the arrow's path and causes the arrow to hit the bullseye, then the bullseye was not a result of the archer's abilities. But unlike Sosa, Turri does not require that our cognitive abilities be reliable (see § 1). This also fits with his account of achievements explained in § 1 above: Achievements involve simply attaining one's intended outcome *through* one's (un)reliable ability. In my own words, Turri holds that S knows or intellectually achieves P iff P is true, and S believing P is the result of or manifests S's (un)reliable cognitive abilities.[14]

One tempting argument to make against any account of knowledge that allows for the possibility of unreliable knowledge is that such accounts would implausibly allow for *lucky* knowledge. Turri's account of knowledge seems especially vulnerable to this objection since it seems that the novice archer who achieves a bullseye on her first try has beginner's luck even though she achieved the bullseye, in some sense, *through* her abilities. In response, Turri agrees that lucky knowledge is implausible but he denies that abilism allows for lucky knowledge:

> The fact that someone cannot reliably produce an outcome does not entail that it's "just luck" when she does produce it. Unreliable

13 The second and third definitions indicate that Turri now holds that neither belief nor truth is a necessary condition for knowledge. I will ignore these aspects of Turri's account in this paper since they do not affect my arguments below.

14 Recall from § 1 that this account of achievement is distinctive because it does not require achievements manifest one's competence which involves manifesting one's abilities *reliably* (see Sosa 2007, 2015; Zagzebski 2009; and Greco 2010).

doi: 10.48106/dial.v74.i3

> performers usually still have *some* ability or power to produce the relevant outcome. *Unreliability* does not equal *inability*. (2015a, 533)

While Turri does not explicate the different kinds of luck at issue here,[15] the ideas are clear enough to be intuitively compelling. The novice archer who hits the bullseye through their unreliable abilities (e.g. through effort and concentration) does not succeed just by luck; while the archer who hits the bullseye because of a gust of wind does succeed by luck. Likewise, for Turri, *intellectual* achievements that issue from one's unreliable cognitive abilities are not lucky in the way that achieving a true belief through, say, guessing is lucky. Despite his poor track-record, when House correctly diagnoses a patient through his great diagnostic ability, he does so in a way that an avid fan of *House M.D.* does not when they guess the correct diagnoses. Because many unreliable processes *manifest one's ability* while lucky processes do not, Turri argues that his account of knowledge does not allow for lucky knowledge.

In essence, Turri is making the following argument:

1. Not all unreliable cognitive processes are lucky.
2. Some of the processes in (1) are non-lucky but unreliable cognitive processes that manifest one's cognitive ability.
3. Some of the processes in (2) can produce knowledge.
4. Thus, unreliable knowledge is possible.

I agree with Turri that unreliability does not equal inability and that, per premise one, we should not think that all unreliable processes are just lucky processes. To deny these claims is to implausibly deny that there are nascent cognitive abilities. I also agree with Turri that, per premise two, his account of knowledge does not allow for lucky knowledge. However, the key issue is whether premise three is true because if it is, then abilism is true and unreliable knowledge is possible.[16]

15 See Pritchard (2005) for an analysis of the different kinds of epistemic luck.

16 For various theoretical reasons, many authors would deny this premise. For example, Pritchard (2012) and Kelp (2013) argue that an ability condition on knowledge should be combined with a safety condition in order to deal with counterexamples typically leveled against ability conditions (e.g. fake barn cases). And since safety is a kind of reliability condition, Turri is presumably committed to rejecting these accounts of knowledge. Others have suggested that when the ability condition on knowledge is properly unpacked it entails an anti-luck condition precisely because it entails a reliability condition. For example, Sosa (2015), Carter (2016), and Beddor and Pavese (2020) have suggested, on different grounds, that the best version of a cognitive ability condition

To see why premise three is false it is important to first realize that the Jessica example in § 3 is one instance of someone who fits Turri's definition of abilist/unreliable knowledge but intuitively fails to have knowledge. Jessica's true belief that the blurry picture in front of her is of a basketball is the result or manifestation of her unreliable cognitive ability to recognize such images (i.e. 25% average accuracy) but she fails to have knowledge until she is told her belief is true. Premise three is false because counterexamples like this can be generalized to show that unreliable/abilist knowledge is impossible. In short, I argue that this unreliable/abilist knowledge is impossible because any agent that is in a sufficiently favorable epistemic position to have unreliable/abilist knowledge will fail to have knowledge. And as was shown in § 3, Jessica is in such a sufficiently favorable epistemic position for unreliable/abilist knowledge but she intuitively fails to have knowledge.

One might object that Jessica is not in a sufficiently favorable epistemic position to have unreliable/abilist knowledge. Firstly, an objector could argue that knowledge can be unreliably achieved only above some threshold of unreliability (e.g. above 40%). So, while Jessica is very reliable in comparison to her peers, she still only has 25% reliability and falls below this threshold for unreliable knowledge. Additionally, one could object that our intuitions about the Jessica case may be compromised by the fact that Jessica's unreliability is caused by her sub-par eyesight or malfunctioning ability to see. Indeed, what makes the House case compelling is that House's unreliability is not caused by a sub-par or malfunctioning ability (since he is the best of the best) but because of the difficulty of his job—i.e. diagnosing unusual patients. So, for these reasons one could argue that the Jessica case is not a convincing counterexample to abilism and the possibility of unreliable knowledge.

In response, I claim that additional examples can be constructed to avoid these pitfalls that nevertheless show that unreliable/abilist knowledge is impossible:

> Ashley is a professional singer. While Ashley does not have perfect pitch, after many years of studying, practicing, and performing she has gained some ability to accurately identify notes played on a piano. Specifically, Ashley is able to accurately identify what single note is played by listening alone with almost 50% average accuracy.

for knowledge entails a safety condition. Presumably Turri is also committed to rejecting all of these arguments. My argument below undermines Turri's third premise directly via novel counterexamples without relying on any of these theoretical considerations.

doi: 10.48106/dial.v74.i3

> In contrast, the average lay person is almost never able to correctly identify the right note since they have no ability to recognize which of the 12 possible notes is played. Those with perfect pitch are able to recognize which note is played with near 100% accuracy. Imagine that you are watching Ashley practice her ability over the period of half an hour. In this time, you see her correctly identify what note is played on average almost 50% of the time. Furthermore, you notice that when Ashley is wrong, she is never more than a musical half-step from the right answer (e.g. if the answer is A#, Ashley answers A; or if the answer is F, Ashley answers E).

Unlike Jessica, Ashley is much more reliable at almost 50% and, like House, does not have a sub-par or malfunctioning ability. You could say that she *nearly* has perfect pitch since her answers indicate that even when she is wrong, she is still tracking the correct pitch. But even with this great ability to identify pitches by auditory means alone, imagine that Ashley is played a Db note on a piano and correctly answers Db, but is not yet told that her answer is correct. At this point, does Ashley *know* that the note is a Db? Intuitively, Ashley does *not* know the answer is Db, and I contend the only explanation for this intuition is that despite her nascent perfect pitch ability she is still unreliable at identifying pitches. Thus, abilism is false because examples like this show that one can have a true belief that manifests one's unreliable cognitive abilities without having knowledge.

So, to reiterate, examples like this also show that unreliable knowledge is impossible since such agents are in sufficiently favorable epistemic conditions to have this kind of knowledge, but intuitively still fail to have knowledge. Furthermore, I contend that many more examples can be constructed to support the intuition that unreliable agents like Jessica and Ashley fail to have knowledge. In summary, I am making the following argument:

1. If those in sufficiently favorable epistemic positions to have unreliable/abilist knowledge fail to have knowledge, then unreliable abilist/knowledge is impossible.
2. Ashley, Jessica, etc., are in sufficiently favorable epistemic positions to have unreliable/abilist knowledge but fail to have knowledge.
3. Thus, unreliable/abilist knowledge is impossible.

In conclusion, Turri has not established that unreliable knowledge is possible and there are decisive reasons for thinking knowledge requires reliability.*

Jonathan Dixon
0000-0003-4040-7289
Wake Forest University
dixonj@wfu.edu

References

Alston, William P. 1995. "How to Think about Reliability." *Philosophical Topics* 23(1): 1–29, doi:10.5840/philtopics199523122.

Beddor, Robert [Bob] and Pavese, Carlotta. 2020. "Modal Virtue Epistemology." *Philosophy and Phenomenological Research* 101(1): 61–79, doi:10.1111/phpr.12562.

Bradford, Gwendolyn. 2015. *Achievement*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198714026.001.0001.

Carter, J. Adam. 2016. "Robust Virtue Epistemology as Anti-Luck Epistemology: A New Solution." *Pacific Philosophical Quarterly* 97(1): 140–155, doi:10.1111/papq.12040.

Dellsén, Finnur. 2017. "Reactionary Responses to the Bad Lot Objection." *Studies in History and Philosophy of Science* 61: 32–40, doi:10.1016/j.shpsa.2017.01.005.

—. 2018. "The Heuristic Conception of Inference to the Best Explanation." *Philosophical Studies* 175(7): 1745–1766, doi:10.1007/s11098-017-0933-2.

van Fraassen, Bas C. 1989. *Laws and Symmetry*. Oxford: Oxford University Press, doi:10.1093/0198248601.001.0001.

Goldman, Alvin I. 1979. "What is Justified Belief?" in *Justification and Knowledge: New Studies in Epistemology*, edited by George Sotiros Pappas, pp. 1–24. Philosophical Studies Series n. 17. Dordrecht: D. Reidel Publishing Co. Reprinted in Goldman (2012, 29–49), doi:10.1007/978-94-009-9493-5_1.

—. 2012. *Reliabilism and Contemporary Epistemology: Essays*. New York: Oxford University Press, doi:10.1093/acprof:oso/9780199812875.001.0001.

Goldman, Alvin I. and Beddor, Robert [Bob]. 2021. "Reliabilist Epistemology." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/sum2021/entries/reliabilism/.

* I would like to thank the audiences at the 3rd Annual Philosurfer Convergence and the 2019 Eastern APA for their valuable feedback on previous versions of this paper. Thanks to Heather Spradley for her insightful commentary at the Eastern APA. I would also like to thank Hilary Kornblith, Christopher Meacham, Sophie Horowitz, Cameron Gibbs, and especially Elizabeth C. Hupfer for providing written comments and discussing these issues at length with me.

GRECO, John, ed. 2008. *The Oxford Handbook of Skepticism*. Oxford Handbooks. New York: Oxford University Press, doi:10.1093/oxfordhb/9780195183214.001.0001.
—. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511844645.
HETHERINGTON, Stephen Cade. 1998. "Actually Knowing." *The Philosophical Quarterly* 48(193): 453–469, doi:10.1111/1467-9213.00114.
—. 1999. "Knowing Failibly." *The Journal of Philosophy* 96(11): 565–587, doi:10.2307/2564624.
—. 2016. *Knowledge and the Gettier Problem*. Cambridge: Cambridge University Press, doi:10.1017/cbo9781316569870.
KELP, Christoph. 2013. "Knowledge: The Safe-Apt View." *Australasian Journal of Philosophy* 91(2): 265–278, doi:10.1080/00048402.2012.673726.
LACKEY, Jennifer. 2007. "Why we Don't Deserve Credit for Everything we Know." *Synthese* 158(3): 345–361, doi:10.1002/9781119420828.ch13.
—. 2009. "Knowledge and Credit." *Philosophical Studies* 142(1): 27–42, doi:10.1007/s11098-008-9304-3.
MCAULIFFE, William H. B. 2015. "How Did Abduction Get Confused with Inference to the Best Explanation?" *Transactions of the Charles Sanders Peirce Society* 51(3): 300–319, doi:10.2979/trancharpeirsoc.51.3.300.
PRITCHARD, Duncan. 2005. *Epistemic Luck*. Oxford: Oxford University Press, doi:10.1093/019928038X.001.0001.
—. 2008."Greco (2008) on Knowledge: Virtues, Contexts, Achievements." *The Philosophical Quarterly* 58(232): 437–447, doi:10.1111/j.1467-9213.2008.550.x.
—. 2009. "Apt Performance and Epistemic Value [on Sosa (2007)]." *Philosophical Studies* 143(3): 407–416, doi:10.1007/s11098-009-9340-7.
—. 2012. "Anti-Luck Virtue Epistemology." *The Journal of Philosophy* 109(3): 247–279, doi:10.5840/jphil201210939.
SARTWELL, Crispin. 1991. "Knowledge is Merely True Belief." *American Philosophical Quarterly* 28(2): 157–165.
—. 1992. "Why Knowledge is Merely True Belief." *The Journal of Philosophy* 89(4): 167–180, doi:10.2307/2026639.
SCHUPBACH, Jonah N. 2014. "Is the Bad Lot Objection Just Misguided?" *Erkenntnis* 79(1): 55–64, doi:10.1007/s10670-013-9433-8.
SOSA, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge. Volume I*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199297023.001.0001.
—. 2015. *Judgment and Agency*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198719694.001.0001.
TURRI, John. 2015a. "Unreliable Knowledge." *Philosophy and Phenomenological Research* 90(3): 529–545, doi:10.1111/phpr.12064.
—. 2015b. "From Virtue Epistemology to Abilism: Theoretical and Empirical Developments." in *Character: New Directions from Philosophy, Psychology, and Theology*,

edited by Christian B. MILLER, R. Michael FURR, Angela M. KNOBEL, and William FLEESON, pp. 315–331. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780190204600.003.0015.

—. 2016a. "A New Paradigm for Epistemology: From Reliabilism to Abilism." *Ergo* 3(8): 189–231, doi:10.3998/ergo.12405314.0003.008.

—. 2016b. "Knowledge as Achievement, More or Less." in *Performance Epistemology: Foundations and Applications*, edited by Miguel Ángel FERNÁNDEZ VARGAS, pp. 124–135. New York: Oxford University Press, doi:10.1093/acprof:oso/9780198746942.003.0008.

—. 2017. "Epistemic Situationism and Cognitive Ability." in *Epistemic Situationism*, edited by Abrol FAIRWEATHER and Mark ALFANO, pp. 158–167. Oxford: Oxford University Press, doi:10.1093/oso/9780199688234.003.0009.

—. 2019. "Virtue Epistemology and Abilism on Knowledge." in *The Routledge Handbook of Virtue Epistemology*, edited by Heather BATTALY, pp. 309–316. Routledge Handbooks in Philosophy. London: Routledge, doi:10.4324/9781315712550-26.

WILLIAMSON, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press, doi:10.1093/019925656X.001.0001.

ZAGZEBSKI, Linda Trinkaus. 2009. *On Epistemology*. Belmont, California: Wadsworth Publishing Co.

Are There Occurrent Continuants?
A Reply to Stout's "The Category of Occurrent Continuants"

Riccardo Baratella

Processes are occurrents that were, are, or will be happening. They endure or they perdure, i.e. they are either "fully" present at every time they happen, or they rather have temporal parts. According to Stout (2016), they endure. His argument assumes that processes may change. Then, Stout argues that, if something changes, it endures. As I show, Stout's Argument misses its target. In particular, it makes use of a notion of change that is either intuitive but illegitimate or technical but question-begging.

In "The Category of Occurrent Continuants," Stout (2016) argues that processes are both occurrents and continuants (i.e. they endure). His argument assumes that processes may change over time and seeks to show that, on this assumption, if something changes, it endures. I argue that such an argument fails: either it makes illegitimate use of an intuitive notion of change, or it makes use of a technical, but question-begging, notion of change.

1 Background Notions

According to Stout, processes are things that are, were, or will be happening. Examples include my writing this article—something that is happening right now—or the concert that was happening yesterday. Processes are described or referred to in answering the progressive question: "What is (was, will be) happening?" The basic feature of expressions describing or referring to processes is the use of the progressive aspect.

Stout contrasts processes with events. Events are things that happened or will happen. Examples include the explosion that will take place next year, and my winning the race that happened yesterday. Moreover, the basic feature

of expressions describing or referring to events is the use of non-progressive aspect.[1]

Events and processes both exist over time—i.e. they persist. There are two main accounts of persistence. The first one is *perdurance theory*—the thesis that things of a certain kind perdure. Intuitively, something perdures if and only if it is extended in time and has different temporal parts at different times—a different temporal part for each moment of time. The other account of persistence is *endurance theory*—the thesis that things of a certain kind endure. Intuitively, something endures if and only if it is "all" there at each moment at which it exists. Events, rather uncontroversially, perdure. However, Stout argues that, in this respect, processes differ from events: processes, he claims, endure.[2]

According to Stout, perduring entities are things that *primarily* have their properties atemporally. Such a characterization can be explained via the perdurance analysis of sentences of the form "*x* has the property of *sitting* at *t*." According to perdurance theory, the temporal qualification "at *t*" is part of the subject of the sentence, "*x* at *t*," which denotes the *t*-temporal part of *x*. In turn, the predication of the property *sitting* has no temporal connotation at all: the property is atemporally exemplified by the temporal part *x*-at-*t*.[3] This means that the exemplification of the property *sitting* by the *t*-temporal part of *x* is not relativized to times: the exemplification involves only that temporal part and the property of *sitting*. According to perdurance theory, the atemporal exemplification is basic and temporal predications, such as "sitting at *t*," are analyzed in terms of it. As a result, a sentence like "*x* has property *P* at time *t*" is true if and only if *x* has atemporally a *t*-temporal part that has atemporally the property *P*.

By contrast, enduring entities are things that *primarily* have their properties at times. Let me clarify such a characterization by considering "*x* has the

1 For some objections to this way of articulating the distinction between events and processes, see Steward (2013). She further develops the framework proposed by Mourelatos (1977), according to whom processes are picked out by "mass-quantified nominalizations" derived from predications with an imperfective aspect, while events are individuated by "count-quantified nominalizations" derived from predications with a perfective aspect. In this article, I won't take a stand on such a dispute, and focus instead on Stout's Argument, assuming for the sake of argument his criterion for distinguishing between events and processes.

2 For a detailed discussion of these notions, see Simons (1987), Hawley (2001), Sattig (2003), and Varzi (2003).

3 For the notion of atemporal exemplification, see Simons (1987, 122), Hawley (2001, 13–14), Sider (2001, 56), and Stout (2016, 46–47).

property of *sitting* at *t*." Within endurance theory, the subject of the sentence is simply "*x*," which denotes a "three-dimensional" entity *x*. The temporal qualification belongs to the predicate which results in "having the property of *sitting* at *t*." Such a predicate must, now, be analyzed—according to Stout, via a notion of exemplification which is fundamentally temporal.[4] In particular, Stout adopts the tensing the copula strategy, according to which the temporal qualification modifies the relation of exemplification (while keeping the subject *not* tensed). As a result, the previous sentence is analyzed as "*x* has-at-*t* the property of *sitting*."[5] In general, sentences containing temporal predications, such as "*x* has property *P* at time *t*," are true if and only if *x* has-at-*t* the property *P*.

2 Stout's Argument

Stout argues that processes persist by enduring rather than by perduring.[6] He asks to consider a fight that went on outside his house between 11.55 p.m. and 12.05 a.m. last night. The fight *was happening* at midnight. So, it is a process. Stout's description of the fight makes it intuitive to maintain that it actually changes:

> At first it was quite brutal, but after a few minutes it became less ferocious, though as if to make up for this, it got gradually more noisy until the police arrived and stopped it. On the face of it it is a thing that continues through time and has different properties at different times. (2016, 50)

Stout's Argument can now be reconstructed as follows.[7] To begin with, it immediately follows from Stout's description of the fight that:

(1) The fight is first brutal at *t*, and it is not brutal at *t**.

4 Endurance theory rejects the notion atemporal exemplification as incomplete or unintelligible. To see this, suppose that *x* is both sitting today and not-sitting tomorrow. Suppose also that *x* endures. If we adopted the notion of atemporal exemplification, we would get that *x* is both sitting and not-sitting. For some concerns against this standard idea, see Hansson Wahlberg (2007).

5 Lewis (2002) argues against the tensing the copula analysis. Again, for argument's sake, I assume with Stout that it is a workable position.

6 Stout (1997) argues for the same thesis; for a reply, see Steward (2013). The analogy between enduring objects and processes has also been recently supported by, e.g. Galton (2006).

7 For Stout's Argument, see Stout (2016, 44–50).

doi: 10.48106/dial.v74.i3

Stout further assumes the following, seemingly intuitive notion of change (2016, 45):

> CHANGE. Something changes if and only if this thing has a property at one time and at a later time the very same thing does not have that very property.

Further, (1) is an intuitive case of change. Indeed, given (1) and (CHANGE), the fight changes—call this latter claim "FIGHTCHANGES."

Now, in order to understand Stout's Argument, we need to be able to interpret its sentences, i.e. to give their truth-conditions. Moreover, it seems plausible to assume that the truth-conditions of sentences involving notions such as continuity over time or persistence require, implicitly or explicitly, the assumption of a *theory of persistence*. Call this assumption "A1." Clearly, in order to establish endurantism, Stout's Argument must go through irrespective of how its key assumptions are interpreted, i.e. irrespective of one's chosen theory of persistence. But this, I will argue, isn't the case.

Since processes persist either by enduring or by perduring, when interpreting (1) we must consider two cases: respectively, the perdurance and the endurance interpretation.[8] Thus, let us first interpret (1) within perdurance theory—the view according to which the fight primarily has its properties atemporally. On this interpretation, (1) intuitively entails (FIGHTCHANGES). However, (CHANGE) is incompatible with the perdurance interpretation of (1):

> *Proof 1*. Given a perdurance reading, (1) boils down to the following situation: the fight has (atemporally) a t_1-temporal part that has (atemporally) the property of *being brutal* and it has (atemporally) a successive t_2-temporal part that does not have (atemporally) the property of *being brutal*. So, the temporal part that has the property of *being brutal* is different from the part that does not have that property. Moreover, since any entity involved in the scenario has its properties and relations atemporally, nothing can have a property or a relation *and then* fail to have it. But, then, given (CHANGE), nothing can change in the previous situation. However, given (1) and the implication from (1) to (FIGHTCHANGES), we get that the fight

8 Following Stout, I narrow down my focus on the two main accounts of persistence, and set aside for present purposes theories such as the Stage View.

> changes. Contradiction. So, we need to reject one of our assumptions. Since (1) and (FIGHTCHANGES) seem unassailable, we must either reject (CHANGE) or the perdurance interpretation of (1).

