

# dialectica

International Journal of Philosophy

## Contents

PHILIPP BLUM, <i>Editorial: Dialectica Goes Open Access</i> .....	1
VINCENT CONITZER, <i>The Personalized A-Theory of Time and Perspective</i> ...	3
NADINE ELZEIN, <i>Determinism, "Ought" Implies "Can" and Moral Obligation</i>	33
LUKE ROELOFS, <i>Consciousness, Revelation, and Confusion: Are Constitutive Panpsychists Hoist by their Own Petard?</i> .....	63
DONNCHADH O'CONNELL, <i>In Defence of Facts: Grounding, Essential Properties and the Unity Problem</i> .....	97
HSENG XIONG JOSHUA GOH, <i>Strevens's Counterexample to Lewis's "Causation as Influence" and Degrees of Causation</i> .....	127
MARC-KEVIN DAoust, <i>Consistency, Obligations, and Accuracy-Dominance Vindications</i> .....	141
FRASER MACBRIDE, <i>Review of Soames (2018)</i> .....	159
MARIO SCHÄRLI, <i>Review of Oppy (2018)</i> .....	165
HAMID TAIEB, <i>Review of Antonelli (2018)</i> .....	173

# 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

## Editorial Board

Jérôme Dokic, EHESS, Paris, France  
Pascal Engel, EHESS, Paris, France  
Manuel García-Carpintero, Universitat de Barcelona, Spain  
Diego Marconi, Università di Torino, Italy  
Carlos Moya, Universitat de València, Spain  
Martine Nida-Rümelin, Université de Fribourg, Switzerland  
François Recanati, Collège de France, Paris  
Marco Santambrogio, Università degli Studi di Parma, Italy  
Peter Simons, Trinity College Dublin, Ireland  
Gianfranco Soldati, Université de Fribourg, Switzerland  
Marcel Weber, Université de Genève, Switzerland

## Editors

Fabrice Correia, University of Geneva  
Philipp Blum, University of Lucerne (also managing editor)

## Review Editors

Stephan Leuenberger and Philipp Blum

## Editorial Committee

Philipp Blum (né Keller), Claudio Calosi, Zoé Christoff, Fabrice Correia, Catharine Diehl, Matthias Egg, Patrik Engisch, Andrea Giananti, Jörg Löschke, Arturs Logins, Giovanni Merlo, Robert Michels, Ryan Miller, Paolo Natali, Donnchadh O'Conaill, François Pellet, Edgar Phillips, Steph Rennick, Maria Scarpati, Mike Stuart, Fabrice Teroni, Daniel Vanello, Lisa Vogt.

## Consulting Board

Johannes Brandl (Salzburg), João Branquinho (Lisboa), Elke Brendel (Bonn), Ingar Brinck (Lund), Eros Corazza (Ikerbasque and Carleton), Josep Corbi (València), Michael Esfeld (Lausanne), Dagfinn Føllesdal (Stanford and Oslo), Frank Jackson (Australian National University, Canberra), Max Kistler (Paris I), Max Kölbel (Wien), Jan Lacki (Genève), Karel Lambert (Irvine), Paolo Leonardi (Bologna), Fraser Macbride (Manchester), Josep Macià (Barcelona), Genoveva Martí (Barcelona), Élisabeth Pacherie (Institut Jean Nicod, Paris), David Piñeda (Girona), Wlodek Rabinowicz (Lund), Barry Smith (Buffalo), Thomas Strahm (Bern), Christine Tappolet (Montréal), Neil Tennant (Ohio State), Mark Textor (King's College London), Achille Varzi (Columbia University, New York), Alberto Voltolini (Torino), Timothy Williamson (Oxford).

March 2020

## Contents

PHILIPP BLUM, <i>Editorial: Dialectica Goes Open Access</i> . . . . .	1
VINCENT CONITZER, <i>The Personalized A-Theory of Time and Perspective</i> . . .	3
NADINE ELZEIN, <i>Determinism, “Ought” Implies “Can” and Moral Obligation</i>	33
LUKE ROELOFS, <i>Consciousness, Revelation, and Confusion: Are Constitutive Panpsychists Hoist by their Own Petard?</i> . . . . .	63
DONNCHADH O’CONNAILL, <i>In Defence of Facts: Grounding, Essential Properties and the Unity Problem</i> . . . . .	97
HSENG XIONG JOSHUA GOH, <i>Strevens’s Counterexample to Lewis’s “Causation as Influence”, and Degrees of Causation</i> . . . . .	127
MARC-KEVIN DAoust, <i>Consistency, Obligations, and Accuracy-Dominance Vindications</i> . . . . .	141
FRASER MACBRIDE, <i>Review of Soames (2018)</i> . . . . .	159
MARIO SCHÄRLI, <i>Review of Oppy (2018)</i> . . . . .	165
HAMID TAIEB, <i>Review of Antonelli (2018)</i> . . . . .	173

PROOF

# Editorial

## *Dialectica* Goes Open Access

PHILIPP BLUM

We are happy to announce that *Dialectica* is now an Open Access journal. Starting with this issue, the journal has adopted the so-called “Platinum” or “Diamond” Open Access model under which we do not charge access fees to readers nor article processing charges to authors. Thanks to the generous support of the Swiss National Science Foundation (CRSK-1-190939) and swissuniversities (OAHUBSP), all new issues are available at [dialectica.philosophie.ch](http://dialectica.philosophie.ch). In the near future, we hope to make freely available a full bibliography and a detailed submission statistics.

We hope that this audacious step will increase our readership, attract more and better submissions and reward our many industrious referees, without whom our journal would not be possible. By making all tools and techniques devised for our OA-transition freely available, by documenting the transition process itself, putting the reasoning behind our decisions out for your scrutiny and disclosing the difficulties encountered in establishing a sustainable financial model, we also hope to convince other well-established philosophy journals to free themselves of the increasingly tight grip of profit-oriented publishers and to turn the ideals of Open Science into action.

The success of our discipline’s transition to Open Access will depend on four factors:

- that established journals, in particularly the most important ones, start the transition to Open Access immediately, with the aim to sever their links to for-profit publishers;
- that the many new journals coming into existence, reflecting in theme and character the diversity of our growing discipline, are all fully (i.e. Platinum) Open Access;
- that philosophers strongly support the transition by making a collective decision to only submit to, referee for, and edit Open Access journals;

- that our funding bodies make their support contingent on publication in OA venues, and that new funding models are devised which allow universities and libraries to directly fund the running costs of OA journals – an “OA coalition” of philosophy journals should be created to press for this cause.

While we hope to have contributed to the first of these objectives, our transition has two further goals. First, to strengthen our institutional basis in Switzerland, notably by recruiting more Swiss philosophers into the Editorial Committee. Second, to make the refereeing process faster and more positive-oriented. In our new “fishpond” model, members of the Editorial Committee pick anonymized papers that they hope to promote and send them out to referees. If, based on the reports, they make a successful case to the committee, the papers are accepted. Submissions are sent back to the authors after one month if they have not been picked up. This is not to be understood as a (desk-) rejection, but simply as an acknowledgment of the limitedness of our resources. We hope that this triple-blind, positive-oriented process will shorten the turn-around time for authors and make the editorial and referee work more attractive.

The ongoing transition process has already benefitted from much help, including much needed technical advice by Denis Maier and Albert Krewinkel. Let me thank, first and foremost, my co-editor Fabrice Correia, the members of the Editorial Board and the Editorial Committee, Julien Dutant, the head of the *Dialectica* OA initiative, the members of the SNSF *Spark* project team (Jonathan Biedermann, Sharon Casu, Thomas Hodgson, Nemo Krüger, Ryan Miller, Sandro Räss, Marco Toscano, Christian Weibel), our library consultants Rebecca Iseli Büchi and Gian-Andri Töndury and the new managing editor of *Dialectica*, Marco Schori – philosophers who sacrificed some of their research time to make a practical impact in our common quest to make the world a little better.

Philipp Blum

 0009-0007-2821-8203

Editor and Project Leader

philipp.blum@philosophie.ch

# The Personalized A-Theory of Time and Perspective

VINCENT CONITZER

A-theorists and B-theorists debate whether the “Now” is metaphysically distinguished from other time slices. Analogously, one may ask whether the “I” is metaphysically distinguished from other perspectives. Few philosophers would answer the second question in the affirmative. An exception is Caspar Hare, who has devoted two papers and a book to arguing for such a positive answer. In this paper, I argue that those who answer the first question in the affirmative—A-theorists— should also answer the second question in the affirmative. This is because key arguments in favor of the A-theory are more effective as arguments in favor of the resulting combined position, and key arguments against the A-theory are ineffective against the combined position.

In a series of unconventional but lucid works, Caspar Hare has laid out and defended a theory of *egocentric presentism* (or, in his more recent work, *perspectival realism*), in which a distinguished individual’s experiences are *present* in a way that the experiences of others are not (2007, 2009, 2010). Closely related ideas appear in the writings of others. One example is Valberg’s notion of the “personal horizon,” especially considering his discussion of “the truth in solipsism” and his insistence that “my” horizon is really “the” (preeminent) horizon (2007). Merlo’s “subjectivist view of the mental” is arguably even more closely related; he argues that “one’s own mental states are metaphysically privileged vis-à-vis the mental states of others” and discusses in detail the relationship of his view to Hare’s (2016). As another example, in a review of “The Character of Consciousness” (Chalmers 2010), Hellie (2013) argues that this work fails to do justice to the *embedded point of view* aspect of consciousness. He illustrates this with what he calls a “vertiginous question”: why, of all subjects, is *this* subject (the one corresponding to the human being Benj Hellie) the one whose experiences are “live”? In other work (2019), I explore whether the “liveness” of one particular perspective is a *further fact*—a

91 fact that does not follow logically from the physical facts of the world—by  
 92 considering the analogy to looking in on a simulated world through a virtual  
 93 reality headset: besides the computer code that determines the physics of  
 94 the simulated world, there must be additional code that determines which  
 95 simulated agent’s perspective to show on the headset.

96 In any case, Hare’s exposition of these ideas is clearest for the present  
 97 purpose, so I will focus on it. In an effort, possibly with limited success, to avoid  
 98 misrepresenting his position, as well as to clarify the relation to other work, let  
 99 me introduce my own terminology. Let us refer to the theory that states that  
 100 there is a metaphysically (rather than merely epistemically) distinguished<sup>1</sup> *I*  
 101 (or *Self*<sup>2</sup>) as the  $\alpha$ -theory. The intent is to emphasize the analogy with how the  
 102 A-theory (McTaggart 1908) states that there is a metaphysically distinguished  
 103 *Now*.<sup>3</sup> Similarly, I will refer to the theory that contradicts the existence of any  
 104 metaphysically distinguished *I* as the  $\beta$ -theory. Hare is thus defending the

- 
- 1 Throughout the paper, I will be deliberately noncommittal about the exact nature of such a metaphysical distinction. The reason is that the arguments presented here do not depend on what this distinction consists in. In the analogous case of a metaphysically distinguished *time* (rather than a metaphysically distinguished subject), by not committing to any particular interpretation, I can simultaneously address all varieties of A-theorists—presentists, moving-spotlight theorists, growing-block theorists, etc.—even though they disagree about the exact nature of the *Now*’s metaphysical distinction. Of course, there is disagreement even about how to define the individual varieties. Deasy (2017) discusses this at length, and proposes to define each of the main varieties as the conjunction of the A-theory (which he takes to mean “There is an absolute, objective present instant”) and a proposition about whether things begin and/or cease to exist. For example, for the growing-block theorist, that proposition is “Sometimes, something begins to exist and nothing ever ceases to exist.” While the distinctions between the various definitions are significant, again, my aim is to steer clear of this debate here and stick to arguments that work for any of these definitions. The same is true for the case of a metaphysically distinguished subject.
- 2 Again, what exactly the distinguished entity is—a human being, a brain, an experience—is not essential to my arguments, so I will remain deliberately noncommittal.
- 3 Is a commitment to a distinguished *Now* what defines the A-theory, or is it a commitment to tensed facts? (And in the latter case, should the  $\alpha$ -theory’s defining commitment instead be to first-personal facts?) To the extent that these commitments are not equivalent, in this paper, I will stick with the commitment to a distinguished *Now* (or *I*), as others have done—e.g., Cameron (2015, 89). For what it is worth, while a detailed analysis is outside the scope of this paper, I believe that they are in fact equivalent. I believe that a distinguished *Now* implies tensed facts, such as the fact that today is July 3, 2019. The other direction is perhaps more controversial, but I believe it holds as well: tensed facts such as the fact that today is July 3, 2019 distinguish a specific time, to which we may refer as the *Now*. A theory such as fragmentalism (Fine 2005a) might be used to dispute the second direction: if we consider *all* tensed facts, including those for other times, then no specific time is distinguished. But, of course, the set of all tensed facts taken together is full of contradictions, as it also contains, for example, the fact that today is not July 3, 2019. Avoiding such contradictions means restricting attention to a consistent fragment—but



105  $\alpha$ -theory. It is not entirely clear to me whether the specific version he defends  
 106 is intended to be analogous to presentism [or actualism—I will refrain from  
 107 discussing modality in this paper, but the parallels between time/subjectivity  
 108 and modality are well recognized; see Prior and Fine (1977); Bergmann (1999)],  
 109 or rather to something like a spotlight theory (or possibilism). In fact, his  
 110 writing suggests different answers to this question in different places, and I  
 111 will not attempt to resolve this small mystery here.

112 Others have commented on the idea of a metaphysically distinguished *I*—or,  
 113 similarly but not equivalently,<sup>4</sup> a metaphysically distinguished *Here*—in the  
 114 context of the philosophy of time. (While the differences between a metaphysically  
 115 distinguished *I* and a metaphysically distinguished *Here* will not matter  
 116 for some of the arguments presented in this paper, it is useful to note that, in  
 117 the context where a distinguished *I* is combined with a distinguished *Now*,  
 118 the combination of these two immediately implies a distinguished *Here* as  
 119 well—namely, the location of the distinguished individual at the distinguished  
 120 time.<sup>5</sup>) However, they have usually dismissed it rather quickly, in order to  
 121 move on with the case of a metaphysically distinguished *Now* (whether or  
 122 not they support the latter). For example, Zimmerman (2005, 422) writes:

123 An egocentric analogue of actualism ('personalism', to steal and  
 124 abuse a term) is very hard to imagine. Perhaps there is some kind  
 125 of not-merely-epistemological solipsism that would qualify. In  
 126 any case, only the maniacally egocentric could be this sort of  
 127 personalist.

128 Further back, Williams (1951, 458) writes:

129 Perhaps there exists an intellectualistic solipsist who grants the  
 130 propriety of conceiving a temporal stretch of events, to wit, his  
 131 own whole inner biography, while denying that the spatial scheme  
 132 is a literal truth about anything. Most of the disparagers of the  
 133 manifold, however, are of opposite bias. Often ready enough to

---

this in turn distinguishes a specific time. For further discussion of problems that fragmentalism faces, see Cameron (2015, 86–102).

4 For a discussion of the differences and their implications, in the related context of the Lewisian and Quinean accounts of centered worlds, see Liao (2012).

5 The combination similarly implies a distinguished observational frame of reference corresponding to the distinguished individual's state of motion. All of this does, of course, require the distinguished individual to be spatially located and to move through time and space, rather than, say, an immaterial soul or something existing for only an instant.

134 take literally the spatial extension of the world, they dispute the  
 135 codicil which rounds it out in the dimension of time.

136 Fine (2005a, 285) treats the case of first-personal realism in detail, but ad-  
 137 vocates for adopting a nonstandard variety of realism, either taking reality  
 138 to be relative to a standpoint, or (his preferred option) considering it to be  
 139 fragmented.<sup>6</sup> He notes:

140 It has seemed evident that, of all the possible worlds, the actual  
 141 world is privileged; it is the standpoint of reality, as it were, and  
 142 the facts that constitute reality are those that obtain in this world.  
 143 On the other hand, if we ask, in the first-personal case whether  
 144 we should be a nonstandard realist (given that we are going to  
 145 be first-personal realists in the first place), then the answer to  
 146 most philosophers has seemed to be a clear ‘yes’. It has seemed  
 147 metaphysically preposterous that, of all the people there are, I  
 148 am somehow privileged - that my standpoint is *the* standpoint of  
 149 reality and that no one else can properly be regarded as a source  
 150 of first-personal facts. The case of time is perplexing in a way that  
 151 these other cases are not.

152 I believe that there is value in exploring the  $\alpha$ -theory more thoroughly, rather  
 153 than dismissing it summarily for being repugnant in one way or another. The  
 154 words “egocentrism” and “solipsism” are both loaded with too much baggage.  
 155 While “egocentric,” taken literally, aptly describes the  $\alpha$ -theory, the common  
 156 interpretation of the word carries various negative connotations, and it is not  
 157 clear to me that these are fair to apply to every possible  $\alpha$ -theorist. Just as  
 158 A-theorists can take great interest in times other than their own (otherwise  
 159 why would they bother to write papers?), the  $\alpha$ -theorist can presumably take  
 160 great interest in people other than herself.<sup>7</sup> The relation to solipsism is also

---

6 Lipman (2015) discusses fragmentalism in more detail.

7 It should be noted here that, on the face of it, Hare (2007, 2009) does introduce his theory to justify placing greater weight on oneself than on others in making decisions. However, he also points out that the (distinguished) presence of an experience is only one factor in making decisions (“It is better that there be present suffering from a hangnail than absent suffering of leg-crushing.”). Perhaps more importantly, key examples that Hare uses in these works to support his theory are preferential in nature, such as an example where one knows that CJH (Hare) and Joe Bloggs have been in a train crash, CJH is about to have a painful operation, the subject knows he is one of these two but cannot remember which one, and so the subject hopes to not be CJH (2007). Such preferential examples are quite helpful to illustrate and motivate these types of theories—similar

161 not obvious. Hare intends for his theory to be only a weak and subtle version  
 162 of solipsism that does not deny the existence of others' consciousness (Hare  
 163 2009, 41–46), and others have granted him as much (e.g., Smith 2011; and  
 164 Mark Johnston in the introduction to Hare 2009).<sup>8</sup>

165 Indeed, a key point is that, just as there are multiple versions of the A-theory,  
 166 there are also multiple versions of the  $\alpha$ -theory, and these vary in the status  
 167 they accord to other individuals. Perhaps more importantly—and this is the  
 168 main focus of this paper—something is lost when attempting to study the A  
 169 vs. B question separately from the  $\alpha$  vs.  $\beta$  question; the two are very much  
 170 interrelated. To illustrate this, consider a theory that allows a distinguished *I*  
 171 that is not alive at the time of the distinguished *Now*, thereby treating the two  
 172 types of distinction as independent. Many of the arguments that I give in what  
 173 follows would do little to support such a theory. Hence, in what follows I will  
 174 not take the  $\alpha$ A-theory—the label that I will use for a view that combines the  $\alpha$ -  
 175 theory with the A-theory—to allow this possibility; what I have in mind is that  
 176 a single (*living-*)*person-stage* is distinguished. This interrelation is relevant to  
 177 the previous point. For example, the  $\alpha$ A-theorist may accord to other persons  
 178 the same metaphysical status as she does to herself in past and future time  
 179 slices.

180 After presenting, for the sake of illustration, some versions of the  $\alpha$ A-theory  
 181 in section 1, I will argue in section 2 that key arguments that have been given  
 182 to support the A-theory support the  $\alpha$ -theory just as well, and in fact sup-  
 183 port the combined  $\alpha$ A-theory especially strongly, placing the onus on the  
 184  $\beta$ A-theorist to explain why she accepts the A-theory but not the  $\alpha$ -theory. (It  
 185 would seem that most A-theorists, at least publicly, are  $\beta$ A-theorists in my  
 186 terminology.) Specifically, in subsection 2.1 I will discuss the argument from  
 187 presence *simpliciter*, and in subsection 2.2 the argument from the appropriate-  
 188 ness of sentiments such as those expressed by “Thank goodness that’s over!” I  
 189 will also argue, in section 3, that some serious challenges that the  $\beta$ A-theorist  
 190 faces are much less problematic for the  $\alpha$ A-theorist. Specifically, in subsection  
 191 3.1 I will discuss the argument from special relativity, in subsection 3.2 the  
 192 argument that the direction of time may be a local matter, in subsection 3.3

---

ones can be given to motivate the A-theory, as Hare does and others have done before him—even if one does not wish to normatively endorse the preferences used in the example. I will also discuss such examples later in this paper.

8 Others have tried to distinguish between more and less defensible versions of solipsism along similar lines; a particularly notable example is Valberg (2007). Similar ideas also appear in Johnston (2010).

193 the argument that asks for the rate at which time passes, and in subsection  
194 3.4 the argument from time travel and Gödelian universes.

195 Overall, my main objective is to argue that the  $\alpha$ A-theory is superior to the  
196  $\beta$ A-theory.<sup>9</sup> I would similarly argue that the  $\alpha$ A-theory is superior to the  $\alpha$ B-  
197 theory, but I do not expect many to defend the latter view.<sup>10</sup> This would leave  
198 the  $\alpha$ A-theory and the  $\beta$ B-theory as the remaining candidates. The reader  
199 might expect that my next step will be simply to argue that the  $\alpha$ -theory is  
200 so unappealing that we should accept the  $\beta$ B-theory, and hence, *a fortiori*,  
201 the B-theory. However, I believe that that conclusion is too hasty; an effective  
202 discussion of the relative merits of the  $\alpha$ A-theory and the  $\beta$ B-theory requires  
203 arguments of a different type than what I will present here. So, I will be  
204 content to let both theories stand for now.

## 205 **1 Some Versions of the $\alpha$ -Theory**

206 The A-theory counts among its supporters presentists, moving-spotlight the-  
207 orists, and growing-block theorists. Can we conceive of similar distinctions  
208 among  $\alpha$ -theorists? Rather than studying this in isolation from the A vs. B  
209 question, it seems more enlightening to ask what natural versions of the  
210  $\alpha$ A-theory there are. (Common versions of the A-theory and the B-theory  
211 can straightforwardly be reinterpreted as versions of the  $\beta$ A-theory and the  
212  $\beta$ B-theory.) I will present some versions in this section. My aim here is not to  
213 defend specific versions or to reach any definitive conclusion about which  
214 version is best. I also make no claim that this list is exhaustive, though I believe  
215 that it includes the versions that are most natural to discuss in the context of  
216 the existing literature on the A-theory. The aim of this exercise is merely to  
217 clarify some aspects of the  $\alpha$ A-theory and prevent overly narrow interpreta-

---

9 Of course, to accept this conclusion, it is not necessary to agree with every single argument presented here.

10 In fact, Hare (2009, 48) writes that “If you think that theories that dignify a slice of history do not survive sustained critical inspection, then you can still be a four-dimensionalist egocentric presentist. Indeed, I find that an attractive position.” This may appear to put him in the  $\alpha$ B-camp. However, on the whole in this section on the relationship to positions in the philosophy of time (Hare 2009, 46–50), he is clear that egocentric presentism does not commit one to a particular view on time, while also stating that the moving-spotlight theory is the most analogous one. Elsewhere (Hare 2010), he writes, “If you find yourself sympathetic to [the central tense realist idea] then I recommend that you consider *going the whole hog*, and becoming a perspectival realist” (emphasis mine), which might be interpreted to imply that perspectival realism is a stronger position than tense realism. In any case, as I hope will become clear from this paper, the  $\alpha$ A-theory does not at all require a dignified *slice* of history.

218 tions of it. Moreover, it will be helpful to refer to some of these versions in  
 219 what follows. I will also contrast these versions with some scenarios from the  
 220 literature.

221 **PERSONALIZED PRESENTISM.** This is the most natural way to adapt  
 222 presentism into an  $\alpha A$ -theory. In this version, there is a single dis-  
 223 tinguished individual whose experience at a single distinguished  
 224 point in time is, in some sense, “present.” (I hope that the intended  
 225 meaning of “presence” is at least somewhat clear at this point; I will  
 226 discuss it in more detail in subsection 2.1.) Beyond this present expe-  
 227 rience, nothing exists. Or, perhaps, some part of the outer world can  
 228 be granted some type of existence; but other experiences do not exist.  
 229 However, presumably, the present experience can change (more on  
 230 this below), just as presentists typically consider it possible for the  
 231 Now to change.

232 **PERSONALIZED MOVING SPOTLIGHT.** As in the classical moving-  
 233 spotlight theory, a spotlight moves over the four-dimensional block  
 234 universe, except now this spotlight shines on a single individual  
 235 (or that individual’s experience) at a single point in time. For the  
 236 personalized moving spotlight, it is less obvious how it moves (more  
 237 on this below).

238 **PERSONALIZED GROWING BLOCK.** In the classical growing block  
 239 theory, time slices are added to the block that contain all the events in  
 240 the universe at that point in time. In the personalized growing block  
 241 theory, only those parts of spacetime are added that are experienced  
 242 by a distinguished individual (and, perhaps, their past light cones).

243 Every one of these versions of the  $\alpha A$ -theory leaves several possibilities for  
 244 *how* the point of present experiences—the “I-Now”—could change or move  
 245 (if it changes or moves at all).<sup>11</sup> These include the following variants:

---

11 The word “I-Now” sounds more mystical than I would like, but we will need such a word. The word “spotlight,” when interpreted as shining on a single individual’s experience at a single point in time, would give the right idea, except it seems to commit the discussion to a view that all of the four-dimensional spacetime block exists, but not all of it is illuminated. While I do not want to dismiss such a view, in what follows we will not require this as an assumption. In contrast, the awkward word “I-Now” does not seem to rule out any of the possibilities. (Similarly, *Hellie 2013* uses “me-now.”)

246 SINGLE INDIVIDUAL OVERALL. The I-Now moves along with a  
 247 single individual throughout his or her lifetime. It is never associated  
 248 with any other individual.

249 CHANGING INDIVIDUAL ( $\alpha$ A-reincarnation). At the end of the  
 250 distinguished individual's lifetime, the I-Now jumps to another in-  
 251 dividual. We can consider various subvariants. For example: (1) the  
 252 I-Now cannot jump backwards in time; (2, a relativistic subvariant)  
 253 the I-Now can jump anywhere that is outside of all the past light  
 254 cones of points in spacetime that the I-Now occupied earlier; (3) the  
 255 I-Now can jump anywhere it has not previously been; (4) the I-Now  
 256 is not constrained in where it can jump.<sup>12</sup>

257 RAPIDLY CHANGING INDIVIDUAL. The I-Now can jump from one  
 258 individual to another even before the former's demise, and then  
 259 jump back to the previous individual as well. We can consider the  
 260 same subvariants as for  $\alpha$ A-reincarnation.

261 It is admittedly odd to propose all these different versions of the  $\alpha$ -theory  
 262 without making any serious attempt to justify them individually or to claim to  
 263 be exhaustive.<sup>13</sup> Again, my goal in doing so is merely to illustrate some of the  
 264 possibilities that the theory leaves open. The availability of multiple distinct  
 265 interpretations should not be surprising given the analogy and interrelation  
 266 with the A-theory. It is also clear that some of these versions are much more  
 267 solipsistic than others, or, at least, fit the negative connotations of solipsism  
 268 more than others.

269 Moreover, in earlier work on theories resembling the  $\alpha$ -theory, scenarios  
 270 are often sketched that fit much better with some of these versions than with  
 271 others. Usually, this is done without much discussion of why the author prefers

---

12 The last two subvariants seem more difficult to reconcile with the personalized growing block theory, and might also have negative implications for free will.

13 For example, perhaps it is not even necessary for the I-Now to change only in a sequential manner as in these variants; perhaps it can change along multiple dimensions, corresponding to changes across time and changes across space or individuals. Skow's (2009) relativistic moving-spotlight theory, in which individual points in spacetime are "lit up" from the perspective of points in *superspacetime*, seems very much in line with such a view. This also raises important questions about how these dimensions interact: Is temporal change objective or subjective? Is subjectivity eternal or temporary? For related questions on the interaction of time and modality, see Dorr and Goodman (2020).

272 such a version or even of what the alternatives might be. This has the effect of  
273 opening up the theory to criticisms that another version of the  $\alpha$ -theory might  
274 have avoided. Consider the following passage by Hare (2009, 51) (discussing  
275 a thought he had as a child), corresponding to a single-individual-overall  
276 theory:

277         Isn't it amazing and weird that for millions of years, generation  
278         after generation of sentient creatures came into being and died,  
279         came into being and died, and all the while there was this absence,  
280         and then one creature, CJH, unexceptional in all physical and  
281         psychological respects, came into being, and POW! Suddenly there  
282         were present things!

283 Later on, Hare (2009, 83) considers a type of reincarnation:

284         Is it necessary that only one person ever have present experiences?  
285         Again, the natural thing is to say no. Egocentric presentism gives  
286         me conceptual resources to imagine being one sentient creature,  
287         and then, later, being another sentient creature. So (recall Nagel's  
288         "fantasy of reincarnation without memory") I can imagine that, af-  
289         ter a lifetime of oblivious egg consumption, I die a happy philoso-  
290         pher, then find myself in a cage eighteen inches tall by twelve  
291         inches wide, my beak clipped to its base. This need not involve  
292         imagining that CJH dies a happy philosopher and then becomes  
293         a battery chicken. It may only involve imagining that after CJH's  
294         death there are again present experiences, and they are the expe-  
295         riences of a battery chicken. Once again this is a real, real nasty,  
296         metaphysical possibility.

297         So "the one with present experiences" is a definite description  
298         that may be satisfied by different things at different times. Like  
299         all such descriptions, it behaves as a *temporally nonrigid referring*  
300         *term*.

301 Similarly, Valberg (2013, 366) writes:

302         We can, however, give sense to the possibility that a human be-  
303         ing other than JV in the past was "me," or that a human being  
304         other JV [sic] might be "me" in the future. That is, it makes sense  
305         experientially (as a way things might be or develop from within  
306         my experience) that, in the past, a human being other than JV

307 occupied the position at the center of my horizon, or that a human  
 308 being other than JV will occupy this position in the future.

309 Again, the main point here is to make clear how many possibilities the  $\alpha$ -  
 310 theory leaves open and thereby to prevent overly specific interpretations. The  
 311 discussions in the remainder of the paper generally apply to all of the above  
 312 versions of the  $\alpha$ A-theory. A reader who wants to keep just a single version in  
 313 mind might focus on, for example, personalized presentism or a personalized  
 314 moving spotlight theory, with a single individual overall.

## 312 **Revisiting Arguments in Favor of the A-theory**

316 In this section, I will revisit some well-known arguments in favor of the A-  
 317 theory. Subsection 2.1 concerns the argument from presence *simpliciter* and  
 318 subsection 2.2 concerns the argument from the appropriateness of sentiments  
 319 such as those expressed by “Thank goodness that’s over!” In both cases, the  
 320 argument will be shown to support the  $\alpha$ A-theory more strongly than the  
 321  $\beta$ A-theory, because the argument supports a distinguished *I* just as it supports  
 322 a distinguished *Now*. Whether these arguments are indeed effective against  
 323 the B-theory is not the topic of this paper, so I will not review responses that  
 324 B-theorists may give to these arguments here.

### 2251 *Presence Simpliciter*

326 Arguably the most basic argument in favor of the A-theory is that of “the presence  
 327 of experience.” Many have made such an argument; a good exposition of  
 328 one is given by Balashov (2005). The argument is that my current experience  
 329 of writing this paper is *present* (or *occurs*<sup>14</sup>) in a way that my going through  
 330 security at the airport yesterday is not present. This is not to be taken as a  
 331 relative statement; everyone will agree that the writing experience at 5:50pm  
 332 on March 18, 2019 is present *at 5:50pm on March 18, 2019* in a way that the  
 333 airport security experience at 8:15am on March 17, 2019 is not present *at*  
 334 *5:50pm on March 18, 2019*. Rather, the writing experience seems present in an

---

14 Balashov (2005) uses “presence” and “occurrence” to refer to different concepts, but it seems to me that others have used “presence” to refer to a concept that is closer to Balashov’s “occurrence”. In any case, this latter concept is what I am after, and I hope that the use of “*simpliciter*” makes this clear.



335 *absolute* sense that does not require the boldface phrases, and this is referred  
 336 to as presence *simpliciter*.

337 I argue that, if we are to entertain such a notion, for it to be at all palatable,  
 338 it must be personalized, for the following reason. Just as my earlier airport  
 339 security experience is not present *simpliciter*, neither is David's experience of  
 340 eating breakfast in Australia present *simpliciter*, even if this event happens to  
 341 take place at the same time.<sup>15</sup> Let me first attempt to explain what I mean by  
 342 this, and then argue for it. In order to clarify what I mean, it is tempting to  
 343 write that David's breakfast experience is not present *simpliciter to me*. But to  
 344 do so would undermine the argument, in the exact same way that it would  
 345 undermine the purely temporal version of the argument to say that my airport  
 346 security experience is not present *simpliciter right now*. In the latter sentence,  
 347 "*simpliciter*" is clearly at odds with the indexical "right now." The exact same  
 348 is true about the juxtaposition of "*simpliciter*" and "to me." If an experience  
 349 takes place *simpliciter*, then to capture this we should not add any relativizing  
 350 indexical phrases.

351 Moreover, it seems that only an experience can be present *simpliciter* in this  
 352 way.<sup>16</sup> For example, it is not at all clear to me what it would mean for a chair  
 353 to *itself* be present *simpliciter*. My *experience* of a chair—visual, tactile, and  
 354 the result of significant cognitive processing—can be present *simpliciter*. Such  
 355 an experience is the kind of thing that can have the "liveness" that past and  
 356 future experiences do not, and that others' experiences do not. But I cannot  
 357 imagine what it would mean for the chair to *itself* be "live" in this way. If we  
 358 are willing to be a bit loose with our language, in most cases it will not cause  
 359 confusion to, as a shorthand, say that the chair is present *simpliciter* when we  
 360 really mean to refer to my experience of the chair. But if we are being strict,  
 361 the experience is not the chair itself. Moreover, it seems that an experience  
 362 can only be had by a single person<sup>17</sup> at a single time,<sup>18</sup> and it does not seem

---

15 There is, of course, the question of what "at the same time" even really means given that in special relativity, simultaneity depends on the frame of reference. I will discuss relativity later; for the purpose of the current argument, we may assume a Newtonian universe.

16 Merlo (2016, 326–327) makes a similar point.

17 I use "person" here, and throughout, in a broad sense; presumably animals and perhaps artificial intelligence can similarly have experiences. Also, in common parlance, of course two people can "share an experience," but I use "experience" here more narrowly in its phenomenological sense.

18 Along the same lines, Hare (2009, 49) describes the distinguished nature of his current experience and emphasizes that it is an easy-to-make "big mistake" to extend this to other current experiences. Hare (2010) presents an argument with strong similarities to the one presented here. Finally, at the end of his paper, Skow (2009) also discusses the vivid nature of present experiences and

363 that two distinct experiences, corresponding to different individuals and/or  
 364 times, can be co-present *simpliciter* in this way. So, if anything, the argument  
 365 would suggest the existence of a metaphysically distinguished (I, Now) pair.

366 Is this argument equivocating between “presence” in the temporal sense  
 367 and “presence” in the experiential sense? Indeed both meanings of the word  
 368 seem to play a role, and I believe that this is revealing rather than misleading.  
 369 Insofar as the current moment in time has a “liveness” that other moments do  
 370 not, it has it only through my own experience; the same moment elsewhere,  
 371 even if experienced by someone else, lacks this liveness just as a past moment  
 372 here, even if experienced by me, lacks it. In this way, the two meanings of the  
 373 word are inextricably linked. Hare (2009, 100) similarly argues that it is in  
 374 fact advantageous that the word “present” has multiple readings.

375 It is also important here not to be misled by how we use language. The sen-  
 376 tence “David is eating breakfast” is, in a sense, simpler than “I went through  
 377 airport security yesterday morning.” Both sentences explicitly refer to their  
 378 subject (“David” and “I”), but only the latter needs to explicitly refer to when  
 379 the event took place (“yesterday morning”) in order to place it in time. So the  
 380 first sentence has a type of simplicity that the second one lacks; we could add  
 381 “now” to the former, but it is not needed. On the other hand, dropping “I” from  
 382 the second sentence leaves it grammatically mangled. From this asymmetry  
 383 between “I” and “now” one might be tempted to conclude that the word “sim-  
 384 pliciter” more naturally corresponds to what is happening *now*—since the  
 385 word “now” is usually not needed for sentences concerning the present—than  
 386 it would correspond to what is happening to *me*—since a word such as “I” or  
 387 “me” is usually needed for a sentence concerning the first person.

388 However, I would argue that the significance of this asymmetry is not  
 389 metaphysical, but rather entirely linguistic. So many of our spoken sentences  
 390 concern the present that, pragmatically, it would be inefficient to require  
 391 adding a word like “now” to all these sentences. On the other hand, usually a  
 392 conversation concerns multiple actors, so it is important to make it clear who  
 393 is the subject in each sentence. To see that this is the driving force behind the  
 394 asymmetry, consider a different context: my planner. In my planner, I write  
 395 entries such as “attend faculty meeting at noon.” It would be an inefficient use  
 396 of my time to add “I” (or “I will”) to the beginning of the sentence, because I  
 397 would have to do so for almost every entry in my planner! In contrast, naturally,

---

argues that a local spotlight shining on a single individual explains this just as well as a global one (though he does not argue that it actually explains it *better*).

398 each of my planner entries *must* have a time associated with it; after all, if  
 399 the event were happening right now, I would not have to add an entry to my  
 400 planner. So, in the context of my planner, the roles that subject and time play  
 401 in the pragmatic issue at hand are reversed: the former is generally implicit  
 402 but the latter is not.<sup>19</sup> This appears to confirm that the asymmetry is due to  
 403 pragmatic reasons.

## 2.2 The Appropriateness of Wanting Things to (not) be Past

405 Another well-known argument (Prior 1959; Zimmerman 2008) in favor of  
 406 the A-theory (and presentism in particular) concerns the appropriateness of  
 407 statements such as “Thank goodness that’s over!” Here, “that” might refer to  
 408 something like a headache the speaker was experiencing. It is often argued  
 409 that the B-theory does not provide the resources to capture the full significance  
 410 of this statement. Prior argues that the meaning of such a statement is not that  
 411 it is good that the headache takes place at a point in spacetime earlier than  
 412 the point at which the statement is uttered; in his words, “Why should anyone  
 413 thank goodness for that?” Instead, what the statement is getting at is that the  
 414 headache is simply *over*, and the A-theory provides the resources to capture  
 415 this. But one might similarly argue in favor of the  $\alpha$ -theory, for example  
 416 appealing to the appropriateness of statements such as “Thank goodness that  
 417 is not happening to me!” This is closely related to the question of whether  
 418 self-bias could be metaphysically justified, as studied by Hare (2007, 2009).  
 419 The  $\beta$ A-theorist is likely to complain that the analogy is not apt, because the  
 420 second statement merely reflects a selfish disposition rather than something  
 421 more fundamental. It is not clear to me why the same could not be said of the  
 422 first statement, that the statement merely reflects the speaker’s callousness  
 423 towards her past self. To avoid this criticism, perhaps one can make the first  
 424 statement about someone else (“Thank goodness John’s headache is over!”),  
 425 but, and I believe this is telling, the argument seems to lose force with this  
 426 move.

427 Let us explore this in a bit more depth. Suppose all headaches last exactly  
 428 one or two days with no ill effects afterwards, and consider the following two  
 429 statements:

---

19 For additional discussion of the linguistic asymmetry between time and space, and how this asymmetry is driven by pragmatic concerns in communication, see Butterfield (1984).

430  $S_1$ : Thank goodness John's headache, which started yesterday, ended yes-  
 431 terday as well, rather than continuing into today.

432  $S_2$ : Thank goodness John's headache, which started the day before yester-  
 433 day, ended the day before yesterday as well, rather than continuing into  
 434 yesterday.

435 Here, we imagine caring a great deal about John and preferring him not to  
 436 suffer. Under the  $\beta A$ -theory, one would expect  $S_1$ : to have a significance not  
 437 shared by  $S_2$ :, as the former concerns a difference in what is happening *now*,  
 438 whereas the latter concerns a difference that is in any case entirely in the past.  
 439 It is not clear to me that such a difference in significance is really there. Is it  
 440 not just as reasonable to appreciate that John did not suffer yesterday, as it is  
 441 to appreciate that he is not suffering today?

442 Yet, one may have an intuition that indeed,  $S_1$ : has a significance that  $S_2$ :  
 443 does not. I believe that the likely grounds for this intuition are not germane to  
 444 the issues under discussion here, and we can modify the scenario to remove  
 445 these grounds. First, in the first situation, if John were still having a headache, I  
 446 might feel compelled to try to *do something* to alleviate his suffering. However,  
 447 this is easily addressed by postulating that it is common knowledge that I can  
 448 do nothing of the sort. Second, if John is in my immediate environment and  
 449 I see him suffering, this may cause me to suffer as well, for example due to  
 450 the mirror neurons in my brain. But this is merely returning us to an example  
 451 where I myself suffer, which is precisely what we were trying to avoid by  
 452 introducing John. Hence, we should postulate that John is somewhere else  
 453 entirely.

454 To make all this concrete, suppose that John has decided to go on a two-  
 455 month retreat in a faraway country. He will not communicate until he gets  
 456 back. Halfway into his retreat, I realize that around this time of year, he  
 457 always gets a headache, which may last one or two days. I care for him and  
 458 so I hope that it is just a one-day headache this time. But I will not find out  
 459 until he comes back and tells me. Imagining this scenario, I do not find myself  
 460 concerned specifically about whether his headache happens to be taking place  
 461 right now, or not.<sup>20</sup>

---

20 In this example, there is nothing to synchronize John's experience with mine; his life is unfolding in parallel to mine and it is hard to see why it would matter which events are contemporaneous. As we will discuss in subsection 3.1, we can make the example even more extreme by having John fly far off into space somewhere, so that, as far as the theory of relativity is concerned, there really is no absolute answer to the question whether his headache is taking place at the same

462 Hence, given that the scenario is set up appropriately, I remain unconvinced  
 463 that there is any significant difference between  $S_1$ ; and  $S_2$ ; and this seems to  
 464 deal a blow to the  $\beta$ A-theory. Naturally, the  $\beta$ B-theory avoids this blow; but I  
 465 believe the  $\alpha$ A-theory also avoids it, in that John today is just as much “outside  
 466 the I-Now” as John yesterday, because I am not John. In fact, compared to the  
 467  $\beta$ B-theory, the  $\alpha$ A-theory does a better job explaining why something about  
 468 the example seems to change when I myself am brought into it. That is, if we  
 469 replace “John’s” with “my” in the statements above to obtain  $S'_1$  and  $S'_2$ , then it  
 470 does seem that  $S'_1$  has a significance that  $S'_2$  does not.  $S'_2$  is not an unreasonable  
 471 statement—it makes sense to appreciate having suffered less than one might  
 472 have, just as it makes sense to appreciate someone else suffering less than  
 473 he might have—but only  $S'_1$  concerns the immediate presence or absence  
 474 of suffering, which is the vivid characteristic that imbues “Thank goodness  
 475 that’s over!” examples with their intended significance.<sup>21</sup>

476 Indeed, both Suhler and Callender (2012) and Green and Sullivan (2015)  
 477 report on an experimental study by Caruso et al. (2008) in which subjects  
 478 were asked what would be fair compensation for a particular task. The study  
 479 found that when subjects were asked to imagine themselves doing the task  
 480 in the future, they felt that they should be compensated significantly more  
 481 than when they imagined themselves doing the task in the past; but this  
 482 effect disappeared when they were asked to imagine someone *else* doing it.  
 483 Suhler and Callender (2012) take this to invalidate the “Thank goodness that’s  
 484 over” argument, and Greene and Sullivan (2015) argue for complete temporal  
 485 neutrality in making decisions. (The argument for temporal neutrality is  
 486 worked out in detail in Sullivan (2018). Hurka (1993, 61) argues that temporal  
 487 neutrality is appropriate for certain non-hedonic goods, but is convinced  
 488 that it is not for avoiding pain, by the example from Parfit (1984, 165) that

---

time as my current experience. If so, caring about simultaneity seems to require a very strong commitment to the  $\beta$ A-theory, as it requires that there be an additional fact about simultaneity, over and above the theory of relativity, that is important for what we should care about, even though no physical measurement could ever tell us whether two events actually were or were not simultaneous in this sense.

21 Some of this is reminiscent of Turri’s (2013) “That’s outrageous!” example. Turri argues that just as the appropriateness of statements such as “Thank goodness that’s over!” can be used to support presentism, the appropriateness of statements such as “That’s outrageous!” can be used to attack it, because it seems perfectly legitimate to be outraged by, say, a past genocide. I consider it telling that “Thank goodness that’s over!” examples typically involve oneself and “That’s outrageous!” examples typically involve others; this may well be what is driving the difference in conclusions from these examples.

we would prefer a more painful operation in the past to a less painful one in the future.) The analysis above suggests that while indeed, the results of the Caruso et al. (2008) study cast doubt on whether the “Thank goodness that’s over” argument effectively supports the  $\beta$ A-theory, they are perfectly consistent with this argument supporting the  $\alpha$ A-theory.

### 3 Revisiting Arguments Against the A-theory

In this section, I will revisit some well-known arguments against the A-theory. Subsection 3.1 concerns the argument from special relativity, subsection 3.2 concerns the argument that the direction of time may be a local matter, subsection 3.3 concerns the argument that asks for the rate at which time passes, and subsection 3.4 concerns the argument from time travel and Gödelian universes. In all cases, the  $\alpha$ A-theory will be shown to avoid most of the bite that these arguments inflict on the  $\beta$ A-theory, roughly because the arguments hinge on the fact that the Now is global in nature—that is, it stretches across all of space. Because the I-Now is local in nature, the arguments are ineffective against the  $\alpha$ A-theory.

#### 3.1 *Special Relativity*

Einstein’s theory of relativity has often been invoked to criticize the A-theory. Unlike in a Newtonian universe, in the special theory of relativity, simultaneity is not absolute; rather, whether two events are simultaneous depends on the reference frame. But if there is no absolute simultaneity, then how can there be an absolute Now? Special relativity can also be used to cast doubt on specific arguments in favor of the A-theory—or at least, the  $\beta$ A-theory. For example, let us modify the example from subsection 2.2 by putting John on a faraway planet, so that whether his headache is earlier or later than our own time depends on the reference frame. This seems to make it difficult to hold the position that, in order to know how we should feel about John’s headache, it is important to know whether it is in the past or in the future. Now, perhaps there may still be a separate, absolute sense in which John’s headache is in the past, even if this is not implied by the theory of relativity. But if there is not, this poses a problem for using the “Thank goodness that’s over!” argument in support of the  $\beta$ A-theory—but, importantly, not for using it in support of the  $\alpha$ A-theory, because, as discussed in subsection 2.2, in that case the argument is only made about one’s own pains rather than those of someone

523 on a faraway planet. Still, we must investigate the implications of relativity  
524 for the  $\alpha$ A-theory more broadly.

525 Some (e.g., Markosian 2004) have argued that, in fact, a philosophically  
526 austere version of the theory of relativity could explain the empirical evi-  
527 dence without implying that there is no absolute simultaneity. The relation  
528 of absolute simultaneity could be added on top of the theory of relativity.  
529 For example, one might suppose that there exists a distinguished frame of  
530 reference that determines which events are absolutely simultaneous. Positing  
531 such a distinguished frame seems a rather awkward and inelegant addition to  
532 the theory, one that is rather contrary to the spirit of the theory of relativity  
533 and perhaps more in line with older theories of a stationary aether. But, Zim-  
534 merman (2008) has argued that such an addition to the physical theory is no  
535 different in kind from the addition of a distinguished Now in the first place.  
536 That may be so, but it is a further addition, and it seems that, for the sake of  
537 parsimony, each addition should at least count against the resulting theory.  
538 The analogy is also imperfect. It can at least be argued that we know when  
539 the Now is; in contrast, it is not clear whether and how we could ever know  
540 what the distinguished frame of reference is. Zimmerman (2011) discusses  
541 and responds to all these concerns in far more detail than I can do here, and  
542 argues well that they are not fatal to the  $\beta$ A-theory, but it is clear that at least  
543 they pose significant challenges.

544 In any case, the above arguments only concern the  $\beta$ A-theory. In the  $\alpha$ A-  
545 theory, there is no need for any observer-independent simultaneity at all.  
546 While the Now in the  $\beta$ A-theory must be global—in the sense that every-  
547 where in the universe, there are events happening Now, thereby introducing  
548 an observer-independent simultaneity relation across all of space—the I-Now  
549 in the  $\alpha$ A-theory is local. The precise nature of this locality—for example,  
550 whether the I-Now is spatially extended—does not matter much for the argu-  
551 ments at hand; what matters is that the I-Now is associated with an observer,  
552 and that that observer can be localized in spacetime. Specifically, this ties the  
553 I-Now to the frame of reference associated with that observer;<sup>22</sup> if so desired,  
554 simultaneity could be determined based on this frame of reference according  
555 to the theory of relativity. For that matter, no notion of simultaneity across  
556 space is even required for the theory to make sense. While the  $\beta$ A-theory  
557 necessitates such a notion—whatever is happening Now across space must

---

22 The definition of what constitutes a frame of reference varies. Here, we consider a frame of reference to be determined purely by its state of motion, rather than to also include a coordinate system.

558 be simultaneous, in an objective sense—it does not seem to pose any problem  
 559 for the  $\alpha A$ -theorist to hold that there is no absolute notion of simultaneity. As  
 560 far as the  $\alpha A$ -theorist is concerned, we can define a notion of simultaneity for  
 561 convenience, for example the one based on the theory of relativity and the  
 562 distinguished frame of reference corresponding to the I-Now as just suggested,  
 563 but none is truly needed. In fact, the problems that the theory of relativity  
 564 poses for the A-theory have already led to at least one proposal similar to the  
 565  $\alpha A$ -theory, namely Skow's (2009) relativistic spotlight theory,<sup>23</sup> in which the  
 566 spotlight shines locally, not globally.<sup>24</sup>

### 3<sup>62</sup>2 *The Direction of Time*

568 For any version of the  $\beta A$ -theory in which time flows, there needs to be an  
 569 objective *direction* in which time flows. Presumably, it flows from what we  
 570 perceive as the past to what we perceive as the future. But if the laws of physics  
 571 are invariant to time reversal, then these laws do not naturally provide such a  
 572 direction. It is commonly held that what we perceive as the direction of time  
 573 is tied to the entropy gradient, and that this entropy gradient may well be  
 574 reversed in other parts of spacetime. If so, we may imagine a Doppelgänger  
 575 being that is otherwise very much like ourselves, living its life in such a part,  
 576 backwards in time from our perspective (Williams 1951; Maudlin 2002). The  
 577 Doppelgänger would presumably think that *we* have it backwards, that the  
 578 direction of time's flow is opposite from what we think it is. So what gives  
 579 us reason to believe that we are the ones to have it right? A key issue here is  
 580 that presumably, the  $\beta A$ -theory requires time to flow in the same direction  
 581 everywhere; the direction should be *globally* consistent.<sup>25</sup> It has been argued  
 582 that we have no reason to believe that the Doppelgänger even has mental

23 In earlier work, Stein (1968, 18) hints at a similar theory when he contemplates what would result from an argument by Putnam (1967) if one tried to preserve a different intuition about the relationship between what is present and what is real. It is not clear whether he intends at all to defend such a theory.

24 Hare (2010) and Hare (2009, 48) also make some of the points that I made in this subsection. Fine (2005a, 2006) similarly gives a detailed discussion of what, for the realist, should replace the role of times when we take into account special relativity, and concludes that most plausibly frame-time pairs should take their role, in combination with a nonstandard type of realism in which either realities are indexed to different frame-times or reality is fragmented.

25 The Now is not localized under the  $\beta A$ -theory, so that there is a single Now across space; but if it moves in one direction in one location and in the opposite direction elsewhere, it is hard to imagine that after moving in these opposite directions it remains the *same* Now across these locations.



583 states at all, by virtue of the fact that the way its life proceeds is so unlike the  
584 way ours proceeds (Maudlin 2002). But this seems a rather odd conclusion,  
585 since we have supposed that, *mutatis mutandis* for the difference in direction,  
586 the Doppelgänger's life is entirely like ours. For a more detailed discussion  
587 of this point and these issues more generally, see Price (2011) and references  
588 cited therein.

589 In contrast, the putative existence of persons living in parts of spacetime  
590 with a reversed entropy gradient, living their lives backwards in time (from our  
591 perspective), poses no problem for the  $\alpha$ A-theory. This is because the I-Now is  
592 inherently *local* (in both a spatial and a temporal sense), so it does not matter if  
593 the entropy gradient is reversed elsewhere; all that matters is what the entropy  
594 gradient is *here* (and *now*), because that is what determines the direction in  
595 which the I-Now moves. If the I-Now actually tracks a Doppelgänger at some  
596 point, it does not appear to pose any problem for the theory for it to then move  
597 in the opposite direction. (This may pose problems for some of the specific  
598 illustrative versions presented earlier in section 1, but it poses no problem  
599 for the other versions.) We can view *external* time as nothing more than a  
600 dimension through which the I-Now travels.

601 Taking this to an extreme, we may even imagine a machine that transports  
602 you to another region of space where the entropy gradient is reversed relative  
603 to ours, and that transforms you into a Doppelgänger there. You will, in some  
604 sense, continue your life there uninterrupted, except moving in the opposite  
605 temporal direction. Of course being transported to another region of space is  
606 likely to be a bit shocking; but, if such scenarios are possible at all, there seems  
607 to be no reason to believe that your experiences will be any different than they  
608 would have been if instead you had been transported to a region of space that  
609 happens to have the same entropy gradient (and not been transformed into a  
610 Doppelgänger). Accommodating this intuition is easy under the  $\alpha$ A-theory;  
611 for example, the I-Now could simply jump along with you and then start  
612 moving backwards (from our initial perspective). On the other hand, this  
613 example appears problematic for versions of the  $\beta$ A-theory that require a  
614 globally defined direction of time, because such a theory would lead to the  
615 conclusion that one of the two halves of your life is lived, in an *absolute* sense,  
616 backward. If we believe Maudlin's (2002) argument, we would then conclude  
617 that you had real mental states in only half of your life. This seems to be an  
618 odd conclusion. If near the end of your life you were transported back to the  
619 original spacetime region, the suggestion that you had not had any real mental

620 states since the original transportation event would seem utterly bizarre to  
 621 you!

### 3.3 *The Rate of Time's Passage*

623 Opponents of the A-theory (or  $\beta$ A-theory) have also criticized it as follows: if  
 624 the Now moves, what is the rate at which it moves? It has been argued that if  
 625 one says that it moves at 1 second per second, this poses a problem for the  
 626 theory, because one can cancel the units of seconds and conclude that the rate  
 627 is simply 1, and (supposedly) 1 is not a rate (e.g., [Olson 2009](#)). Now, the idea  
 628 that a unitless rate is not a rate is simply nonsense. This has been convincingly  
 629 argued elsewhere: Skow ([2011](#)) uses the example of sociologists tracking what  
 630 the “most common birth year” in the population is. One would expect the  
 631 most common birth year to generally increase by roughly 1 year every year,  
 632 though the rate may be higher or lower than 1 depending on demographic  
 633 phenomena. In any case, the rate is unitless (one might just as well say the  
 634 rate is approximately 1 decade per decade). The example is convincing to me,  
 635 and clearly many other examples of sensible unitless rates can be provided.  
 636 One such example is particularly relevant here: due to relativity, satellites and  
 637 astronauts on the International Space Station age at a slightly different rate  
 638 than objects and people on the surface of the Earth. The amount of time that  
 639 such a satellite or astronaut experiences per unit of Earth surface time is a  
 640 unitless rate.<sup>26</sup> This example actually seems to pose a more serious problem for  
 641 the answer that time moves at “1 second per second”—if the idea is to think  
 642 of time as moving globally rather than just locally, then in just *whose* seconds  
 643 are we measuring this rate? In any case, a weaker version of the original  
 644 criticism seems to hold up: the question only allows uninformative answers.  
 645 The answer that it moves at “1 second per second” seems tautological. We

---

26 One might counter that these conditions in fact correspond to different units, namely Earth surface seconds and ISS seconds, so that we in fact do not obtain a unitless rate. But this misses the point that a second denotes the same amount of aging for the people in each condition. The unitless rate indicates how much faster people in one condition age than those in the other, and for this comparison no units are needed. Similarly, we need no units to say that one person is 1.2 times as tall as another. That the rate being unitless is meaningful is further illustrated by the fact that it can be both above and below 1, because of the opposing effects of relative velocity time dilation and gravitational time dilation; there is an orbit, about half the radius of the Earth above the surface, at which the rate is 1 ([Ashby 2002](#)). The rate being 1 at this orbit is not just a meaningless consequence of how we defined the units; it is the orbit at which astronauts age equally fast as those on the surface.

646 could instead introduce the concept of *supertime* to track the Now’s motion  
647 through time, so that at different points in supertime, the Now is at a different  
648 time. (For a detailed discussion of the metaphor of supertime, see Skow 2012.)  
649 Then, we can ask how many seconds pass per supersecond. However, there  
650 seems to be every reason to simply define the supersecond so that the answer  
651 becomes “1 second per supersecond,” which remains uninformative.

652 In the  $\alpha A$ -theory—or, at least, in versions of it where the I-Now moves  
653 along with a person through time (see section 1)—the question of how fast  
654 the I-Now moves does not pose such problems. First, the fact that on a space  
655 station, a different amount of time is experienced to pass no longer poses any  
656 problem, because the I-Now is local, so there is no requirement that time  
657 passes at the same rate everywhere. Moreover, the question of how fast the  
658 I-Now moves can have more interesting answers. In the relativistic example  
659 above, it is natural to respond that the I-Now moves at a different rate when it  
660 is associated with an astronaut in orbit than it does when it is associated with  
661 a person on the surface. Alternatively, let us put relativity aside for a moment  
662 and focus on the I-Now’s experiential aspect instead. One might reasonably  
663 hold that the I-Now moves through external (i.e., clock) time at a different  
664 rate when it is associated with a person who is under anesthesia than it does  
665 when it is associated with someone who is highly alert.

666 If we allow ourselves to speculate, a computational<sup>27</sup> theory might be used  
667 to unify these two examples: consider a person’s “clock speed”—the number  
668 of mental operations, according to some suitable definition, per (Earth sur-  
669 face) second—and take this to determine the rate at which the I-Now moves.  
670 Specifically, let us define a supersecond so that there is always exactly one  
671 mental operation per supersecond. Then, the number of (Earth surface) sec-  
672 onds per supersecond—which is just the reciprocal of the clock speed defined  
673 above—will vary in the different scenarios above, in a way that conforms with  
674 our intuitions. Focusing on Earth surface seconds per supersecond (regardless  
675 of the location of the person) simultaneously addresses both the relativistic  
676 and the experiential components of the scenarios, and also allows us to han-  
677 dle mixed cases, such as a space station inhabitant who is under anesthesia.  
678 In such a case, the number of mental operations per Earth surface second  
679 can be written as the number of mental operations per space station second,  
680 multiplied by the number of space station seconds per Earth surface second,

27 It is important to hold a sufficiently broad view of “computation” here; such broad views are common among those working on the theory of computation. Alternatively, and less ambitiously, the reader may just view this as a suggestive analogy to the clock speed of a computer.

681 thereby separating out the experiential and relativistic components, respec-  
682 tively. This shows that these two components are compatible. Per the theory  
683 of relativity, there is nothing special about Earth surface seconds, as opposed  
684 to space station seconds or Mars surface seconds; they are just different ways  
685 to measure external time.

686 Supertime, so defined, perhaps more naturally corresponds to our sense  
687 of passage, leaving regular time (as tracked by clocks) in the more modest  
688 role of a dimension through which we happen to pass, as noted earlier. That  
689 is, this notion of supertime would allow us to give metaphysical meaning to  
690 the idea of time passing more or less quickly from a subjective viewpoint. Of  
691 course, this view may conflict with other intuitions that we have developed.  
692 In our ordinary experience of time, relativistic issues do not come into play,  
693 and our waking experience of how fast time passes is usually fairly stable.  
694 Given this, we tend to conceive of time as objective, and treat any variance  
695 in how we perceive its passage as a mere error in estimation. For the current  
696 purpose, I believe such intuitions are misleading. The following two examples  
697 are intended to illustrate that it is in fact quite natural to assign primary  
698 importance to the notion of supertime as defined here. In each of them, we  
699 will imagine a choice between two alternatives that result in you having  
700 different amounts of time but equal amounts of supertime left in your life. I  
701 argue that you should be (close to) indifferent between the options in both  
702 scenarios.

703 *Example 1.* It is the year 2400, and you are part of a group of people on a  
704 lifelong space voyage. The group is about to split up into two subgroups that  
705 will take separate spacecraft. It is common knowledge that the two subgroups  
706 will never communicate again, either with each other or with the people left  
707 on Earth. You get to choose in which subgroup you will be. They are indis-  
708 tinguishable, except the two spacecraft will move to orbits around different  
709 massive bodies, with different relativistic time dilations. If you choose to be  
710 on spacecraft 1, your life will therefore be shorter in Earth time than it would  
711 be on spacecraft 2. As a result, your first reaction may be that you would prefer  
712 to be on spacecraft 2. But, I argue, upon closer inspection there is little reason  
713 for this. This is because, to make up for the shorter amount of Earth time in  
714 your life on spacecraft 1, correspondingly more events will happen per unit of  
715 Earth time on spacecraft 1. You would experience entirely similar lives on the  
716 two spacecraft, with equally many interesting events taking place on both. If it  
717 were possible to communicate from Earth to the spacecraft, you might prefer  
718 being on spacecraft 2 because (for example) more papers, books, and movies

719 would be produced on Earth and sent to spacecraft 2 for your consumption  
720 during your life. But we have assumed that such communication is impossible.  
721 As far as I can see, there does not seem to be any compelling reason to have a  
722 preference about on which spacecraft you continue your voyage.

723 *Example 2.* It is again the year 2400, but this time we will stay on the surface  
724 of the Earth. After a long and happy life, you have regrettably contracted  
725 an incurable disease that, if left untreated, will kill you almost immediately.  
726 Unfortunately, the only possible treatments will put you in a type of comatose  
727 state until your death. You will, however, have wonderful dreams in this  
728 state. Due to secrecy issues, your friends and family will never be made aware  
729 of your predicament. There is no chance at all that any new treatment will  
730 become available during the remainder of your life. You have a choice between  
731 medications  $M_1$  and  $M_2$ . Compared to  $M_1$ ,  $M_2$  would keep you alive for twice  
732 as long, but would allow your brain to process at only half the rate. Your  
733 first reaction may be that you would prefer to receive  $M_2$ . But again, I argue,  
734 upon closer inspection there is little reason for this. Because of the difference  
735 in brain processing rates, you would have equally many wonderful dreams  
736 under the two medications. If your friends and family could visit you in your  
737 comatose state, you might prefer for them to have that option for a longer  
738 or shorter period of time, but we have ruled this out. If you had hopes that  
739 scientists could develop a cure, you would prefer  $M_2$  to give the scientists more  
740 time, but we have also ruled this out. As far as I can see, there does not seem  
741 to be any compelling reason to have a preference about which medication  
742 you receive.

743 In summary, to the extent that the question about the rate at which the  
744 Now moves poses a problem for the  $\beta A$ -theory, it does not pose this problem  
745 for the  $\alpha A$ -theory, since for the latter the answer to the question need not be  
746 tautological.

### 344 *Time Travel and Gödelian Universes*

748 A final criticism of the  $\beta A$ -theory is that it does not make much sense of  
749 time travel scenarios. Following Lewis (1976), it seems natural to distinguish  
750 between *external* time and the time traveler's *personal* time. But if one takes  
751 external time seriously in the metaphysical sense, as would be expected of  
752 a  $\beta A$ -theorist, it would appear one cannot simultaneously do the same for  
753 personal time. This, in turn, necessitates unintuitive attitudes towards time  
754 travel. The following passage by Sider (2005, 333) illustrates this perfectly:

755 But if personal time bears little similarity to external time then  
 756 “personal time” is merely an invented quantity, and is mislead-  
 757 ingly named at that. That I will view a dinosaur in my personal  
 758 future amounts merely to the fact that I once viewed a dinosaur,  
 759 and moreover that this is caused by my entry into a time machine.  
 760 Since this fact bears little resemblance to the facts that constitute  
 761 a normal person’s genuine future, I could not enter the time ma-  
 762 chine with anticipation and excitement at the thought of seeing a  
 763 dinosaur, for it is not true that I am *about* to see a dinosaur, nor is  
 764 the truth much *like* being about to see a dinosaur. If anything, I  
 765 should feel fear at the thought of being annihilated by a device  
 766 misleadingly called a “time machine”. The device causes it to be  
 767 the case that I once viewed a dinosaur, but does not make it the  
 768 case in any real sense that I *will* view dinosaurs.

769 Perhaps there is a way out of this conclusion for the  $\beta A$ -theorist, but I cannot  
 770 see it. Or perhaps she is willing to bite the bullet and accept the conclusion  
 771 that (at least backward) time travel is to be avoided at all cost. In any case, the  
 772  $\alpha A$ -theorist avoids this issue. For her, personal time is what is taken seriously,  
 773 and she can legitimately look forward to—if this is in fact something to look  
 774 forward to—her encounter with a dinosaur.<sup>28</sup>

775 Closely related to the issue of time travel is that of Gödelian universes that  
 776 cannot be given a global temporal ordering. The theoretical possibility of such  
 777 universes perhaps poses a problem for some versions of the  $\beta A$ -theory. The  $\alpha A$ -  
 778 theory, however, does not require any global temporal ordering. For versions  
 779 of the  $\alpha A$ -theory with a moving I-Now, one may yet worry if such universes do  
 780 not create different problems. For example, Dieks (2006) discusses an example  
 781 by Reichenbach (1958, 141–142) in which a person loops around to meet his  
 782 earlier self again at a particular point in spacetime. Dieks, who argues for a  
 783 B-theoretic notion of local becoming, argues that this example illustrates that  
 784 even a local type of spotlight is problematic. He argues that when the spotlight  
 785 shines on the region in spacetime where the younger and older versions of the

28 Well, she may still hesitate, to the extent that it is not obvious that the presence of experience, the I-Now, will follow her through the time machine rather than go somewhere else. As an example that illustrates this ambiguity, it may be one of these unmarketable time machines that also leave behind a badly burned body, apparently alive for a few more seconds, where the traveler entered the time machine. (See Hare (2009, 58) for a similar example.) But at least her believing that it will follow her back in time (rather than transitioning to a different person at the same time, or staying with a burnt body) would not cause any inconsistency with her other beliefs.

786 person meet, there must in fact be two distinct spotlights, one that will travel  
787 with the younger version and one that will travel with the older version. Then,  
788 the spotlight associated with the younger version loops around as that version  
789 becomes the older version, eventually reaching the same region again. By  
790 the same reasoning as before, we will again need two spotlights at this point.  
791 But the other spotlight, the one that was initially associated with the older  
792 version, is not available for the task, being meanwhile associated with an even  
793 older version. So we will need a third spotlight, and so on ad infinitum, which  
794 seems problematic.

795 But it is easy to find an escape from Dieks' argument. The fact that the  
796 two versions of the person are (roughly) at the same point in spacetime does  
797 not imply that the spotlight shines on them simultaneously *in the supertime*  
798 *sense*. That is, the "same" spotlight might earlier (in supertime) light up the  
799 younger version only (i.e., that version's experience at that point) and later  
800 (in supertime) the older version only. Hence, there is no need to introduce  
801 additional spotlights when the meeting point is reached. This illustrates one  
802 advantage of associating the spotlight with person-stages (I-Now) rather than  
803 with small regions of spacetime (Here-Now): even though the younger and  
804 the older version are both in (roughly) the same location at the same time,  
805 they correspond to different person-stages. This requires, of course, that in  
806 this type of scenario we associate the I-Now with a person-stage (where a  
807 younger and an older version of the same person at the same time are still  
808 considered separate person-stages), rather than with a pair of a person and a  
809 time, which in this case might pick out both person-stages. This interpretation  
810 of the I-Now in any case aligns better with the other arguments presented in  
811 this paper. For example, it seems hard to imagine the (simultaneous) presence  
812 *simpliciter* of the *combination* of both person-stages. Also, the older person-  
813 stage may think, looking at the younger person-stage, "Thank goodness I am  
814 no longer that immature!" The idea that the spotlight was previously (in the  
815 supertime sense) associated with the younger person-stage and now with the  
816 older person-stage seems to capture the significance of this statement well.  
817 Finally (and more speculatively), if we imagine the brain of the older stage to  
818 have slowed down and no longer to be processing at the rate of his younger  
819 self, associating the I-Now with person-stages would allow us to say that the  
820 I-Now moves at a different rate with respect to external time when associated  
821 with each of these two person-stages.

## 824 Conclusion

823 Upon inspection, key criticisms of the A-theory are only effective as criticisms  
 824 of the  $\beta$ A-theory, and key arguments in favor of the A-theory are much more  
 825 convincing as arguments for the  $\alpha$ A-theory. To the extent I have succeeded in  
 826 showing that A-theorists are rationally compelled to be  $\alpha$ -theorists as well,  
 827 surely many will interpret this as a significant blow to the A-theory because  
 828 they consider the  $\alpha$ -theory implausible. Nevertheless, some philosophers may  
 829 well be willing to adopt some version of the  $\alpha$ A-theory (Hare being an obvious  
 830 example). As I emphasized earlier, a detailed discussion of the relative merits  
 831 of the  $\alpha$ A-theory and the  $\beta$ B-theory is outside the scope of this paper. Such a  
 832 discussion is sure to revisit many familiar arguments in the philosophy of time  
 833 and modality (and mind), and is unlikely to reach a swift conclusion.<sup>29</sup> I do  
 834 hope to have convinced the reader that the  $\alpha$ A-theory will fare better in such  
 835 a comparison than the  $\beta$ A-theory. The former has an internal consistency that  
 836 allows it to escape some of the more damaging criticisms to which the latter  
 837 has fallen prey.\*

Vincent Conitzer

 0000-0003-1899-7884

Duke University

conitzer@cs.duke.edu

## 842 References

- 843 ASHBY, Neil. 2002. "Relativity and the Global Positioning System." *Physics Today* 55(5):  
 844 41–47, doi:[10.1142/9789812700988\\_0010](https://doi.org/10.1142/9789812700988_0010).
- 845 BALASHOV, Yuri. 2005. "Times of Our Lives: Negotiating the Presence of Experience."  
 846 *American Philosophical Quarterly* 42(4): 295–309.
- 847 BERGMANN, Michael. 1999. "(Serious) Actualism and (Serious) Presentism." *Noûs*  
 848 33(1): 118–132, doi:[10.1111/0029-4624.00145](https://doi.org/10.1111/0029-4624.00145).
- 849 BUTTERFIELD, Jeremy. 1984. "Seeing the Present." *Mind* 93(370): 161–176, doi:[10.1093/mind/xciii.370.161](https://doi.org/10.1093/mind/xciii.370.161).
- 850 CAMERON, Ross P. 2015. *The Moving Spotlight: An Essay on Time and Ontology*. Oxford:  
 851 Oxford University Press, doi:[10.1093/acprof:oso/9780198713296.001.0001](https://doi.org/10.1093/acprof:oso/9780198713296.001.0001).

29 This seems all the less likely given that the problem connects to other challenging problems, such as the Sleeping Beauty problem—see e.g., Conitzer (2015).

\* I am thankful to anonymous referees who provided especially thorough and helpful comments, which significantly improved the paper.



- 853 CARUSO, Eugene M., GILBERT, Daniel T. and WILSON, Timothy D. 2008. "A Wrinkle  
854 in Time: Asymmetric Valuation of Past and Future Events." *Psychological Science*  
855 19(8): 796–801, doi:10.1111/j.1467-9280.2008.02159.x.
- 856 CHALMERS, David J. 2010. *The Character of Consciousness*. Oxford: Oxford University  
857 Press, doi:10.1093/acprof:oso/9780195311105.001.0001.
- 858 CONITZER, Vincent. 2015. "Can Rational Choice Guide Us to Correct *De Se* Beliefs?"  
859 *Synthese* 192(12): 4107–4119, doi:10.1007/s11229-015-0737-x.
- 860 —. 2019. "A Puzzle About Further Facts." *Erkenntnis* 84(3): 727–739, doi:10.1007/s106  
861 70-018-9979-6.
- 862 DEASY, Daniel. 2017. "What is Presentism?" *Noûs* 51(2): 378–397, doi:10.1111/nous.1  
863 2109.
- 864 DIEKS, Dennis. 2006. "Becoming, Relativity and Locality." in *The Ontology of Spacetime*,  
865 edited by Dennis DIEKS, pp. 157–176. Philosophy and the Foundations of Physics  
866 Series n. 1. Amsterdam: Elsevier Science Publishers B.V., doi:10.1016/s1871-  
867 1774(06)01008-4.
- 868 DORR, Cian and GOODMAN, Jeremy. 2020. "Diamonds Are Forever." *Noûs* 54(3):  
869 632–665, doi:10.1111/nous.12271.
- 870 FINE, Kit. 2005a. "Tense and Reality." in *Modality and Tense: Philosophical Papers*, pp.  
871 261–320. Oxford: Oxford University Press, doi:10.1093/0199278709.003.0009.
- 872 —. 2005b. "The Reality of Tense." in *Modality and Tense: Philosophical Papers*, pp.  
873 305–313. Oxford: Oxford University Press, doi:10.1093/0199278709.001.0001.
- 874 —. 2006. "The Reality of Tense." *Synthese* 150(3): 399–414. Journal publication of [an  
875 abridged version of] Fine (2005b), doi:10.1007/s11229-005-5515-8.
- 876 GREENE, Preston and SULLIVAN, Meghan. 2015. "Against Time Bias." *Ethics* 125(4):  
877 399–414, doi:10.1086/680910.
- 878 HARE, Caspar. 2007. "Self-Bias, Time-Bias, and the Metaphysics of Self and Time."  
879 *The Journal of Philosophy* 104(7): 350–373, doi:10.5840/jphil2007104717.
- 880 —. 2009. *On Myself, and Other, less Important Subjects*. Princeton, New Jersey: Prince-  
881 ton University Press, doi:10.1515/9781400830909.
- 882 —. 2010. "Realism about Tense and Perspective." *Philosophy Compass* 5(9): 760–769,  
883 doi:10.1111/j.1747-9991.2010.00325.x.
- 884 HELLIE, Benj. 2013. "Against Egalitarianism [on Chalmers (2010)]." *Analysis* 73(2):  
885 304–320, doi:10.1093/analys/ans101.
- 886 HURKA, Thomas. 1993. *Perfectionism*. New York: Oxford University Press, doi:10.109  
887 3/0195101162.001.0001.
- 888 JOHNSTON, Mark. 2010. *Surviving Death*. Princeton, New Jersey: Princeton University  
889 Press, doi:10.1515/9781400834600.
- 890 LEWIS, David. 1976. "The Paradoxes of Time Travel." *American Philosophical Quarterly*  
891 13(2): 145–152. Reprinted in Lewis (1986, 67–80).
- 892 —. 1986. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.109  
893 3/0195036468.001.0001.

- 894 LIAO, Shen-yi. 2012. "What are Centered Worlds?" *The Philosophical Quarterly* 62(247):  
895 294–316, doi:10.1111/j.1467-9213.2011.00042.x.
- 896 LIPMAN, Martin A. 2015. "On Fine's Fragmentalism." *Philosophical Studies* 172(12):  
897 3119–3133, doi:10.1007/s11098-015-0460-y.
- 898 MARKOSIAN, Ned. 2004. "A Defense of Presentism." in *Oxford Studies in Metaphysics*,  
899 volume I, edited by Dean W. ZIMMERMAN, pp. 47–82. Oxford: Oxford University  
900 Press, doi:10.1093/oso/9780199267729.003.0003.
- 901 MAUDLIN, Tim. 2002. "Remarks on the Passing of Time." *Proceedings of the Aristotelian*  
902 *Society* 102: 259–274, doi:10.1111/j.0066-7372.2003.00053.x.
- 903 MCTAGGART, J. McT. Ellis. 1908. "The Unreality of Time." *Mind* 18(68): 457–484,  
904 doi:10.1093/mind/xvii.4.457.
- 905 MERLO, Giovanni. 2016. "Subjectivism and the Mental." *Dialectica* 70(3): 311–342,  
906 doi:10.1111/1746-8361.12153.
- 907 OLSON, Eric T. 2009. "The Rate of Time's Passage." *Analysis* 69(1): 3–9, doi:10.1093/an  
908 alys/ann001.
- 909 PARFIT, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press, doi:10.1  
910 093/019824908X.001.0001.
- 911 PRICE, Huw. 2011. "The Flow of Time." in *The Oxford Handbook of Philosophy of Time*,  
912 edited by Craig CALLENDER, pp. 276–311. Oxford Handbooks. Oxford: Oxford  
913 University Press, doi:10.1093/oxfordhb/9780199298204.003.0010.
- 914 PRIOR, Arthur Norman. 1959. "Thank Goodness That's Over." *Philosophy* 34(128):  
915 12–17. Reprinted in Prior (1976, 78–84), doi:10.1017/s0031819100029685.
- 916 —. 1976. *Papers in Logic and Ethics*. London: Gerald Duckworth & Co. Edited by Peter  
917 Geach and Anthony Kenny.
- 918 PRIOR, Arthur Norman and FINE, Kit. 1977. *Worlds, Times and Selves*. London: Gerald  
919 Duckworth & Co. Edited by Kit Fine; based on manuscripts by Prior with a preface  
920 and a postscript by Kit Fine.
- 921 PUTNAM, Hilary. 1967. "Time and Physical Geometry." *The Journal of Philosophy* 64(8):  
922 240–247. Reprinted in Putnam (1975, 198–205), doi:10.2307/2024493.
- 923 —. 1975. *Mathematics, Matter and Method. Philosophical Papers, Volume 1*. Cambridge:  
924 Cambridge University Press. Second edition: Putnam (1979).
- 925 —. 1979. *Mathematics, Matter and Method. Philosophical Papers, Volume 1*. 2nd ed.  
926 Cambridge: Cambridge University Press. First edition: Putnam (1975), doi:10.101  
927 7/cbo9780511625268.
- 928 REICHENBACH, Hans. 1928. *Philosophie der Raum-Zeit-Lehre*. Berlin: de Gruyter. En-  
929 glish translation: Reichenbach (1958), doi:10.1515/9783111485676.
- 930 —. 1958. *The Philosophy of Space and Time*. Mineola, New York: Dover Publications.  
931 English translation of Reichenbach (1928) by Maria Reichenbach and John Freund.
- 932 SIDER, Theodore. 2005. "Traveling in A- and B-Time." *The Monist* 88(3): 329–335,  
933 doi:10.5840/monist200588326.

- 934 SKOW, Bradford. 2009. "Relativity and the Moving Spotlight." *The Journal of Philosophy*  
935 106(12): 666–678, doi:10.5840/jphil20091061224.
- 936 —. 2011. "On the Meaning of the Question 'How Fast Does Time Pass?'" *Philosophical*  
937 *Studies* 155(3): 325–344, doi:10.1007/s11098-010-9575-3.
- 938 —. 2012. "Why does Time Pass?" *Noûs* 46(2): 223–242, doi:10.1111/j.1468-0068.2010.  
939 00784.x.
- 940 SMITH, Nicholas J. J. 2011. "Inconsistency in the A-Theory." *Philosophical Studies*  
941 156(2): 231–247, doi:10.1007/s11098-010-9591-3.
- 942 STEIN, Howard. 1968. "On Einstein-Minkowski Space-Time." *The Journal of Philosophy*  
943 65(1): 5–23, doi:10.2307/2024512.
- 944 SUHLER, Christopher and CALLENDER, Craig. 2012. "Thank Goodness that Argument  
945 is Over: Explaining the Time Value Asymmetry." *Philosophers' Imprint* 12(15),  
946 <http://hdl.handle.net/2027/spo.3521354.0012.015>.
- 947 SULLIVAN, Meghan. 2018. *Time Biases: A Theory of Rational Planning and Personal*  
948 *Persistence*. Oxford: Oxford University Press, doi:10.1093/oso/9780198812845.001.  
949 0001.
- 950 TURRI, John. 2013. "That's Outrageous." *Theoria* 79(2): 167–171, doi:10.1111/theo.120  
951 05.
- 952 VALBERG, Jerome J. 2007. *Dream, Death, and the Self*. Princeton, New Jersey: Princeton  
953 University Press, doi:10.1515/9780691190181.
- 954 —. 2013. "The Temporal Present." *Philosophy* 88(345): 369–386, doi:10.1017/s0031819  
955 113000016.
- 956 WILLIAMS, Donald Cary. 1951. "The Myth of Passage." *The Journal of Philosophy*  
957 48(15): 457–472. Reprinted in Williams (2018, 159–172), doi:10.2307/2021694.
- 958 —. 2018. *The Elements and Patterns of Being: Essays in Metaphysics*. Oxford: Oxford  
959 University Press. Edited by A.R.J. Fisher, doi:10.1093/oso/9780198810384.001.00  
960 01.
- 961 ZIMMERMAN, Dean W. 2005. "The A-Theory of Time, the B-Theory of Time, and  
962 'Taking Tense Seriously.'" *Dialectica* 59(4): 401–457, doi:10.1111/j.1746-8361.2005.  
963 01041.x.
- 964 —. 2008. "The Privileged Present: Defending an 'A-Theory' of Time." in *Contemporary*  
965 *Debates in Metaphysics*, edited by Theodore SIDER, John HAWTHORNE, and Dean  
966 W. ZIMMERMAN, pp. 211–225. Contemporary Debates in Philosophy n. 10. Malden,  
967 Massachusetts: Wiley-Blackwell.
- 968 —. 2011. "Presentism and the Space-Time Manifold." in *The Oxford Handbook of*  
969 *Philosophy of Time*, edited by Craig CALLENDER, pp. 163–245. Oxford Handbooks.  
970 Oxford: Oxford University Press, doi:10.1093/oxfordhb/9780199298204.003.0008.

PROOF

# Determinism, “Ought” Implies “Can” and Moral Obligation

NADINE ELZEIN

Haji argues that determinism threatens deontic morality, not via a threat to moral responsibility, but directly, because of the principle that “ought” implies “can”. Haji’s argument requires not only that we embrace an “ought” implies “can” principle, but also that we adopt the principle that “ought” implies “able not to”. I argue that we have little reason to adopt the latter principle, and examine whether deontic morality might be destroyed on the basis of the more commonly embraced “ought” implies “can” principle alone. I argue that despite what look like initially compelling reasons why we might suppose that this weaker conclusion is similarly destructive to deontic morality, we actually have good reason to doubt that it has any practical relevance for moral deliberation at all.

While most of the literature on morality and determinism focuses on threats to moral responsibility, determinism might be thought to threaten morality on separate grounds. Haji draws on the popular principle that “ought” implies “can”, in order to show that determinism undermines deontic morality (1998, 1999, 2002, 2019). Similar arguments are presented by Lockie (2018), although Lockie, unlike Haji, does not intend to defend scepticism about obligation, but rather to show that any such scepticism is inherently self-defeating.

By “deontic morality”, Haji has in mind any moral use of the terms “ought” and “ought not”, as well as moral judgements of right and wrong. While he concedes that judgements of moral “good” and “bad” may still make sense within a deterministic framework, he argues that the action-demanding normative terms associated with obligations and prohibitions would be seriously undermined. Determinism precludes moral duty.

However, as Haji himself makes explicit, in order to reach this conclusion, we need not only an “ought” implies “can” principle, but also an “ought” implies “able not to” principle (2002, 28). A similar principle is found in Lockie’s work (2018, 181). I will argue, firstly, that even if we accept the

1001 popular “ought” implies “can” principle, there are good reasons to reject any  
 1002 “ought” implies “able not to” principle. Secondly, without the “ought” implies  
 1003 “able not to” principle, such arguments are limited to establishing a much  
 1004 weaker conclusion; we cannot conclude that there are no moral duties at all,  
 1005 only that there are no *unfulfilled* moral duties. Thirdly, while this weaker  
 1006 conclusion may look similarly problematic at first sight, from a practical  
 1007 perspective it actually makes very little difference to morality.

### 1008 **1 Determinism, Ability, and “ought” Implies “can”**

1009 The principle that “ought” implies “can” has certainly seemed compelling  
 1010 to many,<sup>1</sup> although it’s not uncontroversial.<sup>2</sup> Haji originally calls his “ought”  
 1011 implies “can” principle “K”, and then later “Kant’s Law/Obligation”. But for  
 1012 present purposes, let us simply call this sort of principle “OIC” (so as to match  
 1013 the broader class of principles under discussion). Haji (2002, 14) formulates  
 1014 his version of OIC roughly as follows:

1015 OIC. As of time  $t$ , an agent  $S$ , ought morally to do something  $A$  at  
 1016 time  $t^*$  (where  $t^*$  may either be  $t$  or a time later than  $t$ ) only if  $S$  can,  
 1017 as of  $t$ , do  $A$  at  $t^*$ ; and, as of  $t$ ,  $S$  ought not to do  $A$  at  $t^*$  only if  $S$  can,  
 1018 as of  $t$ , not do  $A$  at  $t^*$ .

1019 According to this principle, an agent only ought to do something if she actually  
 1020 can do it, and ought only to refrain from doing something if she actually can  
 1021 refrain from doing it.

- 
- 1 The principle is commonly thought to originate with Kant, and was famously defended by Moore (1922). Since then it is more often taken to be a basic platitude than explicitly argued for, but there are some explicit defences of the principle: see Sapontzis (1991), Griffin (1992), Streumer (2003, 2007, 2010), and Vranas (2007). For defences of related principles, see Graham (2011) and Kühler (2013).
- 2 For some critiques, see Lemmon (1962), Williams (1965), Brouwer (1969), Trigg (1971), van Fraassen (1973), Heintz (1975), Brown (1977), Sinnott-Armstrong (1984, 1988), Rescher (1987, chaps. 2, 26–54), Saka (2000), and Fischer (2003). Cf. Kekes (1984) and Stern (2004).

## 1221 The Analysis of “can”

1023 Given that there are broad variations in the way that we might interpret “can”,<sup>3</sup>  
 1024 there are also variations in the way that we might interpret OIC. Haji’s (2002,  
 1025 23) most moderate definition is as follows:

1026       MODERATE OIC. Agent *S* ought to do something *A*, only if *S* has  
 1027       the opportunity to do *A*, is physically and psychologically able to do  
 1028       *A*, and *A*’s accomplishment is not “strictly out of *S*’s control”.

1029 While this is taken to be the bare minimum required for ability, Haji adds  
 1030 that it may also require being motivationally able, and having the right sort of  
 1031 “know-how” (2002, 16–24).

1032       Physical and psychological possibility are fairly straightforward notions.  
 1033       Plausibly an agent is only “able” to perform actions that are consistent with  
 1034       their psychological characteristics and their physical abilities. The inclusion  
 1035       of the stipulation that the agent must be “psychologically able” may, however,  
 1036       seem controversial. It means that an agent with a strong aversion, say, may  
 1037       count as unable to do something, even if she could succeed in doing it should  
 1038       she choose to. One reason we might nonetheless endorse this reading, as  
 1039       Haji points out, is that it is natural to suppose that an agent with a serious  
 1040       enough phobia might be excused for her failure to do something that her  
 1041       phobia prevents her from doing. For instance, we would not typically consider  
 1042       an agent “able” to save a drowning child if a severe phobia rendered her  
 1043       incapable of entering the water (Haji 2002, 22).

1044       Moreover, endorsing a relatively strong sense of “can” may prove indis-  
 1045       pensable to the argument as a whole. That is because the argument aims to  
 1046       establish that the ability to do otherwise is ruled out by determinism, where  
 1047       this involves the very *same* sense of ability for which it will be true that “ought”  
 1048       implies “can”. Any weakening of the sense of “can” utilised in the OIC prin-  
 1049       ciple may risk introducing a corresponding weakening of the argument for  
 1050       supposing that determinism rules out the ability to do otherwise in precisely

3 Among other points of contention, there is a long-standing dispute about whether “can” ought to be analysed conditionally (Moore 1903; Ayer 1946; Smart 1961; Schlick 1939; Lewis 1981; Berofsky 2002), non-conditionally (Campbell 1951; Chisholm 1964; Lehrer 1968; van Inwagen 1983, 2000, 2004, 2008; Kane 1999; Clarke 2009; Grzankowski 2014), or dispositionally (Smith 1997, 2003; Vihvelin 2004, 2011, 2013; Fara 2008). Even within these camps there is significant scope for disagreement. For more general discussions, see also Kratzer (1977), Mele (2003), Maier (2015), and Weir (2016).

1051 that sense. For example, Haji notes that if we supposed a merely conditional  
 1052 analysis of “can” would do, according to which the ability to do otherwise  
 1053 simply requires that the agent could do otherwise if she chose to, then this  
 1054 would make it dubious to suppose that determinism rules this ability out  
 1055 (2002, 67–68).

1056 In fact, Haji argues that even if such conditional abilities are present, deter-  
 1057 minism robs us of the opportunity to do otherwise. If any factors, internal or  
 1058 external, prevent an agent from *exercising* some skill they have, then this will  
 1059 constitute a barrier to their having the opportunity to exercise it (2002, 22).<sup>4</sup>

1060 Finally, the “control” requirement is supposed, at the very least, to rule out  
 1061 having the “ability” to do things that happen purely by fluke (Haji 2002, 22). In  
 1062 analysing such control, Haji cites Vihvelin, who states: “We make judgments  
 1063 about ability on the basis of evidence of a reliable causal correlation between  
 1064 someone’s attempts to do a certain kind of act and the success of her attempts.”  
 1065 (2000, 142). This sort of control neither entails nor is entailed by possession  
 1066 of the other senses of “ability”. Plausibly, an agent’s phobia may make her  
 1067 psychologically and motivationally unable to purchase a pet snake, but doing  
 1068 so may not be “strictly out of her control”; were she to try, she could reliably  
 1069 succeed. Similarly, if a golf novice hits a hole in one on her first attempt, this  
 1070 certainly shows that she is physically able to hit a hole in one, but if it is an  
 1071 unrepeatable fluke, then it will still be “strictly out of her control”.

## 1122 *Determinism and Obligations*

1073 Haji and Lockie use rather complex arguments to reach the conclusion that de-  
 1074 terminism rules out all obligations. Moreover, Lockie’s argument incorporates  
 1075 the additional goal of showing that any argument in favour of determinism  
 1076 would be self-defeating, and Haji’s argument incorporates his attempt to show  
 1077 that if nothing is obligatory, then nothing is right or wrong either. I am not  
 1078 going to address the latter part of Lockie’s argument,<sup>5</sup> and I am not going to  
 1079 consider whether Haji is right to suppose that wrongness and rightness de-

4 I am doubtful about the idea that the very same sense of “can” that’s at issue in OIC is also the sense in which the ability to do otherwise might plausibly be ruled out by determinism. We have already noted that if we invoke weaker definitions of “able to” in our OIC principle, it will be difficult to establish that the relevant abilities are threatened by determinism. But for the purposes of this discussion, I will simply grant this point. See Haji (2002, 60–65) for his own arguments to this effect.

5 I have examined Lockie’s transcendental argument in more detail elsewhere (Elzein and Pernu 2019).



1080 pend on obligation. While this claim has been contested,<sup>6</sup> I am happy to grant  
 1081 it. Moreover, in what follows, it is the status of actions as obligatory (rather  
 1082 than right or wrong) that will be the prime focus. So for present purposes,  
 1083 we can work with a simplified version of the argument, which might go as  
 1084 follows:

- 1085 1. If determinism is true, no agent is ever able to act otherwise than they  
 1086 do act. (basic premise)
- 1087 2. If no agent is ever able act otherwise than they do act, then no agent ever  
 1088 has an obligation to act otherwise than they do act. (premise derivable  
 1089 from OIC)
- 1090 3. If determinism is true, no agent ever has an obligation to act otherwise  
 1091 than they do act. (from 1 and 2, via hypothetical syllogism)

1092 While 3 is an interesting conclusion, it is weaker than the the one that is ulti-  
 1093 mately defended by either Haji or Lockie. It does not entail that if determinism  
 1094 is true, there are no obligations, merely that that there are no *unfulfilled* obli-  
 1095 gations. It leaves open that agents sometimes both have and fulfil moral duties.  
 1096 In order to reach the stronger conclusion, that there are no obligations at all,  
 1097 Haji introduces a parallel principle, which he calls “CK” (2002, 28). Lockie  
 1098 (2018, 182) puts forward a similar principle. Elsewhere, Haji gives the same  
 1099 sort of principle different titles, such as “Kant’s Law/Impermissible” (Haji  
 1100 2019, 8) or “Obligation/Alternate” (Haji and Herbert 2018a, 186). Let us sim-  
 1101 ply call this whole class of principles “OIAN T principles” (for “ought” implies  
 1102 “able not to”). Haji (2002, 28) defines the relevant sort of principle, omitting  
 1103 the temporal indices, as follows:

1104 OIAN T. If one ought to do *A*, then one can refrain from doing *A*  
 1105 (and if one ought not to do *A*, then one can do *A*).

1106 If we grant OIAN T, we can also establish that there are no obligations to  
 1107 do what we actually do, given our inability to do otherwise. A simplified  
 1108 argument of this form runs as follows:

- 1109 1. If determinism is true, no agent is ever able to act otherwise than they  
 1110 do act. (basic premise)

---

6 See Pereboom (2001, 141–147) for an objection, and Haji (2002, 51–52) for his defence.

- 1111 2. If no agent is ever able act otherwise than they do act, then no agent  
 1112 ever has an obligation not to act otherwise than they do act. (premise  
 1113 derivable from OIANT)
- 1114 3. If determinism is true, no agent ever has an obligation not to act other-  
 1115 wise than they do act. (from 1 and 2, via hypothetical syllogism)
- 1116 4. If determinism is true, no agent has an obligation to act as they actually  
 1117 do act. (from 3, an equivalence through double negation)

1118 The final step from 3 to 4 is valid provided we grant that “not acting otherwise”  
 1119 entails “acting as one actually does”. For present purposes, “acting as one  
 1120 actually does” should be understood broadly, so as to be fulfilled if the agent  
 1121 does not act otherwise; hence it should include the agent’s inaction, if the  
 1122 agent in question is not actually doing anything. Granted this broad reading, it  
 1123 should be uncontroversial that “not acting otherwise” directly entails “acting”  
 1124 as one actually does. It should be similarly obvious, granted this broad reading,  
 1125 that premise 2 is entailed by OIANT.