Since we have (CHANGE) by assumption, we must reject the perdurance interpretation of (1). Hence, the fight does not perdure and perdurance theory is refuted.

Let us now interpret (1) within endurance theory—the view that the fight primarily has its properties at a time. From (CHANGE) and the endurance reading of (1), we can derive that the fight changes:

> *Proof 2.* Given an endurance reading, (1) boils down to the following situation: the fight has-at-t_1 the property of *being brutal* and it does not have-at-t_2 the property of *being brutal.* Then, the fight satisfies (CHANGE). So, it changes.

Since, intuitively, the fight changes, and given that, with (CHANGE) in place, it can only change on an endurance reading of (1), we must conclude that the fight endures—i.e. endurance theory provides the correct account of persistence. Or so Stout argues.

3 Against Stout's Argument

Stout's Argument is unsound. More specifically, either the argument makes an illegitimate use of an intuitive notion of change, viz. (CHANGE), or it makes use of a theoretical, but question-begging notion of change.[9]

I think it is fair to grant that any adequate theory of persistence must account for intuitive cases of change, such as (1). But is (CHANGE) really incompatible with a perdurantist perspective? Let us consider it again:

> CHANGE. Something changes if and only if this thing has a property at one time and at a later time the very same thing does not have that very property.

9 Crowther (2018) offers a criticism of Stout's Argument different from the one presented here. However, I don't find Crowther's argument convincing for two main reasons. First, to the extent that he concedes that (CHANGE) is incompatible with perdurance theory, he falls prey of the same objection I shall raise against Stout's Argument. Second, his account is an extreme version of Kim's view, according to which occurrences are property-exemplifications. As such, it faces a problem of overgeneration, which makes his view implausible (see Hendrickson 2006).

doi: 10.48106/dial.v74.i3

(CHANGE) is a claim about *persisting* entities. So, by assumption A1, it must be interpreted within a theory of persistence—i.e. its truth-conditions must be interpreted either within perdurance theory or within endurance theory.

Let's consider the *perdurance* interpretation first. Since (CHANGE)'s right hand-side makes temporal predications, we must now interpret it according to the *perdurance* account of temporal predication. Recall, according to this "x has property P at time t" is true if and only if x has atemporally a t-temporal part that has atemporally the property P. Accordingly, "x has property P at time t but lacks it at a later time" is interpreted as "x has a temporal part at time t, x-at-t, that has property P and x has a different temporal part at a later time t^*, x-at-t^*, that does not have that property." Thus, the overall perdurance truth-conditions of (CHANGE) are as follows:

> PERDCHANGE. Something changes if and only if it has a temporal part at a time t that has a property and it has a different temporal part at a later time t^* that does not have that property.[10]

Let us now interpret (CHANGE) within *endurance* theory. Recall, according to the endurance account of temporal predication, "x has property P at time t" is true if and only if x has-at-t the property P. Then, (CHANGE)'s right hand-side is interpreted as "x has-at-t a property and x does not have-at-t^* (with $t < t^*$) that property" and the overall endurance truth-conditions of (CHANGE) are as follows:

> ENDCHANGE. Something changes if and only if it has-at-t a property and x does not have-at-t^* (with $t < t^*$) that property.

Depending on one's theory of persistence, (CHANGE) can be interpreted in one of two ways: (ENDCHANGE) or (PERDCHANGE). However, as I now argue, neither reading supports Stout's conclusion, that processes endure.

To begin with, it is now immediate to show that perdurance theory is compatible with (CHANGE)—contra Stout's Argument—and that it can easily account for (1) as an intuitive case of change.

10 On the perdurance notion of change, see Hawley (2001, 12), Sider (2001, 212) and Wasserman (2006). The reply presented here is already hinted at in a number of places—see e.g. Hawley (2001, 12), Sider (2001, 212) and Hofweber (2009, 303–311).

> *Proof 3.* Given perdurance theory, (1) boils down to the following situation: the fight has (atemporally) a t_1-temporal part that has (atemporally) the property of *being brutal* and it has (atemporally) a successive t_2-temporal part that does not have (atemporally) the property of *being brutal*. Now, under the adoption of perdurance theory, (CHANGE) must be interpreted as (PERDCHANGE). But, then, the fight satisfies the right-side of (PERDCHANGE). Therefore, it changes.

Thus, perdurantism is compatible with all the assumptions in Stout's Argument, i.e. the argument fails to establish that processes endure. Given perdurantism, Stout's Argument is *unsound*: it interprets (1) within perdurance theory without doing the same with (CHANGE). However, once perdurantism is assumed, it has to be applied all the way down—both to (1) *and to* (CHANGE).

Stout might of course object that (CHANGE) is to be interpreted as (ENDCHANGE), i.e. it should be given an *endurance* interpretation. And, he might point out, given (ENDCHANGE), Stout's Argument *is* sound.

However, the endurance interpretation of (CHANGE) is not available to Stout: it begs the question against perdurance theory. To see this, it is sufficient to notice that perdurance theory is incompatible with (ENDCHANGE):

> *Proof 4.* Given perdurance theory, (1) boils down to the following situation: the fight has (atemporally) a t_1-temporal part that has (atemporally) the property of *being brutal* and it has (atemporally) a successive t_2-temporal part that does not have (atemporally) the property of *being brutal*. Now, any entity involved in the scenario has its properties and relations atemporally. Then, nothing can satisfy the right-side of (ENDCHANGE)—according to which the relation of exemplification is temporally modified. So, nothing can change in the previous situation. However, given (1) and the implication from (1) to (FIGHTCHANGES), we get that the fight changes. Contradiction. Hence, given the previous assumptions, we must reject the perdurance interpretation of (1).

To be sure, endurance theory is compatible with (ENDCHANGE) and can account for (1) as an intuitive case of change:

doi: 10.48106/dial.v74.i3

Proof 5. The endurance interpretation of (1) is the following: the fight has-at-t_1 the property of *being brutal* and it does not have-at-t_2 the property of *being brutal*. Now, given (ENDCHANGE)—according to which something changes if and only if it has-at-t a property and x does not have-at-t^* (with $t < t^*$) that property—the fight satisfies the right-side of (ENDCHANGE). Therefore, it changes.

However, (ENDCHANGE) clearly begs the question against perdurance theory. Since (ENDCHANGE) provides the *endurance* truth-conditions for (CHANGE), it presupposes endurance theory and therefore isn't neutral between endurantism and perdurantism. As a result, on such a reading, Stout's Argument is circular: it establishes what it has already assumed, viz. an endurantist account of persistence for processes. That is, Stout's Argument against perdurantism only goes through if one *assumes* that perdurantism is false. The argument is valid but, of course, not very interesting.[11]

It might be objected that (ENDCHANGE) provides the correct characterization of change and that, for this reason, it cannot be plausibly rejected.[12]

However, the objection fails to convince. First off, (CHANGE)—our intuitive notion of change—can be interpreted both within perdurance theory, as

11 I reconstructed Stout's argument by assuming the notion of change Stout explicitly adopted, i.e. (CHANGE) (2016, 44). Then, I showed that either his argument makes illegitimate use of (CHANGE), or it makes use of a technical, but question beginning notion of change—i.e. (ENDCHANGE). It might be objected that Stout's Argument must be reconstructed as conditional whose antecedent is the endurantist reading of (CHANGE), i.e. (ENDCHANGE):

$$\Gamma \vDash (\text{ENDCHANGE}) \rightarrow (\text{Processes}\sim\text{Endure}),$$

where Γ is the set of assumption Stout relies on (which I've granted for argument's sake). However, Stout explicitly claims that he aims to show that processes endure (2016, 42, 50)—not the conditional conclusion that processes endure if we adopt the endurantist truth-conditions of (CHANGE). We must therefore include the antecedent of the above conditional among our assumptions:

$$\Gamma, (\text{ENDCHANGE}) \vDash (\text{Processes}\sim\text{Endure}).$$

Now, this version of Stout's Argument clearly amounts to the case just considered in the main text—viz. a version that includes (ENDCHANGE) together with *Proof 4* and *Proof 5*. As a consequence, (ENDCHANGE) begs the question against perdurance theory. Since (ENDCHANGE) provides the *endurance* truth-conditions for (CHANGE), it presupposes endurance theory and therefore isn't neutral between endurantism and perdurantism. More precisely, such a version of Stout's Argument is circular: it establishes what it has already assumed, i.e. an endurantist account of persistence for processes.

12 Versions of such an objection can be found in e.g. Geach (1972, 304) and Simons (1987, 126).

(PerdChange), and within endurance theory, (EndChange). Pending any argument to the effect that (EndChange), and only it, correctly accounts for change, the objection amounts to mere foot stamping.

Second, I granted that any adequate theory of persistence must account for intuitive cases of change. Now, the specific characterization of change a theory adopts is part of how it explains the required phenomena. If such a characterization helps the theory to account for cases of change, then the characterization is adequate for that theory. In other words, any specific characterization of change is relative to a particular framework of persistence. But, then, it makes poor sense to claim that (EndChange) is the "correct" characterization of change independently from a specific theory of persistence. Hence, the objection must be resisted.[13]

Summing up, Stout's Argument for the thesis that processes endure is based on (1), (Change), and the fact that (1) intuitively entails (FightChanges). However, as I've argued, sentences such as (1), including (Change), must be interpreted within a theory of persistence. And, as we've seen, the only interpretation of Stout's Argument in which the argument goes through is also one in which (1) and (Change) receive endurantist truth-conditions, i.e. they are interpreted on an endurantist semantics that is incompatible with perdurantism. As a result, Stout's Argument is viciously circular: it presupposes precisely what it is meant to establish, viz. that processes endure.[14]*

13 A nontrivial consequence of this reply is that events—such as a football match that was first boring and then exciting—can also change. This is not, however, a problematic result. Indeed, the thesis that events cannot change has been supported by the same kind of arguments as those examined in relation to Stout's Argument—see e.g. Simons (1987, 126) and Stout (2016, 47). These arguments essentially stand, or fall, with Stout's Argument. And, I have argued, they fall.

14 Of course, the fact that Stout's Argument fails does not entail that processes do not endure. However, the question as to whether they endure must be settled by arguments of a different kind than those discussed here.

* I would like to thank Emiliano Boccardi, Claudio Calosi, Pierdaniele Giaretta, Alessandro Giordani, Nicola Guarino, Giancarlo Guizzardi, Julien Murzi, Thomas Sattig, Maria Scarpati, Giuseppe Spolaore, Helen Steward, Achille C. Varzi, and two anonymous referees of *Dialectica* for precious comments and discussions on previous versions of this article. I also would like to thank the participants at the conferences: "ENFA7," "Space and Time: An Interdisciplinary Approach – Universitas Vilnensis," "WIP Seminar – University of Salzburg," "Modal Metaphysics: Issues on the (Im)Possible VII Conference," "OberSeminar – University of Tübingen," "93rd Joint Session of the Aristotelian Society and the Mind Association," "The Joint Ontology Workshops 2019." This work was supported by the DAAD Scholarship under grant number 91645817; the Ernst Mach Grant - worldwide under grant number ICM-2018-09910; a Research Assistant Contract in the project NEXON at the Free University of Bozen-Bolzano.

Riccardo Baratella
0000-0002-4387-2912
Free University of Bozen-Bolzano
riccardo.baratella@unibz.it

References

CROWTHER, Thomas. 2018. "Processes as Continuants and Process as Stuff." in *Process, Action, and Experience*, edited by Rowland STOUT, pp. 58–81. Oxford: Oxford University Press, doi:10.1093/oso/9780198777991.003.0004.

GALTON, Antony. 2006. "On What Goes On: The Ontology of Processes and Events." in *Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS2006)*, edited by Brandon BENNETT and Christiane FELLBAUM, pp. 4–11. Amsterdam: IOS Press.

GEACH, Peter Thomas. 1972. *Logic Matters*. Oxford: Basil Blackwell Publishers. Republished as Geach (1980).

—. 1980. *Logic Matters*. Berkeley, California: University of California Press. Republication of Geach (1972), doi:10.1525/9780520352391.

HANSSON WAHLBERG, Tobias. 2007. "The Problem(s) of Change Revisited." *Dialectica* 61(2): 265–274. Published under the name "Hansson", doi:10.1111/j.1746-8361.2007.01113.x.

HAWLEY, Katherine. 2001. *How Things Persist*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199275434.001.0001.

HENDRICKSON, Noel. 2006. "Towards a More Plausible Exemplification Theory of Events." *Philosophical Studies* 129(2): 349–375, doi:10.1007/s11098-004-1647-9.

HOFWEBER, Thomas. 2009. "The Meta-Problem of Change." *Noûs* 43(2): 286–314, doi:10.1111/j.1468-0068.2009.00707.x.

LEWIS, David. 2002. "Tensing the Copula." *Mind* 111(441): 1–13, doi:10.1093/mind/111.441.1.

MOURELATOS, Alexander P. D. 1977. "Events, Processes, and States." *Linguistics and Philosophy* 2(3): 415–434, doi:10.1007/BF00149015.

SATTIG, Thomas. 2003. "Temporal Predication with Temporal Parts and Temporal Counterparts." *Australasian Journal of Philosophy* 81(3): 355–368, doi:10.1080/713659704.

SIDER, Theodore. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press, doi:10.1093/019924443X.001.0001.

SIMONS, Peter M. 1987. *Parts: A Study in Ontology*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199241460.001.0001.

STEWARD, Helen. 2013. "Processes, Continuants, and Individuals." *Mind* 122(487): 781–812, doi:10.1093/mind/fzt080.

2727 STOUT, Rowland. 1997. "Processes." *Philosophy* 72(279): 19–27, doi:10.1017/s0031819
2728 100056631.
2729 —. 2016. "The Category of Occurrent Continuants." *Mind* 125(497): 41–62, doi:10.109
2730 3/mind/fzv138.
2731 VARZI, Achille C. 2003. "Naming the Stages." *Dialectica* 57(4): 387–412, doi:10.1111/j.
2732 1746-8361.2003.tb00279.x.
2733 WASSERMAN, Ryan. 2006. "The Problem of Change." *Philosophy Compass* 1(1): 48–57,
2734 doi:10.1111/j.1747-9991.2006.00012.x.

Robinson's Regress Argument from Vagueness to Dualism

Critical Notice to Robinson (2016)

Dean Zimmerman

Howard Robinson's *From the Knowledge Argument to Mental Substance* contains two quite different arguments from the vagueness of composite objects to the conclusion that I am not a physical object at all. One of them, developed over the course of several chapters, takes the following form: All composite physical objects (and only composite physical objects are candidates to be a human being) are non-fundamental; non-fundamental things are inevitably vague in various ways; this vagueness shows that we must "make a conceptual interpretation of them," treating them as "artefacts of conceptualisation"; and this in turn precludes our identifying ourselves with any such things. Some interesting morals fall out of close consideration of Robinson's argument; but, in the end, materialists can reasonably resist it.

Howard Robinson and I both find it problematic to identify a person with a vague object. We agree that all the sensible *physical* candidates for being a person are vague, and we do not shrink back from the radical conclusion: that we are immaterial thinking things. Although the arguments we give lead from vagueness to immateriality, they are very different in strategy.[1]

Part II of Robinson's *From the Knowledge Argument to Mental Substance* (2016) contains two main paths that lead from the vagueness of physical objects to the conclusion that I am not one. One important argument has to do with the indeterminacy of identity for physical objects under contrary-to-fact conditions, and the (alleged) determinacy of identity for minds under such conditions. I shall ignore this intriguing argument (found in Chapter 12), and focus on a different one. It is based on considerations that are spread

1 For my statement, and restatement, of an argument from the vagueness of candidate physical objects to substance dualism, see Zimmerman (2010, 2011).

among several chapters; putting the pieces together will take some time and effort. Much as I would welcome additional support for my own conclusion, I discover several plausible ways for materialists to resist Robinson's argument. Unless I am missing something (which I might be—the book is dense, the arguments complex), materialism faces little danger from this quarter.

I am a great admirer of Robinson's work, this book included. Here, however, I focus entirely on the part of the book to which I have the most objections. I show my admiration in classic schoolyard fashion—by punching him on the arm as hard as I can, and then running away.

1 Fundamentality, Nonfundamentality, and Vagueness

Robinson does many things in Part II of *From the Knowledge Argument to Mental Substance*. One of the main threads running through its chapters is support for the premises of the following argument, which shall be my focus: All physical objects (or at least all the ones that are any kind of candidate to be a human being) are "non-fundamental"; they do not belong to the "fundamental level" of reality. Non-fundamental things are inevitably vague in various ways; they are vague in their boundaries, and many of their characteristic properties are vague as well. This vagueness, he says, shows that we must "make a conceptual interpretation of them," treating them as "artefacts of conceptualisation." And this in turn precludes our identifying ourselves with any such things. Robinson's more exact formulations of this argument will be considered shortly. First, I try to clarify what Robinson means by "fundamental" and "fundamental level"; what it is for a thing to be an "artefact of conceptualisation," or for us to "make a conceptual interpretation of" something; and how vagueness is connected with these two ideas.

Robinson talks of "levels" and "ontologies," with different ontologies located at or constituting different levels. There is a "fundamental physical level" which he calls "basic physics" (leaving open what form that might take), and non-basic levels, some of which are the subjects of the special sciences.[2] This talk of levels is flexible; the term "level" (and "ontology") is open to (at least) two interpretations: it can refer either to a theory (in which case it is a "representational ontology"; that is, "a conceptual picture of the world") or to the entities that are the subject matter of the theory (Robinson 2016, 168).

2 See Robinson (2016, 168–169) for introduction of basic and non-basic "ontologies"; and Robinson (2016, 177, 180–181) for discussion of "levels," "basic physics" and the special sciences. "Basic ontology" and "basic level" seem to be equivalent.

In the latter use, "the ontology of the theory" simply refers to the ontological commitments of the theory—the things that would have to exist, were the theory true.

Robinson is using "level" in the first way when physics is described as one level, and the special sciences (biology, geology, meteorology, etc.) are said to represent higher levels (see e.g. 2016, 156–157, 220). Different levels will use a different terminology; higher level theories introduce terms not found at lower levels. The concepts these terms express, and the properties to which they refer, can be called "higher level" as well. The most fundamental theory will include terms for the most fundamental or basic properties. The properties of interest to the special sciences will be less fundamental, less basic, than the properties of interest to physics.

Robinson uses "level" in the second, ontological sense when he asks about "the causal efficacy of the non-fundamental levels" (2016, 181), and whether "at the fundamental level," there are just events in space-time, or enduring objects (2016, 208). If physicalism is true, the level of the "basic ontology" consists only of electrons, quarks, or whatever entities are the subject matter of fundamental physics; the level of biology comprises all the organisms and (at least some of) their parts; the level of geology includes boulders and lava flows, etc.

Although Robinson does not explicitly invoke David Lewis's notion of a "natural property," his views about fundamentality of levels, and the association of vagueness with higher levels, can be fairly explicated in terms of naturalness. Lewis uses natural properties for many purposes, but I invoke them here only in their role as resemblance-makers.[3]

Plato introduced the metaphor of "carving nature at the joints." Natural properties are posited as the joint-carving ones; each natural property represents a respect in which things can objectively resemble one another. But naturalness is not all-or-nothing. Schemes of classification, and the terms used in scientific theories, may be more or less natural. For example, all mammals resemble one another in certain respects, so being a mammal ensures some degree of objective similarity. But it does not ensure exact similarity with respect to any one precise feature. Being a mammal is much less natural than having a certain height or weight.

3 Lewis (1983) is the locus classicus on natural properties. For an overview of what they have been thought to do, see Dorr (2019).

doi: 10.48106/dial.v74.i3

Degrees of naturalness suggest the possibility (some would say the necessity) of a lowest level consisting of the most natural properties—what Lewis called the "perfectly natural" properties.[4] A property is perfectly natural if, and only if, it is responsible for one of the most basic respects in which things can objectively resemble one another—the respects of resemblance that "comprise a minimal basis for characterizing the world completely" (Lewis 1983, 346).

There is a clear connection between vagueness and naturalness. At least some vague terms, such as "bald" or "tall," are used to ascribe a degree of objective similarity among the things to which they are applied. So there are vague natural properties. But vague terms cannot represent *perfectly* natural properties. When things resemble one another in virtue of sharing a vague property, they resemble in a less-than-perfectly-precise way. The degree of similarity imposed by the resemblance must, then, be less than perfect.

2 Fundamental Objects, Garden Variety Objects, and "Artefacts of Conceptualisation"

Robinson's fundamental level, or "basic ontology," conceived of as a set of entities, consists of the "basic constituents of the world, not constituted by anything else" (2016, 168). Robinson's "basic constituents" are in the same line of work as Joshua Brown's "perfectly natural objects." Brown defines perfectly natural objects as all and only those that possess perfectly natural properties (2016, 260). This raises an interesting question for Brown and Robinson: is it true, as Brown assumes, that no composites—nothing that is constituted by further things—can have perfectly natural properties?

All objects, including the perfectly natural ones, will clearly have some less than perfectly natural properties. Suppose that electrons are perfectly natural objects, and that the precise mass and charge of an electron are perfectly natural properties. An electron also has the property of being less than one kilogram, and the property of being negatively charged—properties that clearly do not ensure a *precise* degree of resemblance among the things that

4 Schaffer gives reason to doubt whether the hierarchy of natural properties must have a bottom; see (2003). So far as I can see, everything Robinson wants to say about levels, reduction, etc. makes perfect sense so long as the hierarchy of resemblance-making properties includes a level below which there is no vagueness, and all higher-level resemblances supervene upon the distribution of these non-vague properties. It is not obvious that Schaffer's examples of non-atomistic physical worlds cast doubt on this assumption.

share them. Since electrons have them, it cannot be that perfectly natural objects have *only* perfectly natural properties. But Brown takes it to be at least likely that no composites have *any* perfectly natural properties (2016, 265); and, in particular, that organisms do not have any perfectly natural properties (2016, 259–260). If that is right, and similar things can be said for entities located at all the other levels that Robinson regards as higher (which includes all composites), then Brown's category of perfectly natural objects lines up nicely with Robinson's category of fundamental or basic things.