1126 The first argument shows that, given determinism, no agent has an obli-  
 1127 gation to act otherwise than they do act. The second argument shows that,  
 1128 given determinism, no agent has an obligation to act as they actually do either.  
 1129 Between the two arguments, this rules out all moral obligations.

1130 While the first argument appears compelling, the second argument seems  
 1131 considerably weaker. The principle upon which it rests, OIANT, seems more  
 1132 dubious than the principle invoked by the first argument, OIC. If we reject  
 1133 the argument from OIANT to the conclusion that if determinism were true,  
 1134 no one would be obligated to do what they actually do, then we are left with a  
 1135 weaker conclusion: that if determinism were true, no one would be morally  
 1136 obligated to act otherwise than they do act.

## 1132 2 How Plausible is OIANT?

1138 Haji offers various lines of argument in favour of accepting OIANT: the first  
 1139 is a simple appeal to symmetry between OIC and OIANT. Lockie’s work  
 1140 also draws on the intuition that there ought to be symmetry between such  
 1141 principles. However, even if we doubt that there is any obvious *inherent* reason  
 1142 to suppose that the two principles are symmetrical, we might argue that we  
 1143 ought to accept such symmetry on the basis that both principles are taken to be  
 1144 motivated primarily by a two-way freedom requirement (this seems to be the  
 1145 supposed basis of the symmetry for Haji). Haji also offers a “theory-fuelled”

1146 argument, which appeals to a particular analysis of obligation. I will argue  
 1147 that OIANT is, at least on the face of it, inherently implausible before going  
 1148 on to deal with each of these arguments in turn.

## 2.4.1 *The Prima Facie Implausibility of OIANT*

1150 It has already been noted that psychological ability is crucially included in the  
 1151 definition of “able to” invoked in Haji’s OIC and OIANT principles. In light of  
 1152 this, however, “ought” implies “able not to” has some undesirable implications.  
 1153 Many actions that seem obviously morally prohibited are also psychological  
 1154 impossibilities for most psychiatrically well-adjusted individuals. For instance,  
 1155 my psychology is such that I could not take a chainsaw and use it to saw off  
 1156 the arms of a small child. To be clear, I don’t mean a child that has gangrene,  
 1157 say, and needs those limbs removed urgently on pain of death, but a perfectly  
 1158 healthy child; one whose limbs I have no reason to remove. In fact, I could  
 1159 not do such a thing even if I were offered reasons, if they were of the wrong  
 1160 sort: e.g., I could not saw off the arms of a child for a monetary incentive  
 1161 (even if I were offered a very reasonable market rate). Does this entail that it  
 1162 is not morally obligatory for me to refrain from sawing off the arms of small  
 1163 children?

1164 This conclusion seems counterintuitive. It is the fact that such an action  
 1165 would be morally reprehensible which may well, in this case, explain *both*  
 1166 my irresistible aversion to it *and* my reasons for supposing that it is morally  
 1167 obligatory that one refrains from such behaviour.

1168 Unlike Haji, I think it is plausible to suppose that my inability to do such  
 1169 a thing entails that I cannot be held responsible for not doing it, and hence  
 1170 deserve no praise.<sup>7</sup> The moral expectation that I refrain from dismembering  
 1171 small children is a very easy standard for me to meet. It seems close to the  
 1172 bare minimum you might reasonably expect of me, so I hardly deserve a  
 1173 medal. But it seems one thing to say that I don’t deserve praise, and quite  
 1174 another to say that sawing off the arms of small children would not be morally  
 1175 impermissible. We are usually quite happy to talk about being psychologically

---

7 Haji is persuaded on the basis of Frankfurt’s argument (1969) that, despite the threat to deontic morality, determinism poses no threat to moral responsibility (1998, 2002). See also Haji and McKenna (2004, 2006). Obviously, however, given the threat to deontic morality, determinism entails that there would be no right or wrong actions to actually blame or praise agents for. In contrast, I remain sceptical about whether Frankfurt-style examples really do establish that the ability to do otherwise is irrelevant to moral responsibility (Elzein 2013, 2017).

1176 compelled to do things that we also have a duty to do. We might even suppose  
 1177 that it is the very *fact* that something is perceived as morally prohibited that  
 1178 (at least sometimes) explains an agent's psychological aversion to doing it.

1179 The principle that "ought" entails "able not to" surely seems dubious. We  
 1180 ought to accept it only if we are offered very compelling arguments.

## 2.2 *The Defence from Apparent Symmetry*

1182 The first argument appeals to the apparent symmetry between "ought" im-  
 1183 plies "can" principles and "ought not" implies "can" principles (along with,  
 1184 presumably, the latter's complement stipulation, that "ought" implies "able  
 1185 not to"). Haji argues "that it is difficult to see why control requirements of  
 1186 deontic obligatoriness would differ, in this respect from control requirements  
 1187 of deontic wrongness" (2002, 29). He interprets OIC as postulating an alter-  
 1188 native possibilities condition as a control requirement for obligatory actions,  
 1189 and supposes that similar considerations would count in favour of accepting  
 1190 an alternative possibilities condition on prohibited ones.

1191 Even "ought" implies "can" is controversial, but it has a strong history of  
 1192 philosophical support behind it and it seems highly intuitive. "ought" implies  
 1193 "able not to", in contrast, has nothing like the same standing. As Nelkin notes,  
 1194 the principle is not usually seen as axiomatic, and the alleged symmetry that  
 1195 Haji sees between these sorts of principle is hardly obvious (2011, 102).

1196 In fact, I think there is a plausible basis for "ought" implies "can" that  
 1197 simply has no parallel in the case of "ought" implies "able not to". The ap-  
 1198 peal of "ought" implies "can" principles may in fact *not* rest on any control  
 1199 requirement that involves alternative possibilities. More plausibly, their ap-  
 1200 peal may be grounded in the simple idea that it is unreasonable to demand  
 1201 the impossible. We may well suppose that it is unreasonable to demand the  
 1202 impossible *without* supposing that this rests on a control requirement that  
 1203 involves alternative possibilities.

1204 Any demand that is impossible to meet will, by an obvious logical entail-  
 1205 ment, also be a demand with respect to which the agent lacks two-way control.  
 1206 But there is no entailment in the other direction. There is certainly no logical  
 1207 entailment from the plausible idea that it is unreasonable to demand the

1208 impossible to the far less plausible claim that it is unreasonable to demand  
1209 the unavoidable.<sup>8</sup>

1210 If there are cases in which we are plausibly required to do something that  
1211 we also cannot refrain from doing, then we have good reason to suppose that it  
1212 is the unreasonableness of the demand to do the impossible that is doing all of  
1213 the work in rendering principles like OIC plausible, and that two-way control  
1214 is irrelevant. Of course, we have already examined such a case: the case of  
1215 morally abhorrent actions that an agent is also psychologically incapable of.

1216 Moreover, think about cases in which it is uncertain whether or not one  
1217 is physically capable of committing some wrong. For example, I think that  
1218 it would be morally impermissible for me to leave the house with a kitchen  
1219 knife and stab to death the first person I see. However, I have absolutely no  
1220 idea whether I could physically *succeed* in such an endeavour, even supposing  
1221 I tried my best. It seems absurd to suppose that I should first have to be in  
1222 a position to know whether I could succeed in order to work out whether  
1223 stabbing an innocent bystander is morally impermissible (appeal to some  
1224 theory of normative ethics ought to settle *that* question quite irrespective of  
1225 my abilities).

1226 There is also a clear a disparity here with respect to duty and prohibition.  
1227 Plausibly, I can only be morally required to save the drowning child if I am  
1228 capable of it. If it is uncertain whether I will be physically able to, then we  
1229 might plausibly say that I have a duty to try, even if I could not have a duty to  
1230 succeed in my attempt. In contrast, it barely seems coherent to assert that it  
1231 would be impermissible for me to *try* to stab someone to death while asserting  
1232 at the same time that it would not be impermissible for me to *actually* stab  
1233 someone to death. For one thing, I could hardly *succeed* in such an attempt  
1234 without first *making* the attempt, so if the latter is prohibited, it seems the  
1235 former must be too. Moreover, it seems that the very reason we are prohibited  
1236 from attempting certain things is precisely because it would be wrong to  
1237 actually do those things, so a stand-alone prohibition against attempting would  
1238 typically make very little sense unless coupled with a prohibition against  
1239 actually doing what one is attempting to do.

1240 Moreover, there are obvious reasons why we might expect such an asym-  
1241 metry. In general, having a duty to do something might be thought to depend  
1242 on our having strong moral reasons to do it. One would expect moral rea-

---

8 Granted, the demand may be pragmatically pointless in any situation in which all parties *know* that it will be inevitably met, but this hardly renders it unreasonable.

1243 sons to behave in ways that parallel reasons of any other sort, such as, for  
 1244 instance, epistemic or prudential ones. And reasons of every other sort seem  
 1245 to be asymmetric with respect to our abilities in precisely the way that I claim  
 1246 moral reasons are. Perhaps it cannot be true that an agent ought to believe  
 1247 something if she is incapable of believing it. But it does not seem to follow  
 1248 that she could not have good reason to believe something that she is incapable  
 1249 of doubting (if a belief is indubitable, this is typically thought to be a point in  
 1250 its favour). Or consider prudential reasons. If you are starving hungry (bar-  
 1251 rring any conflicting considerations) you have good reason to eat. If you are  
 1252 incapable of eating, this would undermine those reasons. But it's not at all  
 1253 obvious that if you cannot resist eating, that would in any way weaken the  
 1254 reasons you have in favour of eating.<sup>9</sup>

1255 There are clear grounds for supposing that our reasons are limited to those  
 1256 things that we are able to do, while not being similarly limited to what we are  
 1257 able to avoid. Our reasons are typically based on some sort of independent  
 1258 *value* that's at stake. If a reason for performing some action or believing some  
 1259 proposition is based on some value (e.g., good evidence or a strong moral or  
 1260 prudential case), then insofar as we are capable of sensitivity to that value, we  
 1261 will be sensitive to the reasons it generates. But there would be no point at all  
 1262 in possessing a parallel capacity for *insensitivity* towards those same values.  
 1263 Here's another way to put the point: if we are violating some core value, we  
 1264 had better have a good excuse for doing so. Being incapable of respecting the  
 1265 value certainly *is* a good excuse. If we are instead respecting the value, we  
 1266 need no excuse for doing so, so no parallel ability to do otherwise is called for  
 1267 in order to render our behaviour intelligible. That something is impossible is,  
 1268 in itself, a *reason* for not bothering. In contrast, the fact that we cannot avoid  
 1269 choosing to do something doesn't undermine the rationale for doing it at all.  
 1270 In some cases, it may well be the very strength of the rationale in favour of  
 1271 performing some action or adopting some belief that *explains* why doing so  
 1272 might be irresistible to us.

---

9 This is not entirely uncontroversial. Lockie (2018) argues that prudential and epistemic reasons, as well as moral ones, depend on our ability to avoid doing or believing the thing in question. I am doubtful about OIANT principles in relation to all of these classes of reasons, but I think that Lockie is right in maintaining that there could be little intelligible basis to suppose that moral reasons were unique in this respect, hence if OIANT principles are to be plausible in the moral realm, we should expect them to be defensible in the epistemic and prudential realms too. Though of course, if we accept OIANT principle across the board, including in the epistemic realm, we would then, arguably, need to embrace Lockie's further conclusion: that any argument in favour of determinism would be automatically self-refuting.

1273 Demanding the impossible is unreasonable on the basis that an *inability* to  
1274 do something may render one’s otherwise bad or irrational behaviour perfectly  
1275 reasonable in the circumstances. This is not dependent on any alternative  
1276 possibilities requirement for control, as evidenced by the fact that a person’s  
1277 perfectly decent but unavoidable behaviour may well be entirely reasonable  
1278 and explicable, even if they cannot resist this behaviour, on the basis that  
1279 it is explained by their sensitivity to certain values. Such an explanation  
1280 would more plausibly be *weakened* by introducing the additional ability to be  
1281 *insensitive* to those values as opposed to being *strengthened* by it.

1282 Moreover, whether a demand constitutes a demand for the impossible is  
1283 asymmetric with respect to what the agent *must* do and what the agent *cannot*  
1284 do. While it is unreasonably demanding to expect an agent to do the impossible,  
1285 it is in no way similarly unreasonably demanding to expect an agent to do the  
1286 inevitable. Since the requirement is so easily met, quite the opposite seems  
1287 to be true; the inevitability is, if anything, evidence for the conclusion that  
1288 such a requirement is *undemanding*. But in any case, there is certainly no  
1289 parallel entailment of demandingness. This is precisely why psychiatrically  
1290 well-adjusted individuals don’t deserve medals for not dismembering small  
1291 children.

1292 We cannot support OIANT then, by a simple appeal to the alleged symmetry  
1293 with OIC. Moreover, it is not all all obvious that the insistence on symmetry  
1294 can be propped up with the consideration that both OIC and OIANT depend  
1295 on a two-way freedom.

### 2<sub>463</sub> The “Theory-Fuelled” Defence

1297 The “theory-fuelled” defence draws on Feldman’s analysis of obligation in  
1298 terms of the comparative value of the possible worlds accessible to agents  
1299 (1986). More recently, Haji calls this the “doing the best we can” model  
1300 (DBWC) (2019; see also Haji and Herbert 2018a).

1301 In short, the analysis contends that we are morally obligated to actualise  
1302 the best world that we can actualise of all of those “accessible” to us, where  
1303 “best” is understood in terms of a ranking of the “deontic” or “intrinsic” value  
1304 of worlds, according to whichever theory of normative ethics is endorsed  
1305 (e.g., for a utilitarian it may be the world with the greatest sum of utility, for a  
1306 Kantian it may be the world in which we act in accordance with universalisable  
1307 maxims, whereas for a virtue ethicist it may be the world in which we best  
1308 act in accordance with the virtues).

1309 There needn't be a *unique* best world; perhaps various worlds are tied for  
 1310 first place. But we are obligated to actualise *a* best world. However, some facts  
 1311 may be "unalterable"; there are certain states of affairs that would occur in  
 1312 every possible world accessible to us (e.g., the sun will rise tomorrow, various  
 1313 statements about the past will be true, etc.) If those states of affairs occur in  
 1314 all of the worlds that are accessible to us, then it is trivially true that they  
 1315 will also occur in all of the *best* worlds accessible to us. But now we have a  
 1316 problem: it appears that anything unalterable will automatically be obligatory.  
 1317 We will automatically be obligated to actualise any world that we cannot avoid  
 1318 actualising. Yet this is counterintuitive; it seems intuitively wrong to say that  
 1319 I have a moral duty to actualise a world in which the sun rises tomorrow or  
 1320 to actualise a world in which certain statements about the past are true.

1321 Haji's solution is to appeal to an **OIANT** principle. That is, we assume that  
 1322 further to supposing that we can only be obligated to bring about states of  
 1323 affairs that are accessible to us, we must *also* suppose that we can only be  
 1324 obligated to bring about any particular state of affairs on the explicit condition  
 1325 we are *also* able to actualise a world in which those states of affairs do *not*  
 1326 obtain.

1327 Perhaps this is one way to maintain a DBWC theory consistent with ensuring  
 1328 that the unalterable should not automatically be obligatory. But it is not the  
 1329 only way, and it's hardly obvious that it is the most plausible way. For instance,  
 1330 instead of endorsing **OIANT**, we could instead add the (far more compelling)  
 1331 stipulation that we can only be obligated to bring about any outcome insofar  
 1332 as that outcome is causally dependent on our *intentions*.<sup>10</sup>

1333 In fact, Haji's claim that the relevant sort of ability for duty requires that  
 1334 actions not be "strictly out of one's control" commits to precisely this. More  
 1335 recently, Haji and Herbert have defended the claim that the sort of ability  
 1336 relevant to duty ought to be robust, in the sense that requires, among other  
 1337 things, that it is strongly agentive, where this involves being brought about  
 1338 by an agent *intentionally* (2018a, 2018b). However, if having a duty requires  
 1339 that we are able to fulfil that duty in precisely this robust sense, this already  
 1340 rules out having the duty to bring about *some* unalterable states of affairs; it  
 1341 rules out precisely having the obligation to bring about states of affairs that

---

10 To be clear, I do not mean to suppose that the outcome must be caused by a *prior* intention. Rather, we should include any outcome that could be brought about through the agent's own deliberate efforts. This means, at least, that the agent's *intention in acting* is causally relevant to the outcome.



1342 will occur independently of our intentions, and hence rules out having such  
1343 obligations as seeing to it that the sun rises tomorrow.

1344 Moreover, this plausibly explains why it seems intuitively obvious that  
1345 we *are* obligated to refrain from dismembering small children, even if not  
1346 refraining from such behaviour is a psychological impossibility, consistent  
1347 with the fact that it does *not* seem plausible that we are obligated to see to it  
1348 that the sun rises tomorrow. Since the very point of moral duties is to guide  
1349 our intentions, we should expect those duties to be limited in scope to those  
1350 outcomes that are dependent on our intentional behaviour.

1351 Short of having some independent reason to favour a solution that requires  
1352 us to invoke **OIANT** over the principle that duties are limited to intention-  
1353 dependent states of affairs, it seems we ought to favour the latter. While  
1354 **OIANT** principles seem inherently problematic, the principle that one cannot  
1355 be obligated to bring about a state of affairs that will happen independently of  
1356 one’s intentions seems like a basic truism. Given the ready availability of this  
1357 solution, a state of affairs being unalterable need not make it automatically  
1358 obligatory (even if we explicitly reject **OIANT**). Importantly, however, the  
1359 fact that some state of affairs is unalterable doesn’t *rule out* our having an  
1360 obligation to bring it about either.

1361 Haji and Herbert further note that if we explicitly *presume* that if some-  
1362 thing is unalterable, then it cannot be obligatory, this would also provide a  
1363 basis from which to argue in favour of **OIANT** principles (2018a, 188). But  
1364 I am arguing precisely that we have no good independent reason to accept  
1365 such a presumption. The fact that I am not robustly capable of committing  
1366 certain morally heinous acts may well establish that my avoidance of such  
1367 acts is unalterable. But the point is precisely that we have no good reason to  
1368 suppose that this is inconsistent with it being obligatory that I refrain from  
1369 committing those acts. So while the presumption that unalterability rules  
1370 out obligatoriness could certainly provide a basis for accepting an **OIANT**  
1371 principle (via a fairly obvious entailment), such a presumption is itself no  
1372 more plausible than the **OIANT** principles it is invoked to establish and is no  
1373 less in need of independent justification.

1374 In sum then, it seems that we have no reason to accept **OIANT**. Recall, how-  
1375 ever, that **OIANT** was a crucial component of the argument to the conclusion  
1376 that determinism entails that nobody ought morally to do anything. Without  
1377 it, we are entitled only to the weaker claim that, given determinism, no one  
1378 ought morally to act otherwise than they do. We must now assess whether,

1379 from a practical perspective, this weaker conclusion turns out to be just as  
1380 destructive.<sup>11</sup>

1381 The following section assesses the implications of embracing just the  
1382 weaker conclusion entailed by determinism and **OIC**, given a rejection of  
1383 **OIANT**. In particular, the aim is to question whether this weaker conclusion  
1384 *alone* should be regarded as destructive to deontic morality, even if we follow  
1385 Haji in supposing that no one has a duty to do otherwise.<sup>12</sup>

### 1386 **3 The Lack of Obligation to Act Otherwise**

1387 The conclusion that nobody is obligated to act otherwise than they actually do  
1388 may seem problematic enough. Let us call this claim “Unfulfilled Obligation  
1389 Scepticism” (**UOS**):

1390 UOS. If an agent *S*, as of a time *t*, actualises a world in which state  
1391 of affairs *p* occurs, this entails that *S* had no moral obligation, as of  
1392 *t*, to actualise a world in which state of affairs *p* does not occur.

1393 This means that only our *actual* choices and actions could possibly count  
1394 as obligatory. We may sometimes both have and fulfil moral obligations, but  
1395 we can never have a moral obligation that we contravene. Perhaps this alone  
1396 undermines deontic morality. **UOS** may seem to threaten moral deliberation,  
1397 obligation, or motivation, rendering them practically unintelligible. Let’s  
1398 examine these potential threats in turn.

#### 390 *UOS and Moral Deliberation*

1400 Firstly, it might be argued that **UOS** renders moral deliberation practically  
1401 impossible. By “moral deliberation”, I mean reasoning about what to do in  
1402 advance of deciding, rather than reasoning about how to appraise an action  
1403 that has already occurred.

1404 There are several reasons why **UOS** might look problematic. We always  
1405 know in advance that there is no way that our actions will possibly count

---

11 For illuminating explorations of arguments to this more modest effect, see Nelkin (2011, 100–103) and Jeppsson (2016).

12 Since the following section is premised explicitly on assessing the implications of rejecting **OIANT** and embracing **OIC alone**, any readers who are unpersuaded by the arguments so far, aimed at establishing that we can embrace the latter without the former, can essentially stop reading here.

1406 as "forbidden" at the time that we perform them. Moreover, whether we  
1407 are obligated to perform any action seems closely dependent on whether we  
1408 choose to, so we might suppose that UOS robs us of any intelligible way to  
1409 give rational weight to our purported duties prior to actually making a choice.

1410 Suppose that Ada is a highly rational moral agent, who has recently become  
1411 convinced of the truth of UOS. She believes that she can only be morally  
1412 obligated to do something if she does in fact do it. She now faces the following  
1413 situation: Ada's uncle has arranged in his will for her to receive all of his  
1414 fortune should he die. However, he is planning to change his will when he  
1415 visits the solicitor's office later today. Her uncle has two small children and  
1416 had previously supposed that his wealthy wife's ample income would stand  
1417 them in good stead should he suddenly die, so he had planned to leave his  
1418 fortune to Ada, his favourite niece. However, his wife has just died in a freak  
1419 accident (leaving her fortune to her husband). If he should suddenly die too,  
1420 his children would now be left orphaned and destitute, while Ada would  
1421 receive all of his wealth, including that of his late wife. In contrast, Ada has a  
1422 decent job and a reasonably high income of her own. She will be fine without  
1423 a substantial inheritance. He is therefore planning to change his will, leaving  
1424 the bulk of his fortune to his children and a much more modest sum for Ada.  
1425 She can appreciate the reasonableness of her uncle's decision.

1426 However, while she is alone visiting him, he collapses unconscious, and  
1427 appears to be dying of a heart attack. No one else knows that Ada is visiting.  
1428 She could easily walk away without calling an ambulance. She would then be  
1429 rich enough to buy that Ferrari she always wanted. As a rational moral agent,  
1430 Ada certainly would have supposed that she had a moral obligation to call an  
1431 ambulance *prior* to being persuaded of the truth of UOS. But she must now  
1432 work out what bearing this principle has. Should it change the way that she  
1433 morally deliberates?

1434 I endorse the idea that we ought to do the best we can, where this involves  
1435 being obligated to bring about the best of the intention-dependent states of  
1436 affairs accessible to us. So Ada ought to actualise the best intention-dependent  
1437 state of affairs she can. This only seems to require two abilities: firstly, she  
1438 must be able to compare the deontic value of the worlds that would result  
1439 from various rival intentions, and secondly, she needs to suppose that she can  
1440 actualise the best of them. We ought to ask whether UOS poses any obstacle  
1441 to her doing either of these things.

1442 Firstly, let's think about her ability to assess the value of the intention-  
1443 dependent states of affairs between which she is deliberating. On virtually

1444 any theory of normative ethics, the world in which she calls an ambulance  
 1445 will look superior to the world in which she does not call an ambulance. If she  
 1446 doesn't call an ambulance, she will perform no morally admirable actions, and  
 1447 her greed and cruelty will result in an innocent man dying, and his children  
 1448 being left orphaned and destitute. If she does call an ambulance, she will have  
 1449 done a good deed, and through her fairness and kindness, she would ensure  
 1450 that he survives to care for his children. For deontologists, virtue ethicists and  
 1451 consequentialists alike then, the world in which she calls an ambulance will  
 1452 be ranked morally superior to those in which she refrains.

1453 Does she need assurance of her duty in advance? It seems not. On any  
 1454 plausible DBWC analysis, moral duties are not going to be stand-alone con-  
 1455 siderations that exert their moral pressure on us independently of the other  
 1456 facts about the situation. A world  $w$  that we might actualise does not count  
 1457 as morally superior to some other world  $w'$  on the *basis* that we are morally  
 1458 obligated to actualise  $w$  instead of  $w'$  (that supposition would render the  
 1459 DBWC account entirely vacuous). The explanation is always the other way  
 1460 around: we are morally obligated to favour actualising  $w$  over  $w'$  precisely  
 1461 *because* we have some independent basis to suppose that  $w$  is superior to  
 1462  $w'$ . The obligation arises because one of these worlds has a higher "intrinsic  
 1463 value". Values are conceptually prior to obligations: duties are the conceptual  
 1464 outputs of values.

1465 But the point needn't rest on accepting a DBWC analysis either. Quite  
 1466 independently of whether one accepts that analysis, it is a mistake to think  
 1467 that duty is conceptually prior to moral value. Consider Kant. There can be few  
 1468 theorists who afford duty a more fundamental status. Yet even for Kant, duties  
 1469 are not independent additional substantive reasons for acting; they are derived  
 1470 from considerations about the rational wills of other agents, which confer on  
 1471 them a status as ends in themselves. While Kant encourages us to act "from  
 1472 duty", as opposed to merely "in conformity with duty" (*Groundwork*, 1998,  
 1473 1.10–11), he certainly doesn't suppose that duties exist and exert pressure  
 1474 independently of the values that give rise to them; respecting duty is simply  
 1475 the same thing as respecting other rational beings. It's hard to imagine any  
 1476 plausible system of ethics according to which duties are not derived from  
 1477 some prior moral value.

1478 Perhaps it will be accepted that Ada (as a rational agent with some theory  
 1479 of normative ethics up her sleeve) knows that the world in which she calls  
 1480 an ambulance for her uncle is better than the world in which she refrains  
 1481 from calling an ambulance (i.e., she knows that there are substantive moral

1482 considerations in favour of calling an ambulance). Granted that she knows  
1483 this, she must also know she is obligated to call an ambulance insofar as she  
1484 can. But given determinism, we may worry that she has no reason to think  
1485 that she can.

1486 This concern is misguided. Firstly, we must dispense with any idea that if  
1487 her intention is determined, then her actions are fixed no matter what she  
1488 intends. To reason like this would be to commit the “fatalist’s fallacy”: even  
1489 if her action is predetermined, this does not entail that it isn’t conditional  
1490 upon her intentions. If she is determined to call an ambulance, this will  
1491 be *because* it is determined that her deliberative process culminates in her  
1492 forming an intention to call an ambulance, and this brings it about that she  
1493 calls an ambulance. Determinism does not make our attempts to act causally  
1494 ineffective.

1495 Secondly, she has no reason to suspect, in advance of making up her mind,  
1496 that she cannot call an ambulance. While it is possible that determinism robs  
1497 her of the ability to call an ambulance, it might just as easily rob her of the  
1498 ability to refrain. She has no reason to favour the presumption that her calling  
1499 an ambulance is impossible over the presumption that it is inevitable. The  
1500 only way that she can find out which of these she is determined to do is by  
1501 reaching a *decision*.

1502 From an epistemic perspective, both decisions remain open. As Pereboom  
1503 (2001, 147–148), Fischer (2006), and Jeppsson (2016) have all argued, such  
1504 epistemic openness is all we need in order for it to be rational to make a value-  
1505 driven choice. As Fischer puts the point, if one were asked to choose which of  
1506 two doors to walk through, and told that behind one them is a million dollars  
1507 while behind the other there is a den of rattlesnakes, it would be ludicrous  
1508 to suppose that the truth of determinism might weaken the rational case in  
1509 favour of choosing the door with the money, or that one would be forced to  
1510 just “wait and see what happens” instead of making a value-driven choice  
1511 (2006, 329).

1512 Moreover, suppose we grant that determinism introduces a doubt about  
1513 whether Ada can call an ambulance (we should not grant this, given the  
1514 deliberatively irrelevant nature of the “doubt”, when both options remain  
1515 epistemically open, but suppose we grant it anyway). Doubts about whether  
1516 we can do things do not usually weaken our rationale for *trying* when there is  
1517 something morally significant at stake. Obviously, sometimes failure comes  
1518 with other off-putting risks; you may be reluctant to dive into the river to  
1519 save the drowning child, but it is usually the risk to your own life rather than

1520 the possibility of failing in your attempt that causes such reluctance. There  
 1521 is always *some* risk of failure, even with the simplest actions, regardless of  
 1522 determinism. One is “always at the mercy of the world”, as O’Shaughnessy  
 1523 famously notes (1973, 370). But it would be very strange for anyone to suppose  
 1524 that this should stop us from even attempting to bring about better outcomes.

1525 Suppose that Sofia is in a hospital when the main power supply fails. Luck-  
 1526 ily, there is a short-term emergency power supply that will keep the electricity  
 1527 going for five minutes, during which time the back-up generator can be acti-  
 1528 vated, saving the lives of hundreds of patients whose life-support machines  
 1529 will otherwise fail. Now suppose that Sofia is the only person with access to  
 1530 the button that activates the back-up generator. There would be something  
 1531 seriously wrong with Sofia if she reasoned as follows: “I only ought to activate  
 1532 the back-up generator if I can. But there is no guarantee that this button  
 1533 works, so I don’t know that I can. I therefore see no reason to bother pressing  
 1534 it”. Ordinarily, we do not need a guarantee that we can do something before  
 1535 we attempt to do it when there are morally significant outcomes at stake.

1536 There seems to be no reason to suppose that UOS poses any serious obstacle  
 1537 to moral deliberation. Nonetheless, something emerges from this picture that  
 1538 might seem troubling. Essentially, we can escape being duty-bound to do  
 1539 things simply by choosing not to do them. If Ada does not call an ambulance,  
 1540 it will turn out, once her choice has been made, that she has done nothing  
 1541 wrong. Her choosing not to call an ambulance conveniently establishes that  
 1542 she had no moral obligation to call one. Moral obligations become easily  
 1543 escapable.

1544 On the one hand, it may be argued that there is something conceptually  
 1545 amiss about the idea of a moral obligation that could easily be escaped; we  
 1546 might think that inescapability is an essential condition of moral duty. Hence  
 1547 we would still have a serious threat to deontic morality if it turned out that  
 1548 all of our purported “duties” were easily escapable. On the other hand, the  
 1549 worry may be about motivation; perhaps it will be accepted that we could  
 1550 have duties that were easily escapable, but we might wonder why anybody  
 1551 would comply with them.

### 3<sub>52</sub> *UOS and Moral Obligation*

1553 The problem of easy escapability arises because we seem to have some power  
 1554 over whether we do certain things: even if causal determinism entails that  
 1555 we are unable to do otherwise, it does not entail that our actions are “strictly

1556 out of our control”; there is often a reliable causal correlation between our  
1557 attempts to do things and the success of those attempts. UOS thus seems to  
1558 give us a further power that might seem unpalatable; the power to escape  
1559 being duty-bound to do something merely by choosing *not* to do it.<sup>13</sup>

1560 We may well be aware of the fact that in forming the intention to act as we  
1561 do, we will also be conjuring up proof that we lack any ability to do otherwise,  
1562 and will therefore be actualising a situation in which we have no duty to do  
1563 otherwise. This may appear to leave our moral duties precariously at the mercy  
1564 of our wills. I see two reasons why this implication might look problematic;  
1565 the first appeals to a Kantian notion of obligation, and the second rests on a  
1566 broader conceptual concern about the inescapability of duty.

1567 Firstly, philosophers influenced by Kant may suppose that moral duties  
1568 are necessarily “categorical imperatives”. Kant distinguished hypothetical  
1569 imperatives, which depend on our contingent aims and desires, from cate-  
1570 gorical ones, which apply to us necessarily regardless of our contingent aims  
1571 and desires (*Groundwork*, 1998, 2.25). When one is morally obligated to do  
1572 something, the obligation is inescapable in the sense that one ought to do it  
1573 (insofar as one can) regardless of whether one wants to do it.

1574 Kant’s claim that moral duties are categorical imperatives is controversial.  
1575 While this claim is plausibly at the core of any objectivist analysis of meta-  
1576 ethics, many philosophers favour subjectivism. If moral duties are grounded in  
1577 our subjective aims and desires, they will not be “inescapable” in this Kantian  
1578 sense.<sup>14</sup> But I am inclined to side with Kant here, so I will not pursue this  
1579 line of argument. I doubt that anything without the character of a categorical  
1580 imperative could seriously count as a “moral obligation”.

---

13 In fact, whether such a power will count as making our duties “easily escapable” may depend on one’s view of deterministic agency. Some incompatibilists will suppose that even if an agent can escape a duty merely by intending to do so, this doesn’t make duties “escapable” in any significant sense because agents lack control over which intentions they form in the first place. For someone who takes this view, the problem of easy escapability doesn’t seem to arise at all. But even some incompatibilists will be concerned about the idea that intending not to fulfil a duty suffices to establish that the agent was never subject to a duty in the first place. This may be worrying irrespective of whether we suppose that the intention itself is freely formed.

14 Contemporary subjectivism has its roots in the work of early modern sentimentalists, such as Hutcheson, Hume and Smith, and finds more recent expression in that of 20th century non-cognitivists, such as Ayer (1936), Stevenson (1937, 1944), Hare (1952), and Gibbard (1990). But even those who advocate gentler forms of mind-dependence of morality, like Williams (1979) will struggle to accept that moral duties could be categorical imperatives. See also Foot (1972) and McDowell (1978).

1581 UOS is, however, perfectly consistent with the claim that moral duties are  
 1582 categorical imperatives. The DBWC notion of moral obligation certainly does  
 1583 not entail that moral duties depend on an agent's subjective aims and desires  
 1584 (with the possible exception of certain duties towards oneself, if there are any).  
 1585 The reason why we are morally obligated to actualise certain possible worlds  
 1586 is because they are the most valuable of the ones that we are able to actualise,  
 1587 according to our favoured theory of normative ethics. And the reason why  
 1588 determinism, given Haji's argument, entails that we are never obligated to  
 1589 actualise alternative worlds is not because we do not *want* to actualise those  
 1590 worlds, but because we *cannot* actualise them.

1591 Ada should call an ambulance if she can. This has nothing to do with  
 1592 whether she wants to call an ambulance, and everything to do with the fact  
 1593 that the world in which she calls an ambulance is more valuable than the  
 1594 world in which she does not. It is not more valuable because her own subjective  
 1595 aims and desires deem it to be (perhaps she prefers the world in which she  
 1596 inherits a fortune and buys a Ferrari). It is more valuable because of the  
 1597 comparatively high worth of her character, her actions, and/or the likely  
 1598 outcome of those actions. More generally, whatever your favoured analysis of  
 1599 obligation, I maintain that it is these sorts of substantive moral considerations  
 1600 that ground Ada's duties, and these need not leave her duties precariously  
 1601 contingent on her subjective aims and desires.

1602 While it may be an essential feature of moral obligations that they are  
 1603 categorical imperatives then, this is not inconsistent with UOS. There is,  
 1604 however, a stronger sense in which it might be claimed that moral duties are  
 1605 necessarily inescapable. We might suppose that there is something wrong with  
 1606 the idea that there could be duties that are opted into; duties that we could  
 1607 have escaped being subject to in the first place. This sense does seem plausibly  
 1608 to be threatened by UOS, but it's doubtful that this really is an essential feature  
 1609 of duty at all.

1610 Promise-making is a prime example of a duty that has to be opted into. We  
 1611 typically suppose that we are duty-bound to keep our promises, even if we  
 1612 could have escaped taking on such a duty in the first place. The important point  
 1613 is that we did *not* escape taking on this duty. Consider another example: it is  
 1614 obligatory to feed one's children as opposed to letting them starve. Nonetheless,  
 1615 many of us are under no such obligation, because we have chosen not to have  
 1616 children. While the same means of contraception were presumably available  
 1617 to many of those who chose to have children, citing this fact would hardly get  
 1618 them off the hook for letting their children starve. Again, the fact that they



1619 could, in theory, have escaped the obligation does not usually imply that they  
1620 cannot have a genuine obligation if they did not *actually* escape it.

1621 There seems to be no sense of inescapability such that it both plausibly  
1622 qualifies as an essential feature of moral obligation and is plausibly ruled out  
1623 by UOS.

### 3.3 UOS and Moral Motivation

1625 Perhaps it is not moral obligation that is threatened by UOS, but moral *moti-*  
1626 *vation*. While we may intelligibly have duties that are escapable in the sense  
1627 specified by UOS, the worry may be that this would threaten any basis that  
1628 we might have for complying with them.

1629 Return to Ada: suppose we accept that her ability to easily escape being  
1630 duty-bound does not undermine her duty, so long as she doesn’t in fact escape  
1631 it. We might now worry about what sort of motivational basis Ada could have  
1632 to incur the duty: by merely not bothering to call an ambulance, she can  
1633 ensure that she had no obligation to call one in the first place. She only has  
1634 a duty insofar as she willingly opts into it. Given that she stands to gain so  
1635 much from not opting into it, we might wonder what incentive she could have  
1636 for opting in.

1637 We have already noted that duties do not, however, provide extra reasons  
1638 for action that exert pressure on us independently of the moral considerations  
1639 that give rise to them (see 3.1). I maintain that a competent moral agent acts  
1640 out of duty not merely because it *is* her duty, but because she cares about the  
1641 substantive moral considerations which underpin the duty (in terms of any  
1642 DBWC analysis, these considerations determine the relative values of the rival  
1643 intention-dependent worlds that she might choose to actualise). It is only if  
1644 we accept the dubious assumption that the desire not to contravene a duty is  
1645 the *sole* basis of moral motivation (and that the desire to fulfil duties is always  
1646 curiously absent) that UOS seems to seriously undermine moral motivation.

1647 Putting aside the possibility of determinism and UOS, let’s think about  
1648 ordinary cases that parallel the sort of escapability of duty that we are con-  
1649 templating. Suppose that Aisha believes that she ought to give blood so long  
1650 as she is eligible to. She also knows that she has a blood donation appoint-  
1651 ment in one month’s time. Now suppose that she is planning to go on holiday  
1652 before the appointment, and she is trying to decide where to go. She suddenly  
1653 remembers that if she opts for the destination in sub-Saharan Africa instead  
1654 of the destination in Europe, this will stop her from being eligible to give

1655 blood for at least a year. If it stops her from being eligible to give blood, it will  
 1656 also remove any moral duty that she has to give blood. Should we expect this  
 1657 to motivate her to opt for sub-Saharan Africa instead of Europe? Insofar as  
 1658 Aisha counts as a competent moral agent, I very much doubt that we should  
 1659 expect this. She may even regard it as a reason *not* to opt for the destination  
 1660 in sub-Saharan Africa.

1661 Competent moral agents typically care about their duties because they care  
 1662 about the moral pressures that give rise to them. The reason why Aisha may  
 1663 be willing to incur the duty, even though she has been given an easy way of  
 1664 escaping it is because she cares about people who need blood transfusions.  
 1665 It is because of those people, after all, that she even takes herself to *have* a  
 1666 duty to give blood if she is eligible to; she thinks that the world in which she  
 1667 contributes to the supplies of blood banks is better than the world in which  
 1668 she does not. All those car crash victims and children with leukaemia are not  
 1669 going to just *go away* because she is not personally duty-bound to help them.  
 1670 If she didn't care about these people, she might just as easily contravene the  
 1671 moral duty as escape it.

1672 This brings us to the crux of the issue: the very *same* considerations that  
 1673 count in favour of fulfilling the duty, should you have it, count just as strongly  
 1674 in favour of opting into the duty, if you need to do so in order to fulfil it. And  
 1675 the very *same* considerations that count in favour of opting out of the duty,  
 1676 if you can, count just as strongly in favour of contravening the duty, if you  
 1677 cannot. In no case then, does the fact that the duty can only be fulfilled if  
 1678 opted into (i.e., UOS) change the agent's reasons for deciding either way. Just  
 1679 like Aisha, the reasons that Ada has for fulfilling her duty to call an ambulance  
 1680 for her uncle (should she have such a duty) also count in favour of incurring  
 1681 the duty if she needs to incur it in order to fulfil it. And the same reasons she  
 1682 has to opt out of incurring the duty would count in favour of contravening  
 1683 the duty if its existence did not depend on her opting into it. In no case does  
 1684 it appear rational for her to arrive at a different decision, given UOS, than she  
 1685 would have arrived at without it.

1686 It's unclear why anyone would be keenly motivated not to contravene a  
 1687 duty, while at the same time caring so little about fulfilling one. Such a mind-  
 1688 set seems to be directly inconsistent with the sort of sensitivity to value that  
 1689 characterises competent moral deliberation. What exactly is the imagined psy-  
 1690 chology of an agent who is highly motivated by an aversion to contravening  
 1691 duties while also trying to avoid fulfilling them? Such an agent, despite her  
 1692 thorough commitment to not contravening duties, would be completely indif-

1693 ferent to the moral pressures that actually give rise to duties, as well as being  
1694 positively *averse* to fulfilling a duty if it’s possible to escape it. Even if it were  
1695 possible for an agent to have this bizarre attitude towards moral pressures,  
1696 this certainly does not capture the way most of us morally deliberate.

1697 A competent moral agent typically reasons *from* considerations about the  
1698 respective values of the courses of action between which she is deliberating  
1699 to conclusions about what she ought morally to do. The moral landscape for  
1700 anyone who reasons in this way seems to be largely untouched by UOS. So it  
1701 does not appear to pose a serious threat to moral motivation.

#### 1702 4 Conclusion

1703 Haji argues, similarly to Lockie, that there could be no moral obligations at all  
1704 if determinism were true. In order to establish this conclusion, however,  
1705 we must invoke both an “ought” implies “can” principle and an “ought”  
1706 implies “able not to” principle. In section 1, I argued that without OIANT,  
1707 we could establish only the weaker conclusion that there are no *unfulfilled*  
1708 moral duties. In section 2, I argued that we ought to reject OIANT, and  
1709 hence that only the weaker conclusion has been plausibly established. Finally,  
1710 in section 3, I argued that while this weaker conclusion may initially look  
1711 just as damaging, it actually has surprisingly little practical importance for  
1712 morality. While I believe (contra Haji, and in agreement with Lockie) that  
1713 determinism plausibly threatens moral responsibility, I deny that it poses any  
1714 serious independent threat to deontic morality.

1715 I admit that aspects of this thesis seem paradoxical. It seems odd to suppose  
1716 that if determinism is true, this entails that nobody ever violates a moral  
1717 duty. The air of paradox arises, I think, from two sources. Firstly, from the  
1718 fact that we do not know in advance what we are capable of doing, since  
1719 we do not know in advance which actions are impossible and which are  
1720 inevitable. This means that acting otherwise remains an epistemically and  
1721 pragmatically live option when we contemplate our potential moral duties  
1722 in advance. Secondly, it may well be that the sense of “can” typically used  
1723 in relation to principles like OIC is actually distinct from the sense of “can”  
1724 according to which determinism robs us of the opportunity to do otherwise.

1725 I have granted for the sake of argument that OIC is true and that we can  
1726 use a single sense of “can” both in formulating OIC and in defence of the in-  
1727 compatibilist claim that nobody can do otherwise if determinism is true. This  
1728 has the upshot that nobody can be obligated to act otherwise if determinism

1729 is true, and hence that there are no unfulfilled duties in a deterministic world.  
 1730 If that conclusion seems too counterintuitive to accept, then an alternative  
 1731 strategy would be to question whether we should accept all of the following  
 1732 three theses: (1) that OIC is true, (2) that determinism may well be true, and  
 1733 (3) that no one can do otherwise if determinism is true in precisely the same  
 1734 senses of “can” according to which “ought” implies “can”. My goal has been  
 1735 to argue that even if we *accept* all three, the threat to morality might not be as  
 1736 all-encompassing as it seems. Whether we should accept all three is another  
 1737 question entirely.<sup>15</sup>

1738 Nadine Elzein

1739  0000-0003-0312-2233

1740 University of Warwick

1741 nadine.elzein@warwick.ac.uk

## 1742 References

- 1743 AYER, Alfred Jules. 1936. *Language, Truth and Logic*. 1st ed. London: Victor Gollancz.  
 1744 —. 1946. “Freedom and Necessity.” *Polemic* 5: 36–44. Reprinted in Ayer (1954, 271–284).  
 1745 —. 1954. *Philosophical Essays*. London: MacMillan Publishing Co.  
 1746 —, ed. 1959. *Logical Positivism*. Glencoe, Illinois: Free Press.  
 1747 BEROFSKY, Bernard. 2002. “Ifs, Cans, and Free Will: The Issues.” in *The Oxford Hand-*  
 1748 *book of Free Will*, edited by Robert H. KANE, pp. 181–201. Oxford Handbooks. New  
 1749 York: Oxford University Press. Second edition: Kane (2011), doi:10.1093/oxfordhb  
 1750 /9780195178548.003.0008.  
 1751 BROUWER, Frederick E. 1969. “A Difficulty with ‘Ought Implies Can’.” *The Southern*  
 1752 *Journal of Philosophy* 7(1): 45–50, doi:10.1111/j.2041-6962.1969.tb02040.x.  
 1753 BROWN, James M. 1977. “Moral Theory and the Ought-Can Principle.” *Mind* 86(342):  
 1754 206–223, doi:10.1093/mind/lxxxvi.342.206.  
 1755 CAMPBELL, Charles Arthur. 1951. “Is ‘Free Will’ a Pseudoproblem?” *Mind* 60(240):  
 1756 441–465, doi:10.1093/mind/LX.240.441.  
 1757 CHISHOLM, Roderick M. 1964. *Human Freedom and the Self*. University of Kansas:  
 1758 Department of Philosophy. The Lindley Lecture for 1962, reprinted in Chisholm  
 1759 (1989, 5–17).  
 1760 —. 1989. *On Metaphysics*. Minneapolis, Minnesota: University of Minnesota Press.  
 1761 CLARKE, Randolph. 2009. “Dispositions, Abilities to Act, and Free Will: The New  
 1762 Dispositionalism.” *Mind* 118(470): 323–351, doi:10.1093/mind/fzp034.

---

15 I am very grateful to Yuhan Fu, Ivo Evans Storrie, Dr Anneli Jefferson, and Dr Tuomas K. Pernu, for valuable comments on an earlier draft of this paper. I am also very grateful to Professor Ishtiyaque Haji and to anonymous reviewers for their incredibly helpful feedback.

- 1763 ELZEIN, Nadine. 2013. "Pereboom's Frankfurt Case and Derivative Culpability." *Philosophical Studies* 166(3): 553–573, doi:10.1007/s11098-012-0061-y.
- 1764 —. 2017. "Frankfurt-Style Counterexamples and the Importance of Alternative Possibilities." *Acta Analytica* 32(2): 169–191, doi:10.1007/s12136-016-0305-0.
- 1765 —. 2019. "To Be Able to, Or to Be Able Not to? That is the Question: A Problem for the Transcendental Argument for Freedom." *European Journal of Analytic Philosophy* 15(2): 13–32, doi:10.31820/ejap.15.2.1.
- 1766 —. 2008. "Masked Abilities and Compatibilism." *Mind* 117(468): 843–865, doi:10.1093/mind/fzn078.
- 1767 FELDMAN, Fred. 1986. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Philosophical Studies Series n. 35. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-009-4570-8.
- 1768 FISCHEK, John Martin. 2003. "'Ought-Implies-Can,' Causal Determinism and Moral Responsibility." *Analysis* 63(1): 244–250, doi:10.1093/analys/63.3.244.
- 1769 —. 2006. "Free Will and Moral Responsibility." in *The Oxford Handbook of Ethical Theory*, edited by David COPP, pp. 321–355. Oxford Handbooks. Oxford: Oxford University Press, doi:10.1093/oxfordhb/9780195325911.003.0013.
- 1770 FOOT, Phillipa R. 1972. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review* 81(3): 305–316. Reprinted in Foot (2002, 157–173), doi:10.2307/2184328.
- 1771 —. 1978. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Oxford University Press. Second edition: Foot (2002).
- 1772 —. 2002. *Virtues and Vices and Other Essays in Moral Philosophy*. 2nd ed. Oxford: Oxford University Press. First edition: Foot (1978), doi:10.1093/0199252866.001.0001.
- 1773 VAN FRAASSEN, Bas C. 1973. "Values and the Heart's Command." *The Journal of Philosophy* 70(1): 5–19, doi:10.2307/2024762.
- 1774 FRANKFURT, Harry G. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66(23): 829–839, doi:10.2307/2023833.
- 1775 GIBBARD, Allan F. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Cambridge, Massachusetts: Harvard University Press.
- 1776 GRAHAM, Peter A. 2011. "'Ought' and Ability." *The Philosophical Review* 120(3): 337–382, doi:10.1215/00318108-1263674.
- 1777 GRIFFIN, James. 1992. "The Human Good and the Ambitions of Consequentialism." *Social Philosophy and Policy* 9(2): 118–132, doi:10.1017/s0265052500001436.
- 1778 GRZANKOWSKI, Alex. 2014. "'Can' and the Consequence Argument." *Ratio* 27(2): 173–189, doi:10.1111/rati.12033.
- 1779 HAJI, Istiyaque. 1998. *Moral Appraisability: Puzzles, Proposals, and Perplexities*. New York: Oxford University Press, doi:10.1093/oso/9780195114744.001.0001.
- 1780 —. 1999. "Moral Anchors and Control." *Canadian Journal of Philosophy* 29(2): 175–203, doi:10.1080/00455091.1999.10717510.
- 1781
- 1782
- 1783
- 1784
- 1785
- 1786
- 1787
- 1788
- 1789
- 1790
- 1791
- 1792
- 1793
- 1794
- 1795
- 1796
- 1797
- 1798
- 1799
- 1800
- 1801
- 1802

- 1803 —. 2002. *Deontic Morality and Control*. Cambridge: Cambridge University Press, doi:10  
1804 .1017/cbo9780511498794.
- 1805 —. 2019. *The Obligation Dilemma*. Oxford: Oxford University Press, doi:10.1093/oso/  
1806 9780190050856.001.0001.
- 1807 HAJI, Isthiyaque and HERBERT, Ryan. 2018a. “Ability, Frankfurt Examples, and Obliga-  
1808 tion.” *The Journal of Ethics* 22(2): 163–190, doi:10.1007/s10892-018-9267-3.
- 1809 —. 2018b. “Indeterministic Choice and Ability.” *The Journal of Ethics* 22(2): 191–203,  
1810 doi:10.1007/s10892-018-9268-2.
- 1811 HAJI, Isthiyaque and MCKENNA, Michael. 2004. “Dialectical Difficulties in the Debate  
1812 about Freedom and Alternative Possibilities.” *The Journal of Philosophy* 101(6):  
1813 299–314, doi:10.5840/jphil2004101616.
- 1814 —. 2006. “Defending Frankfurt’s Argument in Deterministic Contexts: A Reply to  
1815 Palmer (2005).” *The Journal of Philosophy* 103(7): 363–372, doi:10.5840/jphil20061  
1816 03715.
- 1817 HARE, Richard M. 1952. *The Language of Morals*. Oxford: Oxford University Press,  
1818 doi:10.1093/0198810776.001.0001.
- 1819 HEINTZ, Lawrence L. 1975. “Excuses and ‘Ought’ Implies ‘Can’.” *Canadian Journal of*  
1820 *Philosophy* 5(3): 449–462, doi:10.1080/00455091.1975.10716123.
- 1821 HEUER, Ulrike. 2010. “Reasons and Impossibility.” *Philosophical Studies* 147(2):  
1822 235–246, doi:10.1007/s11098-008-9285-2.
- 1823 VAN INWAGEN, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- 1824 —. 2000. “Free Will Remains a Mystery.” in *Philosophical Perspectives 14: Action and*  
1825 *Freedom*, edited by James E. TOMBERLIN, pp. 1–19. Oxford: Blackwell Publishers.  
1826 The Eighth Philosophical Perspectives Lecture, doi:10.1111/0029-4624.34.s14.1.
- 1827 —. 2004. “Freedom to Break the Laws.” in *Midwest Studies in Philosophy 28: The*  
1828 *American Philosophers*, edited by Peter A. FRENCH and Howard K. WETTSTEIN,  
1829 pp. 334–350. Boston, Massachusetts: Blackwell Publishers, doi:10.1111/j.1475-  
1830 4975.2004.00099.x.
- 1831 —. 2008. “How to Think about the Problem of Free Will.” *The Journal of Ethics* 12(3/4):  
1832 327–341, doi:10.1007/s10892-008-9038-7.
- 1833 JEPSSON, Sofia. 2016. “Reasons, Determinism and the Ability to Do Otherwise.”  
1834 *Ethical Theory and Moral Practice* 19(5): 1225–1240, doi:10.1007/s10677-016-9721-  
1835 X.
- 1836 KANE, Robert H. 1999. “Responsibility, Luck, and Chance: Reflections on Free Will and  
1837 Indeterminism.” *The Journal of Philosophy* 96(5): 217–240, doi:10.2307/2564666.
- 1838 —, ed. 2002. *The Oxford Handbook of Free Will*. Oxford Handbooks. New York: Oxford  
1839 University Press. Second edition: Kane (2011), doi:10.1093/oxfordhb/978019517  
1840 8548.001.0001.
- 1841 —, ed. 2011. *The Oxford Handbook of Free Will*. 2nd ed. Oxford Handbooks. New York:  
1842 Oxford University Press. First edition: Kane (2002), doi:10.1093/oxfordhb/97801  
1843 95399691.001.0001.

- 1844 KANT, Immanuel. 1998. *Groundwork of the Metaphysics of Morals*. Cambridge Texts in  
 1845 the History of Philosophy. Cambridge: Cambridge University Press. Translated and  
 1846 edited by Mary Gregor; with an introduction by Christine M. Korsgaard; second  
 1847 edition: Kant (2012), doi:10.1017/cbo9780511809590.
- 1848 —. 2012. *Groundwork of the Metaphysics of Morals*. 2nd ed. Cambridge Texts in the  
 1849 History of Philosophy. Cambridge: Cambridge University Press. Translated and  
 1850 edited by Mary Gregor and Jens Timmermann; with an introduction by Christine  
 1851 M. Korsgaard; first edition: Kant (1998), doi:10.1017/cbo9780511919978.
- 1852 KEKES, John. 1984. "Ought Implies Can' and Two Kinds of Morality." *The Philosophical*  
 1853 *Quarterly* 34(137): 459–467, doi:10.2307/2219064.
- 1854 KRATZER, Angelika. 1977. "What 'Must' and 'Can' Must and Can Mean." *Linguistics*  
 1855 *and Philosophy* 1(3): 337–356, doi:10.1007/bf00353453.
- 1856 KÜHLER, Michael. 2013. "Who am I to Uphold Unrealizable Normative Claims?"  
 1857 in *Autonomy and the Self*, edited by Michael KÜHLER and Nadja JELINEK, pp.  
 1858 191–212. Philosophical Studies Series n. 119. Dordrecht: Springer, doi:10.1007/978-  
 1859 94-007-4789-0\_9.
- 1860 LEHRER, Keith. 1968. "Cans Without Ifs." *Analysis* 29(1): 29–32, doi:10.1093/analysis/2  
 1861 9.1.29.
- 1862 LEMMON, Edward John. 1962. "Moral Dilemmas." *The Philosophical Review* 71(2):  
 1863 139–158, doi:10.2307/2182983.
- 1864 LEWIS, David. 1981. "Are we Free to Break the Laws?" *Theoria* 47(3): 113–121.  
 1865 Reprinted in Lewis (1986, 291–298), doi:10.1111/j.1755-2567.1981.tb00473.x.
- 1866 —. 1986. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.109  
 1867 3/0195036468.001.0001.
- 1868 LOCKIE, Robert. 2018. *Free Will and Epistemology: A Defence of the Transcendental*  
 1869 *Argument for Freedom*. London: Bloomsbury Academic, doi:10.5040/9781350044  
 1870 180.
- 1871 MAIER, John. 2015. "The Agentive Modalities." *Philosophy and Phenomenological*  
 1872 *Research* 90(1): 113–134, doi:10.1111/phpr.12038.
- 1873 MCDOWELL, John Henry. 1978. "Are Moral Requirements Hypothetical Imperatives?"  
 1874 *Proceedings of the Aristotelian Society, Supplementary Volume* 52: 13–29. Reprinted  
 1875 in McDowell (1998, 77–94), doi:10.1093/aristoteliansupp/52.1.13.
- 1876 —. 1998. *Mind, Value, and Reality*. Cambridge, Massachusetts: Harvard University  
 1877 Press.
- 1878 MELE, Alfred R. 2003. "Agents' Abilities." *Noûs* 37(3): 447–470, doi:10.1111/1468-  
 1879 0068.00446.
- 1880 MOORE, George Edward. 1903. *Principia Ethica*. Cambridge: Cambridge University  
 1881 Press.
- 1882 —. 1922. "The Nature of Moral Philosophy." in *Philosophical Studies*, pp. 310–339.  
 1883 London: Routledge & Kegan Paul.



- 1884 NELKIN, Dana Kay. 2011. *Making Sense of Freedom and Responsibility*. Oxford: Oxford  
1885 University Press, doi:10.1093/acprof:oso/9780199608560.001.0001.
- 1886 O'SHAUGHNESSY, Brian. 1973. "Trying (As the Mental 'Pineal Gland')." *The Journal*  
1887 *of Philosophy* 70(13): 365–386, doi:10.2307/2024676.
- 1888 PALMER, David D. 2005. "New Distinctions, Same Troubles: A Reply to Haji and  
1889 McKenna (2004)." *The Journal of Philosophy* 102(9): 474–482, doi:10.5840/jphil2  
1890 005102934.
- 1891 PEREBOOM, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University  
1892 Press, doi:10.1017/cbo9780511498824.
- 1893 RESCHER, Nicholas. 1987. *Ethical Idealism: An Inquiry into the Nature and Function of*  
1894 *Ideals*. Berkeley, California: University of California Press, doi:10.2307/jj.5233063.
- 1895 SAKA, Paul. 2000. "Ought Does Not Imply Can." *American Philosophical Quarterly*  
1896 37(2): 93–105.
- 1897 SAPONTZIS, Steve F. 1991. "'Ought' Does Imply 'Can'." *The Southern Journal of Philos-*  
1898 *ophy* 29(3): 382–393, doi:10.1111/j.2041-6962.1991.tb00598.x.
- 1899 SCHLICK, Moritz, ed. 1930. *Fragen der Ethik*. Schriften zur wissenschaftlichen Weltauf-  
1900 fassung n. 4. Wien: Springer Verlag. Reprinted in Schlick (2006, 347–536), doi:10.1  
1901 007/3-211-29791-x\_4.
- 1902 —. 1939. "When is a Man Responsible?" in *Problems of Ethics*, pp. 143–158. Englewood  
1903 Cliffs, New Jersey: Prentice-Hall, Inc. Authorized translation of Schlick (1930) by  
1904 David Rynin.
- 1905 —. 2006. *Lebensweisheit. Versuch einer Glückseligkeitslehre, Fragen der Ethik*. Kritische  
1906 Gesamtausgabe n. I.3. Wien: Springer. Herausgegeben von Mathias Iven, doi:10.1  
1907 007/3-211-29791-X.
- 1908 SINNOTT-ARMSTRONG, Walter. 1984. "'Ought' Conversationally Implies 'Can'." *The*  
1909 *Philosophical Review* 93(2): 249–261, doi:10.2307/2184585.
- 1910 —. 1988. *Moral Dilemmas*. Oxford: Blackwell Publishers.
- 1911 SMART, Jamieson John Carswell [Jack]. 1961. "Free-Will, Praise and Blame." *Mind*  
1912 70(279): 291–306, doi:10.1093/mind/lxx.279.291.
- 1913 SMITH, Michael A. 1997. "A Theory of Freedom and Responsibility." in *Ethics and*  
1914 *Practical Reason*, edited by Garrett CULLITY and Berys GAUT, pp. 293–320. Oxford:  
1915 Oxford University Press. Reprinted in Smith (2004, 84–113), doi:10.1093/oso/9780  
1916 198236467.003.0011.
- 1917 —. 2003. "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and  
1918 Compulsion." in *Weakness of Will and Practical Irrationality*, edited by Sarah  
1919 STROUD and Christine TAPPOLET, pp. 17–38. Oxford: Oxford University Press.  
1920 Reprinted in Smith (2004, 114–135), doi:10.1093/0199257361.003.0002.
- 1921 —. 2004. *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics*.  
1922 Cambridge Studies in Philosophy. Cambridge: Cambridge University Press, doi:10  
1923 .1017/cbo9780511606977.



- 1924 STERN, Robert. 2004. "Does 'Ought' Imply 'Can'? And Did Kant Think It Does?" *Utilitas*  
1925 16(1): 42–61, doi:10.1017/s0953820803001055.
- 1926 STEVENSON, Charles Leslie. 1937. "The Emotive Meaning of Ethical Terms." *Mind*  
1927 46(181): 14–31. Reprinted in Ayer (1959, 264–281), doi:10.1093/mind/xlvi.181.14.
- 1928 —. 1944. *Ethics and Language*. New Haven, Connecticut: Yale University Press.
- 1929 STREUMER, Bart. 2003. "Does 'Ought' Conversationally Implicate 'Can'?" *European*  
1930 *Journal of Philosophy* 11(2): 219–228, doi:10.1111/1468-0378.00184.
- 1931 —. 2007. "Reasons and Impossibility." *Philosophical Studies* 136(3): 351–384, doi:10.1  
1932 007/s11098-005-4282-1.
- 1933 —. 2010. "Reasons, Impossibility and Efficient Steps: Reply to Heuer (2010)." *Philo-*  
1934 *sophical Studies* 151(1): 79–86, doi:10.1007/s11098-009-9422-6.
- 1935 TRIGG, Roger. 1971. "Moral Conflict." *Mind* 80(317): 41–55, doi:10.1093/mind/lxxx.31  
1936 7.41.
- 1937 VIHVELIN, Kadri. 2000. "Libertarian Compatibilism." in *Philosophical Perspectives*  
1938 14: *Action and Freedom*, edited by James E. TOMBERLIN, pp. 139–166. Oxford:  
1939 Blackwell Publishers, doi:10.1111/0029-4624.34.s14.8.
- 1940 —. 2004. "Free Will Demystified: A Dispositional Account." *Philosophical Topics*  
1941 32(1/2): 427–450, doi:10.5840/philtopics2004321/211.
- 1942 —. 2011. "How to Think about the Free Will / Determinism Problem." in *Carving*  
1943 *Nature at Its Joints: Natural Kinds in Metaphysics and Science*, edited by Joseph  
1944 Keim CAMPBELL, Michael O'ROURKE, and Matthew H. SLATER, pp. 313–340.  
1945 *Topics in Contemporary Philosophy* n. 7. Cambridge, Massachusetts: The MIT  
1946 Press, doi:10.7551/mitpress/9780262015936.003.0014.
- 1947 —. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford: Oxford  
1948 University Press, doi:10.1093/acprof:oso/9780199795185.001.0001.
- 1949 VRANAS, Peter B. M. 2007. "I Ought, Therefore I Can." *Philosophical Studies* 136(2):  
1950 167–216, doi:10.1007/s11098-007-9071-6.
- 1951 WEIR, Ralph Stefan. 2016. "Relative Modality and the Ability to do Otherwise." *Euro-*  
1952 *pean Journal of Analytic Philosophy* 12(1): 47–62, <https://hrcak.srce.hr/175168>.
- 1953 WILLIAMS, Bernard Arthur Owen. 1965. "Ethical Consistency." *Proceedings of the*  
1954 *Aristotelian Society, Supplementary Volume* 39: 103–123. Reprinted in Williams  
1955 (1973, 166–186), doi:10.1093/aristoteliansupp/39.1.103.
- 1956 —. 1973. *Problems of the Self*. Cambridge: Cambridge University Press, doi:10.1017/cb  
1957 09780511621253.
- 1958 —. 1979. "Internal and External Reasons." in *Rational Action: Studies in Philosophy*  
1959 *and Social Science*, edited by Ross HARRISON, pp. 17–28. Cambridge: Cambridge  
1960 University Press. Reprinted in Williams (1981, 101–113), doi:10.1017/cbo9780511  
1961 814273.005.
- 1962 —. 1981. *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge Univer-  
1963 sity Press, doi:10.1017/cbo9781139165860.

PROOF

1964  
1965

# Consciousness, Revelation, and Confusion

## Are Constitutive Panpsychists Hoist by their Own Petard?

LUKE ROELOFS

1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981

Critics have charged constitutive panpsychism with inconsistency. Panpsychists reject physicalism for its seeming inability to explain consciousness. In making this argument, they commit themselves to the idea of “revelation”: that we know, in some especially direct way, the nature of consciousness. Yet they then attribute properties to our consciousness—like being constituted out of trillions of simpler experiential parts—that conflict with how it seems introspectively. This seems to pose a dilemma: either revelation is false, and physicalism remains intact, or revelation is true, and constitutive panpsychists are hoist by their own petard. But this is too simplistic. Constitutive panpsychists can say that our minds contain innumerable phenomenal states that are “confused” with one another: immediately present to introspection only en masse, not individually. Accepting revelation does not require ignoring the attentional, conceptual, and interpretive limitations of introspection, and these familiar limitations remove the tension between panpsychism and revelation.

1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989

What is the relationship between being conscious and knowing about consciousness? In answering this question, constitutive panpsychists face a delicate balancing act: their own case against physicalism requires that being conscious reveals something of the metaphysics of consciousness, but the stronger they make this claim of revelation, the stronger becomes an objection to their own view sometimes called “the revelation problem”. In this paper I argue that this balancing act, though delicate, is not impossible: there is a plausible, well-motivated “medium-strength” sort of revelation, strong

1990 enough to bring down physicalism but weak enough to leave constitutive  
1991 panpsychism standing.

1992 In section 1, I lay out the background to the panpsychism-physicalism  
1993 debate; in section 2, I distinguish six “revelation theses”; in section 3 I analyse  
1994 the structure and varieties of the revelation problem; and in section 4 and  
1995 section 5 I outline how to address this problem while retaining as much as  
1996 possible of the theses discussed in section 2.

## 1997 **1 Are Panpsychists Hoist by their Own Petard?**

1998 Panpsychists think all the fundamental physical things are phenomenally  
1999 conscious, where “fundamental physical things” is a placeholder for what-  
2000 ever fundamental entities feature in the true physical theory (particles, fields,  
2001 strings, spacetime, etc.). The “constitutive” part of “constitutive panpsychism”  
2002 describes the relationship between macroexperiences (the experiences of  
2003 humans and other animals) and the postulated microexperiences of the fun-  
2004 damental physical entities.<sup>1</sup> This relationship should be something like the  
2005 relationship between the physical features of human bodies (macrophysics)  
2006 and the physical features of the fundamental entities (microphysics). That  
2007 relationship (which we might call being constituted, being grounded, or be-  
2008 ing nothing over and above) generates no “explanatory gap”: even when the  
2009 details currently elude us, it seems clear that macrophysics is fully accounted  
2010 for by microphysics. When you have the right particles, arranged in the right  
2011 pattern, exerting the right forces on one another, and the right laws governing  
2012 them, there is no further problem about how to get hands, chairs, planets,  
2013 etc.: those “come for free” when the microphysical foundations are there.

2014 The failure of consciousness to fit into this neat picture is the objection to  
2015 physicalism that motivates most contemporary panpsychists. Whereas the  
2016 distribution of and relations among subatomic particles seems to explain  
2017 everything about my body, it leaves unexplained why there is anything it feels  
2018 like to be me, and why it feels the particular way it does. In particular, even

---

1 Some panpsychists would not link “macro” and “micro” (terms conveying size) with “human-like” and “fundamental” in this way. In particular, “cosmopsychists” think that the fundamental physical entity is the cosmos as a whole, which is (obviously) bigger than a human being, not smaller (see Gaudry 2008; Jaskolla and Buck 2012; Shani 2015; Nagasawa and Wager 2017; Goff 2017). Though I am sympathetic to cosmopsychism, I do not believe that it changes the essential contours of the revelation problem, though it requires some re-formulating, as noted in footnotes 11 and 14. For now I will, for convenience, speak as though the fundamental physical entities are very small.

2019 knowing the full story about the particles seems to be compatible with not  
2020 knowing what the experiences are like (this is the “knowledge argument,” cf.  
2021 Jackson 1982; Nemirow 1990; Ball 2009), and it seems that a world might have  
2022 been physically identical and yet differed from ours in respect of conscious-  
2023 ness (the “conceivability argument,” cf. Kripke 1980; Chalmers 2009). There  
2024 is a vast literature on whether these are good reason to reject physicalism  
2025 (see, e.g. Chalmers 1996; Dennett 2007; Stoljar 2006; Díaz-León 2011), but  
2026 here I will assume that they are. What comes next? In particular, is constitu-  
2027 tive panpsychism, often offered as an attractive non-physicalist alternative,  
2028 defensible?

2029 Constitutive panpsychism treats consciousness as a fundamental ingredi-  
2030 ent of nature, but tries to treat it the same as other fundamental ingredients  
2031 (mass, charge, spin, force, location, etc.). Just as those other fundamentals are  
2032 widespread in nature, with human beings as simply one particular arrange-  
2033 ment of them, so is consciousness: human experience is not metaphysically  
2034 special, just a complicated combination of widespread components. Consti-  
2035 tutive panpsychism thus retains the monistic spirit of physicalism despite  
2036 recognising consciousness as fundamental. Importantly, non-constitutive ver-  
2037 sions of panpsychism, on which human consciousness somehow “emerges  
2038 from” or is “caused by” microconsciousness but not literally “made up of” it,  
2039 do not secure this advantage. The macrophysical properties of the brain seem  
2040 to be wholly constituted by the microphysical properties of its parts, so if its  
2041 macroscopic consciousness is not similarly constituted by microconscious-  
2042 ness, the hoped-for reconciliation of mind and matter falls apart.

2043 This imposes an explanatory burden: constitutive explanations of human  
2044 consciousness in terms of microconsciousness have to do better than physi-  
2045 calist explanations. And one major line of criticism has been that they do not:  
2046 there is just as much difficulty in explaining how many simple minds combine  
2047 into complex minds as in explaining how mindless things generate minds.  
2048 This broad objection is often called “the combination problem” (Seager 1995,  
2049 280; Chalmers 2017; Roelofs 2019), and has received much discussion from  
2050 both defenders and critics of panpsychism.

2051 One specific strand of the combination problem is “the revelation prob-  
2052 lem”: macroexperiences do not *seem* introspectively to be built up out of  
2053 microexperiences. And constitutive panpsychists can’t just say: “Well they  
2054 *are*, sometimes things aren’t what they seem.” That would license physicalists  
2055 to likewise say: “Exactly! Consciousness *seems* distinct from purely physical  
2056 facts, but it’s actually not.” If being conscious doesn’t reveal the true nature of

2057 consciousness, the case against physicalism is weakened; if it does, then the  
 2058 truth of constitutive panpsychism should be introspectively obvious, which it  
 2059 is not.

2060 This talk of “seeming” and “obviousness” is not the most precise way of  
 2061 presenting things. Authors articulating the sense that there is a problem here  
 2062 say things like:

2063 [...] it is hard to see how smooth, structured macroscopic phe-  
 2064 nomenology could be derived [from microexperiences isomorphic  
 2065 to microphysics]; we might expect some sort of “jagged,” unstruc-  
 2066 tured phenomenal collection instead. (Chalmers 1996, 306)

2067 It is hard to see how [microexperiences] could somehow add  
 2068 up to the phenomenal properties with which we are familiar—  
 2069 properties with the specific, homogeneous character with which  
 2070 we are all acquainted [...]. (Alter and Nagasawa 2012, 90–91)

2071 [Revelation is] inconsistent [...] with my conscious experience  
 2072 turning out to be, in and of itself, quite different from how it  
 2073 appears to be in introspection: i.e., turning out to be constituted of  
 2074 the experiential being of billions of micro subjects of experience  
 2075 [...]. (Goff 2006, 57; cf. Lee 2019, 290–298)

2076 Similar remarks were made by certain non-reductive mind-brain identity  
 2077 theorists in the last century, writing about a perceived “grain problem”:

2078 [Any experience’s] physiological substrate, presumably, is a highly  
 2079 structured, not to say messy, concatenation of changes in elec-  
 2080 trical potential within billions of neurons in the auditory cortex  
 2081 [...]. How do all these microstructural discontinuities and inho-  
 2082 mogeneities come to be *glossed over* [...]? (Lockwood 1993, 274)

2083 How is it that the occurrence of a smooth, continuous expanse of  
 2084 red in our visual experience can [...] involve particulate, discon-  
 2085 tinuous affairs such as transfers of or interactions among large  
 2086 numbers of electrons, ions, or the like? (Maxwell 1978, 398)

2087 Indeed, Lewis makes a very similar argument, though he rejects the idea that  
 2088 experience reveals its nature and so presents the argument as a *reductio* of  
 2089 this idea:

2090 If we know exactly what the qualia of our experiences are, they  
 2091 can have no essential hidden structure - no “grain” - of which we  
 2092 remain ignorant. If we didn’t know whether their hidden “grain”  
 2093 ran this way or that, we wouldn’t know exactly what they were.  
 2094 [...] if nothing essential about the qualia is hidden, then if they  
 2095 seem simple, they are simple. (Lewis 1995, 142, n.14)

2096 Although I think all the above quotations express a similar sort of concern,  
 2097 they do so with different emphasis and framing, and the exact nature of the  
 2098 problem involved is far from clear. In section 3 I try to identify the problems  
 2099 more precisely, and in section 4 and section 5, I resolve them.

## 2100 **2 The Revelation Problem and the Revelation Thesis**

2101 Before examining the revelation problem for panpsychism, we need to exam-  
 2102 ine the background idea of a “revelation thesis” connecting consciousness to  
 2103 knowledge of consciousness. There are actually several different ideas under  
 2104 the broad heading of “revelation”: I will distinguish a total of six distinct revela-  
 2105 tion theses, resulting from a two-fold distinction permuted with a three-fold  
 2106 distinction.

2107 The two-fold distinction concerns whether the claim says (a) that the full  
 2108 truth about consciousness will always be manifest (a “reality→appearance”  
 2109 direction of implication), or (b) that what is manifest about consciousness is  
 2110 always true (an “appearance→reality” direction of implication).<sup>2</sup> Claims of the  
 2111 first sort rule out any aspect of consciousness being “hidden” from us, while  
 2112 claims of the second sort rule out any sort of “illusion” about consciousness.

2113 The three-fold distinction is about the topic of a revelation thesis - what  
 2114 kind of reality it connects with what kind of appearance. Putting things for  
 2115 now in reality→appearance terms, we can distinguish the claims:

- 2116 1. That someone having an experience can know that they are presently  
 2117 having that token experience;<sup>3</sup>

2 Byrne and Hilbert (2007, 77), draw this distinction for colour properties: they “treat Revelation as equivalent to the conjunction of two theses [...] SELF-INTIMATION [and] INFALLIBILITY”, with the former being reality→appearance and the latter appearance→reality.

3 Different authors speak variously of qualia, experiences, types of experience, and types of conscious state: for clarity I will in what follows speak of *phenomenal properties* as the things which phenomenal concepts capture, and whose natures they reveal, and of *experiences* as

- 2118 2. That someone having an experience can gain a special kind of under-  
2119 standing of that phenomenal property;
- 2120 3. That this understanding reveals “the complete nature” of a certain type  
2121 of experience.

2122 The first thesis is sometimes called “self-presentation” or “luminosity”, as  
2123 distinguished from “revelation” (Stoljar 2006, 223). But in other discussions it  
2124 is presented as an integral part of a broader idea called “revelation.” (e.g., Goff  
2125 2017, 109–110). The second thesis is sometimes put in terms of forming con-  
2126 cepts, sometimes of special sorts (e.g., Chalmers 2003a; Goff 2017, 109–110)  
2127 and sometimes just in terms of “understanding” (e.g., Stoljar 2006, 229). The  
2128 third thesis is sometimes put in terms of knowing a phenomenal property’s  
2129 “essence” or “nature”, or knowing all the essential or necessary truths about  
2130 it.<sup>4</sup> Sometimes the term “revelation” or “revelation thesis” is used specifically  
2131 for one of these theses, or for the set of them together, or for the conjunction  
2132 of the second and third. But they are worth distinguishing because, as I will  
2133 show, they support quite distinct revelation arguments against constitutive  
2134 panpsychism, which need to be addressed in quite different ways.

2135 Moreover, we can distinguish reality→appearance and appearance→real-  
2136 ity directions of each of the three, yielding a total of six revelation theses  
2137 (RT1–RT6), as follows:

Topic	Reality → Appearance direction	Appearance → Reality direction
-------	--------------------------------	--------------------------------

---

instantiations of phenomenal properties. To have an experience is to instantiate a phenomenal property, i.e., to be conscious.

4 Some example formulations: the special understanding of an experience type we gain from undergoing it “reveals the essence of Q [the experience type]: a property of Q such that, necessarily, Q has it and nothing else does” (Lewis 1995, 141–142); “for every essential truth T about E, [the subject] knows, or is in a position to know, T” (Stoljar 2006, 228); “the complete nature of the type to which [the experience] belongs is apparent to the concept user” (Goff 2017, 110). Cf. also colour-revelation theses: “If it is in the nature of the colors that p, then after careful reflection on color experience it seems to be in the nature of the colors that p” (Byrne and Hilbert 2007, 77); “The intrinsic nature of canary yellow is fully revealed” (Johnston 1992, 223). Cf. Lee (2019, 291–293), Liu (2019, 2020).



Instantiation	<b>Revelation Thesis 1:</b> If someone instantiates a phenomenal property, it will introspectively seem to them that they are instantiating that property. (Call this the “luminosity” thesis.)	<b>Revelation Thesis 2:</b> If it introspectively seems to someone that they are instantiating a phenomenal property, then they really are instantiating that property. (Call this the “no illusions” thesis.)
Understanding	<b>Revelation Thesis 3:</b> If someone instantiates a phenomenal property, they will be in a position to form a pure phenomenal concept of it. (Call this the “understanding from experiencing” thesis.)	<b>Revelation Thesis 4:</b> If someone is in a position to form a pure phenomenal concept of a phenomenal property, they must be instantiating that property. (Call this the “no understanding without experiencing” thesis.)
Knowledge of nature	<b>Revelation Thesis 5:</b> If someone has a pure phenomenal concept, reflection upon it can reveal the whole nature of the corresponding phenomenal property. (Call this the “self-intimation” thesis)	<b>Revelation Thesis 6:</b> If someone’s reflection upon a pure phenomenal concept presents some feature as pertaining to the nature of the corresponding phenomenal property, that feature really does pertain to the nature of that property. (Call this the “infallibility” thesis)

2138 I think these six theses, though logically independent, form a fairly natural  
 2139 package together, and I will refer to this package (i.e., the conjunction  
 2140 RT<sub>1</sub>–RT<sub>6</sub>) as “the revelation approach”.<sup>5</sup> This package is particularly impor-  
 2141 tant for undergirding modal arguments against physicalism, a role which it  
 2142 is held to have both by its defenders and its critics (e.g., [Stoljar 2009, 2013](#);  
 2143 [Damnjanovic 2012](#); [Liu 2019, 2020](#)). Lewis, for instance, attributes RT<sub>5</sub> and  
 2144 RT<sub>6</sub> to Kripke, as a presupposition of the latter’s inference from the conceiv-  
 2145 ability of pain without any associated brain state to their separate possibility

5 The component theses are often connected by the idea that subjects stand in a certain special relation of “acquaintance” to their experiences (see [Chalmers 2003a](#); [Goff 2015](#)): being directly acquainted with our experiences is what lets us know of their occurrence, and understand their properties in a way that fully reveals their nature. Acquaintance is often taken to be one species of a broader category of relations, called “awareness”, which likewise enable knowledge of various kinds, but which include more mediated forms of awareness like visual awareness, auditory awareness, etc. I am very happy to accept these claims about acquaintance and awareness, but they will not be distinctively important in the discussion that follows.

(Lewis 1995, 328, n.3). Goff (2017, 74–76, 96–106) likewise argues that the conceivability and knowledge arguments require that phenomenal concepts be “transparent”, effectively meaning that RT5 and RT6 must be true.<sup>6</sup> And Chalmers’ version of the conceivability and knowledge arguments relies on the premise that the primary and secondary intensions of phenomenal concepts are equivalent (Chalmers 2003b, 2009), which implies RT5 and RT6.<sup>7</sup>

Although RT5 and RT6 have the clearest role, the falsity of the other revelation theses would also leave the anti-physicalist arguments on a shaky footing. For instance, if RT3 were false, we could worry whether we possessed the pure phenomenal concepts whose “transparency” drove the arguments; if RT2 were false, we could worry that the properties these concepts expressed were not even instantiated (as argued by, e.g., Pereboom 2016, 2019); and RT4 is essential to the knowledge argument, which relies on the premise that someone who has never experienced colour cannot know what seeing colour is like.<sup>8</sup>

### 2163 **What is the Revelation Problem, Exactly?**

2162 So what exactly is the supposed problem for panpsychists? How is it distinct  
2163 from other aspects of the combination problem? Fundamentally, it concerns  
2164 a perceived incompatibility between three things:

- 2165 • the way human consciousness appears in introspection;

---

6 The arguments might not require going all the way to RT5 and RT6. Stoljar (2006, 229–230) suggests that all that is strictly required is that we have a form of access to the natures of phenomenal properties that allows us to know at least something, if not everything, about these natures. Goff argues against such an intermediate position, saying that for any property whose nature we grasp only part of, we can “split” the property into two components, one with an unknown nature and one with a known nature. The arguments against physicalism can then be run just with respect to “that aspect of phenomenal properties whose nature we know”, and for that sub-property RT5 and RT6 will be true. In this paper I will suppose that Goff is right, and seek to defend RT5 and RT6 in their “whole nature” form.

7 A concept’s primary intension is available to reflection, while its secondary intension is the nature of the property that concept expresses, so the coincidence of these two intensions implies that the natures of the properties expressed by pure phenomenal concepts are available to reflection by those who possess the concepts.

8 The revelation approach also comes up in other places. RT1, the “luminosity” thesis, is sometimes appealed to as a distinguishing feature of consciousness (Rosenthal 1993, 359; Kriegel 2009; Strawson 2015, 9). Other philosophers draw on RT1 and RT2 to develop an epistemology of introspection (Chalmers 1996, 218–219; 2003a; Smithies 2019).

- 2166 • the way human consciousness would be, if constitutive panpsychism  
2167 were true;
- 2168 • revelation: the idea that introspection gives special insight into the  
2169 reality of consciousness.

2170 The third element makes any discrepancy between the first and second seem  
2171 fatal. Yet that third element is also something panpsychists cannot readily  
2172 give up.

2173 How should we spell out these core elements? I think there are actually  
2174 three slightly different arguments to be made here, and then a fourth argu-  
2175 ment which engages with the debate on a different combination problem, the  
2176 “palette problem”. Let us consider the pure revelation arguments first, which  
2177 differ primarily in whether they rely on the appearance→reality or reality→ap-  
2178 pearance direction of implication: the first argument says, “Consciousness  
2179 appears to be X, but panpsychism implies it is not really X,” while the second  
2180 and third say, “Consciousness fails to appear to be X, but panpsychism implies  
2181 it really is X.” The first focuses on some positive introspective appearance, and  
2182 accuses constitutive panpsychists of treating that appearance as an “illusion”.  
2183 The others focus simply on the *absence* of a certain appearance.

2184 We can call the first argument the “no illusions” argument, since its third  
2185 premise is RT<sub>2</sub>, the “no illusions” thesis:

- 2186 1. If constitutive panpsychism is true, then human consciousness is always  
2187 “particulate”.
- 2188 2. Human consciousness (often) appears introspectively to be “smooth”.
- 2189 3. Consciousness can’t appear a way that it’s not. (RT<sub>2</sub>)
- 2190 4. Being “smooth” and being “particulate” are incompatible.
- 2191 5. Human consciousness is (often) smooth. (from 2 and 3)
- 2192 6. Human consciousness is (often) not particulate. (from 4 and 5)
- 2193 7. Constitutive panpsychism is false. (from 1 and 6)

2194 Obviously much turns on the meaning of the terms “particulate” and “smooth”,  
2195 but despite the frequency with which they (and similar terms like “continuous”  
2196 and “fragmented”) appear in statements of the problem, it is unclear how to  
2197 define them, and consequently unclear how plausible premises 1, 2, and 4 are.  
2198 This definitional question will be central to my discussion in the next section.

2199 The second and third arguments (involving a “reality→appearance” impli-  
2200 cation) are both suggested in Chalmers’ formulation of what he calls “the  
2201 revelation argument” (2017, 190). Chalmers notes that although constitutive

panpsychism holds consciousness to be “constituted by a vast array of microexperiences”, this vast array is not revealed to us in introspection. This poses a problem if we think both that introspection reveals the nature of consciousness, and that “whatever constitutes consciousness is part of its nature”.

I distinguish two arguments here because I think talk of “introspection” upon “consciousness” can be taken in two quite different ways. One is that introspection focused on *macroexperiences* doesn’t reveal that they are constituted by microexperiences. The other is that introspection focused on *microexperiences* isn’t even possible. The former appears to violate what I above called RT5, the “self-intimation” thesis: reflection upon a pure phenomenal concept reveals the whole nature of a phenomenal property. The latter appears to violate both what I above called RT3, the “understanding-from-experience” thesis, and RT1, the “self-presentation” thesis: having an experience should allow knowledge of its occurrence and a pure phenomenal concept of it.

Focusing on either macroexperiences or microexperiences yields the following two arguments, which I will call the “macroexperience-focused” and “microexperience-focused” argument. The first runs thus, with RT5 as third premise:

1. If constitutive panpsychism is true, each human experience (“macroexperience”) is constituted by a vast array of microexperiences.
2. A vast array of microexperiences is not revealed by reflection on macrophenomenal concepts (i.e., phenomenal concepts based on macroexperiences).
3. The nature of a phenomenal property is revealed by reflection on phenomenal concepts based on experiences of it. (RT5)
4. Whatever constitutes something is part of its nature.
5. The natures of macroexperiences do not involve vast arrays of microexperiences. (from 2 and 3)
6. Macroexperiences are not constituted by vast arrays of microexperiences. (from 4 and 5)
7. Constitutive panpsychism is false. (from 1 and 6)

Clearly, the soundness of this argument depends crucially on what is meant by talk of a property’s “nature”, since that will affect the meaning of premises 3 and 4; this question will be at the heart of my discussion in the next section.

2237 The third (“microexperience-focused”) revelation argument runs thus, with  
2238 a conjunction of RT<sub>1</sub> and RT<sub>3</sub> as its third premise:

- 2239 1. If constitutive panpsychism is true, consciousness is constituted by a  
2240 vast array of microexperiences.
- 2241 2. We cannot know introspectively about microexperiences, nor form  
2242 microphenomenal concepts (i.e., phenomenal concepts based on mi-  
2243 croexperiences).
- 2244 3. If a subject is having an experience, they can know introspectively that  
2245 they are, and form phenomenal concepts based on it. (RT<sub>1</sub> and 3)
- 2246 4. If experiences constitute a subject’s consciousness, that subject under-  
2247 goes them.
- 2248 5. We are not undergoing a vast array of microexperiences. (from 2 and 3)
- 2249 6. Human consciousness is not constituted by a vast array of microexperi-  
2250 ences. (from 4 and 5)
- 2251 7. Constitutive panpsychism is false. (from 1 and 6)

2252 Finally, there is an interaction between a revelation thesis, specifically RT<sub>5</sub>,  
2253 and another aspect of the combination problem, the “palette problem”. How  
2254 do the huge range of phenomenal qualities that humans experience arise from  
2255 a fundamental base which appears to involve only a quite small number of  
2256 fundamental properties? One solution is the “small palette hypothesis”: there  
2257 are only a few basic phenomenal qualities, corresponding to the fundamental  
2258 physical properties, which are somehow “blended” to generate a plethora  
2259 of different qualities for different macroscopic creatures (see [Roelofs 2014](#);  
2260 [Coleman 2015, 2017](#); [Chalmers 2017, 204–206](#)), whose pattern of similarities  
2261 and differences are explained by their differing proportions of the basic ingre-  
2262 dients. Some critics of the small palette hypothesis object that some of our  
2263 phenomenal qualities are too heterogeneous to be blended out of a small set  
2264 of common elements, because they are *completely* dissimilar, with nothing  
2265 phenomenally in common. Goff ([2017, 195](#)), for instance, claims that, “Minty  
2266 phenomenology and red phenomenology have nothing in common” (cf. a  
2267 similar argument in [McGinn 2006, 96](#)). This line of criticism relies on RT<sub>5</sub>  
2268 to rule out these qualities being similar in a way that we cannot recognise  
2269 ([Goff 2017, 195–197](#)). Call this the “small-palette revelation argument”, the  
2270 full structure of which is very similar to that of the macroexperience-focused  
2271 revelation argument:

- 2272 1. If the small palette hypothesis is true, then any two phenomenal quali-  
 2273 ties experienced by humans have something phenomenal in common.  
 2274 2. Reflection on some pairs of human experiences (e.g., red and minty)  
 2275 does not reveal them to have anything phenomenal in common.  
 2276 3. The nature of a phenomenal quality is revealed by reflection on phe-  
 2277 nomenal concepts based on experiences of it. (RT5)  
 2278 4. The natures of two things determine whether they have anything phe-  
 2279 nomenal in common.  
 2280 5. If a pair of phenomenal qualities has something phenomenal in com-  
 2281 mon, reflection on phenomenal concepts based on experiences of them  
 2282 will reveal this. (from 3 and 4)  
 2283 6. Some pairs of human experiences have nothing phenomenal in com-  
 2284 mon. (from 2 and 5)  
 2285 7. The small palette hypothesis is false. (from 1 and 6)

2286 All four arguments have a similar four-premise form: first, a supposed im-  
 2287 plication of constitutive panpsychism (or small-palette forms of it); second,  
 2288 an introspective datum; third, an epistemological thesis about introspection;  
 2289 and fourth, a metaphysical claim, given which the other three premises entail  
 2290 the falsity of constitutive panpsychism (or small-palette forms of it). But de-  
 2291 spite their common form, I will argue that the arguments go wrong in quite  
 2292 different ways.

#### 2294 **4 Ways of Responding to the Revelation Arguments**

2294 The challenge for constitutive panpsychists is to rebut the above four argu-  
 2295 ments without abandoning the revelation approach, components of which  
 2296 underpin all of them. I will show how to rebut each argument in turn, while  
 2297 keeping the relevant revelation theses as strong as I can.

##### 4.1 *The No-Illusions Revelation Argument*

2299 Consider first the “no illusions” argument, which had the following four  
 2300 premises:

- 2301 1. If constitutive panpsychism is true, then human consciousness is always  
 2302 “particulate”.  
 2303 2. Human consciousness (often) appears introspectively to be “smooth”.

- 2304 3. Consciousness can't appear a way that it's not.  
 2305 4. Being "smooth" and being "particulate" are incompatible.

2306 One option for constitutive panpsychists is to deny premise 1, based on defin-  
 2307 ing "particulate" in such a way that a field-based ontology, or a substance-  
 2308 monist ontology, or some other account of physical reality, renders it false that  
 2309 the material world, and any consciousness isomorphic to it, is particulate (see  
 2310 in particular Nagasawa and Wager 2017, 120–121). If the other three premises  
 2311 (and constitutive panpsychism) are accepted, this implies that the kind of  
 2312 consciousness we enjoy is incompatible with some physical theories (those  
 2313 which make matter "particulate") and that we know introspectively that our  
 2314 world is not any of those ways.

2315 However, I think this approach is a mistake. Even if particles are not ulti-  
 2316 mately real, Lockwood's point still holds: even the simplest experience in-  
 2317 volves billions of neurones, ions, and neurotransmitters. Even if the space  
 2318 containing two sodium ions is ultimately just a set of derivative aspects of the  
 2319 one substance, there is still a striking difference in the electrical properties of  
 2320 different regions of that space. To dismiss the problem because particles are  
 2321 not in the fundamental ontology would be too easy. Consequently, I suggest  
 2322 the following definition of "particulate":

2323 *X is particulate* iff X comprises a very large but finite number of parts  
 2324 which differ significantly (in some properties) and discontinuously  
 2325 (on some dimension).

2326 This definition makes the physical brain particulate whatever the fundamental  
 2327 physics turns out to be. Of course this definition will only be as precise as  
 2328 "very large" and "differ significantly and discontinuously". The vagueness of  
 2329 such terms does not stop us from taking "a trillion or more" as a clear case of  
 2330 "very large", and "the mass and charges differences between a water molecule,  
 2331 a potassium ion, and a region of empty space between them" as a clear case  
 2332 of "differ significantly and discontinuously".<sup>9</sup>

---

9 Note also that the definition requires only that the properties of the parts vary discontinuously in *some* dimension, i.e., on some natural way of ordering them, not on all: intuitively, the salient facts about brain parts like potassium ions are things like the abrupt drop in mass from inside the ion's nucleus to outside it, but this abrupt drop might vanish if we instead consider all parts of the brain in a list ordered by mass. But if we want to define "particulate" in a way that does justice to the no-illusions argument, the possibility of finding some dimension on which all variation is continuous should not disqualify the brain from being particulate.

2333 That leaves three remaining options: deny premise 2 (i.e., contradict the  
 2334 supposed introspective observation), deny premise 3 (i.e., reject this particular  
 2335 revelation thesis), or deny premise 4 (i.e., deny that smoothness and particu-  
 2336 larity are incompatible). But everything depends on what “smooth” means.  
 2337 What is the feature of experience that is being reported by those who feel the  
 2338 pull of this argument?

2339 One option is to define “smooth” by ostension: consider some experiences  
 2340 without discernible internal structure, what Lockwood (1993, 274) calls a  
 2341 “phenomenally flawless” experience, and stipulate that “smooth” means the  
 2342 noteworthy feature of those experiences. That would ensure the truth of  
 2343 premise 2, but would make it hard to adjudicate the truth of premise 4. My  
 2344 preference is to define “smooth” in such a way as to ensure the truth of premise  
 2345 4, e.g:

2346 X is *smooth* iff it is not particulate.

2347 There are then a few different ways for something to be smooth: since being  
 2348 particular requires parts, for instance, simple things would count as smooth  
 2349 by default. Alternatively, something might be smooth if its parts do not differ  
 2350 significantly in any respect, or do not differ discontinuously along any dimen-  
 2351 sion. The panpsychist must then deny either premise 2 or premise 3: either  
 2352 say that experience does not appear smooth, or say that it does but isn't.<sup>10</sup>  
 2353 At first glance, both options look difficult: premise 3 is, after all, part of the  
 2354 Revelation Approach (RT<sub>2</sub>), and if premise 2 is false, why did anyone ever  
 2355 advance the argument in the first place?