There are puzzles for Brown's proposal. Are there really no composites with perfectly natural properties? If so, the net mass of a composite, for example, must be less natural than the mass properties of the simple particles that make it up (Brown 2016, 259–260). Alternatively, one might allow for some perfectly natural composites (e.g. the universe as a whole, which may well have a precise finite mass), while denying that the objects of study in the (intuitively) higher-level special sciences have perfectly precise masses and other basic physical properties.[5] I prefer this more liberal approach: perfectly natural composite physical substances may well exist, but the kinds of things that Robinson regards as higher level entities—most saliently, human bodies and brains—are all vague, and do not have perfectly natural physical properties due to their vagueness.

Robinson will go on to argue that a person must be a perfectly natural object, a conclusion which provides considerable support for substance dualism. After all, the only really plausible physical candidates for being me are not perfectly natural. They are what I elsewhere call "garden variety objects" (GVOs): that is, material objects with spatial boundaries that are defined in terms of detectable physical discontinuities and functional roles that are significant (to us) (Zimmerman 2010, 136–137). To be a reasonable candidate, a GVO should at least include the brain or most of the brain, since it is the organ upon which our minds most directly depend. So, a brain, a nervous system, an entire organism, and perhaps even just one hemisphere of the brain—each is a decent candidate for being me, if I am a GVO. These things are reasonably well-demarcated in terms of physical discontinuity with surrounding matter, and functional unity; and they all include (all or at least half of) my brain. And all such objects are vague in their boundaries, and will not have precise

5 Brown does not completely rule out the possibility of fundamental composites, but argues that they run afoul of some plausible metaphysical principles (2016, 264–265).

doi: 10.48106/dial.v74.i3

masses, shapes, electrical charges, locations, or any other perfectly natural physical properties.

It is conceivable that I be a physical object that is not of the garden variety. Suppose there were some sort of special physical particle in my brain—either unlike all the others or uniquely located in a physical Cartesian theater (a venue designed for social distancing, with but a single seat). Discovering this particle might make me wonder whether I was, in fact, identical with *it*, rather than with some larger material object. On the tiny-particle hypothesis, I would be a material object, but I would not be a GVO.[6] However, we have good reason to doubt the existence of such unique physical things; the brain is made of the same gigantic numbers of a few kinds of fundamental particles as all the other pieces of "middle-sized dry goods" that surround us, and there is no central theater in which one particle could occupy a privileged place. If I am to regard myself as a material object, it had better be a GVO. I will restrict attention to just the most obvious candidate GVOs: namely, brains and entire human organisms (human bodies).

Such things are, Robinson says, non-fundamental, and they have non-fundamental properties. The properties figuring in the special sciences—most relevantly, biology and the human sciences—are vague and therefore far from perfectly natural. And he argues that, since these properties fail to be reducible to non-disjunctive more natural properties, non-fundamental objects and properties must be "perspectival":

> [...] the special sciences are best understood as different perspectives on the physical base, usually with certain interests in mind. They are essentially in the same category as patterns, because, though the concepts they involve are well grounded by the basis physical reality, they do not reflect any reality additional to [the] fundamental physical base, except the interests and other perspectives of the humans who employ them. (2016, 220)

Less-than-perfectly-natural kinds (including *brain* and *organism*) and their higher-level properties (their shape, size, chemical make-up, biological properties, and so on) are vague in ways that generate sorites paradoxes. According to Robinson, this is a sign that things falling under such categories and characterized by such properties should not be taken to exist in a "fully realist sense"

6 Roderick Chisholm and Philip Quinn took the tiny-particle hypothesis relatively seriously; see Chisholm (1978) and Quinn (1997).

(2016, 174). The contrast is with existing in a merely "conceptualist" sense. To say that a brain or organism exists in a merely conceptualist sense is to say two things: (i) "[T]he world [is] so organized that it satisfies this concept," which merely means that paradigmatic cases of someone's identifying a particular brain or human body manage to track *something* about the world. We are not just confused when we apply these concepts in some cases but not others. (ii) Nevertheless, "[i]f there were no conceptualisers around (putting God or Divine minds aside)," there would be no brains or organisms or other vague objects—they are mind-dependent "artefacts of conceptualisation" (Robinson 2016, 179). This is what Robinson means by "making a conceptual interpretation" or "making a CI" of a brain, organism, or other object (2016, 178)—it is to affirm their mind-dependence. To say that brains or organisms would be around, with or without conceptualizers, is to give them (or their existence, or their special science properties) a "realist interpretation."

3 Robinson's Regress Argument

The pieces are in place, then, for Robinson's argument that we are not GVOs. He sketches the argument at the end of Chapter 9 (2016, 159) and in Chapter 11 (2016, 179), referring the reader primarily to Chapter 13 (a criticism of Dennett) and the overlapping essay, "Quality, Thought and Consciousness" (2010). (The relevant material in the essay is included in the book, so I will refer just to the book.)

Much of Chapter 13 is aimed specifically at Dennett's instrumentalism about intentionality. Robinson develops a regress argument against Dennett's instrumentalism which he eventually extends to reach the conclusion that minds are not composite objects—recall that Robinson, like Brown, takes *all* composite objects to be problematically vague, like the things I call GVOs; he therefore assumes they all require a "conceptualist interpretation." As I noted, I am reluctant to say that absolutely *no* composites can have perfectly natural properties; perhaps the level of "basic physics" (whatever that turns out to be) includes some large things with parts. Nevertheless, the dualist in me would be excited enough by a successful regress argument that rules out all GVOs as candidates for being thinking things. Once they are eliminated, all the alternative candidate physical objects are highly problematic. So I shall treat Robinson's argument as targeting just GVOs.

The pithiest statement of Robinson's regress argument for dualism is this:

doi: 10.48106/dial.v74.i3

> [I]f all physical composites are artefacts of conceptualisation, and if the human being, brain, mind etc. are physical composites (and they are certainly not physical simples), then they are products of conceptualisation. What is it that does this conceptualising? Not something that only exists conceptually, on pain of a regress. (2016, 179)

Disentangling the argument from the assumption that *all* physical composites are infected with vagueness, I shall construe it as taking the following form:

1. All GVOs are vague (due to their non-fundamentality).
2. All things that are vague (due to non-fundamentality) are "artefacts of conceptualisation."
3. If we were GVOs, we would be "artefacts of conceptualisation." (From 1 & 2)
4. We cannot ourselves be "artefacts of conceptualisation" ("on pain of regress").

⍰ Therefore, we are not GVOs.

Granting that all GVOs are vague, the remaining premises are 2 and 4. In the remainder of the paper, I shall examine the reasons Robinson gives for accepting these two premises. As shall appear, there is much that materialists can say against them.

4 Support for Premise 2

According to Robinson, physicalists must regard not just mental states but all special science properties as "perspectival"—"different perspectives on the physical base, usually with certain interests in mind [...] This is a form of interpretationalism, which presupposes a mind picking out the *fundamenta* that make the higher-order explanations possible." And so physicalists "cannot avoid assigning an irreducible role to the mind in the creation of the non-basic physical levels" (Robinson 2016, 220). In other words, there would be no GVOs without minds to conceive of them.

This is premise 2, and its plausibility depends upon at least two theses: (a) non-fundamental truths would not be true, were there no minds taking up the perspective required to understand them; and (b) were there no creatures with our perspectives there would be no *other* minds (e.g. no divine mind)

capable of understanding them. If either (a) or (b) is false, premise 2 is in big trouble. I consider each in turn.

In his support of (a), Robinson allows that higher-level predicates (like those of the special sciences) can be appropriately attributed to objects because of "real patterns" that exist "out there."[7] But these "non-basic predicates," because of their vagueness, should be treated in a "conceptualist, rather than a realist, manner. Only basic predicates, and those reducible to basic predicates should be treated in a strictly realist way" (Robinson 2016, 175). If special science (or other higher-level) descriptive terms are not reducible to fundamental physics, they correspond to categories that are picked out from our perspective; which "seems to give the interpreting mind an irreducible role in the creation of these sciences" (2016, 158). This implies, in turn, that the objects that display these patterns only exist "out there" because "they are reified as being of a certain kind by an interpretative act" (2016, 158).

The obvious objection to this line of thought is simple and highly intuitive (and reminiscent of G. E. Moore's "Refutation of Idealism" 1903): Granted, minds are needed if these "patterns" are to be *noticed*. But the patterns themselves could *exist* whether or not they are noticed. The creation of a science—a scientific discipline, with its textbooks and methods—may be impossible without minds. But why think the entities and properties *described* by the science could not exist without minds (unless the subject matter includes minds)? In other words, why must one "make a conceptualist interpretation" of these entities and properties?

Robinson's reasons for thinking "that one should treat non-basic predicates in a conceptualist, rather than a realist, manner" (2016, 175) are spread throughout the second half of the book, and they are interrelated in complicated ways. I am not sure that I have fully disentangled them, or seen all the connections among them, but I can discern five distinct lines of argument. (1) The vagueness of the non-basic levels is supposed to show that they are implicitly describing the world in ways that require the existence of concept users. (2) The proper understanding of the autonomy of the special sciences should lead us to interpret biological kinds, for instance, as anthropocentric in ways that require "making a CI" of them (2016, 186–189). (3) The special (physical) sciences are alleged to describe the world in Newtonian ways that

7 I say "appropriately attributed" because Robinson claims that many terms from the special sciences, along with descriptions in terms of the manifest-image or the macrophysical, do not *truly* apply to anything (in virtue of the alleged fact that they presuppose false Newtonian views of space, time, and matter) (2016, 189–190).

doi: 10.48106/dial.v74.i3

are strictly false; and this requires interpreting the terms and ontologies of these sciences conceptualistically (2016, 190). (4) Conceptualism is justified by the fact that it solves problems of constitution (e.g. difficult questions about the relationship between a statue and the clay that constitutes it) and Unger's problem of the many (2016, 179–180, 190–191). And (5) a sparse theory of universals, such as David Armstrong's, only treats the basic predicates in a realist way, all others being "understood in the conceptualist way," which implies making a CI of non-basic things (2016, 175).

Of these, (1) and (5) strike me as the strongest, and I will devote the next two sections to them. As I mention, briefly, at the end of section 5, (2) seems to me to have the strengths and weaknesses of the argument from vagueness; so my response to (1) provides a response to (2). I shall not say much about (3), and (4). In these arguments, Robinson advocates a conceptualist interpretation of certain ways of talking about the world because they seem to be not strictly true. But suppose he is right: chemistry somehow ascribes Newtonian properties to things, descriptions of sculptures imply that the statue and clay are distinct, and platitudes about cats imply that there are many cats right where Tibbles is located. Robinson's strategy in all these cases is to deny that some higher-level statements are true, though they may be apt or appropriate for certain purposes. But that move is open to anyone bold enough to make it, and does not require the conceptual dependence of the subject matter.[8] So I find these arguments much less convincing than the more straightforward claim that, since vague language is really implicitly about us, we must "make a CI" of higher-level theories and ontologies.

5 Vagueness Requires "Making a Conceptual Interpretation" of GVOs

I take (1) to be the strongest of the five strategies. Robinson uses the paradoxes of vagueness to forge a link between non-fundamentality and "making a

8 For example, Robinson says that the special sciences provide us with concepts that are "workable," but not strictly true of the phenomena they are meant to describe. The properties corresponding to these concepts are "ever so slightly inaccurate and, perhaps, false in their fundamental nature of the objects in question" (2016, 190). If a conceptual interpretation of these sciences does not make non-fundamental claims true (and I nowhere find Robinson claiming that it does), I see no advantages to "making a conceptual interpretation" of these claims rather than a realist one—i.e. holding that they are approximately true, but strictly false and (here is the realist part) would still have been false-but-approximately-true even had there been no concept users.

conceptual interpretation" of something. When paradoxes arise, he says our standard practice is to discard non-fundamental talk for more fundamental talk. This practice is supposed to be a sign that vague terms are implicitly about our willingness to ascribe them. Here is a very brief summary of the long, intricate argument of Chapter 10.

How should one respond to the apparent non-bivalence of claims made in vague language, or to the sorites paradoxes generated by vagueness? Not by adopting a non-classical logic, nor by epistemicism, nor by supervaluationism. Look instead to what we actually do: we use normal, bivalent logic with vague terms

> until the vagueness becomes salient; then we either contrivedly precisify them for the present purpose, or move to another discourse that is not vague under the relevant circumstances. […] When the vagueness intervenes, the discourse is either modified or suspended, so that normal logic can once again be deployed. (Robinson 2016, 171)

In other words, vague talk is sometimes inadequate to the expression of the facts, in which case it is thrown out in favor of something more precise, and therefore closer to the basic level.

This practice shows that natural languages do not constitute what Robinson calls a "Logical Unity." The inferential relations among all the propositions expressible in English, for example, cannot be captured in a single formal system because "there appears to be no canonical way of representing the logic of vague predicates" (Robinson 2016, 165).[9] The truly basic level is presumed to admit of "a characterization […] which is free of inconsistency and which can be regimented according to some canonical form" (and "one might hope that this can be a classical two-valued logic") (2016, 166). Replacing vague terms with more precise language (because of looming paradox or failures of bivalence) is interpreted by Robinson as admission that vague talk is "not to be taken as realistically" as statements that could be made in the perfectly precise language of the basic ontology. When we speak truly using vague terms, we offer "an ontologically sketchy way of seeing the world"—it may be true, but is at best "a view, an appearance, a kind of secondary quality of the underlying reality" (2016, 168).

9 Robinson prefers to think of vague statements as expressing propositions, though he is open to the idea that it is vague whether the thing expressed is a proposition; see (2016, 172–173, note 8).

doi: 10.48106/dial.v74.i3

Why exactly does Robinson think that the "ontological sketchiness" of the language used to describe some subject matter requires that we give descriptions in that language a "conceptualist" interpretation? Why, that is, should the vagueness of the language lead us to conclude that the propositions expressed using that language are really propositions about *the way we view things*—therefore implying that persons exist? I find only one explicit source of support for this connection: it is in his treatment of sorites paradoxes, like the paradox of the heap. In order to resist reaching the noxious conclusion at the end of a sorites argument (that a single grain is a heap), one should understand every occurrence in the argument of sentences like "*n* grains constitute a heap" as meaning the following: "*n* grains can properly be *seen or conceptualized as* a heap" (Robinson 2016, 174). Robinson then provides a way of resisting the argument for the conclusion that one grain constitutes a heap, and his strategy turns upon substitution of this psychological description in place of "heap."

If all vague predicates were, implicitly, about how human beings view things (and if satisfying such predicates implied that some human being exists—an important caveat), then the connection between the vagueness of higher levels and conceptualism about those levels would be reasonably clear. If "heap" means or is otherwise equivalent to "a thing someone conceptualizes as a heap," then there would be no heaps unless there were minds capable of applying concepts. As a meaning equivalence, this does not seem very plausible; "that is a heap" does not seem, even implicitly, to be a statement about concept-users, and "heaps of sand exist that no one conceptualizes as heaps" has the ring of truth.[10] Moore's anti-idealist response seems perfectly reasonable here.

It is not even clear that, given Robinson's proposed meaning for "heap," there could be no heaps without concept users. If "that is a heap" means "that is something that *could* properly be seen or conceptualized as a heap (were there creatures like us around to do so)," then heaps exist in worlds

10 There is also a puzzle about how to understand the suggested meaning equivalence. Suppose one replaces "heap" with the proposed meaning. "That is a heap" then becomes: "That is something that can properly be seen or conceptualized as a heap". Replacing "heap" with its meaning in this sentence yields: "That is something that can properly be seen or conceptualized as something that can properly be seen or conceptualized as a heap"—which again should allow for substitution of the proposed meaning for "heap," if the first usage licensed this. Either it is impossible to fully spell out the meaning of the sentence, or doing so yields something infinitely complex. Clearly, "conceptualized as a heap" will need to be understood in some other way, perhaps by means of semantic ascent.

without concept users; one need not "make a CI" of heaps. For this meaning equivalence to imply the mind-dependence of heaps, the truth of a statement of the form "such-and-such can properly be conceptualized as a so-and-so" must require the actual existence of someone, somewhere, with the mental equipment to apply the concept of a so-and-so. In other words, the "can" here must not be interpreted in a highly abstract, "in principle" way; this form of words does not mean "were there, perhaps *per impossibile*, someone around to contemplate the such-and-such, they ought to (or at least are not obliged *not* to) conceptualize it as a so-and-so."

For the proposed meaning equivalence to show that vague terms imply the existence of concept users, Robinson would have to assume that "can," when used in his definitions of vague terms, satisfies a principle along these lines: if it is true, in a possible world, that a such-and-such can be interpreted as a so-and-so, there must be someone who exists in the world in question and who (in some world or other) conceptualizes a such-and-such as a so-and-so. Applying this principle to heaps, one can conclude that it is false that there would have been heaps of minerals and clouds of gas in any universe utterly hostile to living things or other concept users, even if some matter there is arranged in such a way that we would not hesitate to describe it as a heap or cloud.

Although this may sound like the result Robinson wants, it is not quite enough for the conclusion that GVOs like heaps and clouds cannot exist without minds. Perhaps things which *actually* are heaps could exist without being heaps. That might seem to violate the appealing principle that "heap" is a sortal term, and that such terms pick out the essential kinds to which things belong. But the assumption that "heap" corresponds to an *essential* sortal kind is not so obvious when one realizes that it is (according to Robinson) a very extrinsic predicate, encoding facts about how human beings would react to something. Maybe "electron" corresponds to an essential kind, but "electron humans may someday detect" does not—some electrons that we may someday detect also exist in worlds without humans, in which they do not satisfy this description. On Robinson's proposal, "heap" is more similar to "electron humans may someday detect" than it is to "electron." This provides reason to think that, if Robinson is right, a thing that is actually a heap could exist in worlds without concept users—it just would not then be a heap, though it would be intrinsically just as it is in the actual world.

But how plausible is the original claim about meaning equivalence? What is wrong with saying instead that "is a heap" is equivalent to the alterna-

doi: 10.48106/dial.v74.i3

tive I suggested above: "is something that *could* (perhaps *per impossibile*) be conceptualized as a heap, were there someone around to contemplate the question"? Robinson's strategy for resisting the sorites for heaps does not, so far as I can see, depend upon his proposed meaning equivalence as opposed to this alternative suggestion. He points out that "it is a psychological matter whether making some small change affects a subject's inclination to classify an object in a certain way" (2016, 174). He then describes what would happen in any concrete case involving a single subject being asked repeatedly whether something is a heap, while grains are removed one by one: over the course of many removals, confidence in calling the result a heap will decrease, until at some point the subject will refuse to apply the term. This is meant to falsify one of the crucial premises in a sorites: that removal of a single grain can never make a difference to whether a thing is a heap. But everything he says here could just as well be said if "heap" is understood as "*could* be seen as a heap." And the latter interpretation has this going for it: it does not lead to the counterintuitive conclusion that heaps and clouds cannot exist without minds to apply concepts to them.

To sum up then: (1) has not convinced me of the connection between vagueness and "making a CI" of GVOs. A Moorean response to the claim that vague predicates are implicitly about concept users seems plausible to me. And even if Robinson's strategy for responding to sorites paradoxes is enough to motivate the idea that vague terms are implicitly about concept users, the way in which concept users are involved need not support his conclusion.

These reflections can be applied, *mutatis mutandis*, to strategy (2). Robinson claims that the anthropocentric interests that lead us to pick out special science kinds like "star," "planet," "dog," and "cell" are baked right into the meaning of these terms in such a way that there would be no planets or cells without concept users to identify them (just as there would be no heaps without heap-identifiers). Similar responses seem to me justified: the identification conditions for these kinds are no part of the meanings; but even if they were, it would not imply that the things that have them co-exist with minds essentially, since these terms should no longer be thought to correspond to essential sortal kinds.

6 Sparse Realism Requires "Making a Conceptual Interpretation" of GVOs

Strategy (5) turns on Robinson's nominalism about non-fundamental properties. He seems to agree that the metaphysician needs a realist theory of universals for the perfectly natural properties. But he advocates a "sparse" realism, limited to just the perfectly natural. For non-fundamental predicates like "table," "brain," "organism," etc., there need only be corresponding *concepts*. And concepts are mind-dependent things (Robinson 2016, 159, 183). I conjecture that Robinson further believes that, if the constituent concepts did not exist, the propositions in which such concepts figure would not exist either, and therefore could not be true. (Nor, presumably, could they be false.) This implication of his nominalism is suggested by the following remark about the non-fundamental category of tablehood (I note that this is not just a doctrine about artifactual kinds; Robinson clearly intends his moral to apply to non-fundamental categories from the special sciences, including biology, meteorology, etc.): "If there is no tablehood [i.e. if it is not a universal, but merely a concept], there are no instances of it, and so there are, in the realist sense, no tables" (2016, 183). The last clause—"there are, in the realist sense, no tables"—implies, according to his definition of existing in the "realist sense" (2016, 178), that it would not have been true that there are tables had there been no thinkers to use the concept of a table. Why should tables, organisms, mountain ranges, etc. be able to exist in the actual world, but not in worlds without minds; while quarks, electrons, etc. are not mind-dependent in this way? Robinson here seems to be offering sparse realism as the explanation: the fundamental physical entities are instances of real universals, which are *not* mind-dependent; while the non-fundamental things merely satisfy concepts, which *are* mind-dependent.

For this difference to explain why the non-fundamental things are mind-dependent, Robinson must be assuming that the propositions we grasp using these mind-dependent concepts would not be true in worlds without minds. After all, if the propositions about non-fundamental matters existed without us, then some of them, in some mind-free worlds, would have to be true. For example, suppose all life on Earth had been destroyed 700 million years ago, and no minds evolved elsewhere. There would still have been plenty of things we would call organisms, though none with minds. If basic propositions about cell biology would have *existed* in those circumstances, they would *have* to be true (after all, they correctly describe the organisms that existed 700 million

doi: 10.48106/dial.v74.i3

years ago). So Robinson has to think that these propositions themselves are infected with the mind-dependence of their constituent concepts; and that propositions representing the world entirely in terms of universals are not mind-dependent.