2356 The way out lies in scrutinising the word “appears”, and drawing a dis-  
 2357 tinction between illusions, strictly so-called, and easy misinterpretations.  
 2358 Consider some non-mental examples: at first an act appears noble, an argu-  
 2359 ment compelling, a speech beautiful, and yet then I find that upon giving  
 2360 the matter more thought, this appearance vanishes, and I come to think I  
 2361 was mistaken. The act now appears fanatical, the argument sophistical, the  
 2362 speech saccharine; I think myself foolish for being gullible enough for the act,  
 2363 argument, or speech to ever appear otherwise to me. I might say I was subject  
 2364 to an “illusion”, but all this mean is that the act, argument, and speech were  
 2365 such that they could be very readily misjudged.

---

10 Using the ostensive definition would just translate denial of premise 2 into denial of premise 4: either way, the claim is that there is no property incompatible with particularity that consciousness introspectively seems to have.



2366 Contrast this with a white object seen under pure red light, or a straight  
2367 stick seen half in water, or an ambitious Scottish nobleman hallucinating a  
2368 dagger. The object appears red but isn't, the stick appears bent but isn't, and  
2369 there appears to be a dagger, but there isn't. Here no reflection on the appear-  
2370 ances will change them, and the subject cannot hold themselves rationally  
2371 accountable for being subject to them (perhaps for forming beliefs based on  
2372 them, but not for the appearances themselves). Here we have a stronger sense  
2373 of "illusion": it is not that these perceptions are easy to misjudge, it is that  
2374 their very content is false. Call this the "quasi-perceptual" sense of "appears",  
2375 contrasting with the "ready-interpretation" sense (cf. [Stoljar 2013](#); [Kammerer](#)  
2376 [2018](#)).

2377 Premise 3 (RT<sub>2</sub>) is most plausible if read with the "quasi-perceptual" sense  
2378 of "appears". Plausibly it makes no sense to think that my impression of my  
2379 own experience is an "illusion" in this stronger sense: surely it would be the  
2380 "impression" that deserves to be called my experience, since this is what I am  
2381 immediately aware of. To think that consciousness might appear falsely in this  
2382 way seems to involve forgetting that consciousness is how things appear to  
2383 me (cf. [Liu 2020](#)). Or at least, this thought has some appeal, and panpsychists  
2384 need not disagree with it.

2385 But premise 3 is less plausible if understood in terms of the "ready-  
2386 interpretation" sense of "appears", saying that if consciousness is readily  
2387 interpreted as having some property, it must actually have that property. After  
2388 all, which interpretations come readily depends on the subject's expectations,  
2389 background assumptions, interpretive style, etc. An absolute principle, that  
2390 no false interpretation could come readily to *anyone*, would be very close to  
2391 saying, implausibly, that consciousness was never misinterpreted.

2392 So we should read premise 3 as saying that consciousness cannot appear a  
2393 way it's not, in the quasi-perceptual sense of "appear". For the argument to  
2394 remain valid, premise 2 must also be read in terms of the quasi-perceptual  
2395 sense of "appear", not the "ready-interpretation" sense. But now premise 2 is  
2396 much more deniable. We can deny premise 2, in this strong sense, by taking  
2397 the appearance of smoothness to be a matter of what interpretations come  
2398 readily, and not of how things quasi-perceptually appear.

2399 This is my preferred response to the "no illusions" argument: our consci-  
2400 ousness really is particulate, not smooth, but it is readily misinterpreted as  
2401 smooth. But this misinterpretation demands an explanation - what is it about  
2402 the way consciousness *does* appear, which makes us judge it "smooth"?

2403 One answer appeals to the difference between represented structure and  
 2404 structured representations: that is, experience represents things as being  
 2405 smooth, rather than itself being smooth (versions of this proposal appear in:  
 2406 Clark 1989; Stoljar 2001). Critics have worried that experience itself really  
 2407 does seem to display the relevant sort of smoothness (e.g., Alter and Nagasawa  
 2408 2012, 91), and that representing a smooth expanse may be insufficient for  
 2409 introspectively seeming, even in the weak sense, to be smooth (consider the  
 2410 sentence “space is infinitely divisible”). Another answer is to say that many  
 2411 experiences quasi-perceptually appear to have, and thus (by RT<sub>2</sub>) actually  
 2412 have, some property similar to, but not identical to, “smoothness”. In section  
 2413 5 I flesh out this approach.

#### 4v2 *The Macroexperience-Focused Revelation Argument*

2415 Next, consider the macroexperience-focused argument, whose premises are:

- 2416 1. If constitutive panpsychism is true, each human experience (“macroex-  
 2417 perience”) is constituted by a vast array of microexperiences.
- 2418 2. A vast array of microexperiences is not revealed by reflection on  
 2419 macrophenomenal concepts (i.e., phenomenal concepts based on  
 2420 macroexperiences).
- 2421 3. The nature of a phenomenal property is revealed by reflection on phe-  
 2422 nomenal concepts based on experiences of it.
- 2423 4. Whatever constitutes something is part of its nature.

2424 I see little prospect for denying premises 1 and 2,<sup>11</sup> and premise 3 is one of the  
 2425 revelation theses I want to preserve. Chalmers, when he lays out the argument  
 2426 of which this is a variant, advises panpsychists to attack premise 4: to drive a  
 2427 wedge between something’s nature and what constitutes it. I agree that this is  
 2428 the right tack, but everything turns on what kind of “nature” is in question,  
 2429 which in turn depends on how we read premise 3, the self-intimation thesis.  
 2430 I think there is a plausible and well-motivated sense of “knowing a nature”

---

11 It might look like cosmopsychists can wriggle out of premise 1. But this is illusory: the only way cosmopsychists can deny premise 1 is to commit to an analogous premise that supports a *harder* revelation argument. If they deny that the brain is constituted by neurons, ions, etc., they must instead accept a replacement premise 1\*: “If constitutive panpsychism is true, each human experience (‘macroexperience’) constitutes a vast array of microexperiences.” We then run the same argument, with premise 4 replaced by 4\*: “Whatever something constitutes is part of its nature.” And I think premise 4\* is noticeably *more* plausible than premise 4.

2431 which explains why premise 4 is false, without undermining anti-physicalist  
 2432 arguments.<sup>12</sup>

2433 First consider this common gloss: knowing the nature of a property means  
 2434 being in a position to know a priori every necessary truth about that prop-  
 2435 erty.<sup>13</sup> If I know the nature of squareness, I am in a position to know a priori  
 2436 every necessary truth about squareness (like what squares' internal angles  
 2437 sum to, or what kinds of triangles they can be divided into), though not to  
 2438 know contingent truths about it (like whether it is my sister's favourite shape).  
 2439 Likewise if I know the nature of being water, I can know every necessary truth  
 2440 about being water (like that water is a chemical compound, or its molecular  
 2441 mass), though not every contingent truth about it (like whether it is instan-  
 2442 tiated on Earth). This suggests that we know the natures of mathematical  
 2443 properties, but do not automatically know the natures of chemical properties,  
 2444 though perhaps we do now, given scientific progress. And those results seem  
 2445 plausible.

2446 But this gloss is inadequate. Consider someone who knew the nature of  
 2447 squareness but not the nature of triangularity (if that were possible). They  
 2448 would not be in a position to know a priori that every square can be divided into  
 2449 four right-angled triangles. This suggests a refinement: knowing the nature  
 2450 of some property means being in a position to know *a priori* all the necessary  
 2451 truths about that property which involve only other properties whose natures  
 2452 you also know. To put it another way, to know a priori a necessary truth  
 2453 involving two properties, you need to know the natures of both: just knowing  
 2454 the nature of one is not enough.<sup>14</sup> This implies, in particular, that knowing the  
 2455 nature of a constituted property is not sufficient to know about its constitution  
 2456 relationships to other properties, without also knowing the natures of those  
 2457 other properties.

- 
- 12 The argument discussed in Lee (2019) combines premises 3 and 4 into a single claim, "Structure Luminosity: If a subject introspects an experience, then that subject is in a position to know the phenomenal realizers of that experience" (2019, 292). Lee argues (in my view plausibly) that this is false, but does not clearly identify which elements of it remain true, and whether they are enough for anti-physicalist arguments.
- 13 I am abstracting away from difficulties of memory, attention, and general cognitive skills: in practice, many necessary truths might be just too complicated or subtle for a human mind to entertain, but that should not stop us from saying that someone is in a position to know them if all they would need to do so is an enhancement of their general cognitive skills.
- 14 This is not a retreat from the idea that the phenomenal property's "whole nature" is revealed. There is no part of its nature that is hidden: there are only hidden connections between its nature and other natures, and those connections are hidden for the simple reason that those other natures are hidden.

2458 I think this provides a plausible reading of “knowing a property’s nature”,  
 2459 and thereby of RT<sub>5</sub>, which does precisely what constitutive panpsychists need  
 2460 it to do: substantiate their arguments against physicalism, without substanti-  
 2461 ating the revelation argument against their own view. For on this reading  
 2462 of “knowing a nature”, that we know the natures of macrophenomenal prop-  
 2463 erties implies that for any other set of properties whose natures we know,  
 2464 we are in a position to tell a priori whether those properties are sufficient to  
 2465 constitute macrophenomenal properties. And the case against physicalism is  
 2466 that physical properties do not seem a priori to constitute macrophenomenal  
 2467 properties. Of course, this attack only works if we know the natures of phys-  
 2468 ical properties (e.g., if we think of them as exhausted by what physics says  
 2469 about them, as what Stoljar (2001) calls the “t-physical” properties, and what  
 2470 Strawson (2006) calls “physical” properties). It will not work if we think of  
 2471 physical properties as whatever properties physical things have which in fact  
 2472 account for their satisfying the descriptions given by physics (what Stoljar  
 2473 (2001) calls the “o-physical” properties). But that way out is no use to standard  
 2474 physicalism, which needs physical properties to be well-understood: to say  
 2475 that the reason the conceivability argument fails is that there is some mysteri-  
 2476 ous hidden nature of the physical, which plays some crucial role in accounting  
 2477 for consciousness, is to embrace the kind of “non-standard physicalism” (cf.  
 2478 [Stoljar 2006](#)) that is no longer incompatible with panpsychism.

2479 But why doesn’t knowing the natures of macrophenomenal properties  
 2480 substantiate a parallel argument against constitutive panpsychism? Because  
 2481 panpsychists do not claim that we know the natures of microphenomenal  
 2482 properties, because we are not the microsubjects who instantiate those prop-  
 2483 erties (though see the next subsection for some complications of this claim).  
 2484 Without knowledge of the candidate constituting properties, we cannot de-  
 2485 termine a priori their suitability to constitute macrophenomenal properties.  
 2486 All the constitutive panpsychist is committed to is a conditional claim: *if*  
 2487 we were able to grasp the natures of microphenomenal properties, then we  
 2488 could, in principle, see a priori that, when properly arranged, they constitute  
 2489 macrophenomenal properties.

### 4<sup>93</sup> *The Microexperience-Focused Revelation Argument*

2491 Thirdly, consider the microexperience-focused revelation argument: why can’t  
 2492 we introspect microexperiences like we can macroexperiences? The premises  
 2493 of this argument are:

- 2494 1. If constitutive panpsychism is true, consciousness is constituted by a  
 2495 vast array of microexperiences.  
 2496 2. We cannot know introspectively about microexperiences, nor form  
 2497 microphenomenal concepts.  
 2498 3. If a subject is having an experience, they can know introspectively that  
 2499 they are, and form phenomenal concepts based on it.  
 2500 4. If experiences constitute a subject's consciousness, that subject under-  
 2501 goes them.

2502 Again, I see little hope in denying premises 1 or 2,<sup>15</sup> which leaves three options:  
 2503 deny premise 3 (“we *are* undergoing microexperiences, but cannot introspect  
 2504 them”), deny premise 4 (“microexperiences constitute our consciousness, but  
 2505 we do not undergo them”), or show the argument to be invalid.

2506 Goff's approach in his (2017, 189ff.) is to deny premise 4, to “loosen” the  
 2507 relation between microexperiences and macroexperiences, so that although  
 2508 microexperiences in some sense constitute (or “ground”, “compose”, or “form”)   
 2509 macroexperiences, the phenomenal character of the latter contains nothing  
 2510 of the former. The cost of this is that the constitution relation between mi-  
 2511 croexperiences and macroexperiences is thereby made more mysterious. If  
 2512 this relation were one in which both constituted and constituter were un-  
 2513 dergone by the same subject, it could be akin to familiar relations among  
 2514 macroexperiences. For instance, the relation between my total phenomenal  
 2515 field right now and the component experiences that it subsumes (sounds I'm  
 2516 hearing, colours I'm seeing, twinges of physical discomfort, etc.) is plausibly  
 2517 something like constitution. It would be nice if panpsychists could assimilate  
 2518 the microexperience-macroexperience relation to familiar relations like this,  
 2519 where a single subject undergoes all the experiences involved; without that  
 2520 link it is hard to see why microexperiences should really be said to “consti-  
 2521 tute” a macroexperience, as opposed to somehow giving rise to it as a distinct  
 2522 product.

2523 I think the best approach is to say the argument is invalid *when premise 3 is*  
 2524 *qualified* in certain ways that are independently necessary to make it plau-  
 2525 sible. An unqualified form of premise 3 faces easy counterexamples: ferrets

---

15 Again, though one might think cosmopsychists can deny premise 1, there is no advantage to be gained thereby: the replacement premise 1\* - “If constitutive panpsychism is true, human consciousness constitutes a vast array of microexperiences” - will support a revised version of the argument, when paired with 4\* - “If experiences are constituted by a subject's consciousness, that subject undergoes them.” And again, 4\* seems to me even more plausible than 4.

2526 undergo many experiences, but cannot form phenomenal concepts, or know  
 2527 that they are having experiences. But plausibly this is not a counter-example  
 2528 to what premise 3 was intended to say! The problem is not that ferrets' experi-  
 2529 ences are somehow hidden from them, but just that they lack the conceptual  
 2530 competence to recognise their experiences as such. A qualified version of  
 2531 premise 3 would allow for this: it would say that certain kinds of knowledge  
 2532 and concept-formation are possible whenever a subject undergoes an experi-  
 2533 ence *and* meets various other conditions. Another plausible requirement  
 2534 is attention: one must focus on an experience in order to introspect it, and if  
 2535 one is unable to direct one's attention, introspection will be impossible.<sup>16</sup>

2536 So let us consider a qualified reading of premise 3, that includes these  
 2537 conditions: introspective knowledge is possible whenever a subject undergoes  
 2538 an experience, *and* is capable of conceptualising it, *and* focuses their attention  
 2539 on it. The argument has now become invalid: line 5 (“we are not undergoing a  
 2540 vast array of microexperiences”) no longer follows from 2 and 3. There are two  
 2541 reasons why we might be phenomenally undergoing microexperiences but be  
 2542 unable to know them introspectively, compatibly with this weaker reading of  
 2543 premise 3: if humans cannot conceive of experiences as such, or if they are  
 2544 unable to attend to microexperiences. While the first of these options is clearly  
 2545 false, the second is, I think, the best option for the constitutive panpsychist in  
 2546 rebutting the microexperience-focused argument.

2547 This implies that while microexperiences are phenomenally conscious for  
 2548 us, they are not access-conscious for us. That is, microexperiences are pre-  
 2549 sented to us, “right there”, characterising the phenomenal character of our  
 2550 consciousness, but they are not presented in such a way that we can cog-  
 2551 nitively select, access, and identify them. Our relationship to them is rather  
 2552 like our relationship to elements of our experience that are very faint, which  
 2553 require a lot of effort to focus on and distinguish from their surroundings,  
 2554 and which it is correspondingly easier to distract us from. If something in  
 2555 my peripheral vision is roughly the same colour as its surroundings, it would  
 2556 be hard for me to notice it, and if I were distracted, exhausted, or inebriated  
 2557 I might find attending to it all but impossible. Yet it is still part of my phe-  
 2558 nomenology, not somehow hidden from me. The constitutive panpsychist,  
 2559 I am suggesting, should claim that this near-impossibility of attending to  
 2560 peripheral vision while distracted is intensified to a real practical impossibility

16 Goff's statement of revelation (2017, 109–110) mentions attention explicitly, and Chalmers appeals to inattention as a primary reason for thinking that his principles of “detectability” and “reliability” can only hold for the most part, not absolutely (Chalmers 1995, 326; 1996, 218–219).

2561 with microexperiences. In section 5 I situate this impossibility claim within a  
 2562 broader picture of how the mind is constituted by microexperiences, which  
 2563 will help to motivate this response to the microexperience-focused argument.

#### 4.4 The Small-Palette Revelation Argument

2565 Finally, consider the small-palette revelation argument, whose premises are:

- 2566 1. If the small palette hypothesis is true, then any two phenomenal quali-  
 2567 ties experienced by humans have something phenomenal in common.
- 2568 2. Reflection on some pairs of human experiences does not reveal them to  
 2569 have anything phenomenal in common.
- 2570 3. The nature of a phenomenal quality is revealed by reflection on phe-  
 2571 nomenal concepts based on experiences of it. (RT5)
- 2572 4. The natures of two things determine whether they have anything phe-  
 2573 nomenal in common.

2574 Since this is not an argument against constitutive panpsychism per se, there  
 2575 are technically five options for constitutive panpsychists in responding to it:  
 2576 deny one of the premises, or accept the conclusion. Accepting the conclusion  
 2577 would mean accepting a “large palette” version of constitutive panpsychism,  
 2578 with all human and animal qualities present in the base even though that is  
 2579 more than there are distinct physical roles to play (see, e.g., [Lewtas 2013](#)). The  
 2580 downside is that this sacrifices the appealing parsimony, and isomorphism  
 2581 with physics, that had seemed to set constitutive panpsychism apart from  
 2582 traditional sorts of dualism. Denying premise 3 is also unattractive, since it  
 2583 undermines the case for panpsychism over physicalism.

2584 Denying premise 4 here (as Lee does, [2019, 300–301](#)) is harder than denying  
 2585 premise 4 of the macroexperience-focused argument, that “what constitutes  
 2586 something is part of its nature”. I denied the latter because knowing a prop-  
 2587 erty’s nature is not enough to know necessary truths about it which involve  
 2588 the nature of another property; we would have to know that other property’s  
 2589 nature as well. But when it comes to comparing two qualities that we do  
 2590 experience distinctly, it seems to follow that we should be able, in principle,  
 2591 to discern every necessary truth about how those qualities relate, and that  
 2592 should include their resemblance or common constituents.<sup>17</sup>

---

17 Could we find a more carefully qualified version of RT5, on which knowing the natures of two properties enables us to know whether one suffices to constitute the other, but not whether and

2593 We might deny premise 4 in the same way we might deny premise 4 of the  
 2594 microexperience-focused argument, by saying that although the basic qualities  
 2595 constitute the macroqualities, they do not characterise them - the “blending”  
 2596 leaves no trace of the ingredients at all. But this has the same downsides  
 2597 discussed in the last subsection: if microqualities in no way characterise the  
 2598 macroqualities, the form of constitution involved seems mysterious.

2599 That leaves denying premise 1 or premise 2. Premise 1 might seem un-  
 2600 deniable, due to the “interchangeability” of different neurones: experiences  
 2601 of redness and of mintiness involve neurones made of all the same sorts of  
 2602 subatomic particles, so how can one contain any ingredient missing from  
 2603 the other? Any ingredient of the redness experience comes from electrons,  
 2604 quarks, photons, etc., and those same things are all present in the physical  
 2605 basis of a mintiness experience, so how could they not show up in the latter?  
 2606 But this falsely assumes that each macroexperience should contain every in-  
 2607 gredient present in its neural basis, as though each one were the independent  
 2608 product of one discrete subset of neurones. It might instead be that several  
 2609 macroexperiences are all grounded in the activity of the same neurones, being  
 2610 just different aspects of the complex, differentiated experience produced by  
 2611 those neurones.

2612 Consider a bar magnet, whose macroscopic behaviour displays a “north  
 2613 pole” and “south pole”. The north pole does not arise from one half of the  
 2614 magnet, and the south pole from the other half: both macroscopic features  
 2615 arise from very same microscopic physical things, because those things are  
 2616 themselves internally differentiated and their different aspects add up to what  
 2617 looks, from a macroscopic perspective, like two different things. It would be  
 2618 a mistake to say “since all the particles generating the magnet’s north pole  
 2619 also have south poles, why don’t their south poles show up in the magnet’s  
 2620 north pole?” Perhaps mintiness and redness are likewise different aspects of  
 2621 the same complex experience, itself arising from the combination of a great  
 2622 many internally differentiated microexperiences, combining in different ways  
 2623 depending on such things as firing rates and degrees of neural synchrony. Then  
 2624 they might have nothing phenomenal in common, despite being constituted  
 2625 by the same things.

2626 However, there are limitations to this response. It might allow for a few  
 2627 completely dissimilar pairs to be compatible with the SPH, but not that many -

---

how they resemble each other? Maybe, but this feels ad hoc to me; I see no plausible way to motivate it.



2628 if there are a hundred completely dissimilar qualities experienced by humans,  
2629 saying that they arise from the way that internally differentiated aspects of  
2630 microexperiences are combined starts to load microexperiences with too much  
2631 structure for us to retain the SPH. To keep the palette small, there shouldn't be  
2632 too many completely dissimilar pairs of qualities, which is why this response  
2633 to the argument works best when combined with another: denying premise 2.

2634 Denying premise 2 means denying that redness and mintiness have abso-  
2635 lutely nothing at all in common. After all, our ability to recognise two things  
2636 as akin to one another is usually enhanced by our ability to recognise and  
2637 attend to the features they share, and if we never experience their shared fea-  
2638 tures in isolation, we may take them to be entirely unlike even if they are not.  
2639 Sometimes, of course, two qualities seem inarticulately alike even without an  
2640 identifiable shared feature; this is why we routinely describe qualities of one  
2641 modality using terms drawn from another (warm, harsh, sweet, soft, loud,  
2642 etc.). The SPH and RT5 can both be retained as long as idealised scrutiny of  
2643 these inchoate likenesses would reveal a system of qualitative connections  
2644 over our entire experiential range. This view is defended by Coleman:

2645 [...] just as it's possible to move across the colour spectrum in tiny,  
2646 almost undetectable steps, it must be possible to move from tastes  
2647 to sounds, sounds to colors, and so on, via equally tiny steps. Tip-  
2648 toeing between modalities already seems *conceivable* in certain  
2649 cases, perhaps even actual. We know that what we experience as  
2650 "taste" is really some kind of fusion of qualia sourced from the  
2651 nose and from the tongue [...]. To address qualitative incommen-  
2652 surability we must stretch to conceiving of such continuities as  
2653 the rule rather than the exception. (Coleman 2017, 264, emphasis  
2654 in original; cf. Coleman 2015; Hartshorne 1934, 35ff.)

2655 This claim does not seem to me obviously false, but it is at least dubitable.  
2656 Consequently, the revelation approach may be most threatening to consti-  
2657 tutive panpsychists not through any of the three pure revelation arguments,  
2658 but through intensifying the palette problem. Accepting revelation pushes  
2659 constitutive panpsychists towards either a large-palette solution like Lew-  
2660 tas's, or towards Coleman's very bold and ambitious form of the small-palette  
2661 hypothesis.

## 2665 5 Confusion and Revelation

2663 Identifying a premise of an argument that might be false is often not, by  
 2664 itself, an effective way to persuade critics. For all that I have said so far, this  
 2665 “medium-strength” version of revelation, interpreted and qualified so as to  
 2666 undermine arguments against panpsychism while substantiating arguments  
 2667 against physicalism, might be technically consistent but ad hoc and unmotivated,  
 2668 a dingy corner of logical space which panpsychists can awkwardly retreat to.  
 2669 But in fact, these qualified revelation theses flow from a reasonable picture of  
 2670 the limits of human self-knowledge, on which the introspective ignorance that  
 2671 constitutive panpsychism implies differs only in degree from familiar forms of  
 2672 introspective ignorance.

2673 It is commonplace to say that when two experiences become phenomenally  
 2674 unified, they form a composite experience which subsumes them: they still  
 2675 exist, and are still undergone by the subject, but they are now “undergone  
 2676 together”. We easily recognise this when we can discern introspectively not  
 2677 just the composite experience but also its components: but what if the discernibility  
 2678 of the component experiences is not an automatic consequence of the composite  
 2679 experience being composite? We might consider the idea that it depends instead  
 2680 on having the right structure of informational relations among the components.<sup>18</sup>  
 2681 Perhaps if these relations make the subject’s overall dynamics differentially  
 2682 sensitive to multiple distinct features of the experience, the composite experience  
 2683 will be characterised by contrast among those features: they will stand out as  
 2684 distinct things. If not, those features will be present in the composite experience  
 2685 in an undifferentiated way, as a single element whose phenomenal quality is a  
 2686 seamless blend of its components. In short: the component experiences all go  
 2687 in together, but the way they are present in the composite experience depends  
 2688 on how they are organised.

2689 What explains why experiences should compose in this way is a further  
 2690 question, which I cannot here address (though see [Roelofs 2016; 2019, 123–125, 166–170](#)).  
 2691 But suppose some conditional like this were true: when distinct experiences  
 2692 are unified, they can be distinguished by the subject only if they have the  
 2693 right informational structure. Although the human brain is an exquisitely  
 2694 structured processor of information, it has limits. The overall dynamics of the  
 2695 brain can perhaps be sensitive to whether a neurone fires, but not (as far as  
 2696 we know) to which ions in that neurone played which roles in its firing. Since

18 This is a long-standing idea among panpsychists, though spelling it out in detail is not simple. See [Chalmers \(1996\)](#), 284–292; [Chalmers \(2017\)](#), 209–210; [Gabora \(2002\)](#); [Roelofs \(2019\)](#), 171–176.

2697 individual events at the microscopic level are informationally inaccessible,  
2698 they will be experienced by the whole in a blended way. They each make a  
2699 minute difference to the quality of some element of the whole's experience,  
2700 but they do not stand out as distinct elements of it. To use a term made famous  
2701 by Leibniz, they are "confused" with one another, the way that the sounds of  
2702 each bit of water striking the shore are "confused" in the roar of the sea.<sup>19</sup>

2703 I have elsewhere elaborated more fully on the idea of confusion as I under-  
2704 stand it (2019, 126–129), but the essential idea is captured in the following  
2705 definition:

2706 Two experiences are *confused* with each other, relative to a subject,  
2707 iff that subject cannot distinguish them by attending to one without  
2708 simultaneously attending to the other.<sup>20</sup>

2709 It is important to emphasise that confusion is not a matter of a subject "per-  
2710 ceiving" things outside themselves so poorly that they cannot distinguish the  
2711 parts of that outside thing. Confusion is a matter of how the subject's own  
2712 states are related, not a relation between them and something external. For  
2713 example, someone viewing a pointillist painting, for whom the many dots  
2714 of paint "blur together", is not thereby subject to confusion, if they simply  
2715 have a single experience that is the product of many external objects. A better  
2716 example would be someone with an untrained palate, who drinks coffee and  
2717 experiences (let us stipulate) the same diversity of taste and flavour experi-  
2718 ences as a practiced connoisseur but experiences them together as a single  
2719 blended flavour, without being able to pick out the bitterness from the aroma,  
2720 etc.

2721 Confusion may depend on circumstances. When we are tired, distracted, or  
2722 drunk we often cannot distinguish things which we could under better condi-  
2723 tions. Then our experiences are confused only relative to those circumstances.  
2724 Confusion can also depend on a subject's conceptual repertoire: sometimes  
2725 we cannot distinguish two things using their present concepts, but would

---

19 This idea of the mind as comprising a vast number of "little perceptions", most of which cannot be distinguished from one another by the subject, is arguably present in several early modern writers as well as Leibniz, in particular Spinoza, Wolff, and Kant. For discussion see Wilson (1980), Thiel (2011), Liang (2017), and Indregard (2018). To use a more modern phrasing from Andrew Lee (2019), they make up the non-introspectible "microstructure of experience".

20 In the primary instance confusion is defined over tokens, but we can easily define a secondary sense in which two types are confused for a subject when any token of those types onto which a given subject could direct a given operation would be confused with a token of the other type.

2726 be able to if we learnt new ones. Call confusion which can be removed by  
 2727 adjusting the subject's bodily surroundings or condition, or improving their  
 2728 conceptual repertoire, or in some similarly mild way, "shallow confusion", and  
 2729 call confusion which persists even into ideal conditions, "robust confusion".

2730 In between shallow and robust is confusion which persists until the subject  
 2731 becomes distinctly acquainted with a token of the same type as the confused  
 2732 elements. For example, suppose the sensory component of pain is robustly  
 2733 confused with the unpleasant affect pain involves, except for subjects who  
 2734 have experienced "pain asymbolia", the rare condition of feeling pain with-  
 2735 out finding it at all unpleasant (cf. [Grahek 2007](#); [Klein 2015](#)). If they regain  
 2736 normal pain experiences, they might find themselves newly able to attend  
 2737 to its sensory element in isolation. If this were to happen, we might say that  
 2738 their original confusion was "nearly-robust": removable only by somehow  
 2739 acquainting them with (a token of the same type as) one of the confused  
 2740 elements on its own.<sup>21</sup>

2741 When confusion is shallow, we have an easy way to tell that we suffer  
 2742 from it: we remove it and contrast the resulting distinction with the earlier  
 2743 confusion. With sufficiently robust confusion, we would not have such means  
 2744 of recognising it; we could not tell that we were confused. And if we suffered  
 2745 from confusion that was "nearly-robust", it would be undetectable, except  
 2746 by means of independent acquaintance with elements of the same type as  
 2747 the confused ones. We could, that is, be subject to a lot of confusion without  
 2748 being able to tell, introspectively. And if constitutive panpsychism is true - in  
 2749 particular, if micro-experiences corresponding to all the physical details of our  
 2750 brains were somehow present in our consciousness - then we should expect  
 2751 just that: all the experiences of our microparts would be confused relative to  
 2752 us. Call this the Radical Confusion Hypothesis.

2753 Confusion is defined functionally, but that does not imply that confusion is  
 2754 a purely functional fact that makes no phenomenal difference. My suggestion  
 2755 is that undergoing two confused experiences feels different to undergoing  
 2756 two distinguishable experiences, even if those experiences are the same in all  
 2757 intrinsic respects. When the components of an experience are distinguishable  
 2758 by the subject, they are phenomenally present as discernible, separate, parts -  
 2759 there is an experience of phenomenal contrast, of things standing out against

---

21 In other work ([2019, 128–129](#)), I also distinguish between "strong" and "weak", and "symmetrical" and "asymmetrical" confusion, but this does not substantially affect the argument so I omit it here for simplicity.

2760 other things. But when they are confused, they are present qualitatively, as  
2761 contributions to the total quality of the experience they blend into.

2762 How would the Radical Confusion Hypothesis help with the four revelation  
2763 arguments? Recall that in response to the “no illusions” argument, I denied  
2764 premise 2: that human consciousness positively appears introspectively to be  
2765 “smooth” (there defined as “not particulate”). I maintained that this is false if  
2766 “appears introspectively” is read in a strong, quasi-perceptual sense; it is true  
2767 only if “appears introspectively” is read in a weaker sense, as meaning “it is  
2768 easy and natural to interpret experience this way”.

2769 Now I can say *why* this misinterpretation is easy and natural: because many  
2770 human experiences display something close to “smoothness”, namely, all  
2771 their component experiences are nearly-robustly confused with each other,  
2772 distinguishable only by a subject who already knows what to look for. A  
2773 subject who lacks any distinct acquaintance with the ingredients will be  
2774 unable to distinguish them or discern their internal structure. We might say  
2775 that experiences all of whose components are confused with one another are  
2776 “pseudo-smooth”, and it is true (and introspectively obvious!) that many of  
2777 our experiences are pseudo-smooth. But to infer genuine smoothness from  
2778 pseudo-smoothness is a metaphysical over-interpretation which goes beyond  
2779 the introspective deliverances: it is inferring absence of structure from the  
2780 failure of structure to be manifest in a certain way (it is thus very similar to  
2781 the “headless woman illusion” discussed by Armstrong (1968), where not  
2782 seeing someone’s head gives us the vivid but false impression that they have  
2783 no head). The noticeable quality that some experiences have, which prompted  
2784 the “no illusions” argument, is just what radical confusion feels like.

2785 Second, in response to the macroexperience-focused argument I denied  
2786 premise 4, that whatever constitutes something is part of the “nature” that  
2787 is revealed to us by pure phenomenal concepts. I suggested that a priori  
2788 reflection tells us only those necessary truths that involve *only* properties  
2789 whose nature we know - such as whether one could constitute the other. But  
2790 just knowing the nature of one property does not tell all the things that could  
2791 constitute it, nor what constitutes a particular instance of it.

2792 I can now elaborate on this distancing of constitution from “nature”.  
2793 Macroexperiences are composite experiences composed of many microexperi-  
2794 ences confused with one another. Their phenomenal character is determined  
2795 by combining the phenomenal characters of those component experiences,  
2796 which they subsume in fundamentally the same way that a person’s total  
2797 experience at any one time subsumes the partial experiences they are having

2798 at that time. But just as two composites might end up sharing certain  
 2799 properties despite being constituted by different sets of parts, and despite  
 2800 their properties being mere combinations of the properties of their parts, two  
 2801 composite experiences might have the same phenomenal character, despite  
 2802 being constituted by different sets of microexperiences. The particular parts  
 2803 might be essential to the particular macroexperience, but not to the property  
 2804 that it is an instance of.

2805 I also said, in response to the small-palette revelation argument, that dis-  
 2806 tinct macroexperiences might arise from the same neural basis: we need not  
 2807 assume that each distinguishable element of our consciousness contains the  
 2808 entire phenomenal nature of one discrete subset of physical entities. The  
 2809 radical confusion hypothesis reinforces this point: it says that which experi-  
 2810 ences phenomenally contrast or phenomenally blend with one another  
 2811 in human experience reflects the informational structure of the brain, so a  
 2812 single macroexperience might not correspond to any discrete section of the  
 2813 underlying physical substrate. Instead, it will correspond to a set of features  
 2814 of the substrate such that information about them collectively is extracted  
 2815 and used by the brain, but information about them individually is not. Thus  
 2816 different macroexperiences based in the same brain area might have different,  
 2817 even non-overlapping sets of phenomenal ingredients, because they reflect  
 2818 different features of the same microexperiences.

2819 Finally, in response to the microexperience-focused argument I suggested  
 2820 that our ignorance of microexperiences is compatible with our undergoing  
 2821 them, if we cannot attend to them. Now I can add that our inability to attend  
 2822 to microexperiences is part-and-parcel of their being confused for us. Their  
 2823 radical confusion is explained by the limitations discussed above on how  
 2824 much information about microscopic brain events can be extracted by the rest  
 2825 of the brain.<sup>22</sup> Because radically confused experiences cannot be distinctly  
 2826 attended to, we cannot know them or their natures, even though the experi-  
 2827 ences “present themselves” in the sense that if their subject could attend to  
 2828 them they could know them and their natures by introspection.

2829 An opponent might object that even though attending to particular experi-  
 2830 ences can be harder or easier, depending on, e.g., architectural facts about the  
 2831 brain, it cannot be *strictly impossible* for me to attend to an experience, if it  
 2832 is really is an experience I am undergoing. I reply that distinctly attending

---

22 This allows for a limited sense in which microexperiences *are* accessible: namely that they can be accessed only by acts which are also accessing many other microexperiences at the same time. They cannot be *individually* accessed, but they can be accessed *collectively*.

2833 to microexperiences is *not* strictly impossible, just impossible in practice (as  
2834 discussed in Lee 2019, 296–297). They are manifest in our consciousness, but  
2835 incredibly difficult to pick out. After all, it is very difficult for the large-scale  
2836 dynamics of our brain to be sensitive to changes in a single particle, but there  
2837 is no in-principle impossibility in there being such sensitivity, perhaps using  
2838 advanced technology or strange altered states of consciousness.<sup>23</sup>

## 2839 6 Conclusions

2840 The idea of “revelation”, that having an experience provides a special insight  
2841 into its nature, is a key weapon in the armoury of anti-physicalists. But for  
2842 constitutive panpsychists there is a risk it will blow up in their faces. I have  
2843 argued, however, that a suitably-qualified form of the revelation approach can  
2844 bring down physicalism while leaving panpsychism standing: a form which  
2845 reconciles the profound fallibility of the human mind’s self-knowledge with  
2846 the perfect transparency of its access to its itself. Although nothing does or  
2847 could “conceal” our own experiences from us, we are nevertheless limited  
2848 in our ability to attend to their elements, prone to misinterpret them, and  
2849 consequently unable to tell introspectively just how composite they might  
2850 really be.\*

2851

2852

2853

2854

Luke Roelofs

 0000-0002-2424-8583

New York University

luke.mf.roelofs@gmail.com

---

23 Note that there need not be any sharp boundary between “the simplest experiential element that we can distinguish” and “the most complex experiential element that is radically confused.” For different people, under different conditions, different distinctions among one’s internal states and processes may be possible. Radically confused experiences are not a qualitatively distinct sort of experience from distinguishable ones, any more than “places I can walk to in ten minutes” are a sharply separate set of places from those I can walk to in ten minutes; my walking ability, like my introspective discernment, waxes and wanes as I change and as conditions change.

\* This paper expands on ideas presented over pages 132–137 of Roelofs (2019). Their further development owes a great deal to audiences at the Australian National University and the CEU’s workshop “Russellian Monism: Time for the Details” in Budapest.

## References

- 2855
- 2856 ALTER, Torin and NAGASAWA, Yujin. 2012. "What is Russellian Monism?" *Journal of*  
 2857 *Consciousness Studies* 19(9/10): 67–95. Reprinted in Alter and Nagasawa (2015,  
 2858 422–452).
- 2859 —, eds. 2015. *Consciousness in the Physical World: Perspectives on Russellian Monism*.  
 2860 Oxford: Oxford University Press.
- 2861 ARMSTRONG, David M. 1968. "The Headless Woman Illusion and the Defence of  
 2862 Materialism." *Analysis* 29(2): 48–49, doi:10.1093/analys/29.2.48.
- 2863 BALL, Derek. 2009. "There are no Phenomenal Concepts." *Mind* 118(472): 935–962,  
 2864 doi:10.1093/mind/fzp134.
- 2865 BRADDON-MITCHELL, David and NOLA, Robert, eds. 2009. *Conceptual Analysis and*  
 2866 *Philosophical Naturalism*. Cambridge, Massachusetts: The MIT Press, doi:10.755  
 2867 1/mitpress/9780262012560.001.0001.
- 2868 BYRNE, Alex and HILBERT, David R. 2007. "Color Primitivism." *Erkenntnis* 66(1/2):  
 2869 73–105, doi:10.1007/s10670-006-9028-8.
- 2870 CHALMERS, David J. 1995. "Absent Qualia, Fading Qualia, Dancing Qualia." in *Con-*  
 2871 *scious Experience*, edited by Thomas METZINGER, pp. 309–329. Paderborn: Ferdi-  
 2872 nand Schöningh.
- 2873 —. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford  
 2874 University Press.
- 2875 —. 2003a. "The Content and Epistemology of Phenomenal Belief." in *Consciousness:*  
 2876 *New Philosophical Perspectives*, edited by Aleksandar JOKIĆ and Quentin SMITH,  
 2877 pp. 220–272. Oxford: Oxford University Press, doi:10.1093/oso/9780199241286.00  
 2878 3.0009.
- 2879 —. 2003b. "Consciousness and Its Place in Nature." in *The Blackwell Guide to the*  
 2880 *Philosophy of Mind*, edited by Stephen P. STICH and Ted A. WARFIELD, pp. 102–142.  
 2881 Blackwell Philosophy Guides. Oxford: Blackwell Publishers, doi:10.1002/978047  
 2882 0998762.ch5.
- 2883 —. 2009. "The Two-Dimensional Argument Against Materialism." in *The Oxford*  
 2884 *Handbook of Philosophy of Mind*, edited by Brian P. MCLAUGHLIN, Ansgar BECK-  
 2885 ERMANN, and Sven WALTER, pp. 313–337. Oxford Handbooks. Oxford: Oxford  
 2886 University Press. Republished in expanded form, with an afterword, in Chalmers  
 2887 (2010, 141–191), doi:10.1093/acprof:oso/9780195311105.003.0006.
- 2888 —. 2010. *The Character of Consciousness*. Oxford: Oxford University Press, doi:10.109  
 2889 3/acprof:oso/9780195311105.001.0001.
- 2890 —. 2017. "The Combination Problem for Panpsychism." in *Panpsychism: Contemporary*  
 2891 *Perspectives*, edited by Godehard BRÜNTRUP and Ludwig JASKOLLA, pp. 179–214.  
 2892 Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199359943.003.0008.
- 2893 CLARK, Andy. 1989. "The Particulate Instantiation of Homogeneous Pink." *Synthese*  
 2894 80(2): 277–304, doi:10.1007/bf00869488.



- 2895 COLEMAN, Sam. 2015. "Neuro-Cosmology." in *Phenomenal Qualities: Sense, Perception,*  
2896 *and Consciousness*, edited by Paul COATES and Sam COLEMAN, pp. 66–102. Oxford:  
2897 Oxford University Press, doi:10.1093/acprof:oso/9780198712718.003.0003.
- 2898 —. 2017. "Panpsychism and Neutral Monism: How to Make up One's Mind." in *Panpsy-*  
2899 *chism: Contemporary Perspectives*, edited by Godehard BRÜNTRUP and Ludwig  
2900 JASKOLLA, pp. 249–282. Oxford: Oxford University Press, doi:10.1093/acprof:  
2901 oso/9780199359943.003.0011.
- 2902 DAMNJANOVIC, Nic. 2012. "Revelation and Physicalism." *Dialectica* 66(1): 69–91,  
2903 doi:10.1111/j.1746-8361.2012.01290.x.
- 2904 DENNETT, Daniel C. 2007. "What RoboMary Knows." in *Phenomenal Concepts and*  
2905 *Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited  
2906 by Torin ALTER and Sven WALTER, pp. 15–31. Oxford: Oxford University Press,  
2907 doi:10.1093/acprof:oso/9780195171655.003.0001.
- 2908 DÍAZ-LEÓN, Esa. 2011. "Reductive Explanation, Concepts, and A Priori Entailment."  
2909 *Philosophical Studies* 155(1): 99–116, doi:10.1007/s11098-010-9560-x.
- 2910 GABORA, Liane. 2002. "Amplifying Phenomenal Information: Toward a Fundamental  
2911 Theory of Consciousness." *Journal of Consciousness Studies* 9(8): 3–29.
- 2912 GAUDRY, Justin. 2008. "Does Physicalism Entail Cosmopsychism?" Unpublished  
2913 manuscript, [https://panexperientialism.blogspot.com/2008/05/does-physicalism-](https://panexperientialism.blogspot.com/2008/05/does-physicalism-entail-cosmopsychism.html)  
2914 [entail-cosmopsychism.html](https://panexperientialism.blogspot.com/2008/05/does-physicalism-entail-cosmopsychism.html).
- 2915 GOFF, Philip. 2006. "Experiences Don't Sum." *Journal of Consciousness Studies*  
2916 13(10/11): 53–61.
- 2917 —. 2015. "Real Acquaintance and Physicalism." in *Phenomenal Qualities: Sense, Percep-*  
2918 *tion, and Consciousness*, edited by Paul COATES and Sam COLEMAN, pp. 121–145.  
2919 Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198712718.003.0005.
- 2920 —. 2017. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press,  
2921 doi:10.1093/oso/9780190677015.001.0001.
- 2922 GRAHEK, Nikola. 2001. *Feeling Pain and Being in Pain*. Cambridge, Massachusetts:  
2923 The MIT Press. Second edition: Grahek (2007).
- 2924 —. 2007. *Feeling Pain and Being in Pain*. 2nd ed. Cambridge, Massachusetts: The MIT  
2925 Press. First edition: Grahek (2001), doi:10.7551/mitpress/2978.001.0001.
- 2926 HARTSHORNE, Charles. 1934. *The Philosophy and Psychology of Sensation*. Chicago,  
2927 Illinois: University of Chicago Press.
- 2928 INDREGARD, Jonas Jervell. 2018. "Consciousness as Inner Sensation: Crusius and  
2929 Kant." *Ergo* 5(7): 173–201, doi:10.3998/ergo.12405314.0005.007.
- 2930 JACKSON, Frank. 1982. "Ephiphenomenal Qualia." *The Philosophical Quarterly* 32(127):  
2931 127–136. Reprinted in Jackson (1998, 57–69), doi:10.2307/2960077.
- 2932 —. 1998. *Mind, Method, and Conditionals: Selected Papers*. London: Routledge, doi:10  
2933 .4324/9780203271308.
- 2934 JASKOLLA, Ludwig and BUCK, Alexander J. 2012. "Does Panexperiential Holism Solve  
2935 the Combination Problem?" *Journal of Consciousness Studies* 19(9/10): 190–199.

- 2936 JOHNSTON, Mark. 1992. "How to Speak of the Colors." *Philosophical Studies* 68(3):  
2937 221–263, doi:10.1007/bf00694847.
- 2938 KAMMERER, François. 2018. "Can You Believe It? Illusionism and the Illusion Meta-  
2939 Problem." *Philosophical Psychology* 31(1): 44–67, doi:10.1080/09515089.2017.1388  
2940 361.
- 2941 KLEIN, Colin. 2015. "What Pain Asymbolia Really Shows." *Mind* 124(494): 493–516,  
2942 doi:10.1093/mind/fzu185.
- 2943 KRIEGEL, Uriah. 2009. *Subjective Consciousness: A Self-Representational Theory*. New  
2944 York: Oxford University Press, doi:10.1093/acprof:oso/9780199570355.001.0001.
- 2945 KRIPKE, Saul A. 1980. *Naming and Necessity*. Oxford: Basil Blackwell Publishers.
- 2946 LEE, Andrew Y. 2019. "The Microstructure of Experience." *Journal of the American*  
2947 *Philosophical Association* 5(3): 286–305, doi:10.1017/apa.2019.4.
- 2948 LEWIS, David. 1995. "Should a Materialist Believe in Qualia?" *Australasian Journal of*  
2949 *Philosophy* 73(1): 140–144. Reprinted in Lewis (1999, 325–331), doi:10.1080/0004  
2950 8409512346451.
- 2951 —. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University  
2952 Press, doi:10.1017/cbo9780511625343.
- 2953 LEWTAS, Patrick Kuehner. 2013. "What is it like to be a Quark?" *Journal of Conscious-*  
2954 *ness Studies* 20(9/10): 39–64.
- 2955 LIANG, Yibin. 2017. "Kant on Consciousness, Obscure Representations and Cognitive  
2956 Availability." *The Philosophical Forum* 48(7): 345–368, doi:10.1111/phil.12169.
- 2957 LIU, Michelle. 2019. "Phenomenal Experience and the Thesis of Revelation." in *Percep-*  
2958 *tion, Cognition and Aesthetics*, edited by Dena SHOTTENKIRK, Manuel CURADO,  
2959 and Steven S. GOUVEIA, pp. 227–252. Routledge Studies in Contemporary Philoso-  
2960 phy. New York: Routledge, doi:10.4324/9780429462658-13.
- 2961 —. 2020. "Explaining the Intuition of Revelation." *Journal of Consciousness Studies*  
2962 27(5/6): 99–107.
- 2963 LOCKWOOD, Michael. 1993. "The Grain Problem." in *Objections to Physicalism*, edited  
2964 by Howard ROBINSON, pp. 271–292. Oxford: Oxford University Press, doi:10.1093/  
2965 oso/9780198242567.003.0013.
- 2966 MAXWELL, Grover. 1978. "Rigid Designators and Mind-Brain Identity." in *Minnesota*  
2967 *Studies in the Philosophy of Science, Volume IX: Perception and Cognition: Issues in*  
2968 *the Foundations of Psychology*, edited by C. Wade SAVAGE, pp. 365–404. Minneap-  
2969 olis, Minnesota: University of Minnesota Press.
- 2970 MCGINN, Colin. 2006. "Hard Questions: Comments on Galen Strawson [Strawson  
2971 (2006)]." *Journal of Consciousness Studies* 13(10/11): 90–99.
- 2972 NAGASAWA, Yujin and WAGER, Kai. 2017. "Panpsychism and Priority Cosmopsychism."  
2973 in *Panpsychism: Contemporary Perspectives*, edited by Godehard BRÜNTRUP and  
2974 Ludwig JASKOLLA, pp. 113–129. Oxford: Oxford University Press, doi:10.1093/ac  
2975 prof:oso/9780199359943.003.0005.

- 2976 NEMIROV, Lawrence. 1990. "Physicalism and the Cognitive Role of Acquaintance." in  
2977 *Mind and Cognition*, edited by William G. LYCAN, pp. 490–499. Oxford: Blackwell  
2978 Publishers.
- 2979 PEREBOOM, Derk. 2016. "Illusionism and Anti-Functionalism about Phenomenal  
2980 Consciousness." *Journal of Consciousness Studies* 23(11/12): 172–185.
- 2981 —. 2019. "Russellian Monism, Introspective Inaccuracy, and the Illusion Meta-Problem  
2982 of Consciousness." *Journal of Consciousness Studies* 26(9/10): 182–193.
- 2983 ROELOFS, Luke. 2014. "Phenomenal Blending and the Palette Problem." *Thought* 3(1):  
2984 59–70, doi:10.1002/tht3.113.
- 2985 —. 2016. "The Unity of Consciousness, Within Subjects and Between Subjects." *Philo-*  
2986 *sophical Studies* 173(12): 3199–3221, doi:10.1007/s11098-016-0658-7.
- 2987 —. 2019. *Combining Minds: How to Think About Composite Subjectivity*. Oxford: Oxford  
2988 University Press, doi:10.1093/oso/9780190859053.001.0001.
- 2989 ROSENTHAL, David M. 1993. "State Consciousness and Transitive Consciousness."  
2990 *Consciousness and Cognition* 2(3): 355–363, doi:10.1006/ccog.1993.1029.
- 2991 SEAGER, William E. 1995. "Consciousness, Information and Panpsychism." *Journal of*  
2992 *Consciousness Studies* 2(3): 272–288. Reprinted in Shear (1997, 269–286).
- 2993 SHANI, Itay. 2015. "Cosmopsychism: A Holistic Approach to the Metaphysics of Expe-  
2994 rience." *Philosophical Papers* 44(3): 389–437, doi:10.1080/05568641.2015.1106709.
- 2995 SHEAR, Jonathan, ed. 1997. *Explaining Consciousness: The Hard Problem*. Cambridge,  
2996 Massachusetts: The MIT Press.
- 2997 SMITHIES, Declan. 2019. *The Epistemic Role of Consciousness*. Oxford: Oxford Univer-  
2998 sity Press, doi:10.1093/oso/9780199917662.001.0001.
- 2999 STOLJAR, Daniel. 2001. "Two Conceptions of the Physical." *Philosophy and Phenomeno-*  
3000 *logical Research* 62(2): 253–281, doi:10.1111/j.1933-1592.2001.tb00056.x.
- 3001 —. 2006. *Ignorance and Imagination*. Oxford: Oxford University Press, doi:10.1093/01  
3002 95306589.001.0001.
- 3003 —. 2009. "The Argument from Revelation." in *Conceptual Analysis and Philosophical*  
3004 *Naturalism*, edited by David BRADDON-MITCHELL and Robert NOLA, pp. 113–138.  
3005 Cambridge, Massachusetts: The MIT Press, doi:10.7551/mitpress/9780262012560  
3006 .003.0005.
- 3007 —. 2013. "Qualitative Inaccuracy and Unconceived Alternatives." *Philosophy and*  
3008 *Phenomenological Research* 86(3): 745–752, doi:10.1111/phpr.12030.
- 3009 STRAWSON, Galen. 2006. "Realistic Monism: Why Physicalism Entails Panpsychism."  
3010 *Journal of Consciousness Studies* 13(10/11): 3–31. Reprinted in Strawson (2008,  
3011 53–74).
- 3012 —. 2008. *Real Materialism, and Other Essays*. Oxford: Oxford University Press, doi:10  
3013 .1093/acprof:oso/9780199267422.001.0001.
- 3014 —. 2015. "Self-Intimation." *Phenomenology and the Cognitive Sciences* 14(1): 1–31,  
3015 doi:10.1007/s11097-013-9339-6.

- 3016 THIEL, Udo. 2011. *The Early Modern Subject: Self-Consciousness and Personal Identity*  
3017 *from Descartes to Hume*. Oxford: Oxford University Press, doi:[10.1093/acprof:](https://doi.org/10.1093/acprof:oso/9780199542499.001.0001)  
3018 [oso/9780199542499.001.0001](https://doi.org/10.1093/acprof:oso/9780199542499.001.0001).
- 3019 WILSON, Margaret Dauler. 1980. "Objects, Ideas, and 'Minds': Comments on Spinoza's  
3020 Theory of Mind." in *The Philosophy of Baruch Spinoza*, edited by Richard KEN-  
3021 NINGTON, pp. 103–120. *Studies in Philosophy and the History of Philosophy* n. 7.  
3022 Washington, D.C.: The Catholic University of America Press. Reprinted in Wilson  
3023 (1999, 126–140).
- 3024 —. 1999. *Ideas and Mechanism – Essays on Early Modern Philosophy*. Princeton, New  
3025 Jersey: Princeton University Press, doi:[10.1515/9781400864980](https://doi.org/10.1515/9781400864980).

PROOF

# In Defence of Facts Grounding, Essential Properties and the Unity Problem

DONNCHADH O'CONNAILL

3027 A common conception of facts is as worldly entities, complexes made up  
3028 of non-factual constituents such as properties, relations and property-  
3029 bearers. Understood in this way facts face the unity problem, the problem  
3030 of explaining why various constituents are combined to form a fact. In  
3031 many cases the constituents could have existed without being unified  
3032 in the fact—so in virtue of what are they so unified? I shall present a  
3033 new approach to the unity problem. First, facts which are grounded are  
3034 unified by the obtaining of their grounds. Second, many ungrounded  
3035 facts are such that they must obtain if their non-factual constituents exist  
3036 (e.g., if the property  $F_{\text{ness}}$  is essential to a particular,  $a$ , then if  $a$  exists  
3037 the fact that  $a$  is  $F$  must obtain). In this way the obtaining of these facts  
3038 is explained by the essence of some of their constituents. I also address  
3039 the possibility of facts which are brutally unified (i.e., neither grounded  
3040 nor essentially unified), and compare the account I offer with some of  
3041 the main alternatives.

3042 It is common for facts to be understood as worldly entities, complexes made up  
3043 of non-factual constituents such as properties, relations and property-bearers.  
3044 A number of authors have presented a problem for facts understood in this  
3045 way, the problem of unity. This is the problem of explaining the difference  
3046 between the existence of all the constituents of a fact and the obtaining of that  
3047 fact. For instance, a particular entity  $a$  might exist and a property  $F_{\text{ness}}$  might  
3048 be instantiated, but it does not follow that the fact that  $a$  is  $F$  obtains—so in  
3049 virtue of what does this fact obtain?

3050 In this paper I offer a new line of defence against the unity problem. After  
3051 outlining the compositional conception of facts in section 1, I shall state the  
3052 unity problem in section 2 and possible responses to it in section 3. In section 4  
3053 I outline the first part of my defence, which appeals to the notion of grounding:

3054 if a fact is grounded, its unity is explained by the obtaining of its grounds. This  
 3055 raises the issue of whether there are facts which are not grounded, and if so  
 3056 how the unity of these facts can be explained. In section 5 I consider how the  
 3057 unity problem might be addressed if every fact were grounded. In section 6  
 3058 I propose that many facts which are not grounded are plausibly such that  
 3059 the properties they involve are essential to their property-bearers. Because  
 3060 of this, the constituents of these facts are essentially unified. In section 7 I  
 3061 address the possibility that there could be ungrounded facts which are not  
 3062 essentially unified, facts whose unity is not explained in either of the two ways  
 3063 I propose. In section 8, I briefly compare my account of the unity of facts with  
 3064 alternative views proposed by Arianna Betti and William Vallicella. While  
 3065 the account I offer has certain drawbacks compared to these alternatives, it  
 3066 also has important advantages, and should be taken as seriously as any other  
 3067 account of the unity of facts.

## 3068 1 The Compositional Conception of Facts

3069 On the *compositional conception*, a fact is a complex entity made up of non-  
 3070 factual constituents (hereafter “constituents”). In this section I shall present  
 3071 some key aspects of facts thus understood.<sup>1</sup>

3072 First, facts are *non-representational entities*: they do not have truth- or  
 3073 accuracy-conditions, nor do they refer to or designate anything, and they are  
 3074 not about anything in the sense in which intentional states are about their  
 3075 objects. Some facts will include representational entities among their con-  
 3076 stituents (e.g., the fact that the sentence “Tom is wet” is true). But such facts do  
 3077 not themselves represent anything. Furthermore, facts are not metaphysically  
 3078 posterior to propositions which state them; e.g., the identity of the fact that *a*  
 3079 is *F* is not metaphysically determined by the proposition “*a* is *F*” (in contrast  
 3080 with what Kit Fine terms the propositional conception of facts—see his 1982,  
 3081 51–52).

3082 Second, I take facts to be composed of property-bearers and properties  
 3083 (for the purposes of this paper I include relations among the properties).  
 3084 Both properties and property-bearers are relatively coarse-grained entities:  
 3085 for instance, the property *being water* is identical with the property *being*  
 3086 *composed of H<sub>2</sub>O molecules*, whereas the concepts “being water” and “being

1 A more thorough statement of this conception is offered by Betti (2015, 7, 18–30; see also Vallicella 2016a, 115). In what follows I shall ignore questions concerning states of affairs as distinct from facts.

3087 composed of H<sub>2</sub>O molecules” are distinct. Correspondingly, the fact composed  
 3088 of this property and a certain mass of material (e.g., the fact that this body  
 3089 of liquid is water) is more coarse-grained than the proposition “this body of  
 3090 liquid is composed of water.” The properties which help to make up facts are  
 3091 universals. Property-bearers can be either particulars or universals.<sup>2</sup>

3092 Third, because facts are composed of entities which exist and help to make  
 3093 up the world, I take it that facts themselves help to make up what exists; in  
 3094 this sense, they are *worldly* entities (see Betti 2015, 22–24). This conception of  
 3095 facts thus closely corresponds to one rejected by P.F. Strawson, according to  
 3096 whom a fact “is not something in the world. It is not an object; not even (as  
 3097 some have supposed) a complex object consisting of one or more particular  
 3098 elements (constituents, parts) and a universal element (constituent, part)”  
 3099 (1950, 135).

3100 Fourth, the way in which a fact’s constituents are combined to make up  
 3101 that fact is *non-mereological*. In the present context this can be understood  
 3102 as follows: for a fact to obtain is not the same as for its constituents to exist;  
 3103 rather, it is for its constituents to exist *and* be combined or arranged in some  
 3104 specific way (Betti 2015, 65).<sup>3</sup> For instance, the fact that *a* is *F* (hereafter, “*Fa*”)  
 3105 obtains only if *a* *instantiates* the property *Fness*; the fact that *a* is larger than  
 3106 *b* obtains only if *a* and *b* stand (in a particular order) in the relation *larger*  
 3107 *than*. It is helpful to have a term which allows us to contrast the existence  
 3108 of all the constituents of a fact with the obtaining of this fact. When all the  
 3109 constituents exist, I shall refer to them as forming an *aggregate*, where for an  
 3110 aggregate of entities to exist just is for each entity in the aggregate to exist.  
 3111 One might then say that whereas an aggregate is a mereological sum, a fact

---

2 I assume a sparse view of properties, on which it is not the case that each predicate corresponds to a distinct property or relation. For the most part this will not matter in what follows, but it is worth noting that I do not assume that formal ontological predicates such as “instantiates” correspond to distinct properties. Therefore, I do not accept that facts have so-called “secondary” constituents, e.g., a relation of instantiation or a non-relational tie which binds *a* and *Fness*. I shall mostly use examples of facts containing one property-bearer and one property; however, the compositional conception is not itself committed to this restriction.

3 Alternatively, this form of composition can be understood as involving non-extensional mereology (Bennett 2013, 101–102). The notion of non-mereological composition has been challenged by David Lewis (e.g., 1986), but it is accepted by all proponents of the compositional conception of facts. As I understand it, the problem of unity is based on accepting this conception and raising a challenge concerning facts understood in this way.

3112 is a non-mereological complex (Armstrong 1997, 119–122; Meinertsen 2008,  
3113 3).<sup>4</sup>

3114 It is frequently claimed that because facts exhibit non-mereological unity,  
3115 the existence of the constituents of a fact does not itself suffice for the ob-  
3116 taining of that fact (Vallicella 2000, 246; Betti 2015, 54). I shall question this  
3117 claim later, but for the time being I accept it. It is also often claimed that a  
3118 fact is something *over and above* its constituents. This claim is sometimes  
3119 supported by the contention that the constituents can exist without the fact  
3120 obtaining (Vallicella 2000, 238). It is also sometimes supported by appealing  
3121 to the non-mereological composition of facts: “philosophers who do accept  
3122 facts say that when Hargle is sad, alongside these two things (Hargle and  
3123 sadness) there is also a third thing in the world: a special ‘being together’ of  
3124 these two things in a real unity over and above the two things” (Betti 2015,  
3125 30). I shall return to these claims in section 6.

3126 Fifth, I accept what Gonzalo Rodriguez-Pereyra terms the *structuralist*  
3127 *criterion* of fact identity: “facts are identical if and only if they have the same  
3128 constituents combined in the same way” (1998, 520). That is, fact *A* is identical  
3129 with fact *B* iff (i) the constituents of *A* are all identical with the constituents  
3130 of *B*, and vice-versa; (ii) the mode of combination of the constituents in *A*  
3131 is identical with the mode of combination of the constituents in *B*; (iii) *A*  
3132 obtains at exactly the same time as *B*. By “mode of combination”, I mean the  
3133 specific kind of non-mereological composition which characterises each fact.  
3134 This could be that a particular instantiates a property, or that two entities  
3135 stand in a certain relation. In the case of asymmetric relations, it would also  
3136 include entities standing in a certain order in that relation, so that, e.g., *aRb*  
3137 would involve a different mode of combination than *bRa*.

3138 This criterion suggests the following asymmetry: while the identities of the  
3139 constituents of a fact help to determine its identity-conditions, the reverse does  
3140 not hold (Vallicella 2016a, 117). Therefore, on the compositional conception  
3141 “facts are built up out of ontologically more basic materials” (2016a, 115).  
3142 This view of facts can be contrasted with one in which the constituents of  
3143 facts are abstractions from them, such that the identity of the constituents is  
3144 determined by the identity of the facts to which they belong.

---

4 I do not intend talk of aggregates to be ontologically committing—I use it for convenience and if necessary it could be replaced by plural quantification over the constituents (Betti 2015, 53). For ease of presentation, I shall write as though a property must be instantiated in order to exist—however, the discussion can easily be adapted to accommodate a Platonist conception of properties.



3145 Finally, I assume that whenever the constituents of a fact A are arranged  
 3146 in the mode of combination characteristic of A, A thereby obtains (e.g., if a  
 3147 property-bearer *a* instantiates a property *Fness*, the fact *Fa* obtains).<sup>5</sup> This  
 3148 assumption can be challenged.<sup>6</sup> For instance, in E.J. Lowe’s four-category  
 3149 ontology when a universal property is had by a property-bearer we do not  
 3150 need to posit a fact; rather, the property-bearer is characterised by a particular  
 3151 property or mode (2006). But while one could object to facts in this way, this  
 3152 seems to be a different issue to the problem of unity.<sup>7</sup>

## 3152 2 The Problem of Unity

3154 Given the above conception of facts, the unity problem is relatively easy to out-  
 3155 line. Consider *Fa*. For this fact to obtain, its constituents (*a* and *Fness*) must  
 3156 be combined in a specific, non-mereological manner; only in this way will they  
 3157 achieve the kind of unity characteristic of a fact. This kind of unity between  
 3158 *a* and *Fness* would be absent if, for instance, *a* existed and *Fness* was instan-  
 3159 tiated, but *a* did not instantiate *Fness* (e.g., if some entity *b*, wholly distinct  
 3160 from *a*, instantiated *Fness*). In that scenario, *a* would exist and *Fness* would  
 3161 be instantiated, but they would not exist together in the way characteristic of  
 3162 *Fa*.

3163 The unity problem is simply the problem of explaining *why*, given that  
 3164 specific non-factual entities (e.g., *a* and *Fness*) each exist, they are united  
 3165 to form the fact *Fa*. More generally, it is the problem of explaining for any  
 3166 fact *why*, given that its constituents each exist, they are unified in the specific  
 3167 mode of combination characteristic of that fact. A solution to this problem for  
 3168 a specific fact, A, would be a metaphysical explanation of *why*, given that the  
 3169 constituents of A each exist, they are arranged in the mode of combination  
 3170 characteristic of A. As Betti puts it, the problem is “how to account for the unity  
 3171 of relations with their relata and for the unity of properties with their bearers”  
 3172 (2015, 42). Elsewhere she glosses the problem as the search for “something *in*

5 This formulation sets aside issues to do with the time at which the constituents are arranged.

6 Thanks to Jani Hakkarainen for raising this point.

7 The unity problem could be reframed as a problem concerning the unity of specific property-bearers and properties, without mentioning facts. With regard to Lowe’s ontology, the problem would be that of explaining why a specific property-bearer has the modes which characterise it.

3173 *virtue of which* those constituents form a unity” (2015, 45; see also Vallicella  
3174 2000, 242; Orilia 2006, 214; Meinertsen 2008, 3).<sup>8</sup>

3175 The unity problem thus characterized is relatively straightforward to grasp,  
3176 but there are potentially complicating factors which must be addressed. The  
3177 first is that the problem is often described in such a way that it seems to  
3178 presuppose the possibility that the constituents of a fact might exist and  
3179 the fact not obtain. For instance, Vallicella asks: “What makes it the case  
3180 that a number of constituents of the right kinds—constituents which are  
3181 connectable so as to form a fact but *need not be connected to exist*—are actually  
3182 connected so as to form an actual or existing fact?” (Vallicella 2000, 242, italics  
3183 added). Here the italicized phrase expresses the assumption that the existence  
3184 of the constituents need not entail the obtaining of the fact (see also Dodd  
3185 1999, 159; Wieland and Betti 2008, 510; Betti 2015, 54; Perovic 2016, 144). In  
3186 what follows I shall not make this assumption, though I postpone discussion  
3187 until section 6.

3188 Second, the unity problem as I have characterised it should be distinguished  
3189 from different problems with which it might be confused. For instance, at  
3190 one point in their discussion Betti and Jan Wieland ask, “What grounds the  
3191 difference between mereological and unmereological composition?” (2008,  
3192 513). Wieland and Betti present this as a restatement of the original unity  
3193 problem,<sup>9</sup> but I think it is a different problem. The unity problem is the  
3194 question, for any specific fact, of what it is in virtue of which its constituents  
3195 are unified. An answer to this problem may in principle apply to any fact,  
3196 but it will not itself explain the difference between mereological and non-  
3197 mereological composition. It may be that the difference between mereological  
3198 and non-mereological composition cannot be explained, but it would not  
3199 follow that the unity problem cannot be solved (Vallicella makes a similar

---

8 Metaphysical explanations, the kind of explanations expressed by “in virtue of” claims, encompass grounding explanations but also other forms of explanation (see footnote 36 below). Katarina Perovic suggests that the problem of unity consists of “a cluster of problems that are frequently run together” (2016, 145). I have some sympathy with this view, but I suggest that the problem outlined in the main text does not conflate different issues. In terms of the various problems Perovic distinguishes, the problem of unity corresponds both to what she terms the explanatory problem and the Mereological Problem of Unity (2016, 146–149). I suggest that these are really the same problem. What Perovic calls the Mereological Problem concerns the ontological ground of the difference between a fact and the aggregate of its constituents. In this context, the ontological ground is whatever explains the unity of the constituents in the fact; it is that in virtue of which they together form a fact.

9 They introduce the quoted passage by saying, “We can restate the problem immediately” (2008, 513). The context makes it clear that by “the problem” they mean the unity problem.

point in his 2000, 242). Similarly, Julian Dodd glosses the supposed obscurity of the unity of facts as the “problem of the nature of instantiation” (1999, 156). But we need to distinguish between an account of *what* instantiation is (an answer to the problem concerning its nature) and an account of *why* specific entities instantiate specific universals (an answer to the unity problem).<sup>10</sup> More generally, there is a difference between giving an account of *what it is for*  $a$  and  $F$ ness to form a fact, and explaining *why*  $a$  and  $F$ ness are so combined. It is perfectly legitimate to answer the first question by citing the characteristic unity of the fact. For instance, one might be contrasting the unity of a fact with the unity of parts in a mereological sum, or members in a set, in which case it makes sense to refer to  $a$  instantiating  $F$ ness. But it is not legitimate to answer the second question by citing this very unity: it is no use explaining why  $a$  and  $F$ ness are unified in  $Fa$  by saying that  $a$  instantiates  $F$ ness. This would be to simply re-describe what one was asked to explain.<sup>11</sup>

It is crucial to distinguish the unity problem from these other problems (concerning the nature of instantiation, or the difference between mereological and non-mereological composition). These questions concern the very coherence of a theory of facts; they arise insofar as the very idea of facts is considered obscure. The unity problem, in contrast, is based on the assumption that the idea of a fact is coherent. It is only given a coherent notion of facts that the distinction between a fact and the aggregate of its constituents can clearly be drawn; and it is only given this distinction that the unity problem can be posed. Therefore, in addressing the unity problem we can set these other questions aside.

### 3 Possible Solutions

The unity problem has been developed into an argument against facts by a number of different writers (Dodd 1999, 152; Wieland and Betti 2008, 509; Betti 2015, 51). Though details differ, each version of the argument works in roughly the same way: postulating facts gives rise to the unity problem; there

10 We also need to distinguish each of these accounts from an account of *how it is possible* for distinct entities such as  $a$  and  $F$ ness to be unified (Dodd 1999, 151; Vallicella 2002, 26; Maurin 2015, 212–213), or an account of how non-mereological composition is possible (see Eklund’s characterisation of the problem of unity for facts in his 2019, 1236–1237).

11 For further discussion of the difference between questions concerning what something is and questions concerning why it is (as it is), see Audi (2015).

3229 are a determinate number of possible solutions to this problem available; none  
 3230 of these solutions succeed; therefore, we should not postulate facts.<sup>12</sup>

3231 To structure the discussion I shall refer to the range of possible solutions  
 3232 outlined by Betti (2015, 51). The unity of a fact could be explained by:

- 3233 (A) the constituents of the fact, e.g., *a* and/or *Fness*;
- 3234 (B) one or more additional constituents of the fact, i.e., a constituent which  
 3235 is identical to neither *a* nor *Fness*;<sup>13</sup>
- 3236 (C) something external to the fact, i.e., something numerically distinct from  
 3237 either the fact or any of its constituents; or
- 3238 (D) the fact itself.

3239 Bo Meinertsen outlines a version of (B). Vallicella and Francesco Orilia argue  
 3240 for different versions of (C).<sup>14</sup> Betti interprets David Armstrong as in effect  
 3241 putting forward a version of (D). Dodd, Wieland and Betti argue that none of  
 3242 these options can work, and that the unity problem cannot be solved.

3243 Wieland and Betti offer a dissolution of the problem which in effect rejects  
 3244 the assumption that a fact is something other than the aggregate of its con-  
 3245 stituents. This argument appeals to the notion of *bearer-specific properties*. A  
 3246 property is bearer-specific iff it is in its nature to be had by a specific bearer or  
 3247 bearers (Betti 2015, 90).<sup>15</sup> So if *Fness* is a property specific to *a*, *Fness* is such  
 3248 that necessarily if it exists, it is instantiated by *a*. All tropes are bearer-specific  
 3249 properties, but Wieland and Betti deny that all bearer-specific properties are  
 3250 tropes, since it can be in the nature of some bearer-specific properties to be  
 3251 had by many specific entities (2008, 519).<sup>16</sup> If properties are bearer-specific,  
 3252 then the unity problem would be dissolved: *a* would instantiate *Fness* as soon  
 3253 as *Fness* exists, and therefore there would be no difference between the fact  
 3254 and the aggregate of its constituents (Betti 2015, 92).

3255 The response I shall offer to the unity problem does not fall neatly into any  
 3256 of Betti's options—or rather, different parts of the response fall into different  
 3257 options. I shall begin by outlining a version of option (C), though distinct

12 One possible response to this argument would be to claim that facts can obtain without their unity being explained at all. I shall consider this possibility in section 7.

13 For instance, one might treat the relation of instantiation as a further constituent of the fact.

14 Dixon (2018) can be understood as proposing a version of (C), though he does not specifically discuss the problem of unity.

15 Betti prefers the phrase “relata-specific relations” (2015, 89–90). Since I am treating relations as among the properties, this difference is not important.

16 The notion of bearer-specific properties is criticised by Vallicella (2016b, 237–240). It is defended by Wieland and Betti (2008, 521–522), and by Betti (2015, 93–96). I discuss it in section 8 below.

3258 from those offered by Vallicella or Orilia. Whether or not this version of (C)  
3259 solves the unity problem for all facts depends on further assumptions. If there  
3260 are facts to which it does not apply, then some other response to the problem  
3261 must be offered. I shall offer a further response which can be read as a version  
3262 of Betti's option (A), or as dissolving the problem in a manner similar to her  
3263 appeal to bearer-specific properties. It is also important to note that, on my  
3264 approach, there may be facts to which none of options (A)-(D) applies; in  
3265 section 7 I shall defend the possibility of such facts.

#### 3264 4 Grounding and Unity

3267 In this section I shall outline a specific conception of grounding and argue  
3268 that it can help explain the unity of grounded facts.

3269 The terminology of ground is frequently used in order to set out the unity  
3270 problem (Vallicella 2000, 243; Wieland and Betti 2008, 510–511; Betti 2015, 55).  
3271 It may therefore seem odd to appeal to a notion of grounding in order to solve  
3272 this problem. But the appearance of oddity here is easily explained. The notion  
3273 of “ground” used to state the unity problem simply indicates whatever could  
3274 solve it (i.e., whatever it is in virtue of which a fact is unified, or whatever  
3275 explains its unity). Therefore, theorists writing about the unity problem have  
3276 not needed to say a great deal about this notion. For instance, neither Betti  
3277 nor Vallicella systematically characterize this notion or attempt to relate their  
3278 use of it to the recent literature on metaphysical grounding. In contrast, the  
3279 conception of grounding which I shall outline is in large part drawn from this  
3280 literature. And because I am planning to put this conception to constructive  
3281 use in solving the problem, I will need to say more about it.

3282 Grounding is a form of metaphysical determination often linked to certain  
3283 non-causal “in virtue of” explanations (e.g., mental facts obtain in virtue of  
3284 the obtaining of certain physical facts; entities possess dispositional properties  
3285 in virtue of possessing categorial properties; actions have moral properties  
3286 in virtue of certain of their non-moral properties). I take grounding to be a  
3287 worldly relation which underwrites some of these explanations, in much the  
3288 same way that causation is typically thought of as a worldly relation which  
3289 underwrites causal explanations (Audi 2012, 691; Schaffer 2016a, 84). I assume  
3290 that grounding is irreflexive, asymmetric, transitive, non-monotonic, and that

3291 the full grounds of an entity necessitate the existence of that entity.<sup>17</sup> Many of  
 3292 these assumptions have been questioned in the literature, but together they  
 3293 form a familiar and recognisably orthodox conception of grounding.<sup>18</sup>

3294 To this conception I shall need to add more detail about the relata of ground-  
 3295 ing and the specific way or ways in which they are related. I assume that  
 3296 grounding holds between facts understood on the compositional character-  
 3297 isation. This is not part of the orthodox view: it is common for grounding  
 3298 theorists to speak of facts being grounded, but they are usually non-committal  
 3299 as to the nature of these facts. However, the idea that worldly facts can be  
 3300 related by grounding is at least a familiar one (Audi 2012, 687; Raven 2012,  
 3301 689; Trogdon 2018, 1289). If there are worldly facts then they look like good  
 3302 candidates to be grounded, assuming anything can be grounded at all.

3303 What is it for a worldly fact to be grounded? There is probably no non-  
 3304 circular definition or analysis of grounding,<sup>19</sup> but we can still say something  
 3305 informative about it. Examples of informative but circular accounts are found  
 3306 elsewhere in philosophy. For instance, it is possible to think that knowledge  
 3307 cannot be analysed in a non-circular fashion, and also that it is informative to  
 3308 learn that knowledge must satisfy a safety condition; this can be informative  
 3309 even if the account is circular, i.e., if the relevant notion of safety is itself  
 3310 understood in terms of knowledge (Watzl 2017, 66). More generally, “Someone  
 3311 who accepts that there is an informative but non-reductive account of some *F*  
 3312 thus normally will say something about either the internal structure of *F*s or  
 3313 how being an *F* is related to some other phenomena” (Watzl 2017, 66–67). In  
 3314 the case of grounding worldly facts, I think we can do both of these things.

3315 If a worldly fact *Fa* is fully grounded in other facts, then *Fa* obtains because  
 3316 the other facts obtain (at a specific time, in a specific world).<sup>20</sup> For a fact to  
 3317 obtain just is for its constituents to exist and to be unified in a certain way.  
 3318 Therefore, for the grounds of *Fa* to explain the obtaining of that fact is at the  
 3319 very least a good reason to accept that those grounds explain the unity of *a*  
 3320 and *F*ness. So this conception of grounding suggests a straightforward answer

---

17 On the distinction between full and partial grounds, see Fine (2012, 50). I assume that if a fact is partially grounded, it must be fully grounded.

18 For example, Rodriguez-Pereyra (2015) challenges irreflexivity, transitivity and asymmetry; Raven (2013) defends all three features.

19 Though for a recent proposal see Correia and Skiles (2019).

20 This is a non-causal sense of “because” that tracks grounding relations. The indexing to worlds and times is adapted from Skiles (2015, 719). If *Fa* is on this occasion grounded by, e.g., *Gb* and *Hc*, it is possible that in other circumstances it could have been grounded by different facts. In what follows I shall omit this indexing.

3321 to the unity problem, at least as it concerns grounded facts: the constituents  
 3322 of grounded facts are unified by their grounds.<sup>21</sup>

3323 While I think this is the correct explanation of the unity of grounded facts,  
 3324 it is reasonable to ask for more detail: in particular, *how* do the grounds of *Fa*  
 3325 unify its constituents? Again, it may not be possible to provide a non-circular  
 3326 answer to this question. But we can add more detail by considering that,  
 3327 plausibly, worldly facts can be grounded in different ways, depending on their  
 3328 constituents and the constituents of their grounds. In what follows I adopt  
 3329 the following hypothesis: for each instance of grounding, one or more of the  
 3330 constituents of the grounded fact stand in some specific ontological relation  
 3331 or relations to one or more of the constituents of each of the grounds. Which  
 3332 ontological relations obtain will depend on the constituents of each of the  
 3333 facts. For instance, the properties which help to make up the grounds may be  
 3334 determinates of a determinable property helping to make up the grounded  
 3335 fact (e.g., the fact that *a* is red is grounded in the fact that *a* is scarlet). Or  
 3336 the grounded fact may involve a property-bearer which is composed of the  
 3337 property-bearers in its grounds. Examples of this include many facts about  
 3338 functions (e.g., the fact that a computer is running a specific programme is  
 3339 grounded in facts about different sub-systems of the computer).<sup>22</sup>

3340 The ontological relations holding between these constituents will determine  
 3341 how exactly the grounds unify the constituents of the grounded fact. For  
 3342 example, suppose that the fact that *a* is red is grounded in the fact that *a* is  
 3343 scarlet. In this case, each constituent of the grounded fact stands in a specific

---

21 Strictly speaking, each grounded fact would be unified by its immediate grounds (on the distinction between mediate and immediate grounding see [Fine 2012, 50–51](#)). In what follows I shall omit this qualification.

22 These specific ontological relations are very similar to what Kelly Trogdon terms *grounding mechanisms*, “determination relations of a certain sort holding between constituents of grounding facts and constituents of the facts they ground” (2018, 1290). They are also similar to what Tobias Wilsch terms *linking principles*, principles which, roughly speaking, determine which objects and properties combine to form facts (2015, 3302–3304) (thanks to an anonymous referee for suggesting these sources). For more detailed discussion of the different kinds of ontological relations which can hold between the constituents of different facts, see “Grounding and the Unity of Facts” (unpublished manuscript). I do not regard these specific kinds of ontological relation (determinate-determinable, composition, etc.) as themselves kinds of grounding, i.e., “small-g” grounding relations (Wilson 2014, 540). Nor do I accept Wilson’s criticisms of “big-G” grounding, though I cannot discuss the matter here; but see e.g., Cameron (2016); Schaffer (2016b). I should also add that other conceptions of how grounding works are available (e.g., Schaffer 2016a). I am not claiming that the account I shall sketch in the main text is how grounding *should* be understood; I am merely claiming that grounding *can* be understood in this way, and that doing so allows us to see how the grounds of a fact unify it.



3344 ontological relation to a constituent of the ground: *a* is identical with itself,  
 3345 and the property *being scarlet* is a determinate of the property *being red*.  
 3346 The determinable-determinate relation is such that, necessarily, any entity  
 3347 instantiating a determinate property instantiates its determinable.<sup>23</sup> Therefore,  
 3348 if the fact that *a* is scarlet obtains, this will automatically unify *a* and the  
 3349 property *being red* in the fact that *a* is red. In this way, the unity of *a* and the  
 3350 property *being red* is explained by the obtaining of the fact that *a* is scarlet.

3351 Let us consider a slightly more complicated example: suppose there is a  
 3352 tower, *a*, which is exactly one metre tall and which consists of ten bricks  
 3353 piled on top of each other. Suppose also that the fact that *a* is one metre tall  
 3354 is grounded in facts about the height of each of the bricks which compose it  
 3355 and facts about how these bricks are arranged (i.e., they stand on top of each  
 3356 other). Here *Fness* is the property *being exactly one metre tall*, *Gness* is the  
 3357 property *being exactly ten centimetres tall*, and *H* is the relation *standing on*  
 3358 *top of each other*.

3359 Again each constituent of the grounded fact stands in a specific ontological  
 3360 relation to a constituent of each of its grounds. First, the bricks together  
 3361 *compose a*.<sup>24</sup> Second, the height of the bricks, when the bricks stand in *H*, will  
 3362 *sum* to one metre. Here, the ontological relation holds between the property  
 3363 *being exactly ten centimetres tall* and the property *being exactly one metre tall*  
 3364 (one might say that instances of the first property, i.e., instances of *Gness*, are  
 3365 apt to sum together to form an instance of *Fness* when a certain number of  
 3366 bearers of the instances of *Gness* are suitably arranged).

3367 So the full grounds of *Fa* (the fact that *a* is exactly one metre tall) will  
 3368 include ten facts of the form “*Gb*”, “*Gc*”, etc. (i.e., each brick is ten centimetres  
 3369 in height), plus a collective fact of the form “*b, c, etc. are together H*” (i.e., the  
 3370 bricks are stacked on top of each other). When these facts all obtain together,  
 3371 a fact of the form *Fa* will obtain (something which is composed of the ten  
 3372 bricks will be exactly one metre in height). That is, the property *Fness* and *a*  
 3373 (the tower composed of these ten bricks) will each exist and will be unified in  
 3374 the fact *Fa*.<sup>25</sup> Again, it should be clear how the ontological relations holding

23 As Paul Audi puts it, “Anything maroon is red, and indeed, anything maroon is red *in virtue of* being maroon. So it seems that it is the natures of these properties that are responsible for the grounding relation’s obtaining” (2012, 693).

24 Of course, it is a difficult question as to when one entity is composed by other entities, but I set this issue aside here. What matters for present purposes is that *a* is composed of the ten bricks.

25 If “*a*” is a singular term then it may be objected that the obtaining of all the grounds does not suffice to ground the specific fact *Fa*, because of the possibility of Ship of Theseus-style examples



3375 between  $a$  and the bricks, and between  $F_{\text{ness}}$  and  $G_{\text{ness}}$  and  $H_{\text{ness}}$ , help to  
 3376 explain how the grounds of  $Fa$  can unify its constituents.

3377 Each grounded fact is thus unified by something external, i.e., something  
 3378 identical neither with the fact itself nor with any of its constituents. This is a  
 3379 version of Betti's option (C). The precise details of how the constituents of  
 3380 different grounded facts are unified remain to be worked out, but the outline  
 3381 of the approach is clear: examine the constituents of the grounded fact and  
 3382 the constituents of its grounds, and work out which ontological relations hold  
 3383 between them.<sup>26</sup>

3384 It might be objected that this proposal begs the question: it can only work if  
 3385  $a$  and  $F_{\text{ness}}$  are already unified. Suppose that  $Fa$  is grounded in the fact that  
 3386  $b$  is  $G$  ( $Gb$ ). It seems clear that  $a$ 's being  $F$  is a logical precondition for  $Fa$  to  
 3387 stand in any relation. Therefore, in order for  $Gb$  to ground  $Fa$ ,  $a$  must already

---

(Skiles 2015, 721–723). This is an important issue but I shall not address it here: for present purposes what matters is that the obtaining of the grounds explains why  $F_{\text{ness}}$  is unified with whatever it is which the bricks together compose.

- 26 Could there be instances of grounding which do not feature specific ontological relations holding between constituents of the grounded fact and its grounds? These would be instances of what Trogdon terms *bare grounding*, “grounding relations that aren't instances of grounding mechanisms” (2018, 1295). Trogdon mentions as possible examples cases of logical or conceptual grounding, e.g., the fact that  $a$  is  $F$  and the fact that  $b$  is  $G$  together ground the conjunctive fact that  $a$  is  $F$  and  $b$  is  $G$ . I shall not address these examples in detail, but the following strategy is worth noting. On the compositional conception, a fact is composed of non-factual entities (properties, relations and property-bearers). Take the (proposed) conjunctive fact that  $a$  is  $F$  and  $b$  is  $G$ . What are the entities from which it is composed? It might be thought that this fact includes a conjunctive property, e.g., the property  $x$  being  $F \wedge y$  being  $G$ . I am sceptical that there is such a property, but if it exists then it is plausible that the following is essentially true of it: it is instantiated iff some entity  $x$  is  $F$  and some entity  $y$  is  $G$ . In that case, the grounding of the conjunctive fact is not bare, since a specific ontological relation holds between a constituent of the grounded fact (the conjunctive property) and constituents of its grounds (i.e.,  $F_{\text{ness}}$  and  $G_{\text{ness}}$ ). On the other hand, it might be denied that the conjunctive fact includes a conjunctive property: on this view, the conjunctive fact is composed by  $a$ ,  $b$ ,  $F_{\text{ness}}$  and  $G_{\text{ness}}$ , arranged in a specific way. In that case, this proposed conjunctive fact seems to be an aggregate of the two facts which supposedly ground it. Given the compositional conception, it is not at all clear that any mere aggregate of facts should itself be counted as a *fact*. Rather, it is a collection of distinct facts. Any collection of facts can itself be treated as a fact, but this would be to use a different conception of facts, on which facts are logical or conceptual rather than worldly entities. I am not suggesting that facts understood in this way should not be posited, just that they are not the kind of facts which the problem of unity concerns.

3388 be  $F$  (i.e.,  $a$  and  $Fness$  must be unified). Far from unifying  $a$  and  $Fness$ , any  
 3389 grounding relation in which  $Fa$  stands requires that it already be unified.<sup>27</sup>

3390 An initial worry with this objection is that it threatens to prove too much.  
 3391 For with very little modification, it can be deployed against any proposed  
 3392 explanation of the existence of any entity whatsoever (where “existence”  
 3393 includes, e.g., a fact’s obtaining, an event’s occurring, etc.). Suppose we want  
 3394 to explain the existence of some entity  $x$ , and we appeal to a different entity  $y$ ;  
 3395 we say that  $y$ ’s existing, or something else about  $y$ , explains  $x$ ’s existing (e.g.,  
 3396  $y$  might be an event which causes  $x$ ). For this explanation to be correct, it  
 3397 is necessary that  $x$  exists (if  $x$  did not exist, then its existence would not be  
 3398 explained). So the proposed explanation works only if  $x$  already exists. In this  
 3399 way, it turns out that any proposed explanation of the existence of any entity  
 3400 will be circular. But this is surely not so.<sup>28</sup>

3401 One issue here is with the word “already”: it might be objected that if  $x$   
 3402 and  $y$  are events, and if effects occur after their causes,  $x$  could not already  
 3403 have occurred for its occurring to be explained by  $y$ . But I take it that in  
 3404 the objection to my proposal, the word “already” does not indicate temporal  
 3405 priority, but a logical precondition: for  $Gb$  to ground  $Fa$  logically requires that  
 3406  $Fa$  obtains. When the term “already” is understood in this sense,  $y$ ’s causing  
 3407  $x$  logically requires that  $x$  occurs, just as  $Gb$ ’s grounding  $Fa$  logically requires  
 3408 that  $Fa$  obtains.

3409 This gives us reason to think *that* this objection has gone wrong. As to *how*  
 3410 it goes wrong, the answer lies in distinguishing *explanatory* considerations  
 3411 from *modal* considerations (which include what is logically or metaphysically  
 3412 necessary for something to exist). The basic point is this: for  $x$  to modally  
 3413 depend on  $y$  (so that  $x$  cannot exist unless  $y$  exists) does not preclude that  $x$   
 3414 can itself explain (or help to explain)  $y$ ’s existence.

3415 A couple of examples from the literature can help to clarify this point. On  
 3416 a widely accepted view of sets, the singleton set containing Socrates exists iff  
 3417 Socrates exists. But the existence of the set is widely thought to be explained  
 3418 (at least in part) by the existence of Socrates (see, e.g., Schaffer 2016a, 53).

---

27 This objection was suggested to me by certain passages of Vallicella’s (2000, 243, 254). However, it is not clear that he is putting forward this exact argument in these passages. Thanks also to an anonymous referee for pressing me to develop my response to this objection.

28 Vallicella argues that event causation cannot be causation of existence, precisely since both the cause and its effect must occur for a causal relation to hold between them (2002, 27). But this seems false, at least for instances of what Ned Hall terms productive causation, when an event “helps to generate or bring about or produce another event” (2004, 225).

3419 A second example is the Euthyphro dilemma (which has been discussed  
 3420 in the grounding literature—see, e.g., Raven 2012, 692–693). Whichever way  
 3421 one responds to the dilemma, the “because” statement is true iff both the gods  
 3422 will that  $p$  and  $p$  is good. But this modal dependence does not rule out either  
 3423 explanation one might offer (e.g., that  $p$  is good because the gods will that  $p$ ,  
 3424 or that the gods will that  $p$  because  $p$  is good).

3425 How does this apply to my proposed explanation? It is true that  $Gb$ ’s ground-  
 3426 ing  $Fa$  modally depends on (logically requires) that  $Fa$  obtains. But this, I  
 3427 suggest, is not an explanatory dependence; we are not obliged to say that  $Gb$ ’s  
 3428 grounding  $Fa$  is explained, even in part, by  $Fa$ ’s obtaining. And, as per the  
 3429 examples outlined above, the modal dependence of a proposed explanation  
 3430 on its explanans is not by itself sufficient to generate an explanatory circle.  
 3431 Indeed, the modal dependence of the explanation of  $Fa$ ’s obtaining simply  
 3432 reflects the sufficiency of the proposed explanation: if the obtaining of  $Fa$ ’s  
 3433 grounds are sufficient to explain the unity of  $Fa$ , then  $Fa$  must obtain for the  
 3434 explanation to be correct.

## 3435 5 The Vicious Regress Argument

3436 I have argued that grounding can account for the unity of facts which have  
 3437 grounds. This invites the questions of whether there are ungrounded facts,  
 3438 and if so what could account for their unity. There are two options to consider  
 3439 here:

- 3440 (1) There are no ungrounded facts, and every fact is unified by its grounds;
- 3441 (2) There are ungrounded facts, which are not unified in the way in which
- 3442 grounded facts are unified.<sup>29</sup>

3443 I shall consider (1) in this section, and (2) in section 6.

3444 In scenario (1), there are no facts such that they are not fully grounded in  
 3445 some other facts. Facts can form chains of grounding (a collection of facts  
 3446 where any two members of this collection stand in grounding relations to  
 3447 each other). In scenario (1), each fact will stand in an infinite descending  
 3448 chain or chains of grounds.

3449 Whether such chains are possible and whether they could solve the unity  
 3450 problem are contentious issues. The main reason for thinking that such chains

---

29 As we shall see, there is a third option: there are no ungrounded facts, but chains of grounding terminate in entities which are not themselves facts. I shall briefly consider this option in section 6.

are impossible is that they seem to give rise to a vicious regress.<sup>30</sup> The criteria for deciding when regresses are vicious have been subject to extensive debate (Clark 1988; Nolan 2001; Maurin 2007; Wieland 2013). The criterion which seems most relevant in the present context is what Wieland terms the Failure Schema (2013, 99). This schema is summarised by Simon Blackburn: “A strategy gives rise to a vicious regress if whatever problem it was designed to solve remains as much in need of the same treatment after its use as before” (2005, 313). Examples of such strategies include the homunculus regress and the tower of turtles. In each case, a certain problem must be solved in order that something can be the case; an entity is posited in order to solve this problem; but the positing of this entity creates a problem of exactly the same kind as the problem the entity was posited to solve.

This suggests the following objection to scenario (1): the initial problem was how to explain the unity of some fact; in order to explain the unity of this fact, we posited grounds; but these grounds are facts each of which raises an explanatory demand of exactly the same kind as that which we initially faced. To respond to this further explanatory demand by positing further grounds would simply be to generate further problems of the same kind, and so on. What makes this regress vicious is that it makes no progress on the original question (or any progress it makes at any step in the regress is immediately cancelled out). This is arguably what goes wrong in the homonculous and turtle cases.

While this is a problem for (1), it is not clear that it is decisive. The original question was how to explain the unity of some specific fact, *Fa*. By appealing to the grounds of *Fa*, this question is answered. Granted, the answer generates a problem of the same kind: but *ex hypothesi*, for each new fact introduced, we will be able to appeal to *its* grounds to explain its unity. Given the scenario outlined in (1), there will never be a fact which lacks grounds, so the problem of unity can be answered for every fact posited.<sup>31</sup>

30 This objection is raised by Betti concerning a version of option (B), the idea that *a* and *Fness* are unified by the presence of a further constituent such as a relation of instantiation (2015, 56–57).