Does the mind-dependence of the concepts we use to refer to and describe GVOs really imply that, were there no minds, there would be no GVOs—e.g. that there would have been no trees or continents had there been no users of the concept of a tree or continent? Since the conclusion is quite unintuitive, all ways of avoiding it are worth exploring. Fortunately, there are many stations at which one can disembark before reaching Robinson's destination. The assumption of sparse realism will naturally be questioned by many; as will the mind-dependence of concepts for the non-fundamental. But even if one grants Robinson his preferred metaphysics of properties and concepts, there remain ways to resist the slide from the mind-dependence of the concepts used to think about a certain subject matter to the mind-dependence of the subject matter itself.

Some philosophers of language appeal, in a metaphysically serious way, to propositions as abstract entities expressed by sentences and grasped by means of concepts; and they affirm that all propositions are necessarily existing things. They might, like Frege, take concepts of all sorts to be necessarily existing things as well; but that is not the only way in which one could maintain this view of propositions. Perhaps concepts are like words: words are contingent things that can be used to express propositions that are not themselves about the words used to express them, including propositions that could have been true even if those particular words never existed. Why could concepts not be similarly related to the propositions they enable us to grasp?[11]

But there are ways to resist (a) even if one denies the necessary existence of propositions. Suppose that it is contingent which propositions exist; that existing propositions must be constructed out of existing materials, and that the non-existence of an individual or a concept prevents the construction of propositions explicitly about that individual or explicitly involving that concept. There remains room to make a distinction between the way in a which *we* can truly describe a circumstance, given the resources for constructing propositions that actually exist, and the ways in which the circumstance could have been truly described *had it been the case*. Given Robinson's assumption

11 For a survey of reasons to believe in propositions, and in their mind-independence and necessary existence, see McGrath and Frank (2018).

of the contingent existence of (many) propositions (e.g. the ones that are not just about necessarily existing universals), there may be propositions available *to us* that truly describe a non-actual circumstance, but which simply would not have existed had that circumstance obtained. Consider again the world in which the Earth is rendered uninhabitable before sentience evolves. We naturally still want to use plate tectonics to describe that possible history of our planet; there would still have been, for example, continental drift. Had that alternate history occurred, there would have been no propositions about continents and plates, on Robinson's hypothesis, because no one would have been around to take up the perspective from which plate tectonics can be used as an explanatory theory. Still, one wants to say, our description of this counterfactual circumstance is not *false*.

Singular propositions about non-existent individuals pose the same puzzle, under the supposition that such propositions are existentially dependent upon their constituents. A world without Julius Caesar can truly be described *by us* as lacking that very man; but, had that world been actual, no such description would have been possible; the propositions needed to express it would not have existed.[12] Taking this approach to singular propositions has led some philosophers to distinguish between "inner truth" and "outer truth" with respect to a possible world. The inner truths relative to a world are the propositions that would have existed and been true at that world, had it been actual; the outer truths relative to a world include also propositions available to us that (in some hard to specify sense) truly characterize the world, but that would not have been true had that world been actual due to the non-existence of those propositions.[13] If such moves are needed, and available, for the case of singular propositions, they should suffice to undermine (a) as well.[14]

7 The Divine Mind and Assumption (b)

I have surveyed five of Robinson's reasons for (a): the thesis that, were there no minds around, there would be nothing that satisfies what we mean by "organism" or "brain," and nothing would have the properties we attribute

12 For a famous early defense of the contingency of singular propositions, see Prior (1960).

13 For seminal presentations of the "inner truth–outer truth" distinction, see Adams (1981) and Fine (1982, 66).

14 McGrath and Frank (2018, sec. 7.2) describe the use of the "inner truth–outer truth" distinction to resist the necessary existence of propositions in a way that would invalidate (a). They find this strategy used by Pollock (1985) and King (2007, 80–95).

doi: 10.48106/dial.v74.i3

to such things. I take (1), the argument from vagueness, to be the strongest. But it seems eminently resistible; and the others strike me as less compelling. Insofar as premise 2 depends upon the truth of (a), it appears to be in trouble.

premise 2 depends also upon the plausibility of (b): were there no creatures with our perspectives, there would be no *other* minds (e.g. no divine mind) capable of understanding propositions about the non-fundamental. This is not at all obvious to those—like myself and, for that matter, Robinson—who believe in something like the God of most monotheistic religions: a being endowed with both intellect and necessary existence. Had God not created us, God still *could* have done so, and ought therefore to have known what we would have been like, and what sorts of concepts we would have employed. If God cannot know what any non-fundamental phenomena would be like without knowing it in virtue of creating sufficiently intelligent creatures to apply the concepts of the higher level, then God would have to create blindly, to some extent.

The only rationale I can think of for maintaining (b), while accepting the existence of God, would come from emphasis upon the difference between divine and human intellection: perhaps God lacks the ability to think less than perfectly definite thoughts. As John Hawthorne points out, this has radical consequences:

> […] [S]ince our semantic and psychological concepts—means, refers, believes, loves and so on—are vague, we could not on this view coherently think of God as believing that we mean anything, refer to anything, believe anything, or love anything. (2005, 23, n. 12)

I should not like to go so far as that; and so I trust that a perfect being could understand imprecise thoughts. But then (b) looks clearly false, given theism.

All in all, then, I find many reasons to doubt premise 2.

8 Support for Premise 4

But suppose that premise 2 passed muster. Suppose that minds must exist in order for vague objects to exist. Could premise 4 be resisted?

According to premise 4, if the non-fundamental levels—including theories about the behavior of organisms and brains—are to be interpreted conceptualistically, "mind itself cannot be one of those non-basic levels" (Robinson 2016,

220). The reason mind cannot have an irreducible role in creating the levels and also belong to one of the levels is the viciousness of a certain regress.

> [I]f all physical composites are artefacts of conceptualisation, and if the human being, brain, mind etc. are physical composites (and they are certainly not physical simples), then they are products of conceptualisation. What is it that does this conceptualising? Not something that only exists conceptually, on pain of a regress [...]. (Robinson 2016, 179)

Robinson's idea here is that something cannot have the power to generate a level—of objects and their distinctive higher-level kinds and properties—while belonging to that very level, and exercising this power in virtue of the higher-level properties appropriate to that level. If thinking things are themselves higher-level kinds, they are mere "patterns" which require "mental activity to reify them." Reifying oneself would be a problematic kind of bootstrapping, and being reified by other concept users, who are in turn reified by others, etc., would lead to a vicious regress in which no one is reified.

To see whether the regress is truly vicious, I shall explore Robinson's description of it in greater detail. He claims (2016, 219) that it is the same regress that afflicts Dennett's "interpretationalism" about minds—the thesis that all minds display intentional states only "instrumentally, i.e. by interpretation" (2016, 213). As shall appear, Robinson's anti-Dennett regress is not precisely the same as the regress Robinson invokes in the argument under consideration, which I shall call the "anti-GVO-materialism regress." I will make the case that, although the anti-Dennett regress may be vicious, the anti-GVO-materialism regress is not obviously so.

If thinkers are non-fundamental, the concept of a brain or organism is that of a certain "pattern" in the fundamental physical world; and if the distinctive properties of thinkers are non-fundamental, thinking itself—propositional attitudes and other mental states—must be mere patterns, as well. These biological and psychological categories supervene upon fundamental physics (if the physicalists are right) even if they are not identifiable with something more fundamental. Robinson claims that, since they are supposed to supervene upon the fundamental, they are the kinds of patterns that would not have any distinctive effects, were it not for a mind that recognized them or interpreted them (see 2016, 220); and we should therefore give them all a "conceptualist interpretation." The "grounds" for a pattern may exist, he says;

doi: 10.48106/dial.v74.i3

but if it is not strictly identifiable with something fundamental, it does not *automatically* exist; there is "the need for mental activity to reify [the pattern] on the basis of those grounds."

> This explains why the mind cannot be just a pattern: it is presupposed by patterns as their co-inventor, together with the grounding. If the mind itself [were] just a pattern, then there would be the kind of regress with which we started our discussion, for it would not be reified unless it were seen as a pattern, and so on. (Robinson 2016, 219)

The anti-Dennett regress and the anti-GVO-materialism regress are, I think, importantly different. The target of the former is the intentionality of a system. Dennett denies that a brain, for instance, can be "intrinsically intentional." Robinson assumes that there are only two ways to become an intentional system, either by intrinsically being one or by extrinsic interpretation; and that something "cannot have the capacity [to interpret something as an intentional system] solely in virtue of being itself interpreted by something else" (2016, 213). The resulting regress does seem to me to be problematic, given Dennett's full position. However, it is not quite the same as the regress confronting the defender of a "conceptualising interpretation" of GVOs who takes thinkers themselves to be GVOs. I shall argue that the regress (or bootstrapping circularity) involved in supposing that GVOs generate the concepts on which they existentially depend is not so obviously vicious.

Dennett (on Robinson's reading) denies the intrinsic intentionality of physical systems for very special reasons—he is averse to the "magic" of intrinsic "aboutness."[15] "Aboutness" must then come from an extrinsic source. If that source is an interpreter, who must already be able to think about things, it seems we may well be off to the races on a vicious regress. For X to be intentional, someone else S must interpret X's states; but for S to be able to interpret states, S must itself be an intentional system; and so someone else must interpret S's states. And so on, either in a regress or a circle, neither of which seems promising. The buck of "bad voodoo magic" is being passed, but it is never transformed into "good physicalist mojo."

In Robinson's anti-Dennett regress, X cannot give itself a certain property, but must rely upon something else's being related to it in a certain way—and

15 For Dennett's position, Robinson refers us to Dennett (1987, 13–35); I take no position here on whether Dennett has the means to fend off Robinson's criticism.

in a way that requires that further thing to have the same kind of property, but again only in virtue of someone else standing in the same sort of relation to it, and so on. In Robinson's anti-GVO-materialism regress, the problem is quite different. A whole family of things, the Xs, (e.g. all organisms) are supposed to depend for their existence upon a further thing, Y (in this case, the concept of an organism). But the existence of Y itself is supposed to depend upon some of the Xs (at least some of them must have the concept). A direct circle of dependence is supposed to be vicious, and turning it into a chain is supposed to generate a vicious regress. There is a similarity between the two regresses, in that each involves the having of concepts; but it does not seem to me to go much deeper than that.

One of the big differences between the regresses is that, as Robinson understands the anti-Dennett regress, each X, in order to qualify as an intentional system, must get its intentionality from something; since it cannot get it from itself, it must get it from a distinct Y which is itself an intentional system interpreting X as intentional. In the anti-GVO-materialism regress, however, what X needs, in order to exist, is the existence of a thing, Y; Y itself is dependent, and (on pain of circular dependence) is supposed not to be able to depend upon X; so it must depend upon something else, some other Z which deploys the concept Y—but Z need not stand in any interesting relation to X. Z does not need to "interpret X as a Y," or interact in any way with X. If X is a human brain, then what it needs in order to exist is not that any particular thing think about X (or think about the fundamental stuff grounding X) in a certain way, nor that X (or the fundamental stuff grounding X) be in any significant way related to a mind capable of categorizing it; nor that X itself be capable of thinking of itself, or thinking at all. When Robinson says that there would have been no brains or organisms or species had there been no minds to take our sort of perspective on the world, he is not saying that we have to think about each brain or organism or species in order to bring it into existence; he may be an idealist, but he is not what one might call a "Truman show" idealist—someone who disbelieves in all but the GVOs that individual human beings have actually encountered and conceptualized. Robinson nowhere endorses such an extreme position.

Nor need he. The five arguments I surveyed, above, for "making a CI" of organisms (and other higher-level entities) do not support the conclusion that each individual organism (or what-have-you) must be recognized and "thought into existence" by some human concept-user—that each animal must be paraded before some Adam, somewhere, on pain of non-existence. They

doi: 10.48106/dial.v74.i3

are arguments that humans must exist *somewhere* in order for organisms to exist *anywhere*. Take strategy (1), which requires that vague kinds, like heaps, only exist if there are concept-users who can recognize a heap: so long as there are minds capable of taking our sort of perspective on the world, some matter piled in a heap *could* be recognized as such, and so *is* a heap—even if that particular parcel of matter is never in fact recognized as a heap by anyone. Or take strategy (5), which turns on the contingent existence of propositions involving higher-level concepts: so long as the concepts of organisms and brains exist, propositions about organisms and brains are available to be true—including propositions that truly describe organisms and brains no one happens to notice. So all that is required for *all* the brains and organisms to exist is for someone, somewhere, to have the concept.

The regresses may be different, but they might both be vicious nevertheless. However, there is reason to be suspicious of the form of the anti-GVO-materialism regress, since some respectable metaphysical positions imply that it is benign. The doctrine resembles a less radical and quite popular view about universals—defended, for example, by David Armstrong—according to which there can be no uninstantiated universals.[16] Suppose electronhood is a genuine universal, present in all electrons; and that every electron is essentially an electron. It is at least tempting to say that the electron depends for its existence upon electronhood, since it could not exist without exemplifying it. But, according to Armstrong's theory, the existence of the electron is also sufficient for the existence of the universal; and, if it were the only electron ever to exist, the existence of the property would depend counterfactually upon the existence of this electron. This sets up exactly the kind of circularity that is supposed, by Robinson, to be vicious and to generate a regress which is itself vicious.

How does Armstrong deal with the apparent two-way dependence of properties on things that have them, and the dependence of things on their essential properties? I cannot find him directly addressing the question in these terms, but there are some suggestive passages, and some obvious moves available. For one thing, the kinds of (alleged) dependency seem quite different; so there may be no circularity at all, or only circularity of a benign sort. For example, he might well say that, although an electron cannot exist without being an electron, that does not mean it depends for its existence upon the

16 See Armstrong (1989, 75–82). Robinson himself seems attracted to Armstrong's sparse theory of universals (2016, 175, 183).

property of being an electron. Not everything that must exist, if I exist, is something I am dependent upon; some such things are dependent upon me, but accompany me necessarily (e.g. my unit set, which automatically shows up if I do, but arguably depends upon me, and not the reverse). Armstrong could then say that, necessarily, electronhood exists if any electron exists; but that the property is dependent upon its instances.[17]

Alternatively, when a circle of dependence threatens, one may posit shared dependence upon a further thing. Necessary connections exist between the two entities, but these are signs not of a circle of dependence, but of mutual dependence upon something more fundamental yet. Armstrong's later work on particulars and universals treats both as abstractions dependent upon—or, to use a phrase suggesting both dependence and greater fundamentality, *grounded in*—things of a further ontological category: states of affairs. Proposing that universals are "state-of-affairs types [...] brings out the dependence of universals upon states of affairs. As such, it should at least incline us to accept the primary position of states of affairs and to be sceptical about the reality of uninstantiated universals" (Armstrong 1997, 29).

I shall suggest that the anti-GVO-materialism regress can be defused in the second of these two ways: the existence of both organisms and the concept of an organism are dependent upon more fundamental facts. Take some organized fundamental physical activity that is sufficient for the existence of an organism—at least, activity that is sufficient in a world in which someone, somewhere has the concept of an organism. According to premise 2, the activity in question has the status of (constituting) an organism in virtue of someone, somewhere, having this concept. Now, if this "someone, somewhere" is itself an organism, does it not exist in virtue of its own ability to deploy the concept of an organism? If it is the first and only organism employing the concept, it would seem so. But even if there are many organisms with the concept, take them all as a group; are they not pulling themselves up by their own bootstraps?

Because both concept possession and organism status are—on physicalist assumptions—dependent upon (or grounded in) more fundamental physical facts, this does not seem to me to be anything like the kind of bootstrapping needed to escape the anti-Dennett regress. Why could not *both* concept pos-

17 If inability to exist on one's own is a mark of dependence, this response could be read into some of Armstrong's remarks, such as: "a property is a way that a thing is," and "[a] way that things are could hardly exist on its own" (1989, 96–97).

doi: 10.48106/dial.v74.i3

session and organism status be dependent, simultaneously, on the same more fundamental facts?

To undermine premise 4, a critic can maintain that it is intrinsic to the matter making up a properly organized human brain or body that it grounds the existence of a GVO and also that it grounds the exemplification of concepts for GVOs, like the concepts of a brain or animal body. *Something's* having such concepts is, on Robinson's view, necessary for the existence of brains and bodies. But it is the mere existence of the concept of a GVO that is required for there to be truths about that kind of GVO; the possessors of the concept need not be applying it to themselves or other concept-users in order to "bring themselves into existence." So long as the concept is available and the matter of the world is arranged in a way sufficient for application of the concept, there will be truths about such GVOs, whether or not someone applies them. This seems very different from Dennett's intentionality regress—no act of recognizing myself as an organism, nor any other interpretative act aimed at myself, is required for the possibility of the truth that I am an organism, only my (or someone else's) ability to deploy the concept—to take up the human perspective. And this seems sufficiently grounded in the physical—at least on physicalist assumptions about what it is to have a concept.[18]

An analogous case of an aesthetic property will help illustrate how such simultaneous grounding can occur. The example is particularly apt, because the aesthetic property is meant to be very much like the concept of an organism, given a conceptualist interpretation of organisms. The property exists

18 In conversation, Robinson has suggested that he would resist at this point, arguing as follows: For a concept to exist it must actually be deployed by a thinker. It may be possible to describe, in physicalistically respectable functional terms (for example), what it is to take the presence of an organism "on board" in one's practical and theoretical reasoning. And one might think that if some system of particles or hunk of matter satisfies such a description, then someone has the thought we would express as "There's an organism!"—and that the concept *organism* would therefore exist. But Robinson thinks that it is not a necessary truth that, when a collection of particles or a hunk of matter satisfies these kinds of physical-functional descriptions, there exists something that actually thinks, or deploys concepts. (Establishing this claim is one of the goals of the first half of *From the Knowledge Argument to Mental Substance*.) I am inclined to agree, out of a shared antipathy toward functionalism and other forms of physicalism about the mental. It is a nice, and difficult, question whether Robinson is right about this: whether denying physicalistic accounts of thinking and concept possession saves premise 4, rendering the anti-GVO-materialism regress vicious. Even if the physical-functional description of the matter making up my nervous system does not entail the existence of a thinker, it might nevertheless be *causally* sufficient for some further element—say, phenomenal consciousness—which, together with the purely physical facts serves as sufficient grounds for both the existence of a thinker and that thinker's possession of concepts.

contingently; it can, arguably, be exemplified essentially by something, and also depends (at least counterfactually) for its existence upon that very thing; and it also applies to things (namely, passages of literature) that could have existed. In these ways, it is more like *organism* according to Robinson than *electronhood* according to Armstrong. Collections of material parts like those in an amoeba could have been arranged just as they are but fail to be an organism due to the non-existence of certain other organisms, namely, human beings. Nothing like that could happen with electronhood; anything intrinsically just like an electron must be an electron.

In "Kafka and His Precursors," Jorge Luis Borges argues for the contingency of a certain aesthetic property, which could be called *Kafkaesque*. After detecting Kafka's "voice, or his practices" (1964, 199) in a number of literary characters, themes, and passages that predate Kafka's own work, Borges draws a conclusion about the property they have in common:

> If I am not mistaken, the heterogeneous pieces I have enumerated resemble Kafka; if I am not mistaken, not all of them resemble each other. This second fact is the more significant. In each of these texts we find Kafka's idiosyncrasy to a greater or lesser degree, but if Kafka had never written a line, we would not perceive this quality; in other words, it would not exist. (1964, 201)

According to Borges, without Kafka, a host of passages throughout the history of world literature would not have had the property of being Kafkaesque. The words in the passages to which he draws our attention would still have been there, but they would not have been marked by this characteristic. In order for them to resemble one another in this particular way, Kafka (or someone with Kafka's sensibility) had to write a body of literature sufficient to bring the property into existence.

The relevance of Borges's theory for premise 4 can be seen by noting that the body of literature that is responsible for bringing this property into existence (*The Trial*, *The Castle*, etc.) would itself have the property. Is there anything circular about supposing that what Kafka did—the writing of the words he wrote—both brought passages of literature into existence *and* created the property of being Kafkaesque which they exemplify? It does not seem so to me—even if those passages (in *The Trial*, *The Castle*, etc.) are *essentially* Kafkaesque, so that they could not have existed without the property.

The example seems perfectly analogous to the supposition that both the existence of a human organism and the existence of the concept of an organ-

doi: 10.48106/dial.v74.i3

ism be grounded in further, more fundamental physical facts. The complex activity within a brain is like Kafka's putting pen to paper, and the concept of a brain or organism is like the property of being Kafkaesque. The structure and functioning of the cells that make up a particular human brain is enough to ensure the existence of certain GVO concepts, like that of a brain or an organism (so long as the brain in question subserves the activity of thinking about brains and organisms). That same arrangement of living cells is sufficient to ensure the existence of a GVO that falls, essentially, under one of those concepts (namely, the concept of a brain). This is no more paradoxical than the idea that Kafka's writing creates the property of being Kafkaesque and also creates a piece of literature that is essentially Kafkaesque.

I am considering the possibility of resisting premise 4 by agreeing that the existence of organisms and other GVOs depends upon some organism (or brain) using these concepts; call this position "conceptual dependence." The view gives the concept-using organisms a special role among all the instances of things that fall under the concept: had the concept-users not existed, and no replacement thinkers been introduced to cook up the concepts instead, all the other instances of GVOs would not have existed either, even had the matter that makes them up remained just as it is. I grant that there is something vertiginous about this supposition—and a precisely parallel sense of vertigo is created by Borges's theory of the Kafkaesque (which is, of course, part of its typically Borgesian charm). It is illuminating to consider how Borges's aesthetic property can be made to seem less paradoxical: doing so will shed light on how this way of denying premise 4 could be made to seem less strange as well.

It is natural to think of aesthetic properties as intrinsic to the things that have them.[19] If several works of literature resemble one another in a certain intrinsic respect, any one of them could have ceased to be without the others losing the property in virtue of which they resemble—that is part of what it is for the property to be intrinsic. How is it that Kafkaesque is not like this? If it were covertly relational, involving similarity to an aspect of *this particular body of work* (*The Trial*, *The Castle*, etc.), then the dependence would no longer be surprising; the aesthetic property would not be entirely intrinsic after all. It would be partly relational, depending upon *both* the intrinsic features of various works that have it (their proto-Kafkaesque elements, one might say),

19 "Kafka and His Precursors" is not the only place where Borges produces a counterexample to the intrinsicality of aesthetic properties; see also "Pierre Menard, Author of Don Quixote" (1962, 45–55).

and their relation to a property of Kafka's works that is unique to Kafka (the particular ways in which his work is Kafkaesque). Someone who accepts the doctrine of conceptual dependence (for GVOs) thinks that the matter actually constituting an organism, brain, or other GVO could exist in an intrinsically similar state without constituting such things (so long as the matter does not *also* constitute a thinker with the concepts in question). A person who holds this view seems to regard being an organism, brain, etc. as covertly relational; it requires not only an intrinsic ground for the application of the concept (matter arranged as it is in living things), but also the existence (somewhere, at some time) of at least one instance of a person utilizing the relevant concept. The contingently existing concept plays the role of the unique property of Kafka's writing.