31 This is an important difference between the regress of facts to which the truth of (1) would commit us and what Eklund terms the *constitution regress* (2019, 1227–1229). The constitution regress very plausibly is vicious, because no step in this regress explains the fact with which the regress started. So rather than a different problem of the same type occurring at each step, as is the case with the regress generated by accepting (1), in the constitution regress the very problem we started with is never solved.

3480 It may be objected that this strategy is vicious insofar as it explains the  
 3481 unity of facts by assuming the very possibility of any fact being unified. This  
 3482 objection, or something like it, crops up occasionally in the literature:

3483 Even if, assuming there can be facts, facts may depend on each  
 3484 other in never-ending chains of dependence, postulating such  
 3485 chains of dependence does not help when it comes to the very  
 3486 possibility of there being facts to begin with. (Eklund 2019, 1228)

3487 But in the context of discussing the problem of unity, this objection seems to  
 3488 change the subject.<sup>32</sup> We began by asking a local question (what explains the  
 3489 unity of some specific fact or facts); now we are considering a global question,  
 3490 what is required for the possibility of any fact whatsoever.<sup>33</sup> But the chain of  
 3491 grounds was introduced to answer a series of local questions, e.g., why each  
 3492 specific fact is unified: “To claim that an infinite regress is vicious because it  
 3493 doesn’t allow us to answer the global question is to have accused it of having  
 3494 failed to carry out a task it was not designed to complete” (Bliss 2013, 408).<sup>34</sup>

3495 There is more to be said on these specific points and on other ways of  
 3496 characterizing regresses as vicious, but I shall not explore these issues here.  
 3497 My provisional conclusion is that while the regress argument is a problem for  
 3498 the proponent of (1), it is not clearly decisive: that is, it is not obvious that the  
 3499 regress argument renders (1) untenable. That said, it is worth asking how the  
 3500 unity problem might be solved for ungrounded facts.

---

32 This is not to suggest that Eklund himself is guilty of this. In the section where the passage I quoted appears, he is discussing the constitution regress, which is different to the regress which the truth of (1) would set up (see footnote 31 above).

33 Vallicella makes a similar point: the problem of unity “does not concern the nature of fact-unity in general, but the existence of fact-unity in particular cases” (2000, 242).

34 Orilia offers a solution to the unity problem which also appeals to an infinity of facts, and he responds to the threat of a vicious regress in a similar way (2006, 233). I shall not discuss Orilia’s position in detail, but it is worth mentioning two differences between it and my own. The first is that the facts to which Orilia appeals, facts which contain instances of an exemplification relation (what I have termed “instantiation”) are ad hoc; they are posited solely in order to solve the unity problem, without any independent reason to accept them. In contrast, the conception of grounding I have outlined can limit itself to facts which are relatively uncontroversial. Of course there are controversies surrounding grounding claims, but in general such claims are introduced as a way of ordering facts which we have independent reasons to accept. Second, Orilia’s view commits one to a necessarily infinite regress and a necessary infinity of facts given the obtaining of any fact (2006, 230). Even if such a regress is metaphysically possible, considerations of parsimony would favour not positing an infinity of facts if it can be avoided. As we shall see, the grounding response to the unity problem does not by itself commit one to positing an infinity of facts.

## 350 6 Essentially Unified Facts

3502 I propose that at least some ungrounded facts are *essentially unified*.<sup>35</sup> These  
 3503 facts are such that the properties which make them up are essential to their  
 3504 property-bearers, and so the property bearer could not exist without instanti-  
 3505 ating that property.<sup>36</sup> For instance, suppose that the property *being negatively*  
 3506 *charged* is essential to any electron. In that case, a specific election *e* could  
 3507 not exist without instantiating this property, i.e., without the fact that *e* is  
 3508 negatively charged obtaining. More generally, the thesis that some facts are  
 3509 essentially unified entails rejecting the following assumption: “Even if *a* and  
 3510 *F*ness cannot exist except in some state of affairs or other, there is nothing  
 3511 in the nature of *a* and nothing in the nature of *F*ness to require that they  
 3512 combine with each other to form *a*'s *being F*” (Vallicella 2000, 238).

3513 Essential unity can be usefully compared with bearer-specific properties  
 3514 (see section 3). A bearer-specific property is such that if instantiated, it is  
 3515 necessarily instantiated by some specific entity or entities. In essential unity,  
 3516 it is the property bearer which is such that if it exists, it necessarily instantiates  
 3517 a certain property. In each case, one of the constituents of a fact is such that its  
 3518 existence (or instantiation) requires that it combine with the other constituent  
 3519 or constituents.

3520 There are two ways in which essential unity might be said to explain the  
 3521 unity of some ungrounded facts. One way to understand facts which are  
 3522 essentially unified is that there is no ontological difference between them and  
 3523 the aggregate of their constituents. Since there is no difference, there is no  
 3524 need for any explanation of this difference, and the unity problem dissolves.  
 3525 This reasoning mirrors Betti's explanation for why bearer-specific properties  
 3526 dissolve the problem: “If *R* is relata-specific, and thus it is in the nature of *R*

---

35 In the next section I shall consider ungrounded facts which are not essentially unified. It is worth noting that there may be grounded facts which are essentially unified. The unity of these facts would be over-determined. But it is plausible that the vast majority of grounded facts are not essentially unified. Indeed, as noted in section 2, it is often assumed in the literature on the problem of unity that the constituents of a fact could all exist without together composing that fact.

36 The relevant notion of “essence” is the non-modal conception made familiar by Kit Fine. In particular, I have in mind Fine's notion of *constitutive essence* (1995, 276). Note also that I am not suggesting that essentially unified facts are grounded in essential facts about their constituents. Rather, the proposed explanation of the unity of essentially unified facts is an *essentialist explanation* (Glazier 2017, 2872).

3527 to relate  $a$  and  $b$ , then  $aRb$  exists as soon as  $R$  exists. So, there is simply no  
 3528 difference between  $a + R + b$  and  $aRb$ " (2015, 92).

3529 This way of dissolving the unity problem might be thought to face the  
 3530 following objection: it removes any motivation to think of essentially unified  
 3531 facts as genuinely *facts*, as entities over and above their constituents. This  
 3532 is Betti's own conclusion: bearer-specific properties not only dissolve the  
 3533 problem of unity, but also remove the need for the ontological category of  
 3534 compositional facts (2015, 106).<sup>37</sup>

3535 Even if this point is correct, it is compatible with a way of solving (rather  
 3536 than dissolving) the unity problem. Consider the aggregate of entities which  
 3537 we wrongly took to form an essentially unified fact. Let us term this aggregate  
 3538 a *quasi-fact*. Each quasi-fact will include a number of property-bearers and  
 3539 properties or relations such that each property or relation is essential to the  
 3540 property-bearers. We can then adjust the notion of grounding as follows: a fact  
 3541 can be grounded by another fact, or by a quasi-fact, or by some combination of  
 3542 facts and quasi-facts. Every grounded fact will be unified by its grounds (either  
 3543 facts or quasi-facts); and since there is no difference between a quasi-facts  
 3544 and the aggregate of its components, the problem of unity will not arise for  
 3545 quasi-facts.

3546 That said, I am drawn towards the other way in which essential unity can  
 3547 solve the problem. First, I think there is an ontological difference between  
 3548 essentially unified facts and the aggregate of their constituents, even though  
 3549 the existence of the constituents suffices for these facts to obtain. The aggregate  
 3550 of  $e$  and *being negatively charged* just is  $e$  and this property considered together.  
 3551 It involves nothing other than these two entities; there is nothing more to  
 3552 the aggregate's existence than the existence of these entities. In contrast,  
 3553 the fact that  $e$  is negatively charged involves the instantiation by  $e$  of this  
 3554 property; that is, the fact involves these entities being arranged in a specific  
 3555 way. As it happens, these entities are such that when they exist they are  
 3556 automatically arranged in this way. But this does not entail that there is no  
 3557 ontological difference here. The fact still involves a way of being unified which  
 3558 the aggregate does not.<sup>38</sup>

---

37 This conclusion is questioned by Vallicella (2016a, 236).

38 It might be objected that the difference I am positing between essentially unified facts and the aggregates of their constituents appeals to non-mereological composition, and so begs the question in favour of facts. But it is important to be clear on what is at issue here. As was argued in section 2, we can distinguish between explaining *what it is* for constituents to form a fact (e.g., clarifying the distinctive way in which the constituents must be unified so as to form a fact), and

3559 The problem of unity for an essentially unified fact is solved by some of its  
 3560 own constituents. The problem of unity, recall, is the problem of explaining  
 3561 why, given that each of a fact's constituents exist, they are combined in the  
 3562 way characteristic of this fact. In an essentially unified fact, some of its con-  
 3563 stituents are such that necessarily, if they exist they must instantiate certain  
 3564 properties or stand in certain relations. Therefore, given that each of the fact's  
 3565 constituents exist, it is necessary that they are unified so as to form this fact.  
 3566 For instance, it is in virtue of the essence of *e* that it is unified with the property  
 3567 *being negatively charged*.

3568 This account is in effect a version of Betti's option (A): the explanation of  
 3569 why the constituents are unified lies in the essence of one of the constituents  
 3570 itself. Betti herself rejects this option. Since the unity problem presupposes that  
 3571 the constituents in the aggregate are numerically identical to the constituents  
 3572 of the fact, it seems impossible for the difference between the fact and the  
 3573 aggregate to be explained by reference to any of these constituents (2015, 56).  
 3574 But as argued above, the difference between the fact and the aggregate just is  
 3575 the non-mereological unity of the constituents in the fact, and this unity is  
 3576 explained by the essence of the property bearer.

### 3577 **Brutely Unified Facts**

3578 I have outlined an account of the unity of grounded facts, and of ungrounded  
 3579 facts where the properties are essential to the property-bearers. However, it  
 3580 is plausible that if ungrounded facts obtain, not all of them are essentially  
 3581 unified: for instance, the fact that a fundamental particle stands in a certain  
 3582 spatiotemporal location (Dasgupta 2014, 579), or the fact that a simple entity  
 3583 *a* is *F* (where *F*ness is, e.g., a maximally determinate shade of colour). That is,  
 3584 in addition to grounded facts and facts which are ungrounded and essentially  
 3585 unified, there is at least logical space for a third category of facts, facts such  
 3586 that there is nothing in virtue of which their components are unified. Let us  
 3587 term these *brutely unified facts*.<sup>39</sup>

---

explaining *why* a fact obtains given that its constituents exist. The discussion in this paragraph of the main text concerns the first of these issues, not the second. And as mentioned earlier, in addressing the first of these issues it is legitimate to appeal to non-mereological composition, e.g., to instantiation.

39 Thanks to Francesco Spada and to an anonymous reviewer for drawing my attention to the possibility of such facts. It may be that there are facts which do not belong to any of the three categories I have distinguished (i.e., facts which are ungrounded and not essentially unified, but



3588 The possibility of brutally unified facts raises two issues for the position I  
 3589 wish to defend. The first is the general question of whether such facts are  
 3590 possible; the second is whether allowing for such facts weakens my position  
 3591 compared to other responses to the problem of unity. I shall consider the  
 3592 second issue in the next section; for the remainder of this section, I shall  
 3593 discuss the first.

3594 An assumption made by some in the literature is that if the unity of a  
 3595 (supposed) fact cannot be explained, then we have reason to think that this  
 3596 fact cannot exist (e.g., Vallicella 2000, 248; Betti 2015, 103; Maurin 2015, 201).  
 3597 I do not share this assumption. I think it is true of any fact that we can ask  
 3598 why it obtains or why its constituents are arranged as they are, but if it turns  
 3599 out that a positive answer cannot be provided for certain facts, this does not  
 3600 in itself give us reason to doubt that such facts obtain.

3601 Vallicella offers three arguments against the possibility of brutally unified  
 3602 facts. First, he claims that the view that there are such facts leads to “the  
 3603 contradiction that a fact both is and is not a whole of parts” (2002, 20), i.e., an  
 3604 aggregate of its constituents. The argument is as follows:

3605 A fact *is* a whole of parts in that there is nothing ‘in’ it but its parts.  
 3606 For a fact is a complex, and a complex is composed of constituents.  
 3607 Analysis of *aRb* can yield nothing beyond *a*, *R*, and *b*. A fact is *not*  
 3608 a whole of parts in that the existence of the parts does not entail  
 3609 the existence of the whole. Thus a fact is more than the mere sum  
 3610 of its parts. This ‘more’ is something real, and yet it cannot be, or  
 3611 be grounded in, any further constituent of the fact. [...] it seems to  
 3612 be a contradiction to say of a whole that it is an entity in addition  
 3613 to its parts when it is composed of them. (Vallicella 2002, 20)

3614 The problem with this argument is that it equivocates on the first claim, that  
 3615 “there is nothing ‘in’ a fact but its parts” (i.e., its constituents). This claim could  
 3616 be interpreted as meaning “a fact has no constituent other than its parts, e.g.,  
 3617 *a*, *R* and *b*.” But it could also be interpreted as meaning “a fact is reducible to or  
 3618 nothing over and above its parts”, where this would entail, among other things,  
 3619 that a fact obtains if its parts all exist. Interpreted in the first way, the first  
 3620 claim would be accepted by the proponent of the compositional conception;  
 3621 but interpreted in this way, the first claim does not lead to a contradiction with

---

which are not brutally unified either). That said, it is not obvious what would unify such facts, and so I shall set aside this possible further category.

3622 the second claim, that a fact is more than the aggregate of its parts. Interpreted  
 3623 in the second way, the first claim would lead to a contradiction with the second  
 3624 claim; but interpreted in this way, the first claim would not be accepted by the  
 3625 proponent of the compositional conception. Thus, Vallicella's first argument  
 3626 is either a non-sequitur or it begs the question against the proponent of the  
 3627 compositional conception (by assuming that a fact is nothing over and above  
 3628 its constituents).

3629 Vallicella's second argument starts with two facts, *Fa* and *Gb*, which ex hypo-  
 3630 thesi have no constituent in common. Vallicella notes "each fact is precisely  
 3631 a *fact*, which suggests that they have the universal *being a fact* (facthood) in  
 3632 common" (2002, 21–22). But if they have no constituent in common "then  
 3633 facthood is not a common constituent; how then do we explain the circum-  
 3634 stance that they are *both* facts? How do we explain the common categorical  
 3635 status?" (2002, 22). Since Vallicella thinks it cannot be a brute fact that both  
 3636 are facts, nor can either of these facts itself be a brutally unified fact.

3637 Given a sparse conception of properties (see footnote 2), it is not at all clear  
 3638 that there is any good reason to accept that there is a property *being a fact*,  
 3639 or that there are facts of the form: *Fa* is a fact. There are certainly *truths* of  
 3640 the form "*Fa* is a fact." What explains their being true is precisely the factual  
 3641 ontological structure of *Fa*, i.e., *a*'s instantiating *Fness*.

3642 But assume that there is such a property as the property *being a fact*. Pre-  
 3643 sumably this property will be instantiated by all and only facts, and therefore  
 3644 will be something which all and only facts have in common. But why assume  
 3645 that it must be a *constituent* of every fact? On the contrary, it seems obviously  
 3646 mistaken to assume that a fact such as the apple's being red must be partly  
 3647 composed of the property *being a fact*. The proponent of the compositional  
 3648 conception of facts has no need to assume that properties are constituents  
 3649 of the entities which instantiate them.<sup>40</sup> And this is true even if the entities  
 3650 which instantiate properties are themselves facts.

3651 Vallicella's third argument is as follows:

- 3652 (i) if the difference between a fact and its constituents is a brute  
 3653 fact, then it is possible that two facts share all constituents. (ii) But  
 3654 it is not possible that two facts share all constituents. Therefore,

---

40 This is a well-known view of properties, (e.g., Armstrong 1989, 77), but it is not one which I accept, and more importantly it is not one to which the proponent of the compositional conception is committed.

(iii) the difference between a fact and its constituents is not a brute fact; it has an ontological ground. (2002, 22)

The proponent of the compositional conception will accept neither (i) nor (ii). As regards (i), if a fact is brutally unified then the difference between this fact and the aggregate of its constituents is simply that the fact is a *fact*, that is, it consists of the constituents arranged in a certain way. Vallicella claims,

if a fact's being a fact is what distinguishes it from its constituents, then a fact's being a fact is what ultimately distinguishes it from other facts even if there also happens to be a difference in constituents. Each fact, just in virtue of its being a fact, differs from every other fact. (2002, 23)

But on the compositional conception, this is false. That  $Fa$  is a fact, i.e., a complex of constituents arranged in a specific way, is not what distinguishes it from  $Gb$  (which, after all, is just as much a fact). What distinguishes the two is precisely that they have different constituents. More generally, what distinguishes each fact from the aggregate of its constituents is different to what distinguishes each fact from any other fact (the latter is given by the identity-conditions of facts outlined in section 2).

As regards (ii), it is widely thought possible for certain distinct facts to share the same constituents (as with facts including asymmetric relations—see section 2 above). Vallicella dismisses this response as question-begging against (ii), but this claim is highly doubtful.<sup>41</sup> The general point behind rejecting (ii) is that a fact is composed of constituents unified in some specific way (e.g., a particular instantiating a universal, or two particulars being related in a certain order), and that in certain cases the same constituents can be unified in more than one way, giving rise to distinct facts.<sup>42</sup>

41 In a footnote, Vallicella clarifies that what is question-begging is to appeal to  $aRb$  and  $bRa$ 's being distinct facts in support of the claim that facts obtain (2002, 24, n.51). This specific dialectical move might beg the question, but what I am discussing in this paragraph in the main text is not whether facts obtain, but whether we should accept claim (ii), that it is not possible that distinct facts share all constituents.

42 An alternative counterexample to (ii) appeals to a plausible condition on the criteria of identity for facts, that these criteria are time-indexed (see section 2). If  $a$  instantiates  $F$ ness at  $t$  1, ceases to instantiate it at  $t$  2, and instantiates it again at  $t$  3, it seems perfectly reasonable to say that there are two distinct facts composed of  $a$  and  $F$ ness; one obtained at  $t$  1 and ceased to obtain at  $t$  2, the other obtained at  $t$  3.

## 3688 **8 Comparing Different Accounts of Unity**

3682 The account I offer of the unity of facts has two significant limitations compared to alternatives such as those offered by Betti or Vallicella. First, it is  
 3683 a disunified account, proposing different answers to the problem of unity  
 3684 for different facts (e.g., grounded versus ungrounded); in contrast, Betti and  
 3685 Vallicella each offer a unified account.<sup>43</sup> Second, the account I propose is  
 3686 limited in scope, if it is accepted that there can be ungrounded facts which  
 3687 are not essentially unified (this is the second issue mentioned at the start of  
 3688 the previous section). My account does not provide a positive answer to the  
 3689 question of what unifies these facts, whereas the positions defended by Betti  
 3690 and by Vallicella promise to do so.

3692 Each of these limitations is important, but I do not think that they are  
 3693 decisive. While all facts share the ontological structure described in section 2,  
 3694 there are important differences between, e.g., facts which are grounded and  
 3695 facts which are ungrounded, and between facts which are essentially unified  
 3696 and facts which are not. Once these differences are made clear, the cost of a  
 3697 disunified account is diminished; or, to put it another way, once the differences  
 3698 between facts are made clear, it is less obvious that we should expect to find a  
 3699 single account which explains the unity of each fact.

3700 Furthermore, it seems to me to be a mistake to assume from the outset that  
 3701 the problem of unity can be solved for every fact. Once we acknowledge that  
 3702 there are different types of fact, the possibility is opened that there are facts  
 3703 for which no positive answer can be given to the question “Why does this fact  
 3704 obtain?”. Granted, it is methodologically preferable to be able to explain the  
 3705 unity of each fact. That is, all things being equal, we ought to prefer a theory  
 3706 which allows for a positive answer to each question of this form to one which  
 3707 does not. But are all things equal?

3708 I contend they are not; the account I offer has advantages over the main  
 3709 alternatives. My account relies on grounding and on certain properties being  
 3710 essential to their bearers. While grounding and essential properties are by  
 3711 no means uncontroversial, each is a relatively familiar and well-developed  
 3712 idea, and there are reasons for accepting each idea which are independent  
 3713 of any role they might play with regard to the unity of facts. In contrast, the  
 3714 accounts offered by both Vallicella and Betti rely on ontological posits which

---

43 Though this may not be true of Vallicella’s proposal (see, e.g., 2000, 258, n.45).

3715 have not been widely discussed or systematically clarified, each of which is  
 3716 ad hoc, and each of which faces independent considerations against it.

3717 To develop this point, consider first some of the problems facing Betti's  
 3718 ontological posit, bearer-specific properties. First, on Betti's view the identities  
 3719 of properties are implausibly fragile. For instance, consider two entities, *a* and  
 3720 *b*, each of which instantiates the property *being the determinate shade of red*  
 3721 *x*. Now consider a counterfactual situation where *a* does not exist. On Betti's  
 3722 view, in this counterfactual situation *b* would not instantiate the property  
 3723 *being the determinate shade of red x*, since that property can only exist if it  
 3724 is instantiated by *a*. Rather, in that situation *b* would instantiate the distinct  
 3725 (though presumably qualitatively identical) property, *being the determinate*  
 3726 *shade of red x\**. This is surely the wrong result; it seems to me that I can  
 3727 understand what it would be for that very property, *being the determinate*  
 3728 *shade of red x*, to exist and to be instantiated in a situation where *a* did not  
 3729 exist.<sup>44</sup>

3730 Second, Betti's position entails that our knowledge of what properties are  
 3731 is constrained to an implausible degree. Her view requires that one can only  
 3732 be said to know which property is in fact instantiated if one knows each  
 3733 and every entity which bears it (e.g., we can only know we are dealing with  
 3734 the property *being the determinate shade of red x* and not the property *being*  
 3735 *the determinate shade of red x\** if we know that *a* exists). Again, this seems  
 3736 implausible.<sup>45</sup>

3737 Vallicella appeals to a single entity, *U*, to unify all contingently unified facts.  
 3738 He assumes that *U* cannot necessarily unify these facts, as this would mean  
 3739 they were not contingently unified. Therefore, *U* must contingently unify  
 3740 them. As Vallicella puts it,

3741 *U* must have the power of contingent self-determination: it must  
 3742 have the power to contingently determine itself as operating upon  
 3743 its operand. In other words, if *U* is the ground of the contingent  
 3744 unity of a fact's constituents, then *U* contingently *grounds its*  
 3745 *grounding* of the unity of the fact's constituents. (2000, 255)

---

44 This point is even clearer if one accepts that there are determinate quantitative properties, e.g., *being the determinate length x*.

45 Bearer-specific properties would not be so controversial if they were assumed to be tropes; however, as mentioned in section 3 Betti rejects this assumption. Furthermore, ruling out any universal properties or relations brings its own problems (see Armstrong 1989; Lowe 2006).

3746 Vallicella's model for this contingent self-determination is our own free will.  
 3747 Specifically, he suggests that the contents of our thoughts are unified in con-  
 3748 scious acts, as when one judges that *a* is *F* (2000, 255). Indeed, the entities  
 3749 Vallicella proposes as candidates to play the role of *U* are God and transcen-  
 3750 dental consciousness (2000, 252–253).

3751 There is an ambiguity in this account as it stands. Consider one's unifying  
 3752 the contents of a specific thought (say, that *a* is *F*) in an act of judging. What  
 3753 is the unifier here? One's ability to judge is not sufficient to explain the unity  
 3754 of the contents of this thought, since one could exercise this ability without  
 3755 thinking that very thought. Alternatively, the unifier could be a particular  
 3756 exercise of this capacity, e.g., a particular act of judging (that *a* is *F*). But this  
 3757 answer immediately leads to a further problem. A particular act of judging  
 3758 will involve either oneself standing in a relationship to something, e.g., the  
 3759 contents of one's act of judging, or it will involve one instantiating a specific  
 3760 property, e.g., the property *judging that a is F*. Either of these will involve  
 3761 the obtaining of a fact, and furthermore this fact will be contingent. What  
 3762 explains the obtaining of such facts? (Note that the answer cannot be "one's  
 3763 power to freely judge"—what is being asked for is an explanation of one's  
 3764 exercising this power on a specific occasion.) Similarly, *U* may have the power  
 3765 of contingent self-determination, but its having this power is not sufficient to  
 3766 explain the unity of each fact; what is also needed is an explanation of why  
 3767 this power is exercised as it is.<sup>46</sup>


3768 None of this is to suggest that the accounts offered by Betti or by Vallicella  
 3769 cannot work, or that their posits cannot be ultimately defended. But each of  
 3770 their accounts faces serious theoretical problems. The account I offer, though  
 3771 limited in important respects, relies on more familiar and well-established  
 3772 ontological ideas. For this reason, it deserves to be taken as seriously as any  
 3773 other proposed solution to the problem of unity.\*

---

46 Vallicella notes that if *U* is God, then on a standard conception God necessarily has His attributes, e.g., He is necessarily omniscient (2000, 258, n.45). But God presumably does not necessarily have the property *judging that a is F*; nor does He necessarily stand in a unifying relation to the fact that *a* is *F* (and if He did, the fact that *a* is *F* would not be contingent).

\* Thanks to Francesco Spada, Jani Hakkarainen, Antti Tiainen, Arianna Betti, Tuomas Tahko, Henrik Rydén, Michael O'Sullivan and a number of referees for their comments on various drafts of this paper. Thanks also to audiences at the Université de Neuchâtel philosophy colloquium (December 2016) and the Dynamis workshop at the University of Tampere (March 2017) for discussion. My work on this paper was supported by the Academy of Finland (Kulttuurin ja Yhteiskunnan Tutkimuksen Toimikunta, grant number 274715) and the Swiss National Science Foundation (grant number 166320).

Donnchadh O'Conaill

 0000-0003-2747-6693

Université de Fribourg

donnchadh.oconaill@unifr.ch

3774  
3775  
3776  
37773778 **References**

- 3779 ARMSTRONG, David M. 1989. *Universals: An Opinionated Introduction*. Boulder, Col-  
3780 orado: Westview Press.
- 3781 —. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press, doi:10  
3782 .1017/cbo9780511583308.
- 3783 AUDI, Paul. 2012. "Grounding: Toward a Theory of the *In-Virtue-Of* Relation." *The*  
3784 *Journal of Philosophy* 109(12): 685–711, doi:10.5840/jphil20121091232.
- 3785 —. 2015. "Explanation and Explication." in *The Palgrave Handbook of Philosophi-*  
3786 *cal Methods*, edited by Christopher John DALY, pp. 208–230. London: Palgrave  
3787 Macmillan, doi:10.1057/9781137344557\_9.
- 3788 BENNETT, Karen. 2013. "Having a Part Twice Over." *Australasian Journal of Philosophy*  
3789 91(1): 83–103, doi:10.1080/00048402.2011.637936.
- 3790 BETTI, Arianna. 2015. *Against Facts*. Cambridge, Massachusetts: The MIT Press, doi:10  
3791 .7551/mitpress/9780262029216.001.0001.
- 3792 BLACKBURN, Simon. 2005. *Truth: A Guide for the Perplexed*. London: Penguin Books.
- 3793 BLISS, Ricki Leigh. 2013. "Viciousness and the Structure of Reality." *Philosophical*  
3794 *Studies* 166(2): 399–418, doi:10.1007/s11098-012-0043-0.
- 3795 CAMERON, Ross P. 2016. "Do we need Grounding?" *Inquiry* 59(4): 382–397, doi:10.108  
3796 0/0020174x.2015.1128848.
- 3797 CLARK, Romane. 1988. "Vicious Infinite Regress Arguments." in *Philosophical Per-*  
3798 *spectives 2: Epistemology*, edited by James E. TOMBERLIN, pp. 369–380. Atascadero,  
3799 California: Ridgeview Publishing Co., doi:10.2307/2214081.
- 3800 CORREIA, Fabrice and SKILES, Alexander. 2019. "Grounding, Essence, and Identity."  
3801 *Philosophy and Phenomenological Research* 98(3): 642–670, doi:10.1111/phpr.124  
3802 68.
- 3803 DASGUPTA, Shamik. 2014. "The Possibility of Physicalism." *The Journal of Philosophy*  
3804 111(9/10): 557–592, doi:10.5840/jphil20141119/1037.
- 3805 DIXON, Scott. 2018. "Plural Slot Theory." in *Oxford Studies in Metaphysics*, volume XI,  
3806 edited by Karen BENNETT and Dean W. ZIMMERMAN, pp. 193–223. New York:  
3807 Oxford University Press, doi:10.1093/oso/9780198828198.003.0006.
- 3808 DODD, Julian. 1999. "Farewell to States of Affairs." *Australasian Journal of Philosophy*  
3809 77(2): 146–160, doi:10.1080/00048409912348901.
- 3810 EKLUND, Matti. 2019. "Regress, Unity, Facts, and Propositions." *Synthese* 196(4):  
3811 1225–1247, doi:10.1007/s11229-016-1155-4.



- 3812 FINE, Kit. 1982. "First-Order Modal Theories III – Facts." *Synthese* 53(1): 43–122,  
3813 doi:10.1007/BF00500112.
- 3814 —. 1995. "Ontological Dependence." *Proceedings of the Aristotelian Society* 95: 269–290,  
3815 doi:10.1093/aristotelian/95.1.269.
- 3816 —. 2012. "Guide to Ground." in *Metaphysical Grounding: Understanding the Structure of*  
3817 *Reality*, edited by Fabrice CORREIA and Benjamin Sebastian SCHNIEDER, pp. 37–80.  
3818 Cambridge: Cambridge University Press, doi:10.1017/cbo9781139149136.002.
- 3819 GLAZIER, Martin. 2017. "Essentialist Explanation." *Philosophical Studies* 174(11):  
3820 2871–2889, doi:10.1007/s11098-016-0815-z.
- 3821 HALL, Ned. 2004. "Two Concepts of Causation." in *Causation and Counterfactuals*,  
3822 edited by John David COLLINS, Ned HALL, and Laurie A. PAUL, pp. 225–276.  
3823 Cambridge, Massachusetts: The MIT Press, doi:10.7551/mitpress/1752.003.0010.
- 3824 LEWIS, David. 1986. "Against Structural Universals." *Australasian Journal of Philoso-*  
3825 *phy* 64(1): 25–46. Reprinted in Lewis (1999, 78–107), doi:10.1080/00048408612342  
3826 211.
- 3827 —. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University  
3828 Press, doi:10.1017/cbo9780511625343.
- 3829 LOWE, Edward Jonathan. 2006. *The Four-Category Ontology: A Metaphysical Founda-*  
3830 *tion for Natural Science*. Oxford: Oxford University Press, doi:10.1093/0199254397  
3831 .001.0001.
- 3832 MAURIN, Anna-Sofia. 2007. "Infinite Regress – Virtue or Vice?" in *Hommage à*  
3833 *Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*, edited  
3834 by Toni RÖNNOW-RASMUSSEN, Björn PETERSSON, Jonas JOSEFSSON,  
3835 and Dan EGONSSON. Lund: Lunds Universitet, Filosofiska Institutionen,  
3836 <https://www.fil.lu.se/hommageawlodek/site/papper/MaurinAnnaSofia.pdf>.
- 3837 —. 2015. "States of Affairs and the Relation Regress." in *The Problem of Universals in*  
3838 *Contemporary Philosophy*, edited by Gabrielle GALLUZZO and Michael J. LOUX,  
3839 pp. 195–214. Cambridge: Cambridge University Press, doi:10.1017/cbo978131618  
3840 1539.010.
- 3841 MEINERTSEN, Bo Rode. 2008. "A Relation as the Unifier of States of Affairs." *Dialectica*  
3842 62(1): 1–19, doi:10.1111/j.1746-8361.2007.01127.x.
- 3843 NOLAN, Daniel Patrick. 2001. "What's Wrong with Infinite Regresses?" *Metaphilosophy*  
3844 32(5): 523–538, doi:10.1111/1467-9973.00206.
- 3845 ORILIA, Francesco. 2006. "States of Affairs: Bradley vs. Meinong." in *Meinongian*  
3846 *Issues in Contemporary Italian Philosophy*, edited by Venanzio RASPA, pp. 213–238.  
3847 Meinong Studies / Meinong Studien n. 2. Heusenstamm b. Frankfurt: Ontos Verlag,  
3848 doi:10.1515/9783110321128.213.
- 3849 PEROVIC, Katarina. 2016. "Neo-Armstrongian Defence of States of Affairs: A Reply to  
3850 Vallicella (2016b)." *Metaphysica* 17(2): 143–161, doi:10.1515/mp-2016-0010.
- 3851 RAVEN, Michael J. 2012. "In Defence of Ground." *Australasian Journal of Philosophy*  
3852 90(4): 687–701, doi:10.1080/00048402.2011.616900.



- 3853 —. 2013. “Is Ground a Strict Partial Order?” *American Philosophical Quarterly* 50(2):  
3854 193–201.
- 3855 RODRIGUEZ-PEREYRA, Gonzalo. 1998. “Searle’s Correspondence Theory of Truth and  
3856 the Slingshot.” *The Philosophical Quarterly* 48(193): 513–522, doi:10.1111/1467-  
3857 9213.00119.
- 3858 —. 2015. “Grounding is Not a Strict Order.” *Journal of the American Philosophical*  
3859 *Association* 1(3): 517–534, doi:10.1017/apa.2014.22.
- 3860 SCHAFFER, Jonathan. 2016a. “Ground Rules: Lessons from Wilson.” in *Scientific Com-*  
3861 *position and Metaphysical Ground*, edited by Kenneth AIZAWA and Carl GILLET, pp.  
3862 143–170. New Directions in the Philosophy of Science. London: Palgrave  
3863 Macmillan, doi:10.1057/978-1-137-56216-6\_6.
- 3864 —. 2016b. “It is the Business of Laws to Govern.” *Dialectica* 70(4): 577–588, doi:10.111  
3865 1/1746-8361.12165.
- 3866 SKILES, Alexander. 2015. “Against Grounding Necessitarianism.” *Erkenntnis* 80(4):  
3867 717–751, doi:10.1007/s10670-014-9669-y.
- 3868 STRAWSON, Peter Frederick. 1950. “Truth.” *Proceedings of the Aristotelian Society,*  
3869 *Supplementary Volume* 24: 129–156. Reprinted in Strawson (1971, 190–213) and  
3870 Strawson (2004, 147–164), doi:10.1093/aristoteliansupp/24.1.111.
- 3871 —. 1971. *Logico-Linguistic Papers*. London: Methuen & Co. Reprinted as Strawson  
3872 (2004), doi:10.4324/9781351153607.
- 3873 —. 2004. *Logico-Linguistic Papers*. 2nd ed. London: Routledge, doi:10.4324/97813152  
3874 50250.
- 3875 TROGDON, Kelly. 2018. “Grounding-Mechanical Explanation.” *Philosophical Studies*  
3876 175(6): 1289–1309, doi:10.1007/s11098-017-0911-8.
- 3877 VALLICELLA, William F. 2000. “Three Conceptions of States of Affairs.” *Noûs* 34(2):  
3878 237–259, doi:10.1111/0029-4624.00209.
- 3879 —. 2002. “Relations, Monism, and the Vindication of Bradley’s Regress.” *Dialectica*  
3880 56(1): 3–36, doi:10.1111/j.1746-8361.2002.tb00227.x.
- 3881 —. 2016a. “Facts: An Essay in Aporetics.” in *Metaphysics and Scientific Realism: Essays*  
3882 *in Honor of David Malet Armstrong*, edited by Francesco Federico CALEMI, pp.  
3883 105–132. EIDE – Foundations of Ontology n. 9. Berlin: de Gruyter, doi:10.1515/97  
3884 83110455915-007.
- 3885 —. 2016b. “The Case against Facts [review of Betti (2015)].” *Metaphysica* 17(2): 229–241,  
3886 doi:10.1515/mp-2016-0017.
- 3887 WATZL, Sebastian. 2017. *Structuring Mind: The Nature of Attention and How it Shapes*  
3888 *Consciousness*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/97801996  
3889 58428.001.0001.
- 3890 WIELAND, Jan Willem. 2013. “Infinite Regress Arguments.” *Acta Analytica* 28(1):  
3891 95–109, doi:10.1007/s12136-012-0165-1.

- 3892 WIELAND, Jan Willem and BETTI, Arianna. 2008. "Relata-Specific Relations: A Re-  
3893 sponse to Vallicella (2002)." *Dialectica* 62(4): 509–524, doi:[10.1111/j.1746-8361.200](https://doi.org/10.1111/j.1746-8361.2008.01167.x)  
3894 [8.01167.x](https://doi.org/10.1111/j.1746-8361.2008.01167.x).
- 3895 WILSCH, Tobias. 2015. "The Nomological Account of Ground." *Philosophical Studies*  
3896 172(12): 3293–3312, doi:[10.1007/s11098-015-0470-9](https://doi.org/10.1007/s11098-015-0470-9).
- 3897 WILSON, Jessica M. 2014. "No Work for a Theory of Grounding." *Inquiry* 57(5/6):  
3898 535–579, doi:[10.1080/0020174x.2014.907542](https://doi.org/10.1080/0020174x.2014.907542).

PROOF

3900  
3901

# Strevens's Counterexample to Lewis's "Causation as Influence", and Degrees of Causation

HSENG XIONG JOSHUA GOH

3902 Sungho Choi has criticised Michael Strevens's counterexample to David  
3903 Lewis's final theory of "token" causation, causation as "influence." I argue  
3904 that, even if Choi's points are correct, Strevens's counterexample remains  
3905 useful in revealing a shortcoming of Lewis's theory. This shortcoming is  
3906 that Lewis's theory does not properly account for *degrees* of causation.  
3907 That is, even if Choi's points are correct, Lewis's theory does not capture  
3908 an intuition we have about the *comparative* causal statuses of those  
3909 events involved in Strevens's counterexample (we might, for example,  
3910 intuit that Sylvie's ball-firing is *as much/more/less* a cause of the jar's  
3911 shattering as/than is Bruno's ball-firing).

3912 Sungho Choi (2005, 106–113) has criticised Michael Strevens's (2003, 4–7,  
3913 11–17) counterexample to David Lewis's (2000) final theory of "token" causa-  
3914 tion, causation as "influence" (hereafter, "CaI"). I argue that, even if Choi's  
3915 points are correct, Strevens's counterexample remains useful in revealing a  
3916 shortcoming of CaI. This shortcoming is that CaI does not properly account  
3917 for *degrees* of causation. This paper proceeds as follows. Section 1 articulates  
3918 CaI. Section 2 articulates Strevens's counterexample to CaI, and Choi's criti-  
3919 cism of Strevens's counterexample. Section 3 argues that, even if Choi's points  
3920 are correct, CaI does not capture an intuition we have about the *comparative*  
3921 causal statuses of those events involved in Strevens's counterexample (we  
3922 might, for example, intuit that Sylvie's ball-firing is *as much/more/less* a cause  
3923 of the jar's shattering as/than is Bruno's ball-firing).

3921 **CaI**

3925 CaI involves three ideas. The first idea is the “alteration” of an event. Consider  
 3926 this event  $E$ : the vase’s shattering. Lewis defines an “alteration” of  $E$  as “either  
 3927 a very fragile *version* of  $E$  or else a very fragile *alternative event* that is similar  
 3928 to  $E$ , but numerically different from  $E$ ” (2000, 188, emphasis mine).

3929 To elucidate, an event is considered “fragile” if we impose stringent condi-  
 3930 tions for its occurrence (if we say that any change in one of its details turns  
 3931 it into a numerically different event) (Lewis 2000, 185–186). One alteration  
 3932 of  $E$  is  $E$ ’s *actual* alteration: exactly when and how the vase shattered. The  
 3933 other alterations of  $E$  are un-actualised (one example: the vase shattering one  
 3934 millisecond later, and into more pieces).

3935 The second idea is “influence.” Let  $C$  and  $E$  be two single, distinct, actual  
 3936 events. Lewis holds that  $C$  “influences”  $E$  iff

3937 there is a substantial range  $C_1, C_2, \dots$  of different not-too-distant  
 3938 alterations of  $C$  (including the actual alteration of  $C$ ) and there is  
 3939 a range  $E_1, E_2, \dots$  of alterations of  $E$ , at least some of which differ,  
 3940 such that if  $C_1$  had occurred,  $E_1$  would have occurred, and if  $C_2$   
 3941 had occurred,  $E_2$  would have occurred, and so on. (Lewis 2000,  
 3942 190)

3943 Idea three concerns the relationship between influence and causation. Ac-  
 3944 cording to Lewis,  $C$  is a cause of  $E$  iff  $C$  directly influences  $E$ , or there is  
 3945 a chain of stepwise influence (hereafter, “I-CHAIN”) leading from  $C$  to  $E$  (that  
 3946 is, a sequence of (actual) events  $C, D_1, D_2, \dots, D_n, E$ , such that  $C$  influences  
 3947  $D_1$ ,  $D_1$  influences  $D_2$ , ...,  $D_{(n-1)}$  influences  $D_n$ , and  $D_n$  influences  $E$ ) (Lewis  
 3948 2000, 191; see also Lewis 1973, 563).

3949 Let’s observe CaI in action. Consider this scenario: Sylvie throws a rock  
 3950 at a vase. Beside her, Bruno laughs. Here, CaI delivers the intuitive result  
 3951 that Sylvie’s throw is a cause of the vase’s shattering, while Bruno’s laughter  
 3952 is not. This is because Sylvie’s throw has substantial direct influence on the  
 3953 vase’s shattering. That is, there are many different, not-too-distant alterations  
 3954 of Sylvie’s throw (e.g. her throwing one millisecond later/with slightly more  
 3955 force) upon which alterations in the vase’s shattering (i.e. the vase’s shattering  
 3956 one millisecond later/into more pieces) counterfactually depend. Bruno’s  
 3957 laughter, however, has no substantial direct influence on the vase’s shattering.  
 3958 *Maybe* one distant alteration of Bruno’s laughter is so infectious that it delays  
 3959 Sylvie’s throw (and hence, the vase’s shattering) by a second. Nevertheless,

3960 no not-too-distant alteration of Bruno's laughter appears to alter the vase's  
 3961 shattering.<sup>1</sup> Moreover, one cannot identify any I-CHAIN leading from Bruno's  
 3962 laughter to the vase's shattering.

## 3962 2 **Strevens's counterexample to CaI; Choi's criticism**

3964 Here is Strevens's counterexample to CaI (2003, 4–7, 11–17):

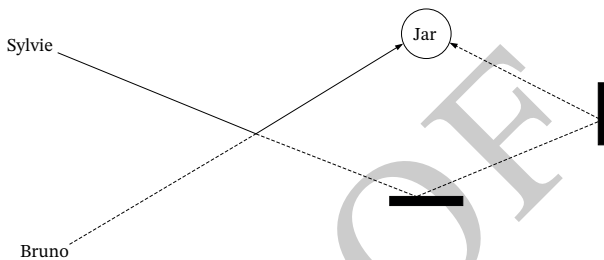


Figure 1: Solid line: actual trajectory of Sylvie's ball. Dotted line: actual trajectory of Bruno's ball.

3965 SCE. At time  $t_1$ , and using identical rifles, Sylvie and Bruno fire  
 3966 at a jar intrinsically identical, minute lead balls. Sylvie, who never  
 3967 misses, shoots so that her ball will ricochet two times prior to striking  
 3968 the jar. Bruno shoots directly at the jar. The balls, however, collide in  
 3969 mid-air at time  $t_c$ . Consequently, they *perfectly* exchange trajectories  
 3970 and spin (we thus take the motion of the balls to be that of two point  
 3971 particles; this admittedly requires something like a fortuitous gust  
 3972 of wind at  $t_c$ ) (2003, 5, fn. 2). Stipulate moreover that the speeds of  
 3973 the two balls are always identical (and extremely high). Ultimately,  
 3974 Sylvie's ball shatters the jar, and Bruno's ricochets, then flies through  
 3975 thin air.<sup>2</sup>

1 Unfortunately, Lewis is vague about what it takes for an alteration of an event to qualify as “not-too-distant.” He says that, for some particular alteration of an event, whether or not we think it to be “not-too-distant” may be a matter of “mood” (2000, 197).

2 Strevens, I think, mistakenly calls SCE a case of “late cutting” pre-emption (2003, 17, fn. 11). Standard late cutting involves the following: an effect; one pre-empting cause; one (non-causal)

3976 Let SF stand for Sylvie’s firing, BF for Bruno’s firing, and JS for the jar’s  
 3977 shattering. For two reasons, Strevens argues that CaI delivers this *unintuitive*  
 3978 result: SF is not *at all* a cause of JS. First, SF has no substantial direct influence  
 3979 on JS (2003, 4–5, 12–13). After all, hold fixed BF, and consider an alteration of  
 3980 SF in which Sylvie fires one millisecond earlier/later, or one in which her rifle  
 3981 points one degree to the left/right. Given the properties of both balls, these  
 3982 alterations result in: no collision → Bruno’s ball striking the jar (before Sylvie’s  
 3983 ball finishes ricocheting) → *no* alteration to JS. Second, there appears no I-  
 3984 CHAIN leading from SF to JS (2003, 5–7, 13–14). This second point, however,  
 3985 is where Choi (2005, 110–113) most seriously disagrees.

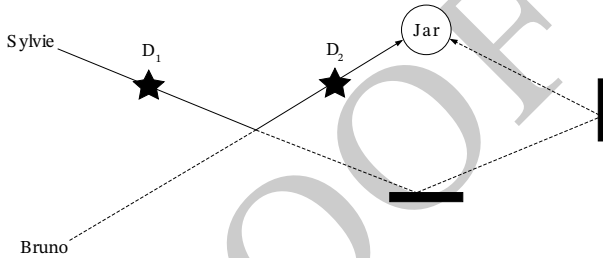


Figure 2:

3986 Referring to Figure 2, and using both Choi’s and Lewis’s terminology (Choi  
 3987 2005, 110–111; Lewis 1986a, 244–249), let  $D_1$  and  $D_2$  be the (fragile) events  
 3988 whose occurrence conditions consist of all the intrinsic and spatio-temporal  
 3989 properties satisfied by the region that Sylvie’s ball occupies at, for  $D_1$ , time  $t_2$   
 3990 before  $t_c$ , and for  $D_2$ , time  $t_3$  after  $t_c$ .

3991 Strevens claims that  $D_1$  has no substantial influence on JS. After all, alter,  
 3992 say, the spatio-temporal properties of Sylvie’s ball at  $t_2$ . This results in: no  
 3993 collision → no alteration to JS. Strevens also claims: SF has no substantial  
 3994 influence on  $D_2$ . After all, alter, say, the timing, or direction of SF. This results  
 3995 in: no collision → the occurrence condition of  $D_2$  being satisfied by Bruno’s  
 3996 ball (Strevens notes that, on Lewis’s metaphysics, it isn’t a violation of the  
 3997 occurrence condition of  $D_2$  if the ball at  $D_2$ ’s spatio-temporal region loses the  
 3998 property of “belonging to” Sylvie (2003, 7); said property, after all, is extrinsic).

pre-empted alternative (see Lewis 2000, 182–184). SCE involves an effect that has, intuitively,  
*two causes*.