I do not find the doctrine of conceptual dependence at all attractive, for reasons that will be obvious from my criticism of premise 2. However, those who accept it are, in effect, saying that *being an organism* is covertly extrinsic and relational in much the same way as Borges's *Kafkaesque*.

I see no inherent instability, then, in holding that an arrangement of matter might be sufficient both to generate a concept of a certain kind and the intrinsic grounds for the application of that concept—in much the same way that Kafka's words could be arranged in such a way as to generate both a new property shared among many literary works, and also an instance of a work that essentially has that property.

9 Conclusion

As I said at the outset, I agree with Robinson that there is something deeply problematic about supposing oneself to be a vague object. However, Robinson's attempt to pinpoint the problem will not, I fear, convince many materialists that the problem is real. Most analytic philosophers would doubtless deny premise 2, adopting the Moorean response to the claim that non-fundamental kinds are mind-dependent. The conclusion that a world without minds could not contain mountains or trees is shocking enough to require considerable defense; and, as I showed, there appear to be plausible ways to resist Robinson's arguments for "making a conceptual interpretation" of the non-fundamental. But, even if these arguments held up, there appear to be plausible lines of resistance to premise 4.

That may not be the end of the story, of course. I treated the five lines of argument in support of premise 2 as independent strands. But I greatly simpli-

doi: 10.48106/dial.v74.i3

fied them. They deserve closer attention, and may well have been intended as interwoven parts of a larger argument that I have not fully grasped. I may also have missed some arguments for premise 2 altogether—the book is complex, and its arguments have many moving parts. My criticism of premise 4 is also far from conclusive. (For one thing, in conversation, Robinson has helped me to see a potential response to my attack upon premise 4.[20]) Often enough, a critique of some philosopher's argument will seem devastating, until the target of the critique has the chance to respond. I hope Robinson returns to these topics soon, and sheds more light on his proposed path from vagueness to the denial of garden-variety materialism.*

Dean Zimmerman
0000-0001-8209-2997
Rutgers University
dwzimmer@philosophy.rutgers.edu

References

Adams, Robert Merrihew. 1981. "Actualism and Thisness." *Synthese* 49(1): 3–41, doi:10.1007/BF01063914.

Armstrong, David M. 1989. *Universals: An Opinionated Introduction*. Boulder, Colorado: Westview Press.

—. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511583308.

Borges, Jorge Luis. 1944. *Ficciones*. Buenos Aires: Sur.

—. 1962. *Ficciones*. New York: Grove Press. Translation of Borges (1944), edited and introduced by Anthony Kerrigan.

—. 1964. *Labyrinths*. New York: New Directions.

Brown, Joshua D. K. 2016. "Natural Objects." *Journal of the American Philosophical Association* 2(2): 254–271, doi:10.1017/apa.2016.6.

Chisholm, Roderick M. 1978. "Is there a Mind-Body Problem?" *Philosophic Exchange* 9: 25–34. Reprinted in Chisholm (1989, 119–128), http://hdl.handle.net/20.500.12648/3333.

—. 1989. *On Metaphysics*. Minneapolis, Minnesota: University of Minnesota Press.

Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, Massachusetts: The MIT Press.

20 The rough outlines of the response, and my tentative defense, can be gleaned from footnote 18, above.

* THANKS

DORR, Cian. 2019. "Natural Properties." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/fall2019/entries/natural-properties/.

FINE, Kit. 1982. "First-Order Modal Theories III – Facts." *Synthese* 53(1): 43–122, doi:10.1007/BF00500112.

HAWTHORNE, John. 2005. "Vagueness and the Mind of God." *Philosophical Studies* 122(1): 1–25. Reprinted in Hawthorne (2006, 165–183), doi:10.1007/s11098-005-0310-4.

—. 2006. *Metaphysical Essays*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199291236.001.0001.

KING, Jeffrey C. 2007. *The Nature and Structure of Content*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199226061.001.0001.

LEWIS, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61(4): 343–377. Reprinted in Lewis (1999, 8–55), doi:10.1080/00048408312341131.

—. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511625343.

MCGRATH, Matthew and FRANK, Devin. 2018. "Propositions." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/spr2018/entries/propositions/.

MOORE, George Edward. 1903. "The Refutation of Idealism." *Mind* 12(48): 433–453. Reprinted in Moore (1922, 1–30) and in Moore (1993, 23–44), doi:10.1093/mind/XII.4.433.

—. 1922. *Philosophical Studies*. London: Routledge & Kegan Paul.

—. 1993. *Selected Writings*. London: Routledge. Edited by Thomas Baldwin, doi:10.4324/9780203769812.

POLLOCK, John L. 1985. "Plantinga on Possible Worlds." in *Alvin Plantinga*, edited by James E. TOMBERLIN and Peter van INWAGEN, pp. 121–144. Profiles n. 5. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-009-5223-2_3.

PRIOR, Arthur Norman. 1960. "Identifiable Individuals." *The Review of Metaphysics* 13(4): 684–696. Reprinted in Prior (2003, 81–92).

—. 1968. *Papers on Time and Tense*. Oxford: Oxford University Press.

—. 2003. *Papers on Time and Tense*. Oxford: Oxford University Press. New edition of Prior (1968) by Per Hasle, Peter Øhrstrøm, Torben Braüner and Jack Copeland, doi:10.1093/oso/9780199256068.001.0001.

QUINN, Philip L. 1997. "Tiny Selves: Chisholm on the Simplicity of the Soul." in *The Philosophy of Roderick M. Chisholm*, edited by Lewis Edwin HAHN, pp. 55–67. The Library of Living Philosophers n. 25. LaSalle, Illinois: Open Court Publishing Co.

ROBINSON, Howard. 2010. "Quality, Thought and Consciousness." in *The Metaphysics of Consciousness*, edited by Pierfrancesco BASILE, Julian KIVERSTEIN, and Pauline PHEMISTER, pp. 203–216. Royal Institute of Philosophy Supplements n. 67. Cambridge: Cambridge University Press, doi:10.1017/s1358246110000135.

—. 2016. *From the Knowledge Argument to Mental Substance: Resurrecting the Mind*. Cambridge: Cambridge University Press, doi:10.1017/cbo9781316092873.

SCHAFFER, Jonathan. 2003. "Is there a Fundamental Level?" *Noûs* 37(3): 498–517, doi:10.1111/1468-0068.00448.

ZIMMERMAN, Dean W. 2010. "From Property Dualism to Substance Dualism." *Proceedings of the Aristotelian Society, Supplementary Volume* 84: 119–150, doi:10.1111/j.1467-8349.2010.00189.x.

—. 2011. "From Experience to Experiencer." in *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by Mark C. BAKER and Stewart C. GOETZ, pp. 168–201. London: Continuum International Publishing Group.

How to Test the Ship of Theseus

Marta Campdelacreu, Ramón García-Moya,
Genoveva Martí & Enrico Terrone

The story of the Ship of Theseus is one of the most venerable conundrums in philosophy. Some philosophers consider it a genuine puzzle. Others deny that it is so. It is, therefore, an open question whether there is or there is not a puzzle in the Ship of Theseus story. So, arguably, it makes sense to test empirically whether people perceive the case as a puzzle. Recently, David Rose, Edouard Machery, Stephen Stich and forty-two other researchers from different countries have undertaken that task. We argue that their tests do not provide any evidence that bears on the question as to whether the Ship of Theseus case is a genuine puzzle. In our discussion we address also what should be taken into account if one wishes to test the puzzling, or not puzzling, status of the Ship of Theseus story.

1 The Test

The story of the Ship of Theseus (SoT from now on) is one of the most venerable conundrums in philosophy. Some philosophers consider it not just a conundrum, but a genuine puzzle (Scaltsas 1980, 152; Wiggins 1980, 97). Others deny that it is so (Smart 1972, 148; 1973, 27). It is, therefore, an open question whether there is or there is not a puzzle in the SoT story, and also whether the case is considered puzzling across different cultures. Recently, David Rose, Edouard Machery, Stephen Stich and forty-two other authors from different countries (RMS from now on) have undertaken the task of conducting empirical tests with a view to provide an answer to that open question.[1]

According to RMS, a puzzle is a thought experiment fulfilling a "provocative function" (2020, 159), which they characterize in terms of two conditions: *ambivalence* and *universality*.

1 RMS's study is part of a larger project made possible through the support of a grant from the Fuller Theological Seminary / Thrive Center in concert with the John Templeton Foundation.

The ambivalence condition is stated as follows: "Readers should feel inclined to assert two prima facie inconsistent propositions" (RMS 2020, 159). As regards universality, RMS point out that a puzzle "[…] must elicit an ambivalent state of mind in readers of all demographic, particularly of all cultural, backgrounds" (2020, 159).

The story of the SoT that RMS presented to participants in their study is adapted from Rose (2015), and it contains the usual elements of the story, namely, a ship whose planks are gradually replaced through maintenance until no original plank remains ("Replacement") and the ship that results from putting together the original planks that were preserved ("Original Parts"). The story was translated into 17 languages and presented to 2,426 people in 22 countries. The participants in the experiment were asked to read the story and to answer whether, in their view, Replacement or Original Parts was the original ship. Their degree of confidence was also measured.

64% of the participants in the study thought that Replacement was the original ship. However, RMS note that, although there was a sharp majority in favor of Replacement, there was "quite a sizable minority—in the 30%–40% range—who thought that Original Parts was the original ship" (2020, 167), a minority that expressed high confidence in their judgment. In any case, regardless of their answer, participants reported, in general, a high level of confidence.[2] Moreover, with slight differences, the disagreement was universal across countries and cultures.

So, RMS conclude:

> Our results do indeed suggest that the Ship of Theseus case is a puzzle: People across cultures are ambivalent about what to say in response to the case. But they do not suggest it is one that feels unsolvable or that it is "irreclaimably paradoxical", placing us in a permanent state of indecision. If this were the case, then we should have found that people were divided on whether Replacement or Original Parts was the Ship of Theseus and that they were not very confident in the option they ultimately settled on. But this is not at all what we found. The majority of sites offered a clear verdict and did so quite confidently. (2020, 168)

Ultimately, according to RMS, "the Ship of Theseus is a genuine puzzle but one that people can solve to their satisfaction" (2020, 169).

2 68%–87% for Replacement and 63%–90% for Original Parts (RMS 2020, 166).

2 The Role of Ambivalence

In our view, the experiment conducted by RMS does not grant any conclusion on the puzzling nature of the SoT story. To see this, let us first reflect on two, very different, puzzles: the Liar and the Trolley Problem.

When we are asked whether the sentence "this sentence is false" is true or false, we can soon perceive the circle that leads to contradiction. And when we face the choice of either pushing the lever killing the one person or refraining from doing anything (thus letting five people die), both choices seem problematic, in spite of the fact that both courses of action are supported by ethical principles that we rely on in ordinary situations.

Indecision and ambivalence are felt when one is confronted with these cases: for different reasons in each case, *we simply do not know what to say*. Arguably, the psychological reaction, the indecision and ambivalence that each of us can feel, is not what makes a given case a genuine puzzle, although it is a good indicator of the existence of a puzzle.[3] That is why we think it is worthwhile to test, as RMS set to do, whether people are ambivalent about the story of the SoT.[4]

However, there is an important confusion in their procedure. The principle of ambivalence, as RMS state it, is ambiguous. The claim "readers should feel inclined to assert two prima facie inconsistent propositions" (RMS 2020, 159) can be understood as requiring *interpersonal* disagreement (among different readers) or *intrapersonal* conflict or indecision, felt by each reader. Only the latter form of clash is arguably a good indicator of the presence of a puzzle. The paradigmatic cases of philosophical puzzles, such as the Liar and the Trolley Problem, do reveal such intrapersonal conflict.

What RMS show is that there is sharp interpersonal disagreement among different readers: 64% of participants thought that Replacement was the original ship whereas 36% thought that Original Parts was the original ship (2020, 163). And the disagreement is indeed sharp because in both cases partici-

3 If we attend to RMS's definition of a puzzle as a thought experiment that fulfills a provocative function, it would seem that ambivalence is for them a constitutive condition of a puzzle. We are more cautious, although we do consider that ambivalence is a good indicator.

4 We are mentioning here the Liar and the Trolley problem because we do think that they are paradigmatic cases of puzzles. A referee has suggested that the story of the statue and the piece of clay might be a better case. That might be so, but our point here does not depend on which particular cases are used as examples of genuine puzzles. Our point is an abstract one about the fact that ambivalence is an indicator of the presence of a puzzle, and that is independent of the examples chosen to illustrate it.

doi: 10.48106/dial.v74.i3

pants were quite confident in their judgment (2020, 166). But the presence of sharp interpersonal disagreement does not qualify as evidence that we are confronted with a genuine puzzle.[5]

If interpersonal disagreement were the mark of a philosophical puzzle then any disagreement that can generate philosophical discussion would constitute a puzzle. But, in general, studies that show that there is interpersonal disagreement about a subject matter are not presented as studies that reveal the puzzling nature of that subject matter.

For instance, Edouard Machery, Ron Mallon, Shaun Nichols and Stephen Stich (2004) conducted an experiment using Kripke's Gödel case (Kripke 1980). The results of that experiment, they argued, show that East-Asians are inclined to think that the man who proved incompleteness and was found dead in mysterious circumstances is the referent of the name "Gödel," whereas Westerners were not at all inclined to this response. Subsequently Edouard Machery, Christopher Olivola and Molly DeBlanc (2009) conducted a similar test in different countries that showed divisions within each culture. In each case, the authors did not present their results as providing evidence for the existence of a puzzle. They simply argued that those results constituted proof that substantial segments of the population do not agree with Kripke's intuitions on the Gödel's case.[6]

These two studies purport to show that there is interpersonal disagreement as to who "Gödel" refers to.[7] And, if the authors do not present the disagreements as providing evidence for the existence of a puzzle, we think, it is precisely because their study is not designed to show intrapersonal disagreement.[8]

The SoT story is often presented as giving rise to a conflict with the transitivity of identity. One feels inclined to say that the SoT is Replacement and also that the SoT is Original Parts, but clearly Original Parts and Replacement are different. In general, showing that some people (perhaps a majority) think

5 It might be even argued that RMS's results militate against the conclusion that the SoT story constitutes a genuine puzzle, precisely because the participants reveal a high degree of confidence, incompatible with intrapersonal ambivalence (namely, it is not the case that *they do not know what to say*). We will address this issue in Section 3.

6 In fact, the divisions reported by Machery, Olivola and De Blanc in India, Mongolia and France are very similar to those reported in the test of the SoT story. For instance, in Mongolia, 66% lean one way and 34% the other, close to the 64% and 36% reported in the SoT test.

7 There has been a long and lively discussion as to what the studies do show, but the issue is of no relevance for the purposes of this paper.

8 Neither set of authors even ask participants for the degree of confidence in their answers.

that, say, A is B and some other people (a substantial minority) think that A is C does not create any contradiction with the principle of transitivity of identity. Some people think that the author of the bestseller *My Brilliant Friend* (published under the name or nom de plume "Elena Ferrante") is the contemporary historian Marcella Marmo and some other people think that the author is the writer Domenico Starnone.[9] Both groups have a claim to being right, for there is evidence pointing in both directions. Clearly, Marcella Marmo is not Domenico Starnone, yet no one would conclude that this disagreement threatens the principle of transitivity of identity.

Although these interpersonal disagreements may be part of interesting philosophical discussions, they surely do not indicate the existence of puzzles. Likewise, the evidence that RMS collect as regards the story of the SoT is not an indicator of the presence of a puzzle.

Now, the results of RMS's test show that people disagree about the right answer. Indeed, they show that such disagreement occurs with high levels of confidence and without indication of intrapersonal conflict. Thus, one might ask: do RMS show (unbeknownst to them) that the SoT story *does not* constitute a genuine puzzle after all? Not quite.

3 The Story and its Presentation

Let us think what would be a good presentation of the SoT story, the kind that we might easily find discussed in an undergraduate course in Philosophy. Ideally, the discussion proceeds in three steps. First, some story is told that invokes the principle that gradual replacement does not affect the identity of an object. For instance, a wall can have its bricks gradually replaced and still remain the same wall. Second, some other story is told that invokes the principle that disassembling and reassembling an object does not affect its identity. For instance, a watch can be disassembled and reassembled in order to clean it and yet remain the same watch. When the SoT story is then presented, readers are in an adequate position to consider whether their answers to the previous two stories entail that both the gradually replaced ship and the reassembled ship have a claim to being the original ship. That would violate the transitivity of identity.

Pickup (2016) underscores that the three steps are fundamental if one is to see a problem in the story of the SoT: in a situation in which an object is

9 Many more people are suspected of being Elena Ferrante. See Davies (2014) and Scammell (2016).

doi: 10.48106/dial.v74.i3

disassembled and reassembled the identity of the object in question seems unproblematic; a situation in which parts of an object are gradually replaced seems entirely unproblematic, too. But then, in a situation that contains the previous two situations as parts, a problem seems to arise.

We are not claiming that the Ship of Theseus story is a genuine puzzle—in fact, the authors are divided on that issue.[10] Our point is that the SoT story should be told in a way in which the alleged conflict between two principles that justify plausible answers in ordinary cases (a conflict that, if it exists, would make the SoT a puzzle) can come to the surface. Asking the question RMS ask without the three-step presentation does not place the subject in an adequate situation to be able to consider whether preservation of identity under gradual replacement, and preservation of identity under disassembly and reassembly conflict.

It might be argued that readers of RMS's vignette will put two and two together and gauge the potential conflict. That may be right. But RMS include no measure to indicate that this is the case, nor an acknowledgment that they are counting on readers making the connections.[11] More importantly, RMS do not allow readers who have gauged the conflict, and feel intrapersonal ambivalence, to express it. The reason is that readers of their vignette have only two options: they have to choose the reassembled ship or the gradually replaced one. But for the reader to be able to express intrapersonal ambivalence, options such as "both," "neither" and "I do not know" should be offered as possible answers as well.[12]

Interestingly, it might be argued that it is an open question whether there is a hierarchical order between the principles that govern identification and reidentification of objects. One might even wonder if such a hierarchy would be sensitive to cultural background. Perhaps, one might argue, this is the reason RMS obtain the result that the majority of people are inclined towards a certain answer and with little hesitation. If the SoT story had been tested in the three-step way suggested here, and if the results had been the same that RMS obtained (namely, interpersonal disagreement and high levels of confidence), then it could be argued that there is a hierarchy of principles

10 See, for instance, García-Moya (2021).

11 We are grateful to a referee for urging us to clarify these points.

12 Adding options has been proposed in conversation with Vilius Dranseika. Also, verbal expression is not the only way to capture indecision or ambivalence. Eye-tracking, for instance, has been used in other experimental studies. See Cohnitz and Haukioja (2015) and Shtulman and Valcarcel (2012). We thank Eugen Fischer for bringing that to our attention.

and that people disagree as regards which principle is prior. If that were the case, the SoT story would be interesting and challenging, but perhaps not a genuine puzzle. Yet, it is important to stress that the way RMS tell the story of the SoT is not useful as a test in that regard either. Testing the presence of a hierarchy requires collecting data about whether certain principles are used happily on some occasions and are overridden in other occasions. Both the happy application of principles and the possibly overriding application must be tested. That could be done by testing the story in the step-by-step way suggested here, but it cannot be achieved by the one-step story presented by RMS.

4 Conclusions

We conclude that RMS's test does not show that the story of the SoT is a puzzle because the data collected is data about interpersonal disagreement which, unlike intrapersonal conflict, is not a good indicator of the presence of puzzles.

In fact, the high level of confidence reported by the participants in the experiment might suggest that the story of the SoT constitutes no puzzle at all. However, the story that RMS present is simply not adequate to test the puzzling nature of the SoT.

Hence, the test conducted by RMS has no bearing on the question as to whether the SoT constitutes a genuine philosophical puzzle, and it does not advance in any way the traditional discussion about this venerable story.

Finally, we think that there is a general lesson to be learned about puzzles and philosophical experiments. A lot of work in experimental philosophy has consisted in highlighting clashes of intuitions between groups of people (e.g. cultures, genders, general public vs. experts). All these studies rely crucially on the existence of interpersonal disagreement, as they should, since their purpose is to highlight disagreements among different people or groups. But RMS take that very same methodology and apply it to test the presence of a puzzle. That is a mistake: testing the presence of a puzzle should focus on intrapersonal conflict and therefore requires a different methodology.*

* Versions of this paper were presented at the 2020 Online Conference in Experimental Philosophy and at the Logos Seminar. We thank the audiences for their comments. We acknowledge the support of the European Union and the Spanish Ministerio de Ciencia e Innovación, through grants FFI2016-80588-R and 2019PID-107667GB-100.

doi: 10.48106/dial.v74.i3

Marta Campdelacreu
0000-0002-2759-4326
Universitat de Barcelona
marta_campdelacreu@ub.edu

Ramón García-Moya
0000-0003-2135-2016
Universitat de Barcelona
ramongarciamoya@gmail.com

Genoveva Martí
0000-0003-1269-4655
ICREA
Universitat de Barcelona
gmarti@ub.edu

Enrico Terrone
0000-0002-1975-8319
Università degli Studi di Genova
enrico.terrone@unige.it

References

COHNITZ, Daniel and HAUKIOJA, Jussi. 2015. "Intuitions in Philosophical Semantics." *Erkenntnis* 80(3): 617–641, doi:10.1007/s10670-014-9666-1.

DAVIES, Lizzy. 2014. "Who Is the Real Italian Novelist Writing as Elena Ferrante?" *The Guardian*, 15.10.2014, http://www.theguardian.com/books/2014/oct/15/who-italian-novelist-elena-ferrante.

GARCÍA-MOYA, Rámon. 2021. "Disassembling the Ship of Theseus." Unpublished manuscript.

KRIPKE, Saul A. 1980. *Naming and Necessity*. Oxford: Basil Blackwell Publishers.

MACHERY, Edouard, MALLON, Ron, NICHOLS, Shaun and STICH, Stephen P. 2004. "Semantics, Cross-Cultural Style." *Cognition* 92(3): B1–B12, doi:10.1016/j.cognition.2003.10.003.

MACHERY, Edouard, OLIVOLA, Christopher Y. and DE BLANC, Molly. 2009. "Linguistic and Metalinguistic Intuitions in the Philosophy of Language." *Analysis* 69(4): 689–694, doi:10.1093/analys/anp095.

PICKUP, Martin. 2016. "A Situationalist Solution to the Ship of Theseus Puzzle." *Erkenntnis* 81(5): 973–992, doi:10.1007/s10670-015-9777-3.

ROSE, David. 2015. "Persistence Through Function Preservation." *Synthese* 192(1): 97–146, doi:10.1007/s11229-014-0555-6.

—. 2020. "The Ship of Theseus Puzzle." in *Oxford Studies in Experimental Philosophy*, volume III, edited by Tania LOMBROZO, Joshua KNOBE, and Shaun NICHOLS, pp. 158–174. Oxford: Oxford University Press. Coauthored by many philosophers from all over the world, doi:10.1093/oso/9780198852407.003.0007.

SCALTSAS, Theodore. 1980. "The Ship of Theseus." *Analysis* 40(3): 152–157, doi:10.2307/3327668.

SCAMMELL, Rosie. 2016. "Who Is Elena Ferrante? Novelist Issues Denial as Guessing Game Goes On." *The Guardian*, 13.3.2016, http://www.theguardian.com/books/2016/mar/13/who-is-elena-ferrante-novelist-issues-denial-as-guessing-game-goes-on.

SHTULMAN, Andrew and VALCARCEL, Joshua. 2012. "Scientific Knowledge Suppresses but Does Not Supplant Earlier Intuitions." *Cognition* 124(2): 209–215, doi:10.1016/j.cognition.2012.04.005.

SMART, Brian. 1972. "How to Reidentify the Ship of Theseus." *Analysis* 32(5): 145–148, doi:10.2307/3327760.

—. 1973. "The Ship of Theseus, the Parthenon, and Disassembled Objects." *Analysis* 34(1): 24–27, doi:10.2307/3327980.

WIGGINS, David. 1980. *Sameness and Substance*. Oxford: Basil Blackwell Publishers.

Mereology is not a Guide to (In)conceivability

A Reply to Giberman (2015)

Mahmoud Morvarid

A sophisticated version of the zombie argument due to David Chalmers runs roughly as follows: a zombie world is ideally primarily conceivable, and whatever is ideally primarily conceivable is primarily possible. Thus, a zombie world is primarily possible, which implies, in turn, that either physicalism is false or Russellian monism is true. Appealing to some plausible mereological considerations, Daniel Giberman presents a novel argument to the effect that zombies are not ideally primarily conceivable. I shall argue, firstly, that a main premise of Giberman's argument is ill-supported, as it trades on a confusion between the primary and the secondary intensions of the "actually" operator. I then consider two lines of reasoning, which might be extracted from Giberman's text, in favour of another chief premise of his argument. I shall argue that the first line of reasoning is flawed, and the second one, in effect, will transform Giberman's argument into a kind of "parity argument" in which his mereological considerations play no role.

Perhaps the most famous objection to physicalism—the thesis that the mental supervenes on the physical—is the zombie argument. Roughly, the idea is that zombies, which are supposed to be complete physical duplicates of you and me that however lack phenomenal consciousness, are conceivable, and whatever is conceivable is possible. So, zombies are possible, which entails that physicalism is false. Yet there remains a controversy as to in which sense of "conceivability" zombies are conceivable, and whether being conceivable in that sense is a good guide to possibility.

Perhaps the most sophisticated version of the argument is due to David Chalmers (1996, 2009), and is heavily based on his two-dimensional semantic framework. The relevant notion of conceivability used in his argument is that

of *ideal primary conceivability*. A sentence S is said to be ideally primarily *negatively* conceivable if and only if it could not be ruled out a priori upon ideal reflection. Moreover, a sentence S is ideally primarily *positively* conceivable when an ideal thinker can imagine a coherent situation that verifies S (Chalmers 2009, 146). Based on these notions, Chalmers (2009, 147–148) presents his thesis regarding the relationship between conceivability and possibility:

> CONCEIVABILITY – POSSIBILITY THESIS (CP). If a sentence is ideally primarily conceivable (whether negatively or positively) then it is primarily possible, that is, its primary intension is true in a metaphysically possible world. (See also Chalmers 2002b, 171–172)

According to Chalmers, the distinction between negative and positive conceivability does not matter for many purposes. His conceivability argument, accordingly, is in fact an argument scheme which can be understood as generating two different arguments, depending on which of these two notions is appealed to therein (we shall see the argument in a moment).

Now, let P stand for the complete physical description of the actual world in the language of complete microphysics. Let Q abbreviate an arbitrary positive truth about phenomenal consciousness. For example, Q could be the truth that some entity is phenomenally conscious, or that there are pains, etc. Chalmers's more recent version of the argument runs as follows (2009, 152):

(C1) $P \wedge \neg Q$ is ideally primarily conceivable.
(C2) If $P \wedge \neg Q$ is ideally primarily conceivable, then $P \wedge \neg Q$ is primarily possible, that is, its primary intension is true in a metaphysically possible world. (From (CP))
(C3) If $P \wedge \neg Q$ is primarily possible, then either $P \wedge \neg Q$ is secondarily possible (that is, its secondary intension is true in a metaphysically possible world) or Russellian monism is true.
(C4) If $P \wedge \neg Q$ is secondarily possible, then physicalism is false.
(C5) Physicalism is false or Russellian monism is true.

Although the conclusion of this argument is weaker than the falsity of physicalism, it is still an important achievement in that, as Chalmers puts it, Russellian monism has so much in common with property dualism that many physicalists would want to reject it (2009, 152). Clearly, if Q is the truth that

there is some conscious being, then $P \wedge \neg Q$ describes a zombie world, in which case the above argument is strictly speaking an argument from the conceivability of a zombie world.

In a recent article, Giberman (2015) proposes a novel objection to the first premise of the above argument. He concentrates on a version of the argument which employs the *positive* notion of ideal primary conceivability (unless otherwise specified, "conceivability" hereafter picks out ideal primary positive conceivability. Other similar phrases should be understood in this way as well). Moreover, he mainly focuses on that version of the argument which takes Q to be the truth that there is some conscious being, which turns the first premise into the claim that a zombie world is ideally primarily positively conceivable (hereafter (C1) should be understood in this way). Giberman takes the second premise and its basis, (CP), for granted. He then tries to show that (CP) and some mereological considerations jointly entail that zombies are not ideally primarily positively conceivable, and consequently that (C1) is false.[1]

In what follows, I first consider Giberman's argument for the inconceivability of zombies. I then argue that his argument suffers from a basic problem, as one of its main premises trades on a confusion between the primary and the secondary intensions of the "actually" operator. Turning to another main premise of his argument, I shall consider two lines of reasoning, which might be extracted from Giberman's text, in favour of that premise. I then argue that the first line of reasoning is flawed, and the second one, in effect, will transform Giberman's argument into a kind of "parity argument" in which his mereological considerations are entirely redundant. The upshot is that Giberman's mereological argument for the inconceivability of zombies cannot get off the ground.

1 Giberman's Argument Explained and Criticized

In constructing his argument, Giberman employs rather complex machinery. He begins by stating a plausible mereological constraint on actual bearers of phenomenal consciousness: "paradigmatic actually conscious objects are mereologically complex, and capable of losing some parts while retaining consciousness" (2015, 122). By "paradigmatic actually conscious objects" he means those mereologically complex conscious objects that are not unde-

1 Throughout his paper, Giberman speaks of the "conceivability of zombies" and the like. Given that he is targeting Chalmers's argument, I take it that Giberman means the *conceivability of a zombie world*. Accordingly, I use these two phrases interchangeably.

doi: 10.48106/dial.v74.i3

tached proper parts of ordinary objects, such as human beings and other conscious animals. Moreover, "conscious" in his usage refers to the familiar intrinsic phenomenal property that you, I, and other conscious animals enjoy. It does not pick out, therefore, some proto-phenomenal property that would give rise to full-blown phenomenal experiences for complex objects under appropriate circumstances (Giberman 2015, 122–123).

As Giberman plausibly argues, I have many proper parts which are either conscious or would be if detached, such as *me-minus-an-arm*. Following Giberman, let us call the disjunctive property of being either conscious simpliciter or conscious-if-detached "consciousness-capability." He proceeds to introduce for any paradigmatic conscious object, x, a "mereological spectrum" from the conscious whole of x, at one pole, to its (presumably consciousness-incapable) most basic physical proper parts, at the other pole. Here is an example:

> Take my case as an example. At one pole of the spectrum (the 'whole pole') will be me and at the other (the 'simple pole') will be my most basic physical spatiotemporal-cum-mereological unit [...]. In between will be all my undetached proper parts. This is an expansive and varied lot. It includes the bearers of such descriptions as 'me-minus-a-quark', 'me-minus-a-neuron', 'me-minus-an-arm', 'a fusion of the easternmost half of my left ear and three cells in my right big toe' [...] and the like. (2015, 123–124)

Assuming that basic proper parts are never actually conscious, Giberman observes, it would follow that somewhere along x's mereological spectrum is a region containing *mereologically minimal consciousness-capable undetached proper parts*. These are the proper parts of x that satisfy the following two conditions:

(a1) If they were to become detached, they would be conscious.
(a2) If they then lost even a single basic part (without replacement by another one), they would be no longer conscious.

Moreover, there must also be a region containing some *mereologically maximal consciousness-incapable undetached proper parts*, that is, proper parts, y, which satisfy the following two conditions:

(b1) If *y* were to become detached it would then need to gain as a part only one additional mereologically basic part (properly placed) in order to be conscious.

(b2) No parts that satisfy the condition described in (b1) are more complex than *y*. (Giberman 2015, 124)

In the next step, Giberman introduces a notion which plays a crucial role in his argument:

> Call the segment on a given object's mereological spectrum that is bounded by these two points [that is, mereologically minimal consciousness-capable undetached proper parts and mereologically maximal consciousness-incapable undetached proper parts] its 'mereological threshold for consciousness' (MTC) since all the object's parts that lie beyond one end of the threshold are consciousness-capable and all the parts that lie beyond the other end are consciousness-incapable. (2015, 124)

So far Giberman has plausibly shown that the above-mentioned mereological constraint implies the existence of an MTC for any paradigmatic conscious object. The object's MTC divides its mereological spectrum into two factions: its consciousness-capable parts and its consciousness-incapable parts.

I am now in a position to present the gist of Giberman's argument for the inconceivability of zombies. The argument is based on a dilemma: for an arbitrary conscious creature, either its MTC is only contingently located on its mereological spectrum, or not. The first horn of the dilemma, Giberman argues, leads to the possibility of what he calls "physical panpsychism," which in turn entails the inconceivability of zombies. The second horn, on the other hand, renders zombies inconceivable. Either way, zombies turn out to be inconceivable. Before going through the full statement of this argument, it is important to see exactly what Giberman means by "physical panpsychism." We can formulate physical panpsychism, as introduced by Giberman, as a conjunction of three statements, the second of which is modified by the "actual" operator:

> Physical Panpsychism (Pan). (i) Phenomenal consciousness is an intrinsic categorical property of mereologically basic particulars, and (ii) it is actually the case that (*T*) phenomenal consciousness plays a constitutive, underwriting role in the fundamental properties

doi: 10.48106/dial.v74.i3

of "final" physics, and (iii) phenomenal consciousness plays a constitutive, underwriting role in the exemplification of consciousness by more complex structures. (See Giberman 2015, 128) [2]

Let us now consider the argument in full detail (2015, 129–130):

(1) If zombies are ideally primarily positively conceivable then physical panpsychism is not primarily possible. [I shall consider Giberman's defence of this premise below]

(2.1) For arbitrary actually conscious physical structure x, either it is ideally primarily positively conceivable that x's MTC could have been different or it is not. (Tautology)

Suppose it is not. Then x's MTC is guaranteed to have an upper bound (which is less complex than x itself) in every conceivable state of affairs in which x exists. So, since every point on x's mereological spectrum beyond the MTC's upper bound contains consciousness-capable parts of x, it is inconceivable for x itself not to be conscious. Since x is an arbitrary actually conscious physical structure, it follows that zombies are inconceivable. For the sake of continuing the argument, then, the present supposition is to be rejected. So:

(2.2) It is ideally primarily positively conceivable that x's MTC could have been different.

(3) If it is ideally primarily positively conceivable that x's MTC could have been different then it is ideally primarily positively conceivable that physical panpsychism is true. [I shall consider Giberman's defence of this premise below]

(4) If it is ideally primarily positively conceivable that physical panpsychism is true then physical panpsychism is primarily possible. (CP)

(5) Physical panpsychism is primarily possible. ((2.2), (3), (4), modus ponens)

2 Here is Giberman's own wording: "Physical panpsychism is the thesis that phenomenal consciousness is an intrinsic categorical property of mereologically basic particulars, which property plays a constitutive, underwriting role in (i) the fundamental properties of 'final' physics at the actual world and (ii) the exemplification of consciousness by more complex structures" (2015, 128).

4186 So:

4187 (6) Zombies are not ideally primarily positively conceivable.
4188 ((1), (5), modus tollens)

4189 The crucial premises of Giberman's argument are (1) and (3), which I shall
4190 examine in turn. Let us firstly consider how Giberman endeavours to back up
4191 (1). He begins with the following remark:

4192 > [P]hysical panpsychism has actuality built in: it is a thesis about
4193 > actual final physics. So even the primary possibility of physical
4194 > panpsychism would entail that actual physics presupposes con-
4195 > sciousness. (2015, 128)

4196 Here, Giberman draws on the specific characterization he provided for physi-
4197 cal panpsychism, which I previously formulated as (PAN). As we have seen,
4198 the second part of (PAN) has actuality built in: "(ii) it is actually the case that
4199 (*T*) phenomenal consciousness plays a constitutive, underwriting role in the
4200 fundamental properties of 'final' physics." He consequently maintains that
4201 "even the primary possibility of physical panpsychism would entail that actual
4202 physics presupposes consciousness." But why does Giberman think that the
4203 mere primary possibility of (PAN) should carry some implication about what
4204 is going on in the actual world? The *only* viable answer seems to be that he is
4205 appealing to a familiar fact about the semantics of the "actually" operator: for
4206 any sentence *S*, the possibility of "actually *S*" entails that *S* is true. Applying
4207 this semantic rule to (PAN), one might arrive, as Giberman seemingly does, at
4208 the following contention:

4209 (G1) If (PAN) is primarily possible then (*T*) is true in the actual world.[3]

4210 Giberman then continues:

4211 > Consequently one cannot coherently conceive of a state of af-
4212 > fairs that is physically indiscernible from the actual world—as
4213 > required by the primary conceivability of zombies—unless either
4214 > the physical structures in that state of affairs are conscious or

3 (G1) is in fact a rephrasing of Giberman's quoted sentence "the primary possibility of physical panpsychism would entail that actual physics presupposes consciousness."

doi: 10.48106/dial.v74.i3

> physical panpsychism is assumed primarily impossible. (2015, 128)

For a better grip on what is going on here we can rephrase Giberman's idea in this passage as follows:

> (*) For any ideally conceivable state of affairs A, if A is physically identical with the actual world then either the physical structures in A are conscious or else (PAN) is not primarily possible.

But how does Giberman move from (G1) to (*)? The requisite assumption for such a transition can be stated in the following way:

> (G2) If (T) is true in the actual world then any ideally conceivable state of affairs which is physically identical with the actual world would be a state of affairs in which the physical structures are conscious.

Clearly, the conjunction of (G1) and (G2) implies (*).[4] The idea behind (G2), presumably, is that if we assume that (T) is actually true then (T) would be a part of the physical characterization of the actual world. Thus, any arbitrary primarily conceivable state of affairs which is physically identical to the actual world must be a state of affairs in which (T) is true. Given (T)'s content, however, it follows that the physical structures in that state of affairs must be conscious as well. Bearing all this in mind, we can formulate Giberman's argument for (1) in the following way, which I shall call ARGUMENT G.

ARGUMENT G.

> (G1) If (PAN) is primarily possible, then (T) is true in the actual world.
> (G2) If (T) is true in the actual world, then any ideally primarily conceivable state of affairs which is physically identical with the actual world would be a state of affairs in which the physical structures are conscious.
> (G3) If any ideally primarily conceivable state of affairs which is physically identical with the actual world is a state of affairs in which the physical structures are conscious, then the zombie world is ideally primarily inconceivable.

4 To see this, notice that the conjunction of $(P \rightarrow Q)$ and $(Q \rightarrow (\forall x\, Fx \rightarrow Gx))$ entails $(\forall x\, Fx \rightarrow (Gx \vee \sim P))$. Moreover, no proposition weaker than (G2) can, in conjunction with (G1), result in (*). Thus, although Giberman does not explicitly state (G2), it is fair to take him as relying on (G2) in his argument for (1).

(G4) If (PAN) is primarily possible, then the zombie world is ideally primarily inconceivable. ((G1), (G2) and (G3))

(G5) If the zombie world is ideally primarily conceivable, then (PAN) is not primarily possible. (Contraposition of (G4))

This, it seems, is how Giberman endeavours to substantiate (1). Unfortunately, (G1) is ill-motivated. As noted above, (G1) is based on the familiar rule about the "actually" operator: for any sentence *S*, the possibility of "actually *S*" entails that *S* is true. But this rule is valid only if the notion of possibility involved therein is that of *secondary* possibility, not *primary* possibility. Admittedly, the secondary possibility of "actually *S*" implies that the secondary intension of "actually *S*" is true in some (metaphysically) possible world, which leads, in turn, to *S*'s being true in the actual world. Now consider cases where primary possibility is involved. The primary possibility of "actually *S*" entails that the primary intension of "actually *S*" is true in some (metaphysically) possible world. But the primary intension of "actually *S*" is the same as the primary intension of *S* (generally speaking, the "actually" operator is redundant when the primary intension of the actualized sentences is appealed to). The reason is that (a) *S* and "actually *S*" are a priori equivalent, that is, the bi-conditional "*S* is true if and only if 'actually *S*' is true" is knowable a priori, and (b) a priori equivalent sentences have the same primary intensions (remember that primary intension is that aspect of meaning which captures a priori relations between sentences).[5] Thus, the mere fact that the primary intension of "actually *S*" is true in a possible world is not sufficient, by itself, to show that *S* is true in the actual world.

We may put the point rather differently. The secondary intension of the term "the actual world" in any possible world picks out the actual world; it has a constant secondary intension. But its primary intension in a given possible world picks out that possible world itself. Thus, the secondary possibility of

5 See, for example, Chalmers (2006, 64). One way to see that a priori equivalent sentences have the same primary intensions is as follows. Suppose that an arbitrary sentence, $S1$, is a priori equivalent to $S2$. Take a possible world, W, in which the primary intension of $S1$ is true. According to the standard characterization of primary intensions, therefore, it is knowable a priori that if D then $S1$ is true, where D is a "canonical description" of W (for such a characterization, see, for example, Chalmers 2002a, 611). Thus, given the a priori equivalence of $S1$ and $S2$, it is knowable a priori that if D then $S2$ is true, which means that the primary intension of $S2$ is true in W as well. Likewise, every possible world in which the primary intension of $S2$ is true is also a possible world in which the primary intension of $S1$ is true. The upshot is that $S1$ and $S2$ have the same primary intension.

doi: 10.48106/dial.v74.i3

"*S* is true in the actual world" (or equivalently, "actually *S*"), not its primary possibility, entails that *S* is true in the actual world. I submit, consequently, that Giberman's support for (1) trades on a confusion between primary and secondary intensions of actualized sentences, and therefore does not get off the ground.

Let us now turn to the other main premise of the argument, (3):

(3) If it is ideally primarily positively conceivable that *x*'s MTC could have been different then it is ideally primarily positively conceivable that physical panpsychism is true.

Given Giberman's characterization of physical panpsychism, (3) can be rewritten as follows:

(3′) If it is conceivable that *x*'s MTC could have been different then it is conceivable that (i) phenomenal consciousness is an intrinsic categorical property of mereologically basic particulars, and (ii) it is actually the case that phenomenal consciousness plays a constitutive, underwriting role in the fundamental properties of "final" physics, and (iii) phenomenal consciousness plays a constitutive, underwriting role in the exemplification of consciousness by more complex structures.

It seems that Giberman tries, at least in one place, to support (3) by utilising the method of conditional proof: he first assumes the antecedent of (3), and then tries to show that given such an assumption, we have good reason to think that its consequent is also the case, that is, to think that it is conceivable that all three components of physical panpsychism are the case.[6] Regarding the first component, (i), he argues that there is no criteria other than imaginability and coherence upon ideal rational reflection for determining where *x*'s MTC could conceivably fall on its mereological spectrum (after all, we are working with ideal *primary* positive conceivability). On the other hand, the antecedent of (3) forces no specific commitment as to where *x*'s MTC could conceivably fall: it merely puts forward the assumption that *x*'s MTC could conceivably fall elsewhere than where it actually does. Thus, once it is assumed that *x*'s MTC could conceivably be different, "it follows that it is conceivable that

6 This is in fact the method Giberman *seems* to employ in (2015, 130–131) to back up (3). As we will shortly see, he tersely alludes to a different motivation for (3) in a subsequent passage on page 138. At any rate, I shall examine both lines of reasoning for (3) which might be extracted from these passages.

the lower bound of an object's MTC could fall anywhere on its mereological spectrum, including the simple pole" (Giberman 2015, 130), which means that consciousness could conceivably be an intrinsic categorical property of mereologically basic particulars. In the next step, Giberman contends that once the conceivability of the first component of (PAN), (i), is granted (under the assumption that *x*'s MTC could conceivably be different), there remains no bar to the conceivability of the rest of (PAN): there is no obvious reason to deny that the conjunction of (i), (ii), and (iii) could conceivably be the case (under the same assumption) (2015, 131).

Giberman's reasoning here is far from convincing. Let us assume that the antecedent of (3′) is the case, that is, it is conceivable that *x*'s MTC could have been different. But it does not follow from this assumption alone that *x*'s MTC could conceivably fall *anywhere* on its mereological spectrum, including the simple pole. Perhaps, given that *x*'s MTC could conceivably be different, it would be only conceivable that *x*'s MTC was nearer to *x*'s whole pole, not to its simple pole. Or perhaps, given that assumption, although *x*'s MTC could conceivably be nearer to *x*'s simple pole, it is not conceivable that it could have fallen exactly at the simple pole. Giberman does not provide any reason to rule out such possibilities, and therefore the support he offers for (3) is insufficient.

Having said this, there is a passage in Giberman's paper where he succinctly hints at a different motivation for (3):

> The first part of the argument for premise (3) of argument (1)–(6) is that there is as good a conceivability argument for conscious mereologically basic physical items as there is for zombies. This is a problem for friends of zombies because conceivability arguments are the primary basis for zombie endorsement. […] So, while it is a problem for friends of zombies that their conceivability standards lead equally to physical panpsychism and zombies, it is not a problem for physical panpsychists. (2015, 138)

The idea is presumably that (3) might be supported by exploiting the very maneuver usually utilized by the anti-physicalists to show that zombies are conceivable. For example, Chalmers argues that "the zombie hypothesis is at least prima facie coherent and imaginable." Thus, to reject its (ideal) conceivability, he continues, "one needs to find something that undermines the prima facie coherence and imaginability, such as some sort of a priori incoherence, contradiction, or unimaginability in the hypothesis that emerges

doi: 10.48106/dial.v74.i3

on reflection" (2009, 154). In a similar manner, one might argue that since physical panpsychism is prima facie coherent and imaginable we are justified in believing that it is (ideally) conceivable, unless we find something that undermines its prime facie coherence and imaginability, which is supposedly not the case.[7]

It is noteworthy that if (3) is to be supported on a ground similar to that which is typically exploited to back up the conceivability of zombies, then Giberman's main argument should be regarded, not as a case *for* the inconceivability of zombies, but rather as a *parity argument* which seeks, at best, to neutralize the anti-physicalist argument for the conceivability of zombies. In other words, given such a support for (3), Giberman should be taken as aiming, in effect, to show that there must be a problem with the typical line of reasoning for the conceivability of zombies, as similar resources employed by that line of reasoning can be exploited to construct an (otherwise successful) argument, namely (1)–(6), for the inconceivability of zombies.

Moreover, if this is the real basis for (3), it would support not only (3), but also the following non-conditional stronger claim

(3″) It is ideally primarily conceivable that physical panpsychism is true

which would render some steps of Giberman's original argument redundant. This way we would arrive at the following, much simpler argument:

(1) If zombies are ideally primarily positively conceivable then physical panpsychism is not primarily possible.
(3″) It is ideally primarily positively conceivable that physical panpsychism is true.
(4) If it is ideally primarily positively conceivable that physical panpsychism is true then physical panpsychism is primarily possible. (CP)
(5) Physical panpsychism is primarily possible. ((3″), (4), modus ponens)
(6) Zombies are not ideally primarily positively conceivable. ((1), (5), modus tollens)

This argument, again, is to be understood as providing a parity argument against the typical line of reasoning for the conceivability of zombies. Notice, however, that no mereological consideration plays any role in this argument. One worry about the above argument, which I shall not peruse here, is that it

7 I thank Daniel Giberman and an anonymous referee of the journal for calling to my attention the passage just quoted in the text.

might well be vulnerable to replies proposed by Chalmers to other analogous parity arguments put forward to neutralize his conceivability argument.[8] The main problem with the above argument, nonetheless, is that (1) is ill-motivated, as I have already defended.

2 Conclusion

I have argued that the main problem with Giberman's mereological argument for the inconceivability of zombies is that the support he provides for premise (1) of his argument is flawed, as it trades on a confusion between the primary and the secondary intensions of the "actually" operator. I have also examined two different lines of reasoning which might be extracted from Giberman's text in favour of premise (3) of his argument. It seems that the first line of reasoning is wanting, and the second one will transform Giberman's argument to a kind of parity argument, which makes no use of his mereological considerations, and which may suffer from (alleged) deficiencies of other parity arguments proposed against Chalmers's conceivability argument. I conclude, therefore, that Giberman's mereological argument for the inconceivability of zombies is too ambitious to have any chance of success.*

Mahmoud Morvarid
0000-0002-3013-5630
School of Analytic Philosophy, IPM, Tehran, Iran
morvarid@ipm.ir

References

Balog, Katalin. 2012. "In Defense of the Phenomenal Concept Strategy." *Philosophy and Phenomenological Research* 84(1): 1–23, doi:10.1111/j.1933-1592.2011.00541.x.

Brown, Richard O. 2010. "Deprioritizing the A Priori Arguments Against Physicalism." *Journal of Consciousness Studies* 17(3/4): 47–69.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

8 For various versions of such parity arguments, see e.g. Marton (1998), Yablo (1999), Sturgeon (2000), Frankish (2007), Brown (2010), Balog (2012). For Chalmers's responses, see (2009, 178–180).

* I would like to thank Mahrad Almotahari, Daniel Giberman, Yasser Pouresmaeil, Hashem Morvarid, Amir Saemi, Ben Young, and anonymous reviewers for the journal for helpful comments on earlier drafts of this paper.

—. 1997. "Moving Forward on the Problem of Consciousness." *Journal of Consciousness Studies* 4(1): 3–46. Reprinted in Shear (1997, 379–422).

—. 2002a. "The Components of Content." in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. CHALMERS, pp. 608–633. Oxford: Oxford University Press.

—. 2002b. "Does Conceivability Entail Possibility?" in *Conceivability and Possibility*, edited by Tamar Szabó GENDLER and John HAWTHORNE, pp. 145–200. New York: Oxford University Press, doi:10.1093/oso/9780198250890.003.0004.

—. 2006. "The Foundations of Two-Dimensional Semantics." in *Two-Dimensional Semantics: Foundations and Applications*, edited by Manuel GARCÍA-CARPINTERO and Josep MACIÀ, pp. 55–140. Oxford: Oxford University Press, doi:10.1093/oso/9780199271955.003.0004.

—. 2009. "The Two-Dimensional Argument Against Materialism." in *The Oxford Handbook of Philosophy of Mind*, edited by Brian P. MCLAUGHLIN, Ansgar BECKERMANN, and Sven WALTER, pp. 313–337. Oxford Handbooks. Oxford: Oxford University Press. Republished in expanded form, with an afterword, in Chalmers (2010, 141–191), doi:10.1093/acprof:oso/9780195311105.003.0006.

—. 2010. *The Character of Consciousness*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780195311105.001.0001.

FRANKISH, Keith. 2007. "The Anti-Zombie Argument." *The Philosophical Quarterly* 57(229): 650–666, doi:10.1111/j.1467-9213.2007.510.x.

GIBERMAN, Daniel. 2015. "Is Mereology a Guide to Conceivability?" *Mind* 124(493): 121–146, doi:10.1093/mind/fzu142.

MARTON, Peter. 1998. "Zombies vs. Materialists: The Battle over Conceivability." *Southwest Philosophy Review* 14(1): 131–138, doi:10.5840/swphilreview199814115.

SHEAR, Jonathan, ed. 1997. *Explaining Consciousness: The Hard Problem*. Cambridge, Massachusetts: The MIT Press.

STURGEON, Scott. 2000. *Matters of Mind: Consciousness, Reason and Nature*. London: Routledge, doi:10.4324/9780203424551.

YABLO, Stephen. 1999. "Concepts and Consciousness: Comments on Chalmers (1997)." *Philosophy and Phenomenological Research* 59(2): 455–463, doi:10.2307/2653683.

Review of Willaschek (2018)

Andrew Stephenson

Willaschek, Marcus. 2018. *Kant on the Sources of Metaphysics: The Dialectic of Pure Reason*. Cambridge: Cambridge University Press.

Marcus Willaschek has written an excellent book on Kant's account of reason as the source of metaphysical speculation in the Transcendental Dialectic of the *Critique of Pure Reason* (CPR). There are insights on every page and it will be essential reading for Kant scholars, especially but not only those who work on the theoretical philosophy. Willaschek's writing and presentation make for an exceptionally clear, accessible read, so the book will also be useful for students. It will be of interest to those working on the history of metaphysics and metametaphysics more generally, and it may also be of interest to contemporary metaphysicians and metametaphysicians. The book should become a standard in its field.

In this short review I briefly introduce the topic of the book, its core structure and content, and some selected points of interest.

The *Critique of Pure Reason* is an investigation into the nature, scope, and limits of pure reason. Such an investigation is necessary, according to Kant, because there is a problem. As he puts it in the opening sentence of the A-edition preface:

> Human reason has the peculiar fate in one species of its cognitions that it is burdened with questions which it cannot dismiss, since they are given to it as problems by the nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason. (CPR Avii)

Kant's claims about the limits of pure reason are well-known, well-studied, and have been generally well-received. Many have found devastating his attack on traditional speculative metaphysics concerning the soul, the world, and God. The same cannot be said of Kant's claims about the source and inevitability of metaphysical speculation as arising from the nature of reason

itself. Willaschek's book puts these latter claims front and centre. His aim is to develop a novel and detailed interpretation of them, as well as a partial defence.

Willaschek calls Kant's account of how metaphysical speculation arises naturally and inevitably from the nature of reason the Rational Sources Account. It consists of three distinct theses (e.g. pp. 5, 157):

> RS-1. Rational reflection on empirical questions necessarily raises *metaphysical questions* about "the unconditioned."
>
> RS-2. Rational reflection (by "pure reason") on these metaphysical questions necessarily leads to *metaphysical answers* that appear to be rationally warranted.
>
> RS-3. The rational principles that lead from empirical to metaphysical questions and from there to metaphysical answers are principles of "*universal human reason*"; that is, they belong to rational thinking as such.

According to Willaschek, we can see this account operating at four different "levels" in the *Critique*, each roughly corresponding to one of the four main parts of the Transcendental Dialectic: the Introduction, on the transition from the logical to the real or transcendental use of reason; Book One, on the system of the transcendental ideas of the soul, the world, and God; Book Two, on the dialectical inferences of reason that purport to provide knowledge of the soul, the world, and God; and the Appendix, on the legitimate regulative use of reason's principles in contrast to their illegitimate constitutive use. On Willaschek's reading, Kant lays out the general framework of his Rational Sources Account in the Introduction to the Transcendental Dialectic, before applying this framework and filling in its details in the parts that follow. It is this structure that provides the organizing principle of the book. It divides into two main parts. Part I, chapters 1–5, offers a detailed interpretation of the "first level," general framework of the Rational Sources Account, while Part II, chapters 6–9, moves to the subsequent three levels in which this framework is applied and fleshed out. There is a very useful general introduction, as well as a fascinating, and I hope promissory, postscript on Kant's practical metaphysics.

Chapter 1 gives a terrific overview of Kant's conceptions of reason and metaphysics by way of background and stage-setting. Chapter 2 then concerns the logical use of reason and the Logical Maxim "to find the unconditioned for the conditioned cognitions of the understanding" (CPR A307/B364). Willaschek argues, controversially but forcefully, that the logical use of reason aims at comprehending the systematic unity of *nature*, not merely the hierarchical ordering of cognitions according to generality, and he gives a detailed account of the content of the Logical Maxim, arguing in particular that it concerns both inferential and epistemic conditioning. The chapter closes with a partial defence of Kant's claim, as Willaschek sees it, that the Logical Maxim is a legitimate regulative principle of universal human reason. It "normatively guides the way rational beings (qua scientists) organize their body of cognitions" and is "valid for rational beings as such" (p. 65). This was one of the few parts of the book that I found somewhat unsatisfying, if only because I found myself wanting more.

Willaschek focuses on two issues that he thinks might make Kant's account seem problematic. First, Kant's foundationalist conception of epistemic justification (in the scientific context) and his view that genuinely scientific knowledge must be certain. Second, the tension between the sheer demandingness of the Logical Maxim and the idea that it binds reasoners per se. In the first case, Willaschek simply points out that the supposedly outmoded aspects of Kant's conception of science can be detached from the basic idea that there is a rational requirement to look for general principles from which specific cognitions can be derived. In the second case, we are told that "the Logical Maxim does hold for everyone, but vacuously so for most, since a necessary condition of its making substantive requirements on us and our cognitive activity is not satisfied in most cases" (p. 70). For the Logical Maxim is a hypothetical rather than a categorical imperative, so that "we are rationally required to pursue [systematic unity in the unconditioned] only when doing so is morally permitted and pragmatically feasible" (p. 64), and the important point is that, precisely because of the demandingness of the Logical Maxim, doing so will only very rarely be pragmatically feasible. In the first case, one might worry that Kant's view has been defended only by stripping it of anything especially Kantian. In the second case, one might worry that such a move makes the Logical Maxim rather too hypothetical, to the point that metaphysical speculation starts to seem somewhat less than *inevitable*. But Willaschek makes a number of philosophically and exegetically interesting

doi: 10.48106/dial.v74.i3

points here and what he says is good as far as it goes. I'm just not sure it goes far enough.

What I thought was missing at this point was any general account of the more foundational issue of what it really means for reasoners as such to be "normatively guided" by a principle like the Logical Maxim, for it to be a "valid rational requirement." There are a number of deep issues here, and a number of very different ways to cash out such claims. Kant of course has interesting, controversial things to say. But it wasn't entirely clear to what extent the Rational Sources Account depends on a uniquely Kantian conception of reason. This matters for what is required to defend it. In any case, I would certainly have welcomed seeing Willaschek's expertise brought to bear on the matter.

Chapter 3 moves from the logical to the real use of reason and from the Logical Maxim to the Supreme Principle: "when the condition is given, then the whole series of conditions subordinated one to the other, which is itself unconditioned, is also given" (CPR 307–308/B364). Willaschek argues for an ontological reading of what Kant means by "given" in the Supreme Principle—when the conditioned *exists*, so too must the unconditioned totality of its conditions *exist*—and he defends philosophically and textually sophisticated accounts of the real conditioning relation, the unconditioned, and the relation of the Supreme Principle to the Principle of Sufficient Reason. This is a complex, tightly argued chapter that well repays the close critical scrutiny it requires.

Chapters 4–5 conclude Part I by giving an original and powerful account of the transition from the Logical Maxim to the Supreme Principle. Chapter 4 focuses on a situated textual analysis of what Willaschek calls the Transition Passage (A307–308/B263), a short, one-sentence paragraph on which Willaschek relies heavily and out of which he teases a lot. Chapter 5 lays out the core philosophical account of the transition that we will see play out at different levels in the chapters that follow. Crucially, Willaschek understands this transition as involving two stages, first the transition from the Logical Maxim to the regulative use of the Supreme Principle, and second the transition from the regulative to the constitutive use of the Supreme Principle. Only the first stage is rationally necessary, in a nutshell because the Logical Maxim, concerning as it does on Willaschek's reading the systematic unity of *nature*, necessarily presupposes the regulative use of the Supreme Principle, which recall concerns the *existence* of the unconditioned. The second stage, by

contrast, merely appears rationally necessary under the (supposedly natural but ultimately spurious) assumption of transcendental realism.

Willaschek gives an intriguing account of exactly how transcendental realism is the "key" to transcendental illusion. Starting with Kant's basic definition of transcendental realism as the view that empirical objects are identical to things in themselves, he argues—via a discussion of noumena and the intuitive intellect—that transcendental realism ultimately comes down to the view that "[t]here is a necessary correspondence between the principles of reason and the principles of reality" (p. 144). From there he proposes that such a view can plausibly be thought a tacit background assumption of everyday rational thinking or common sense, and that this explains the way in which there is a natural tendency, that will forever assert its pull, towards transcendental illusion.

Willaschek gives bivalence as an example of a principle of reason that, on his reading, Kant thinks it would be a mistake to treat as a principle of reality (p. 149). Willaschek's reasoning here seems based on a misreading of the Antinomies. The problem is not that Kant doesn't reject the principle that Willaschek takes him to reject, namely:

> BIVALENCE$_{\text{W}}$. Of the two cosmological claims "The world is finite in magnitude" and "The world is infinite in magnitude," precisely one is true and one is false.

Kant does of course reject BIVALENCE$_{\text{W}}$. But his whole point in the Antinomies, it seems to me, is that claims like BIVALENCE$_{\text{W}}$ don't really follow from the principle of bivalence because the propositions in question, such as those concerning the magnitude of the world, are not really contradictories. They only appear so under some false presupposition. This is clear, I think, from Kant's example of a body that has no smell and thus smells neither good nor not good (A503–504/B531–532). He says that the analogy holds for the other antinomies (A505/B533), the difference being that while the (apparently but not really contradictory) propositions of the mathematical antinomies can both be false, the (apparently but not really contradictory) propositions of the dynamical antinomies can both be true. Nowhere does Kant suggest that there are propositions here that are something other than either true or false.

Now, it may well be that doubts about bivalence somehow follow from Kant's transcendental idealism. But that won't help Willaschek. He later suggests (chapter 9) that it's a philosophical benefit of the Rational Sources

doi: 10.48106/dial.v74.i3

Account that it depends only on a rejection of transcendental realism and not on the acceptance of transcendental idealism. This is coherent, he argues contra Kant, because the two views are contraries rather than contradictories: transcendental realism says that there is a necessary correspondence between the principles of reason and the principles of reality, while transcendental idealism insists that this correspondence only holds for appearances, not things in themselves. One might well deny that there is any such correspondence. Put to one side the plausibility of attributing to Kant such a basic misunderstanding of the relation between transcendental realism and transcendental idealism, especially in a context in which he is so acutely attendant to scope ambiguities and the contrary/contradictory distinction. If Kant presupposes realm-spanning principles of reason in diagnosing and rejecting transcendental realism, that puts pressure on Willaschek's interpretation of the doctrine.

The discussion of "levels" two through four in Part II goes by much more quickly than that of the first level in Part I, as the first level has provided the general template that is then applied and fleshed out at the three subsequent levels. This shift in gear between Part I and Part II enhances rather than detracts from the book, which thereby manages to bring out superbly the often elusive structural similarities between Kant's treatment of the different areas of traditional speculative metaphysics in this long, labyrinthine part of the *Critique*.

Chapter 6 concerns the "second level" transcendental ideas of the soul, the world, and God, and how they are supposed to arise necessarily from rational reflection. They do not, according to Willaschek (and as he admits Kant seems to suggest), arise from the *mere forms* of rational inferences, but rather from rational inferences about specific subject matters in psychology, cosmology, and theology. Their derivation or metaphysical deduction, then, does not itself take place until the "third level," that of the dialectical inferences of reason treated in the Paralogisms, the Antinomies, and the Ideal. This level is the concern of chapters 7–8, with chapter 8 also including Willaschek's account of the "fourth level" regulative-constitutive transition treated in the Appendix. Finally, chapter 9 rounds out the discussion by relating the Rational Sources Account—Kant's account of reason as the source of inevitable metaphysical speculation—back to Kant's more famous account of the limits of reason, i.e. his critique of traditional speculative metaphysics. It is here that Willaschek argues that Kant's general diagnosis of what goes wrong in trying to gain knowledge of the unconditioned or supersensible, namely the tacit

assumption of transcendental realism, is independent of any commitment to transcendental idealism, since the two views are contraries rather than contradictories. The concerns I noted above notwithstanding, this is another particularly excellent chapter. It, along with parts of chapters 5 and 7, will be of special interest to those working on Kant's signature doctrine.

In addition to the general introduction, Part I and Part II each have their own introductions and conclusions, and the same is more or less true of each of the nine individual chapters. All this signposting is welcome. It aids comprehension and it makes the book eminently *usable* for scholars and students alike, as does the fact that the book is extremely well situated in the literature with extensive references throughout. Willaschek's book is an extremely welcome addition to the literature on the Transcendental Dialectic and Kant's metametaphysics more generally.

Andrew Stephenson
0000-0003-4590-1307
University of Southampton, UK
Andrew.Stephenson@soton.ac.uk

doi: 10.48106/dial.v74.i3

Review of Esfeld and Deckert (2018)

Alastair Wilson

Esfeld, Michael & Dirk-André Deckert. 2018. *A Minimalist Ontology of the Natural World*. Routledge Studies in the Philosophy of Mathematics and Physics. London: Routledge.

An aesthetic preference for the minimal is widespread within contemporary metaphysics: consider Sider's mereological nihilism, Paul's mereological bundle theory, Lewis's Humean Supervenience, Quine's desert landscapes. Even in this climate, the ontology proposed in *A Minimalist Ontology of the Natural World* is strikingly austere. It is also strikingly ambitious: Michael Esfeld and Dirk-André Deckert (along with their collaborators Dustin Lazarovici, Andrea Oldofredi, and Antonio Vassallo) propose a systematic metaphysics of nature and argue that it will be adequate to underwrite any possible future physics. The basic elements of the system, which they call Super-Humeanism, are indestructible featureless objects (referred to as "matter points," though there is little recognisably material about them) that are related together in (and individuated by) a changing pattern of spatial distances.[1] Everything else reduces to, or is grounded in, that pattern. The view is squarely in an atomist tradition that the authors trace back via Feynman and Newton to Democritus and Leucippus.

The book proceeds through dauntless construction of the positive Super-Humean view. Arguments for the view are offered, but theory-building is the main focus. In chapter 2, the authors set out a sparse framework of primitive metaphysical ingredients, and outline a general recipe for interpreting physical theories in terms of that framework. This recipe is illustrated by application in chapter 3 to Newtonian gravitational theory, classical electrodynamics and non-relativistic quantum particle dynamics, and—most courageously—in

1 The book's cover evokes one such pattern, featuring a sparse network of yellow and blue nodes linked by yellow and blue lines. Even this sparse image overstates the content of the super-Humean ontology, however; Esfeld and Deckert allow only for one kind of basic entity standing in one kind of basic relation.

chapter 4 to quantum field theory. Chapter 5 explains how relativistic physics is handled.

As already noted, the Esfeld-Deckert framework has a strikingly minimalist aesthetic. This is not, however, the same sense of "minimalist" intended in the book's title. Super-Humeanism is represented as minimal in a more literal way: it is argued that there is no way to reduce the ontology further while remaining empirically adequate. As I shall explain, however, this claim is less ambitious than it may first appear. The proposed framework is put forward as a local minimum in theory space rather than as a global minimum. Moreover, the very idea of a global minimum is of dubious coherence, since different dimensions of minimality may be incommensurable. I shall return to these points below.

What then is the world like according to Esfeld and Deckert? It consists of "matter in space and time, being subject to certain laws, explain[ing] the observable phenomena" (p.1). Matter is taken to be basic: no reduction of it to anything else (such as quantum field configurations) will be countenanced, and matter can neither be created nor destroyed. Spatial relations are likewise taken to be basic: no reduction of them to anything else (such as spin networks in loop quantum gravity) will be countenanced. Temporal relations are not quite basic, though change is. The choice of spatial relations and change as distinct basic notions renders the view non-relativistic at the deepest level, though relativistic physics is emulated at a non-fundamental level. This is a significant point of departure from other versions of Humeanism in the literature (such as those of David Lewis and Barry Loewer) that generally take spatio-temporal relations as basic, and are formulated in a timeless (eternalist) fashion. Super-Humeanism, in contrast, incorporates a form of presentism.

Why assume changing spatial configurations of objects as the basic ingredients? The authors' primary reason lies in a distinctive conception of our evidence: "We adopt an empiricist attitude in insisting on the fact that all the experimental evidence consists in relative particle positions and motion" (p. 12). The idea is that spatial distance has a unique epistemological role: any evidence (or at least, any evidence that we can obtain) ultimately boils down to evidence about spatial arrangements of particles. The idea is that when we take readings from an analogue instrument, what we are really doing is comparing the position of a pointer with the position of markings on a background scale, and that when we take readings from a digital instrument we are really comparing the relative positions of bright and dark pixels. Even observation of the colours of things is explained in terms of positions

of particles: "the frequencies that we usually identify with red light or blue light or green light are taken to refer directly to accelerations of particles" (p. 136). In prioritising the epistemic role of position, Esfeld and Deckert align themselves with a tradition that has come to prominence in recent philosophy of physics, inspired in large part by the work of the physicist John Bell: the so-called *primitive ontology* programme. According to primitive ontologists, the physical world consists of what Bell called "local beables"—localized definite states of affairs, independent of goings-on at other locations. Bell endorsed the epistemic primacy of position measurements, and this line of thought has influenced the work of philosophers such as Valia Allori, Sheldon Goldstein, and Tim Maudlin.

I am sceptical of the arguments offered, in this book and elsewhere, for the epistemic primacy of position. Obviously, once we accept the Super-Humean ontology all evidence becomes evidence of relative position of particles: if everything is made of spatially-arranged particles, then there is nothing else for evidence to be about. Just as obviously, it would be problematically circular to use this point to argue for the Super-Humean ontology. For those who don't antecedently accept a particle-only ontology, it seems that our evidence might in principle take many forms: an example which sometimes comes up in conversation is a measurement device which encodes its readings in the frequency of light that it emits. The epistemic primacy of position requires that we reconceptualize all such evidence as being evidence of position in disguise (as in the quote above from p. 136); however, the same trick could equally be turned to prioritize other physical properties. When we give priority to evidence about position, what we are doing is picking out a particular feature of a long causal chain from measurement to conscious perception and identifying that particular feature—the positions of some key particles—as what the evidence really consists in. But other features of the chain might be prioritized instead, features more congenial to non-particle ontologies. For instance, essentially every causal chain that results in a conscious perception involves electromagnetic interactions at the boundaries of some neurons. Why not identify the electromagnetic field in some suitable region as what our evidence really consists in? While this line of objection could be pursued further, I shall set it aside for the remainder of this review. It is, after all, commendable that the authors should be so clear about the intended epistemological foundation for their metaphysical constructions.

Another component of the Esfeld-Deckert framework seems to have purer metaphysical motivations. The framework is relationist, in the sense that

doi: 10.48106/dial.v74.i3

space and time themselves do not exist as independent entities. What we have is a pattern of spatial relations connecting objects. But the objects thus connected are not conceived as existing independently of the relations they stand in; rather, their existence and identity depend upon their position in the network of relations. The authors thus defend the "moderate structural realism" familiar from Esfeld's previous work with Vincent Lam (2008). According to moderate structural realism, objects and relations are mutually dependent: the basic objects are dependent upon relations (which individuate them), while the basic relations are dependent upon those very objects (in order to have something to hold between). This view, while interesting, is considered speculative even within the open-minded field of metaphysics of science. It has some peculiar consequences, including the exclusion—as metaphysically impossible—of globally symmetrical patterns of spatial relations (mirror universes). Most of the physical theories which the authors consider would seem to have models corresponding to these mirror scenarios, and ruling them out as metaphysically impossible (essentially on grounds of convenience) strikes me as rather ad hoc or at least as a departure from the naturalistic outlook. There is also a hint of a double standard when proposals by Belot and Barbour are criticized for excluding the apparent physical possibility of the universe having non-zero angular momentum (p. 67). Fortunately, moderate structural realism only plays a limited role in the overall framework, and it can be factored out of the view relatively straightforwardly if desired, as George Darby (2018) has argued.

It is no great surprise that an ontology of persistent particles fits well with Newtonian mechanics: Newton imagined a world composed of particles (or corpuscles), and presented his mechanics as a theory of such a world. Although the application of a particle-only ontology to non-relativistic quantum mechanics might seem much more surprising (what about the quantum state?), it will be familiar to philosophers of physics from versions of Bohmian mechanics which regard the wavefunction as nomological in nature.[2] It is even more of a surprise to find a particle-only ontology paired with quantum field theory, and indeed the interpretation of QFT that is offered by Esfeld and Deckert (building on the Wheeler-Feynman absorber theory) has some very strange features. There are no such things as photons, or Higgs bosons, or indeed any bosons at all; there are only distinctive patterns of motions of

2 Indeed some extant versions of Bohmianism come very close to the Super-Humean position—see for example Miller (2014) and Bhogal and Perry (2017).

fermions. Fermions are never in fact created or annihilated: they only become detectable or undetectable. The number of particles is fixed and finite; making it variable or infinite introduces deep pathologies into the theory. Here is not the place for a thorough assessment of the proposed particle-based approach to QFT; for critical discussion, see Caulton (2018). Still, it is worth noting how radically QFT is here being reimagined. If Esfeld and Deckert are correct then the large majority of foundational work on QFT (in regarding particles as emergent and quantum fields as basic) is misconceived, and the intuitive pictures used by working quantum field theorists are deeply mistaken. Unfamiliar though the particle-based approach is, the application of the primitive ontology picture to QFT is developed in an admirably clear and thorough way by Esfeld and Deckert, and it is one of the most significant contributions of the book.

I return now to the question of ontological minimality and how to assess it. Distinguish two senses in which a philosophical proposal can be minimalist: minimalist in the design sense and minimalist in the literal sense. Minimalist in the design sense is the meaning of the term that will be familiar to a general readership: an aesthetic preference where less is more. Minimalist in the literal sense, by contrast, is a bold theoretical claim: to say of a metaphysical system that it is minimal in this sense is to say that no system is objectively more parsimonious than it is, that no system entails the existence of objectively less stuff than it does. While it is undeniable that the Esfeld-Deckert view is design-minimalist—there is austere beauty in their image of the world as an intricate dance of particle motions, a silent choreography of changes in relative position of pointlike elements—it is open to question to what extent it is literal-minimalist. On closer inspection, it emerges that the positive claim that the authors wish to make is more restricted than some of the very ambitious claims that might be attributed to them based on the book's title. Esfeld and Deckert argue that subtracting any elements from their view renders it inadequate (which is plausible) and that taking their view and adding additional elements won't help (which I think is doubtful,[3] but which I will grant for present purposes). Even if they are correct about both these points, though, what this shows is that Super-Humeanism is a local minimum

3 My main reason for doubt is that the book contains no sustained argument that the Super-Humean system is adequate to support the explanatory needs of higher-level science. Indeed, there is prima facie reason to suspect that it is unable to recapture the full explanatory role of physical state spaces. I say more about this problem in Wilson (2018), a companion paper to the present review; see also Lazarovici (2018).

doi: 10.48106/dial.v74.i3

of complexity in the space of empirically adequate fundamental theories. It does not establish the stronger claim that the view is the global minimum of that space—if such a minimum even exists. In discussions with the authors, they have clarified that only the local-minimum claim is intended; but for purposes of fundamental ontology, the global-minimum claim is of primary interest. To make a case for Super-Humeanism as a global minimum, it would have to be compared in some way with rival frameworks. This task is not attempted in the book; Esfeld and Deckert justify this via the claim that no fully-developed field-theoretic alternative fundamental ontology has been set out in the literature. Whether or not this is true, at least the general shape of such views is familiar from foundational discussions, and they can be compared in schematic ways with particle-based approaches. Nor do the authors provide any substantive discussion about what it takes to be minimal, and in particular they say nothing about how to compare minimality for theories that employ different kinds of entities. Super-Humeanism is undoubtedly more minimal than a view that encompasses everything asserted by Super-Humeanism, and in addition recognizes seventeen further scalar fields over spacetime, none of which interact with anything else. But the most interesting questions in the vicinity are not about how these two theories compare, but about how Super-Humeanism compares with other systematic proposals with a wholly different fundamental ontology—with a field-theoretic realism along the lines of Wallace and Timpson's "spacetime state realism" (2010), for example.

Overall, this book is a significant achievement and it will be a standard reference point in the literature on fundamental ontology. The Super-Humean view is set out with clarity, precision and honesty, and new ground is broken in the application of the primitive ontology programme to quantum field theory. The natural world as Esfeld and Deckert conceive it may seem a barren place to live, but careful attention to their vision is still likely to bear philosophical fruit.

Alastair Wilson
0000-0003-3027-5016
University of Birmingham
Monash University
a.j.wilson@bham.ac.uk

References

Bhogal, Harjit and Perry, Zee R. 2017. "What the Humean should Say About Entanglement." *Noûs* 51(1): 74–94, doi:10.1111/nous.12095.

Caulton, Adam. 2018. "A Persistent Particle Ontology for Quantum Field Theory [review of Esfeld and Deckert (2018)]." *Metascience* 27(3): 439–441, doi:10.1007/s11016-018-0323-1.

Darby, George. 2018. "A Minimalist Humeanism? [review of Esfeld and Deckert (2018)]." *Metascience* 27(3): 433–437, doi:10.1007/s11016-018-0324-0.

Esfeld, Michael and Deckert, Dirk-André. 2018. *A Minimalist Ontology of the Natural World*. Routledge Studies in the Philosophy of Mathematics and Physics. New York: Routledge, doi:10.4324/9781315142272.

Esfeld, Michael and Lam, Vincent. 2008. "Moderate Structural Realism About Space-Time." *Synthese* 160(1): 27–46, doi:10.1007/s11229-006-9076-2.

Lazarovici, Dustin. 2018. "Super-Humeanism: A Starving Ontology." *Studies in History and Philosophy of Science. Part B: Studies in History and Philosophy of Modern Physics* 64: 79–86, doi:10.1016/j.shpsb.2018.07.001.

Miller, Elizabeth. 2014. "Quantum Entanglement, Bohmian Mechanics, and Humean Supervenience." *Australasian Journal of Philosophy* 92(3): 567–583, doi:10.1080/00048402.2013.832786.

Wallace, David and Timpson, Christopher G. 2010. "Quantum Mechanics on Spacetime I: Spacetime State Realism." *The British Journal for the Philosophy of Science* 61(4): 697–727, doi:10.1093/bjps/axq010.

Wilson, Alastair. 2018. "Super-Humeanism: Insufficiently Naturalistic and Insufficiently Explanatory [review of Esfeld and Deckert (2018)]." *Metascience* 27(3): 427–431, doi:10.1007/s11016-018-0326-y.

Review of MacBride (2018)

Chris Daly

MacBride, Fraser. 2018. *On the Genealogy of Universals: The Metaphysical Origins of Analytic Philosophy*, Oxford: Oxford University Press.

This is a remarkable and far-reaching book written with impeccable scholarship and considerable acumen. The timeline of its study begins in the late 1890s with Russell and Moore's declaration of their "New Philosophy." It continues through developments in their thought and the contributions of Stout, Whitehead and Wittgenstein. Ramsey's writings in the 1920s, at once synoptic and iconoclastic, conclude the study. The connecting thread concerns how the early analytic philosophers' evolving conception of the particular-universal distinction both influenced and was influenced by their evolving conception of analytic philosophy. It was a process of mutually beneficial illumination in which Cambridge was the crucible of the analytic enterprise. The agenda facing these philosophers was to select between the options provided by two orthogonal distinctions. Ontological pluralism and ontological monism differ about the number of entities there are, whereas categorial monism, categorial dualism and categorial pluralism differ about the number of categories.

One of the many refreshing elements of the book is how it upends much conventional wisdom about the origins of analytic philosophy. For instance, it selects Kant rather than Frege as the progenitor. By raising the question of what the categories are, the forms of representation essential for thought, Kant called into question the concepts of substance and attribute and the relation between them. His Metaphysical Deduction was an ill-fated rescue mission to save these and other "pure concepts of understanding." What remains from Kant, and was taken up by Moore, was the conviction that the particular-universal distinction was indissoluble: either both sides obtain or neither. Switching to the formal mode, a predicate is what is predicated of a subject whereas a subject is the subject of predication. Again, Moore is well known for rejecting Hegel's ontological monism but MacBride convincingly shows that Moore's "The Nature of Judgement" (1899) endorses categorial monism: that in taking concepts to comprise the only category, he rejected

the particular-universal distinction. Russell bolstered this endorsement by drawing on Bradley and arguing that the idea of a substance is misconceived since the parent idea of something determinately referred to and described by a subject-predicate judgment is itself untenable. MacBride very effectively mines Russell's unduly neglected *The Philosophy of Leibniz* (1900) for these and other early statements in the "New Philosophy" of how and why the subject-predicate framework is to be abandoned. Perhaps Russell's most striking argument runs: The surface form of language is misleading, for we can as well say "Humanity belongs to Socrates" as "Socrates is human" and in each case we express the same judgment. "Humanity" may not be the grammatical subject of the second quoted sentence, but humanity belongs just as much to the subject matter of that sentence as Socrates does. Russell's philosophical development involved considerable turbulence: in *The Principles of Mathematics* (1903) he rejected this argument against categorial dualism, only for the argument subsequently to be refurbished by Ramsey and deployed against him. MacBride sees the "New Philosophy" as having a revolutionary phase followed by a reactionary one, as Russell and Moore lapsed from advocacy of categorial monism to apostasy. What ensued was an extended episode of whack-a-mole in which periodic resurgences in Cambridge of the particular-universal distinction, often in a non-standard form, were subjected to strenuous criticism by Stout, Whitehead and Ramsey.

I noted that MacBride displaces Frege with Kant as the *ur*-source of analytic philosophy. MacBride also makes clear how much progress Russell and Moore made independently of Frege. This is especially evident in their appreciation of the structural significance of relations. Entities stand in different relations only if those entities are distinct and so, Russell and Moore inferred, numerical distinctness is not to be accounted for in terms of relational difference. Moore went further: if there are only universals, from this f being over here and that f being over there, it does not follow that there are two fs. It follows only that f recurs—that f is over here and over there. To secure bearers for such universals, and to safeguard our ordinary judgment that there are two fs, Moore invoked the category of particulars, understood as instances of universals. Categorial dualism was thereby reinstated. Ordinary objects were identified with clusters of property-instances, allowing the possibility of distinct property-instances that are themselves exactly alike.

By 1911, however, Moore found the particular-universal distinction to be unclear. One consideration is that the supposed category of universals is gerrymandered, containing monadic properties, relational properties and

relations. This realization opened up the possibility of categorial pluralism: that there is no *a priori* limit to the number of categories. Beginning in 1905 and working independently, Whitehead also drew this conclusion. By rejecting a bifurcation of nature between primary and secondary qualities or of what is observable and what is an instrumental posit, he permitted a diverse ontology, a host of entities that apparently share no common features and so belong to a multiplicity of categories.

During the 1910s and 1920s Stout embraced ontological pluralism and categorial monism on *a posteriori* grounds: what we perceive is a wealth of property-instances (where property-instances form a single primitive kind of entity) that are unified in concrete or distributive ways. These concrete ways yield ordinary particulars and the distributive ways yield determinate or determinable qualities. Universals are eschewed. Stout and Moore subsequently locked horns over the nature of property-instances. In MacBride's opinion, Moore had the better of the argument. First, ordinary speakers do not draw upon a grasp of the theory of classes to understand predicative sentences. *A fortiori*, they do not understand "the glass is fragile" in terms of the glass having a property-instance that is a member of the class of fragile things. Second, to say that the glass and the spider web are fragile is to predicate the same thing of them. Distinct particulars, however, cannot have the same property-instance, although they can have the same universal.

From 1903 Russell's understanding of the particular-universal distinction evolved as his thoughts changed about both propositions and relations. Russell abandoned realism about propositions because of the problem of accounting for false propositions. He came to treat talk of propositions as a mere way of speaking, thanks to his multiple relation theory of judgment and his conjecture that judgment relations have significant higher-order structure. MacBride argues that Russell retained that theory up until 1919 (through his logical atomism phase), reviving his earlier view that each non-symmetric relation has a "direction." To deflect Wittgenstein's famous 1913 criticism of the theory, Russell reverted to a quasi-Fregean view that concepts (i.e. universals) were exclusively predicative and incapable of serving as logical subjects. Notably, Russell recognized that belief ascriptions have different logical forms according to what is believed. There is then no constraint on the number of categories that might be involved in the content of a belief and the way to categorial pluralism is open.

In 1906–1907 Russell devised a proto-picture theory of language that inspired the *Tractatus*; Russell's doubts about his own theory were addressed by

doi: 10.48106/dial.v74.i3

Wittgenstein's more developed efforts. Moreover, the theory that propositions are pictures of reality and that complex propositions are truth-functions of elementary propositions dispenses with the particular-universal apparatus. This liberation movement reached its apogee with Ramsey's incisive contributions. For Ramsey, there is no *a priori* reason why the language required for expressing and characterising atomic facts will be anything like predicate calculus. For example, there might be not just two but three or four or *n* different modes of basic grammatical combination; the overlapping capacities of individual expressions to combine with one another may confound any binary distinction.

MacBride's book investigates a microcosm and a macrocosm. The microcosm is the particular-universal distinction. The macrocosm is a debate between monism, dualism and pluralism: a debate about how many ontological categories there are. Could the macrocosmic debate have been engaged with even if a different microcosmic debate had arisen? Maybe, instead of the particular-universal dualism, a different venerable dualism could have been dominant in early analytic philosophy: the dualism of mind and body. Is it then a historical accident that the debate about the particular-universal distinction assumed the significance that it did? If Russell had written *The Analysis of Mind* (1921) some thirty years earlier, would the debate between Cartesian dualism, idealism and neutral monism have secured foundational status? The particular-universal distinction was central because it is related to questions about the forms of judgement, questions emphasised by Kant, and also to questions about the forms of sentences, questions elevated by the linguistic turn. Nevertheless, had Kant not usurped Descartes, the nature of mind could have remained pivotal. Perhaps the fact that philosophy of mind took centre stage in philosophy in the 1970s and 1980s was redressing an imbalance, albeit resulting in an imbalance of its own.

MacBride has written an absolutely first rate study of early analytic philosophy. The clarity of his writing, the carefulness of his elucidations, the brilliance of his metaphysical discussions, as well as his sympathetic approach to the writing of the Cambridge philosophers, mark out this important and profound work.

Chris Daly
University of Manchester
christopher.daly@manchester.ac.uk

References

MOORE, George Edward. 1899. "The Nature of Judgment." *Mind* 8(30): 176–193. Reprinted in Moore (1993, 1–19), doi:10.1093/mind/viii.2.176.

—. 1993. *Selected Writings*. London: Routledge. Edited by Thomas Baldwin, doi:10.4324/9780203769812.

RUSSELL, Bertrand Arthur William. 1900. *A Critical Exposition of the Philosophy of Leibniz. With an Appendix of Leading Passages*. Cambridge: Cambridge University Press. Second edition: Russell (1937b).

—. 1903. *The Principles of Mathematics*. London: Taylor & Francis. Second edition: Russell (1937a), third edition: Russell (2020).

—. 1921. *The Analysis of Mind*. London: George Allen & Unwin.

—. 1937a. *The Principles of Mathematics*. 2nd ed. London: George Allen & Unwin. Second edition of Russell (1903), with a new introduction; third edition: Russell (2020).

—. 1937b. *A Critical Exposition of the Philosophy of Leibniz. With an Appendix of Leading Passages*. 2nd ed. London: George Allen & Unwin. Second edition of Russell (1900), with a new preface; republication: Russell (1992).

—. 1992. *The Philosophy of Leibniz*. London: Routledge. Reprint of Russell (1937b); with a new introduction by John G. Slater, doi:10.4324/9780203987292.

—. 2020. *The Principles of Mathematics*. 3rd ed. London: Routledge. Third edition of Russell (1903), doi:10.4324/9780203822586.

Published by *Philosophie.ch*

Verein philosophie.ch
Fabrikgässli 1
2502 Biel/Bienne
Switzerland
dialectica@philosophie.ch

https://dialectica.philosophie.ch/

ISSN 1746-8361

Dialectica is supported by the Swiss Academy of Humanities and Social Sciences.

Abstracting and Indexing Services

The journal is indexed by the Arts and Humanities Citation Index, Current Contents, Current Mathematical Publications, Dietrich's Index Philosophicus, IBZ — Internationale Bibliographie der Geistes- und Sozialwissenschaftlichen Zeitschriftenliteratur, Internationale Bibliographie der Rezensionen Geistes- und Sozialwissenschaftlicher Literatur, Linguistics and Language Behavior Abstracts, Mathematical Reviews, MathSciNet, Periodicals Contents Index, Philosopher's Index, Repertoire Bibliographique de la Philosophie, Russian Academy of Sciences Bibliographies.

Contents