3999 Choi, however, claims that Strevens is twice mistaken. (i)  $D_1$  *does* influence  
 4000 JS. After all, alter the *mass*, or *shape*, of Sylvie’s ball at  $t_2$ . Admittedly, if  $t_2$  were,  
 4001 say, right after  $t_1$ , then these alterations result in: Sylvie’s ball taking a different  
 4002 post- $t_2$  trajectory (balls of different mass/shape encounter different amounts  
 4003 of air resistance)  $\rightarrow$  no collision. However, stipulate that  $t_2$  is *right before*  $t_c$ .  
 4004 Then, neither alteration prevents the balls’ collision. Both, however, alter the  
 4005 manner of the collision, and resultantly the manner of JS. Furthermore, (ii)  
 4006 SF *does* influence  $D_2$ . After all, alter the *surface properties*, or *electrical charge*,  
 4007 of the ball Sylvie fires. Neither alteration prevents the balls’ collision. Both,  
 4008 however, in altering an intrinsic property of Sylvie’s ball at  $t_3$ , alter  $D_2$ .

4009 Combining (i), the fact that  $D_1$  influences JS, with the (safe) claim that SF  
 4010 influences  $D_1$ , and combining (ii), the fact that SF influences  $D_2$ , with the  
 4011 (safe) claim that  $D_2$  influences JS, Choi concludes that there are (at least) two  
 4012 I-CHAINS leading from SF to JS—one “via”  $D_1$  (I-CHAIN<sub>1</sub>), and one “via”  $D_2$   
 4013 (I-CHAIN<sub>2</sub>). Thus, CaI delivers the intuitive result that SF is a cause of JS, and  
 4014 “[SCE] spells no trouble whatsoever for [CaI]” (2005, 113).

### 4015 3 CaI, SCE, and Degrees of Causation

4016 I think, however, that even if Choi’s points are correct, SCE still spells some  
 4017 trouble for CaI. In what follows, I argue that, even if Choi’s points are cor-  
 4018 rect, CaI does not capture an intuition we have about the *comparative* causal  
 4019 statuses of SF and BF. Thus, insofar as my argument succeeds, SCE remains  
 4020 useful in revealing the failure of CaI to properly account for *degrees* of causa-  
 4021 tion.<sup>3 4</sup>

4022 Here is the intuition I have in mind:

- 
- 3 In the contemporary literature, there exists the idea that CaI can account for, or at least play a role in our understanding of, degrees of causation. Lewis himself, for example, thinks that degrees of causation track degrees of influence (2000, 191). Another example is found in Woodward (2010). Woodward doesn’t find CaI promising as an analysis of “causation *simpliciter*” (2010, 304). Nevertheless, he suggests that CaI can play a role in “distinguish[ing] [...] *among* causal relationships” (2010, 304). In more detail, Woodward connects the “specificity” of causal relationships in biological contexts to influence (2010, 301–308). And while he doesn’t explicitly state that degrees of causation track degrees of “specificity”, he does state that where  $C_1$  and  $C_2$  are both causes of some effect  $E$ , if the causal relationship between  $C_1$  and  $E$  is more “specific” as compared to the causal relationship between  $C_2$  and  $E$ , then possibly we are justified if we “single out or ‘privilege’ the causal role of [ $C_1$ ]” (2010, 316). See also Braham and van Hees (2009, 331, n16), who discuss one point of similarity between their measure of degrees of causation, and CaI.
- 4 There is another scenario in which, even if Choi’s points are correct, SCE spells trouble for CaI. Say we modify SCE so that both balls detect and decimate balls that aren’t intrinsically similar to

COMPARATIVE INTUITION. SF is (*at least*) as much a cause of JS as is BF.<sup>5</sup>

I think that Comparative Intuition is, and should be, held as strongly as is the (absolute) intuition that SF is a cause of JS. A question arises: what buttresses our intuition in Section 1 that Sylvie’s rock-throw is a cause of the vase’s shattering, while Bruno’s laughter is not? One answer is the following: informed (only) of Sylvie’s rock-throw, I can *predict*, *explain*, and *blame* someone for the vase’s shattering. Informed (only) of Bruno’s laughter, I can do none of these things. However, and to use Jonathan Schaffer’s terminology, note that “the core *epistemic*, *explanatory*, and *ethical* connotations of causation” (2001, 12–13, emphasis mine) are no *more* present in the claim that “BF caused JS,” than they are in the claim that “SF caused JS.” Suppose the jar were a national treasure. First, and to endorse Lewis’s view that we *don’t* ordinarily consider events fragile (2000, 185–186; 1986b, 198), comparing a scenario in which I’m informed (only) of BF with one in which I’m informed (only) of SF, it’s not as if I can only predict JS (here taken as a non-fragile event) in the former. Second, consider the question, “Why did the jar shatter?” It is likely that most would find the answer “Because Sylvie fired” to be no more lacking than the answer “Because Bruno fired.” Third, it’d be surprising if Judge blamed Bruno more than she did Sylvie. More likely, liability for the jar’s damages would be apportioned equally.

Nevertheless, two considerations might motivate

COUNTER INTUITION. BF is more a cause of JS than is SF.

Consideration<sub>1</sub> is this asymmetry: had Sylvie not fired, nothing about JS would have changed. However, had Bruno not fired, the jar would’ve shattered slightly later, and in a slightly different manner. Consideration<sub>2</sub> is that JS occurred at a time, and in a manner more (and, in fact, exactly) in line with Bruno’s, rather than Sylvie’s, intention.

If, however, Consideration<sub>1</sub> and Consideration<sub>2</sub> are what motivate Counter Intuition, then Counter Intuition is misleading. Consider this scenario:

---

them. Then,  $D_1$ ’s influence on JS, and SF’s influence on  $D_2$ , are eliminated. Consequently, CaI must deliver the *unintuitive* result that SF is not *at all* a cause of JS.

5 One may worry that, as stated, Comparative Intuition (absurdly) implies that JS was caused twice over (once by SF, and once by BF). If so, one may read Comparative Intuition as saying that SF and BF contributed to *the* causing of JS to the same degree. On this reading, “degrees of causation” should be read as “degrees of causal contribution” (see [Kaiserman 2016, 387–389](#)).



4053 UNLUCKY PRESIDENT. At time  $t_1$ , Assassin<sub>H</sub> and Assassin<sub>R</sub> poison  
 4054 President's coffee. Assassin<sub>H</sub> uses poison  $H$ , which will induce heart  
 4055 failure at time  $t_4$ . Assassin<sub>R</sub> uses poison  $R$ , which will induce res-  
 4056 piratory failure at time  $t_5$ . At time  $t_2$ , President drinks her coffee.  
 4057 At time  $t_3$ , however, poison  $H$  and poison  $R$  interact in President's  
 4058 system—poison  $H$  neutralises the respiratory-failure-inducing ele-  
 4059 ments of poison  $R$ ; poison  $R$  neutralises the heart-failure-inducing  
 4060 elements of poison  $H$ . But President isn't so lucky—she happens to  
 4061 be fatally allergic to some other element  $e$  of poison  $H$ . Element  $e$   
 4062 induces in President respiratory failure at  $t_5$ , and she dies.

4063 Considerations parallel to Consideration<sub>1</sub> and Consideration<sub>2</sub> are present in  
 4064 Unlucky President. In Unlucky President, we have Consideration<sub>1</sub><sup>\*</sup>, which  
 4065 is this asymmetry: had Assassin<sub>H</sub> not poisoned President's coffee, nothing  
 4066 about President's death would have changed. However, had Assassin<sub>R</sub> not  
 4067 poisoned President's coffee, President would've succumbed to heart failure  
 4068 at  $t_4$ , and not respiratory failure at  $t_5$ . In Unlucky President, we also have  
 4069 Consideration<sub>2</sub><sup>\*</sup>: President's death occurs at a time, and in a manner more  
 4070 (and, in fact, exactly) in line with Assassin<sub>R</sub>'s, rather than Assassin<sub>H</sub>'s, in-  
 4071 tention. However, does either Consideration<sub>1</sub><sup>\*</sup> or Consideration<sub>2</sub><sup>\*</sup> push us  
 4072 to think that "Assassin<sub>R</sub>'s poisoning caused President's death"? No. Most in-  
 4073 tuitively, Assassin<sub>H</sub>'s poisoning caused President's death. This shows that  
 4074 considerations like Consideration<sub>1</sub> and Consideration<sub>2</sub> aren't substantially rel-  
 4075 evant to causation. Thus, if Counter Intuition is motivated by Consideration<sub>1</sub>  
 4076 and Consideration<sub>2</sub>, then Counter Intuition should be suppressed.

4077 Comparative Intuition, then, is justifiably strong. But I now argue that CaI  
 4078 violates this intuition: it counts SF as (significantly) *less* a cause of JS than is  
 4079 BF.

4080 What determines how much a cause BF is of JS? On CaI, it is (roughly) the  
 4081 amount of influence that BF has on JS (Lewis 2000, 92). What determines  
 4082 this amount? Centrally, it is the size of the range of alterations to BF that lead  
 4083 to changes in JS. Accounting for those types of alterations that Strevens *and*  
 4084 Choi consider, there are (at least) *four* types of alterations to BF that lead to  
 4085 said changes: alterations to the *timing* and *direction* of BF, and to the *mass*  
 4086 and *shape* of the ball Bruno fires.

4087 What determines how much a cause SF is of JS? Because SF has no sub-  
 4088 stantial direct influence on JS,<sup>6</sup> CaI must appeal to I-CHAIN<sub>1</sub>/I-CHAIN<sub>2</sub>. For  
 4089 each of these I-CHAINS, however, CaI is silent on whether the determinant is  
 4090 (A) the amount of influence that SF has on  $D_1/D_2$  (the amount of influence  
 4091 present in “link”<sub>1</sub> of the I-CHAIN), (B) the amount of influence that  $D_1/D_2$   
 4092 has on JS (the amount of influence present in “link”<sub>2</sub> of the I-CHAIN), or (C)  
 4093 some weighted average of [(A)+(B)]. Nevertheless, let’s first determine (A)  
 4094 and (B):

4095 “LINK”<sub>1</sub> OF I-CHAIN<sub>1</sub> (At least). *six* types of alterations to SF lead to  
 4096 changes in  $D_1$  (alterations to the *timing* and *direction* of SF, and to  
 4097 the *mass, shape, surface properties* and *electrical charge* of the ball  
 4098 Sylvie fires);

4099 “LINK”<sub>2</sub> OF I-CHAIN<sub>1</sub> (At least). *two* types of alterations to  $D_1$  lead to  
 4100 changes in JS (alterations to the *mass* and *shape* of the ball at  $D_1$ ’s  
 4101 spatio-temporal region);

4102 “LINK”<sub>1</sub> OF I-CHAIN<sub>2</sub> (At least). *two* types of alterations to SF lead  
 4103 to changes in  $D_2$  (alterations to the *surface properties* and *electrical*  
 4104 *charge* of the ball Sylvie fires);

4105 “LINK”<sub>2</sub> OF I-CHAIN<sub>2</sub> (At least). *four* types of alterations to  $D_2$   
 4106 lead to changes in JS (alterations to  $D_2$ ’s *spatio-temporal* properties  
 4107 (this counts for two), and to the *mass* and *shape* of the ball at  $D_2$ ’s  
 4108 spatio-temporal region).

4109 Let the “strength” of an I-CHAIN “link” be the amount of influence present  
 4110 in that “link.” I now claim that, for I-CHAIN<sub>1</sub> and I-CHAIN<sub>2</sub>, CaI must say  
 4111 that what determines how much a cause SF is of JS is the strength of the  
 4112 I-CHAIN’s *weaker* “link.” This follows from my next, more general, claim that  
 4113 if an event  $C$  is a cause of another event  $E$  because there is a (two-“link”)  
 4114 I-CHAIN leading from  $C$  to  $E$ , then how much a cause  $C$  is of  $E$  supervenes  
 4115 upon the strength of said I-CHAIN’s weaker “link.” I will now evidence the  
 4116 just-mentioned general claim by constructing one (two-“link”) I-CHAIN in

6 Admittedly, if Sylvie fires early enough, her ball will ricochet and shatter the jar before Bruno’s ball can. We can, however, all but eliminate this small amount of influence by adding to SCE that the jar is placed at its location right before it actually shatters.

4117 each of two causal scenarios. I will then show that, in these I-CHAINS, varying  
 4118 the strength of the stronger “link” (while holding fixed that of the weaker  
 4119 “link”) *doesn't* vary our intuitions about how much *C* is a cause of *E*. Varying  
 4120 the strength of the weaker “link” (while holding fixed that of the stronger  
 4121 “link”), however, *does*. The first I-CHAIN I construct will possess I-CHAIN<sub>1</sub>'s  
 4122 *strong-weak* pattern of influence (i.e. *C* (SF) has no substantial direct influence  
 4123 on *E* (JS); *C* strongly influences some intermediate event *D* (*D*<sub>1</sub>); *D* weakly  
 4124 influences *E*). The second will possess I-CHAIN<sub>2</sub>'s *weak-strong* pattern of  
 4125 influence (i.e. *C* (SF) has no substantial direct influence on *E* (JS); *C* weakly  
 4126 influences *D* (*D*<sub>2</sub>); *D* strongly influences *E*).

4127 *Scenario 1. Divorce.* Only two things elicit in Wife hatred for Husband  
 4128 (the first significantly more so than the second): (1) the memory  
 4129 of their first fight, which occurred in the rain; (2) the memory of  
 4130 their second fight, which occurred in the fog. Wife, nevertheless,  
 4131 has fallen for Paramour. Thus, she has decided that she will file  
 4132 for divorce from Husband on Thursday afternoon. On Wednesday  
 4133 afternoon, Husband goes on a drinking binge. Late Wednesday night,  
 4134 Husband arrives home. His drunkenness annoys Wife, and the two  
 4135 fight in their driveway. Because fog happens to descend, the fight is  
 4136 so serious to Wife that it (temporarily) lays her thoughts of Paramour  
 4137 to rest, and independently drives her to file for divorce on Thursday  
 4138 afternoon.

4139 We can construct a *strong-weak* I-CHAIN<sub>Divorce</sub> with these three events: (*C*)  
 4140 Husband's drinking binge on Wednesday afternoon; (*D*) the fight late Wednes-  
 4141 day night; (*E*) Wife's filing for divorce on Thursday afternoon. (1) *C* has no  
 4142 substantial direct influence on *E*—altering whether or not/how/what/how  
 4143 long Husband drinks changes nothing about Wife's filing for divorce. (2) *C*  
 4144 strongly influences *D*—altering whether or not/how long Husband drinks  
 4145 changes whether or not/at what time the fight occurs. (3) *D* weakly influences  
 4146 *E*—altering whether or not/how long Wife and Husband fight changes nothing  
 4147 about Wife's filing for divorce. However, if the fight had occurred in the  
 4148 rain, then Wife would've filed for divorce, say, earlier.

4149 Does strengthening I-CHAIN<sub>Divorce</sub>'s stronger “link” (*C*'s influence on *D*)  
 4150 make us intuit that *C* is more a cause of *E* than before? No. Add to Divorce that  
 4151 the fight's topic is sensitive to the type of alcohol that Husband consumes—  
 4152 this *doesn't* make us intuit that Husband's drinking binge is more a cause of

4153 Wife's filing for divorce than before. But what if we strengthen I-CHAIN<sub>Divorce</sub>'s  
 4154 weaker "link" (*D*'s influence on *E*)? Add to Divorce that the timing of Wife's  
 4155 filing for divorce is sensitive to whether or not (but not the extent to which<sup>7</sup>)  
 4156 Husband is drunk during the fight (perhaps Wife takes sober fights most seri-  
 4157 ously, and would've filed for divorce earlier if Husband had been sober during  
 4158 the fight<sup>8</sup>)—contrary to before, this *does* make us intuit that Husband's drink-  
 4159 ing binge is more a cause of Wife's filing for divorce on Thursday afternoon  
 4160 (and not, say, early Thursday morning).

4161 *Scenario 2. Resolve.* Colonel is testing Recruit's resolve. Recruit pos-  
 4162 sesses a button which, if pressed, activates a light which Gunman  
 4163 takes as a signal to shoot Prisoner. Gunman will only ever shoot at  
 4164 time  $t_2$ . Also, iff Recruit doesn't press the button by time  $t_1$ , Colonel  
 4165 will shoot Prisoner at  $t_2$ . The following three events occur: (*C*) Re-  
 4166 cuit presses the button at  $t_1$ ; (*D*) Gunman fires at  $t_2$ ; (*E*) Prisoner  
 4167 dies at  $t_3$ .

4168 *C-D-E form weak-strong I-CHAIN<sub>Resolve</sub>:* (1) *C* has no substantial direct in-  
 4169 fluence on *E*—altering whether or not/how/when Recruit presses the but-  
 4170 ton changes nothing about Prisoner's death at  $t_3$ . (2) *C* weakly influences  
 4171 *D*—altering how Recruit presses the button changes nothing about Gunman's  
 4172 firing at  $t_2$ . And neither does having Recruit press the button *before*  $t_1$ . How-  
 4173 ever, if Recruit hadn't pressed the button (by  $t_1$ ), Gunman wouldn't have  
 4174 fired. (3) *D* strongly influences *E*—altering whether or not/how Gunman fires  
 4175 changes whether or not/how Prisoner dies.

4176 Consider these two possible additions to Resolve: (1) Gunman possesses  
 4177 many rifles to choose from, each of which inflicts death differently; (2) Recruit  
 4178 possesses another button which, if pressed, *prevents* Gunman's firing (Colonel  
 4179 will nonetheless shoot Prisoner at  $t_2$  if this button is pressed<sup>9</sup>). Again, only  
 4180 that addition which strengthens the I-CHAIN's weaker "link" (addition (2))  
 4181 makes us intuit that *C* is more a cause of *E* than before.

4182 There is evidence, then, that in (two-"link") I-CHAINS, how much *C* is  
 4183 a cause of *E* supervenes upon the strength of the I-CHAIN's weaker "link."  
 4184 Consequently, unless one (a) reasonably explains why this doesn't apply to

7 This stipulation denies the substantial direct influence of *C* on *E*.

8 I think that an alteration of the fight in which Husband is sober requires no bigger a Lewisian "miracle" (1979, 468–469) than do those alterations of  $D_1$  that Choi appeals to.

9 This stipulation denies the substantial direct influence of *C* on *E*.

4185 I-CHAIN<sub>1</sub> and/or I-CHAIN<sub>2</sub>, or (b) denies that the causal status of *C* has some-  
 4186 thing to do with I-CHAINS (or counterfactual dependence in general), then  
 4187 how much SF is a cause of JS supervenes upon the strength of “link”<sub>2</sub>, for  
 4188 I-CHAIN<sub>1</sub>, and “link”<sub>1</sub>, for I-CHAIN<sub>2</sub>.

4189 This result, however, likely forces CaI to (counterintuitively) count SF as  
 4190 (significantly) *less* a cause of JS than is BF. After all, *four* types of alterations to  
 4191 BF count towards the influence that BF has on JS. Only *two* types of alterations  
 4192 to *D*<sub>1</sub> count towards the influence that *D*<sub>1</sub> has on JS. And only *two* types of  
 4193 alterations to SF count towards the influence that SF has on *D*<sub>2</sub>. Certainly, it  
 4194 remains possible that for, say, I-CHAIN<sub>2</sub>, the *total number* (as opposed to the  
 4195 number of *types*) of alterations to SF that lead to changes in *D*<sub>2</sub> is greater than  
 4196 the *total number* of alterations to BF that lead to changes in JS. But this would  
 4197 be surprising. Why think, for example, that there are (significantly) more  
 4198 surface properties that Sylvie’s ball might have had, than there are angles at  
 4199 which Bruno might have fired? It also remains possible for the defender of  
 4200 CaI to try to identify more *types* of alterations to SF that lead to changes in *D*<sub>2</sub>.  
 4201 This strategy, however, can only be a stopgap, unless it can be shown that, for  
 4202 each such newly-identified type of alteration to SF, there is no not-too-distant,  
 4203 hitherto-unidentified, type of alteration to BF that leads to changes in JS.  
 4204 Showing this would be difficult. After all, there appear many examples of the  
 4205 latter (e.g. altering properties like the muzzle velocity and barrel length of  
 4206 Bruno’s *rifle* will affect the travel of his ball).

4207 I end by blocking one last maneuver that the defender of CaI might perform.  
 4208 Consider:

4209 “THRESHOLD” OPERATION OF CAI. Causation isn’t a scalar relation.  
 4210 That is, there are no degrees of causation—either an event *C* is  
 4211 a cause of another event *E*, or it isn’t. Thus, if the strength of the  
 4212 weaker “link” of I-CHAIN<sub>1</sub>/I-CHAIN<sub>2</sub> determines anything, it’s simply  
 4213 *whether or not* SF is a cause of JS. That said, in both I-CHAINS, said  
 4214 strength meets that minimum amount of influence *x* required to  
 4215 establish causation. So there is a sense in which CaI *does* capture  
 4216 Comparative Intuition—SF is “as much” a cause of JS as is BF in  
 4217 that neither firing can be said to be more or less a cause than the  
 4218 other. (On “Threshold” Operation, then, any influence that *C* has  
 4219 on *E* exceeding *x* is ignored.)

4220 Besides its diverging from Lewis's writing<sup>10</sup>, there are (at least) two reasons  
4221 to reject "Threshold" Operation.

4222 First, causation *is* plausibly a scalar relation. After all, this appears to be the  
4223 "common sense", or "ordinary", view. For one thing, Hitchcock and Knobe  
4224 offer experimental evidence for their claim that "ordinary causal judgments  
4225 of subjects" come in degrees (2009, 602). For another thing, Michael Moore  
4226 argues that the *law* treats causation as scalar (2009, 71, 118–123; see also  
4227 Brahm and van Hees 2009, 324). Thus, in tort law, the idea of "degrees of  
4228 causal contribution" is both taken as sensible, and employed widely. We see  
4229 this especially in negligence cases in which the doctrine of *divisible harm* is  
4230 invoked so as to apportion liability amongst several defendants according to  
4231 the degree of causal contribution each makes to some *indivisible harm* (Moore  
4232 2009, 118–119). In one such case<sup>11</sup>—*Moore v. Johns-Manville Sales Corp* 781 F  
4233 2d 1061 (5th Cir 1986)—liability for each plaintiff's asbestosis was apportioned  
4234 according to the degree to which each (defendant) manufacturer's (asbestos-  
4235 containing) products caused the plaintiff's asbestosis (i.e. each defendant's  
4236 "degree of relative causation"). Therefore, if we think that our concept of  
4237 causation should accord with how causation is employed "ordinarily," we  
4238 should also think that causation *is* a scalar relation.

4239 Second, determining the value of  $x$  appears impossible. After all,  $x$  cannot  
4240 be some one particular value. This is because we can easily conceive of one  
4241 pre-emption case in which (the event intuited as) the pre-empting cause  
4242 *doesn't* exhibit  $x$  amount of influence on the effect, and another pre-emption  
4243 case in which (the event intuited as) the (non-causal) pre-empted alternative  
4244 *does* (see Dowe 2000, 6–7). One may then suggest that one determine  $x$  on  
4245 a case-by-case basis. This, however, would require one to establish some  
4246 standard set of case features relevant to determining  $x$  (so as to ensure that  
4247 our determinations of  $x$  are not *ad hoc*). At this point, however, I simply  
4248 cannot see what these features might be.\*

4249 Hseng Xiong Joshua Goh

4250  0000-0002-5733-3653

4251 University College London

4252 hseng.goh.14@ucl.ac.uk

10 (2000, 191) indicates that Lewis thinks causation *is* a scalar relation; (2000, 188–189) sees Lewis establish causation with reference to comparative, and not absolute, standards.

11 Moore (2009, 119, fn. 36) contains more case examples.

\* For invaluable input, thanks to Arif Ahmed, Luke Fenton-Glynn, two anonymous referees from the University of Cambridge, and three anonymous referees for *Dialectica*.

## References

- 4253
- 4254 BRAHAM, Matthew and VAN HEES, Martin. 2009. "Degrees of Causation." *Erkenntnis*
- 4255 71(3): 323–344, doi:10.1007/s10670-009-9184-8.
- 4256 CHOI, SungHo. 2005. "Understanding the Influence Theory of Causation: A Critique
- 4257 of Strevens [of Strevens (2003)]." *Erkenntnis* 63(1): 101–118, doi:10.1007/s10670-
- 4258 005-0607-x.
- 4259 DOWE, Phil. 2000. "Is Causation Influence?" Unpublished manuscript, [http://fitelson](http://fitelson.org/269/Dowe_ICI.pdf)
- 4260 [.org/269/Dowe\\_ICI.pdf](http://fitelson.org/269/Dowe_ICI.pdf).
- 4261 HITCHCOCK, Christopher R. and KNOBE, Joshua. 2009. "Cause and Norm." *The Journal*
- 4262 *of Philosophy* 106(11): 587–612, doi:10.5840/jphil20091061128.
- 4263 KAISERMAN, Alex. 2016. "Causal Contribution." *Proceedings of the Aristotelian Society*
- 4264 116(3): 387–394, doi:10.1093/arisoc/aow013.
- 4265 LEWIS, David. 1973. "Causation." *The Journal of Philosophy* 70(17): 556–567. Reprinted,
- 4266 with a postscript (Lewis 1986b), in Lewis (1986a, 159–172), doi:10.2307/2025310.
- 4267 —. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13(4): 455–476.
- 4268 Reprinted, with a postscript (Lewis 1986c), in Lewis (1986a, 32–51),
- 4269 doi:10.2307/2215339.
- 4270 —. 1986a. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.109
- 4271 [3/0195036468.001.0001](https://doi.org/10.1093/0195036468.001.0001).
- 4272 —. 1986b. "Postscript to Lewis (1973)." in *Philosophical Papers, Volume 2*, pp. 172–213.
- 4273 Oxford: Oxford University Press, doi:10.1093/0195036468.003.0006.
- 4274 —. 1986c. "Postscript to Lewis (1979)." in *Philosophical Papers, Volume 2*, pp. 52–66.
- 4275 Oxford: Oxford University Press, doi:10.1093/0195036468.003.0002.
- 4276 —. 2000. "Causation as Influence." *The Journal of Philosophy* 97(4): 181–197, doi:10.2
- 4277 [307/2678389](https://doi.org/10.2307/2678389).
- 4278 MOORE, Michael S. 2009. *Causation and Responsibility: An Essay in Law, Morals, and*
- 4279 *Metaphysics*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199256
- 4280 [860.001.0001](https://doi.org/10.1093/acprof:oso/9780199256860.001.0001).
- 4281 SCHAFFER, Jonathan. 2001. "Causation, Influence, and Effluence." *Analysis* 61(1):
- 4282 11–19, doi:10.1093/analys/61.1.11.
- 4283 STREVENs, Michael. 2003. "Against Lewis's New Theory of Causation: A Story with
- 4284 Three Morals." *Pacific Philosophical Quarterly* 84(4): 398–412, doi:10.1046/j.1468-
- 4285 [0114.2003.00182.x](https://doi.org/10.1046/j.1468-0114.2003.00182.x).
- 4286 WOODWARD, James F. 2010. "Causation in Biology: Stability, Specificity, and the Choice
- 4287 of Levels of Explanations." *Biology and Philosophy* 25(3): 287–318, doi:10.1007/s1
- 4288 [0539-010-9200-z](https://doi.org/10.1007/s10539-010-9200-z).

PROOF



4289

# Consistency, Obligations, and

4290

# Accuracy-Dominance Vindications

MARC-KEVIN DAoust

4291 Vindicating the claim that agents ought to be consistent has proved to  
4292 be a difficult task. Recently, some have argued that we can use accuracy-  
4293 dominance arguments to vindicate the normativity of such requirements.  
4294 But what do these arguments prove, exactly? In this paper, I argue that  
4295 we can make a distinction between two theses on the normativity of  
4296 consistency: the view that one ought to be consistent and the view that  
4297 one ought to avoid being inconsistent. I argue that accuracy-dominance  
4298 arguments for consistency support the latter view, but not necessarily the  
4299 former. I also argue that the distinction between these two theses matters  
4300 in the debate on the normativity of epistemic rationality. Specifically, the  
4301 distinction suggests that there are interesting alternatives to vindicating  
4302 the strong claim that one ought to be consistent.

4303 The normativity of the following formal coherence requirements is con-  
4304 tentious:

4305 BELIEF CONSISTENCY. If A believes that  $p$ , it is false that A believes  
4306 that  $\neg p$ .<sup>1</sup>

4307 CREDAL CONSISTENCY. If A has a credence of  $X$  in  $p$ , then A has a  
4308 credence of  $(1-X)$  in  $\neg p$ .

4309 Do we fall under an obligation to satisfy these requirements?<sup>2</sup> Many philoso-  
4310 phers like John Broome (2013, ch. 13) are convinced that the above require-  
4311 ments are normative, but cannot find a satisfactory argument in favour of  
4312 such a conclusion. Other philosophers are less optimistic. For instance, Niko

---

1 This requirement is sometimes called “Pairwise Consistency”, as in Easwaran (2016).

2 See Way (2010) for an overview of this debate. See Fitelson (2016) on epistemic teleology and coherence requirements. See de Bona and Staffel (2018) on accuracy and approximation of Bayesian requirements of probabilistic coherence. See also Pettigrew (2013, 2016a).

4313 Kolodny (2005; 2007a; 2007b, 230–231) has argued that there is no reason  
 4314 to be consistent. According to him, what matters from an epistemic point of  
 4315 view is acquiring true beliefs (or acquiring beliefs that are likely to be true on  
 4316 the evidence) and avoiding false beliefs (or avoiding beliefs that are likely to  
 4317 be false on the evidence). However, a perfectly consistent system of beliefs  
 4318 (or credences) can be entirely false, inaccurate or improbable on the evidence.  
 4319 So, consistency requirements are not normative, in the sense that one does  
 4320 not necessarily have a reason to be consistent.

4321 Recently, a new strategy has emerged to vindicate the normativity of Consis-  
 4322 tency. This strategy relies on accuracy-dominance principles, which roughly  
 4323 say that if state *Y* is better than state *X* at every possible world, one ought to  
 4324 avoid state *X*. However, there is a weak and a strong interpretation of what  
 4325 is entailed by the accuracy-dominance arguments. According to the strong  
 4326 interpretation, accuracy-dominance arguments entail that one ought to be  
 4327 consistent. Joyce, for instance, argues that:

4328 It is thus established that degrees of belief that violate the laws of  
 4329 probability are invariably less accurate than they could be. Given  
 4330 that an epistemically rational agent will always strive to hold  
 4331 partial beliefs that are as accurate as possible, this vindicates the  
 4332 fundamental dogma of probabilism [according to which degrees  
 4333 of belief must make conformity to the axioms of probability].  
 4334 (1998, 600)

4335 According to the weak interpretation, accuracy-dominance arguments merely  
 4336 entail that ought not to be inconsistent. Easwaran, for instance, says that “we  
 4337 can use dominance to *eliminate*” the inconsistent doxastic options (2016, 826,  
 4338 emphasis added). In other words, dominance is here used to argue against  
 4339 inconsistency. Thus, we can make the following distinction between two  
 4340 views:

4341 **NORMATIVITY+**. Given the accuracy-dominance arguments, *A*  
 4342 ought to be consistent.

4343 **NORMATIVITY–**. Given the accuracy-dominance arguments, *A*  
 4344 ought not to be inconsistent.

4345 This paper argues that, while accuracy-dominance arguments can vindicate  
 4346 Normativity–, they do not necessarily vindicate Normativity+. Specifically,

4347 accuracy-dominance arguments vindicate Normativity+ when supplemented  
 4348 with a contentious hypothesis concerning the relationship between reasons for  
 4349 and reasons against. Hence, accuracy-dominance arguments do not vindicate  
 4350 Normativity+ *on their own*.

4351 In Section 1, I clarify the debate on the normativity of Consistency. In Sec-  
 4352 tions 2 and 3, I present two important arguments in the debate surrounding the  
 4353 normativity of Consistency: accuracy-dominance arguments and Kolodny's  
 4354 objection from truth-conduciveness. Both arguments are veritistic: They as-  
 4355 sume that only true beliefs bear final epistemic value, and only false beliefs  
 4356 bear final epistemic disvalue. I argue that, under the assumption that veritism  
 4357 is true, the only way to make sense of both arguments is to make a distinc-  
 4358 tion between Normativity+ and Normativity- (i.e. to deny that both views  
 4359 are coextensive). Then, I argue that accuracy-dominance arguments fail to  
 4360 vindicate Normativity+.

4361 This is not necessarily bad news. In conclusion, I explain why this might  
 4362 be an occasion to adjust our expectations in the debate on the normativity  
 4363 of formal coherence requirements. Many people think that there is some-  
 4364 thing bad or suboptimal with inconsistent combinations of attitudes. The  
 4365 mistake might have been to try to explain this assumption in terms of *an*  
 4366 *obligation to be consistent*. Being in a position to vindicate Normativity- while  
 4367 remaining neutral on Normativity+ could be advantageous in the debate on  
 4368 the normativity of formal coherence requirements.

## 4369 1 The "Why-Be-Consistent?" Challenges

4370 There are many putative explanations of why one ought to have *some* con-  
 4371 sistent combinations of beliefs. They stem from the normative authority of  
 4372 truth, knowledge or reasons, as in the following:

4373 TRUTH VINDICATION. One ought to believe *p* if and only if *p*. Truth  
 4374 is consistent (or: Inconsistent propositions cannot be true simulta-  
 4375 neously). So, one ought to have some consistent combinations of  
 4376 beliefs (e.g. the true ones).

4377 KNOWLEDGE VINDICATION. One is epistemically permitted to  
 4378 believe *p* if and only if one is in a position to know that *p*. Knowledge  
 4379 is consistent (or: Propositions that one is in a position to know

4380 cannot be inconsistent with each other). So, one is only epistemically  
4381 permitted to believe consistent combinations of beliefs.

4382 REASONS VINDICATION. One is epistemically permitted to believe  
4383 *p* if and only if one has sufficient epistemic reason to believe *p*.  
4384 One never has sufficient epistemic reason to believe *p* and suffi-  
4385 cient epistemic reason to disbelieve *p* simultaneously. So, one is  
4386 only epistemically permitted to believe consistent combinations of  
4387 beliefs.<sup>3</sup>

4388 Philosophers like Broome (2013) and others are worried that the above puta-  
4389 tive vindications do not fully vindicate the normativity of Consistency. Some  
4390 consistent combinations of beliefs may include some false, unjustified or  
4391 unreasonable beliefs. Even if consistent agents sometimes believe proposi-  
4392 tions that are false, unjustified or unreasonable, it seems that they satisfy a  
4393 distinct obligation to have consistent beliefs (e.g. an obligation that does not  
4394 boil down to truth, knowledge or reasons). In other words, perhaps the agent  
4395 is unjustified, mistaken or unreasonable, but one could still say: *At least he or*  
4396 *she is consistent*. Here, the putative obligation to be consistent will not come  
4397 from truth, knowledge or reasons.<sup>4</sup>

4398 So, according to some philosophers, the above vindications are somehow  
4399 incomplete. Perhaps we can easily argue that agents ought to have *some*  
4400 consistent combinations of beliefs, but finding a vindication of Consistency  
4401 that covers all the possible consistent combinations of beliefs has proved to  
4402 be a difficult task.

4403 It should also be noted that the normativity of Consistency is part of a  
4404 broader debate on the normativity of *structural rationality*. Structural rati-  
4405 onality allegedly requires of agents not to be incoherent—for example, not to  
4406 be akratic, not to have intransitive preferences, and so forth (Worsnip 2018b,  
4407 2018a). So, in addition to Consistency, there are other putative structural  
4408 requirements of rationality, like:

---

3 Kolodny (2007a) endorses this view. See Daoust (2020) for discussion.

4 In fact, Broome (2013, ch. 11) is interested in the stronger claim that rationality is a *source* of normativity. So, he is not interested in offering a derivative vindication of consistency requirements, that is, a vindication of these requirements on other grounds (like truth, knowledge, or reasons). By contrast, dominance principles are often tied to rationality (see e.g. Joyce 1998).

4409 INTER-LEVEL COHERENCE. Rationality requires that, if A believes  
 4410 that he or she has sufficient epistemic reason to believe  $p$ , then A  
 4411 believes that  $p$ .<sup>5</sup>

4412 INSTRUMENTAL PRINCIPLE. Rationality requires that, if A intends  
 4413 to  $\phi$ , and A believes that  $\psi$ -ing is a necessary means to  $\phi$ -ing, then  
 4414 A intends to  $\psi$ .<sup>6</sup>

4415 Broome and others have tried to find compelling arguments for the claim that  
 4416 *structural rationality* has normative authority. However, structural rationality  
 4417 is neutral on whether one's beliefs should be true, reasonable or amount to  
 4418 knowledge. Some entirely false and unreasonable belief systems can satisfy  
 4419 the requirements of structural rationality. So, at least given the agenda of these  
 4420 philosophers, a good vindication of the normativity of Consistency should  
 4421 cover the cases in which one's beliefs are false or unreasonable.

4422 An interesting feature of accuracy-dominance arguments is that they re-  
 4423 main neutral on whether one's beliefs should be true, reasonable or amount  
 4424 to knowledge. They focus on what is wrong with having some combinations  
 4425 of beliefs, regardless of the substantive properties of such beliefs.

## 4426 2 Accuracy-Dominance and Consistency

4427 Accuracy-dominance arguments for vindicating the normativity of Consis-  
 4428 tency come from decision theory and rely on the following principle:

4429 STRONG DOMINANCE. If an available state  $X$  is strongly dominated  
 4430 by an available state  $Y$  at every possible world, in the sense that state  
 4431  $Y$  is better or has more value than state  $X$  at every possible world,  
 4432 one ought to avoid state  $X$ .

4433 Strong Dominance has been used to vindicate probabilism, the view roughly  
 4434 stating that an agent's rational credences should satisfy the probability ax-

5 Coates (2012) and Lasonen-Aarnio (2020) have argued that responding correctly to one's evidence sometimes entail believing "P, but I am irrational to believe P", which is an incoherent combination of attitudes. They conclude that such incoherence is not necessarily irrational. See Greco (2014), Horowitz (2014), Kiesewetter (2016), Littlejohn (2018), Titelbaum (2015) and Worsnip (2018b) for various responses to this view.

6 See, among others, Broome (2013, sec. 9.4), Kiesewetter (2017, ch. 10) and Way (2013) on the Instrumental Principle.

4435 ioms. With respect to some inaccuracy measures such as the Brier score,  
 4436 probabilistically inconsistent agents have access to a credence function that  
 4437 is less inaccurate (and thus less epistemically disvaluable) at every possible  
 4438 world (Joyce 1998; Leitgeb and Pettigrew 2010; Pettigrew 2016a).

4439 For the sake of simplicity, I will leave aside dominance for credence and  
 4440 focus on dominance for belief (these arguments have the same structure, but  
 4441 dominance arguments for belief are more accessible).

4442 There is a plausible explanation of why inconsistent combinations of beliefs  
 4443 are strongly dominated. An agent can take different doxastic attitudes towards  
 4444  $p$ , as in the following:

- 4445 (i) Believing  $p$  and not disbelieving  $p$ ,
- 4446 (ii) Disbelieving  $p$  and not believing  $p$ ,
- 4447 (iii) Neither believing nor disbelieving  $p$ ,
- 4448 (iv) Believing  $p$  and disbelieving  $p$ .

4449 The question is whether (iv) is strongly dominated. To answer this question,  
 4450 we need to determine the epistemic value of (iv) at every possible world. In  
 4451 veritistic frameworks, only true beliefs have final epistemic value and only  
 4452 false beliefs have final epistemic disvalue. Accordingly,  $T$  is the epistemic  
 4453 value of having a true belief (for  $T > 0$ ),  $F$  is the epistemic disvalue of having  
 4454 a false belief (for  $F < 0$ ), and the epistemic value of not believing  $p$  (or not  
 4455 disbelieving  $p$ ) is 0.<sup>7</sup> Finally, assume that  $T \leq -F$ , which amounts to endors-  
 4456 ing a conservative account of epistemic value. The conservative constraint on  
 4457 epistemic value is plausible.<sup>8</sup> As Dorst says:<sup>9</sup>

4458 [An epistemically rational agent] will be doxastically conserva-  
 4459 tive... Why? Well here's a fair coin—does she believe it'll land  
 4460 heads? Or tails? Or both? Or neither? Clearly neither. But if she  
 4461 cared more about seeking truth than avoiding error, why not be-  
 4462 lieve both? She'd then be guaranteed to get one truth and one

7 I'm glossing over some inessential subtleties here. It is possible to assign a value to not believing  $p$  (or to withholding judgment on whether  $p$ ), but ultimately, we would get exactly the same results. See Easwaran (2016, sec. C) and Dorst (2019, 10, n. 12).

8 But this constraint might not stem from accuracy-first epistemology. See Steinberger (2019) and the next footnote.

9 In addition to Dorst's argument, see Easwaran (2016), Easwaran and Fitelson (2015) and Pettigrew (2016b) for similar arguments in favour of the conservative account of epistemic value. See Steinberger (2019) on why alternatives to conservatism are compatible with accuracy-first epistemology.

4463 falsehood, and so be more accurate than if she believed neither...  
 4464 Upshot: we impose a *Conservativeness* constraint to capture the  
 4465 sense in which Rachael has ‘more to lose’ in forming a belief than  
 4466 she does to gain. (2019, 11)

4467 Then, we can determine the possible values of each option at every possible  
 4468 world. Since the value of these options is solely determined by  $p$ ’s truth value,  
 4469 we need to consider the worlds in which  $p$  is true and the worlds in which  $p$   
 4470 is false, as in Table 1.

Table 1: An agent’s doxastic options with respect to  $p$

Doxastic options / possible world	$p$ is true	$p$ is false
Believing $p$ and not disbelieving $p$	$T$	$F$
Disbelieving $p$ and not believing $p$	$F$	$T$
Neither believing nor disbelieving $p$	$0$	$0$
Believing $p$ and disbelieving $p$	$T + F$	$T + F$

4471 Finally, in accordance with Table 1, we can conclude that inconsistent  
 4472 combinations of beliefs are strongly dominated. The following reasoning  
 4473 supports such a conclusion:

- 4474 (1)  $T \leq -F$  (conservative assumption). Accordingly,  $T + F < 0$ .  
 4475 (2) Following (1) and Table 1, believing  $p$  and disbelieving  $p$  simultaneously  
 4476 has an epistemic value of less than 0 at every possible world.  
 4477 (3) However, following Table 1, neither believing nor disbelieving  $p$  has an  
 4478 epistemic value of 0 at every possible world.  
 4479 (C) Therefore, following (2) and (3), inconsistent combinations of beliefs  
 4480 such as believing  $p$  and disbelieving  $p$  are strongly dominated: another  
 4481 available option (neither believing nor disbelieving  $p$ ) is more valuable  
 4482 at every possible world.<sup>10</sup>

4483 Hence, one ought to avoid being inconsistent.

<sup>10</sup> Similar arguments can be found in Easwaran (2016§B) and Pettigrew (2016b, 256). Dorst (2019, 31, esp. proposition 3) argues for a similar but contextualist view.

### 3 Truth-Conduciveness, Reasons For and Reasons Against

#### 3.1 *Kolodny's Objection From Truth-Conduciveness*

The above argument states that inconsistent combinations of beliefs are dominated, which means that one ought not to be inconsistent. Naturally, this seems to suggest that one ought to be consistent. But this equivalence is less obvious than it seems.

To see why, consider Kolodny's argument against the normativity of Consistency. According to him, one does not necessarily have an epistemic reason to be consistent. Rather, what matters from an epistemic point of view is having true beliefs and avoiding false beliefs, and satisfying Consistency does not guarantee a better ratio of true to false beliefs. In fact, some perfectly consistent sets of beliefs are entirely false (or improbable on the evidence). Kolodny summarizes his argument in the following way:

From the standpoint of theoretical deliberation—which asks ‘What ought I to believe?’—what ultimately matters is simply what is likely to be true, given what there is to go on. [...] [But] formal coherence may as soon lead one away from, as toward, the true and the good. Thus, if someone asks from the deliberative standpoint ‘What is there to be said for making my attitudes formally coherent as such?’ there seems, on reflection, no satisfactory answer. (2007b, 231)

In other words, if one merely satisfies Consistency, one is not more likely to end up forming true beliefs and avoiding false beliefs. So, the mere satisfaction of Consistency does not improve one's ratio of true to false beliefs. In view of the foregoing, Kolodny thinks that it is false that one falls under an obligation to be consistent.<sup>11</sup>

#### 3.2 *Comparing the Objection from Truth-Conduciveness and Accuracy-Dominance Arguments*

Kolodny argues that there is no reason to be consistent. His argument relies on the fact that being consistent does not guarantee a good ratio of true to false

---

<sup>11</sup> Elsewhere, Kolodny (2005) raises some objections against the normativity of other structural requirements, such as Inter-Level Coherence.



4514 beliefs. By way of contrast, accuracy-dominance arguments suggest that there  
4515 is good reason not to be inconsistent. If one is inconsistent, one is strongly  
4516 dominated, in the sense that one has access to a better option at every possible  
4517 world. For instance, if one believes  $p$  and disbelieves  $p$  simultaneously, one  
4518 will necessarily improve one's situation by neither believing nor disbelieving  
4519  $p$ .

4520 Accuracy-dominance arguments and Kolodny's objection from truth-  
4521 conduciveness are both veritistic.<sup>12</sup> Indeed, they presuppose that only true  
4522 beliefs bear final epistemic value, and only false beliefs bear final epistemic  
4523 disvalue. Nevertheless, such arguments apparently support incompatible  
4524 conclusions concerning the normativity of Consistency: Kolodny argues  
4525 that veritism entails the denial of the normativity of Consistency, whereas  
4526 accuracy-dominance arguments support the normativity of Consistency. This  
4527 is puzzling.

4528 Perhaps Kolodny and accuracy-dominance theorists do not endorse the  
4529 same version of veritism. Veritism says that only true beliefs have final epis-  
4530 temic value, and only false beliefs have final epistemic disvalue. However,  
4531 when it comes to epistemic obligations and permissions, these assumptions  
4532 concerning epistemic value might translate in many different ways. For in-  
4533 stance, perhaps agents ought to maximize their *total* epistemic score (e.g. the  
4534 total balance of epistemic value they get from their doxastic states), or perhaps  
4535 agents ought to maximize their *expected* epistemic score. For clarity, consider  
4536 the following example: Suppose  $p$  is very likely relative to a body of evidence  $E$ .  
4537 But as it happens,  $p$  is false. Then, believing  $p$  (or having a high credence in  $p$ )  
4538 might maximize expected epistemic value with respect to  $E$ . But disbelieving  
4539  $p$  (or having a low credence in  $p$ ) will maximize epistemic value *tout court*.

4540 Yet, it is implausible that a difference in how Joyce and Kolodny under-  
4541 stand veritism is the reason why they disagree. Kolodny's argument can be  
4542 reformulated in many different ways. Consider the following possibilities: (i)  
4543 Suppose agents ought to maximize *expected* accuracy. Then, Kolodny could  
4544 say: Some consistent combinations of beliefs can minimize expected accuracy  
4545 (believing the most improbable propositions can be consistent). (ii) Suppose  
4546 agents ought to optimize their ratio of true to false beliefs. Then, Kolodny  
4547 could argue that some agents with a very bad ratio of true to false beliefs  
4548 are consistent. (iii) Suppose agents ought to maximize *total* accuracy. Then,  
4549 Kolodny could say: Some consistent combinations of beliefs can minimize

---

12 See notably Goldman (2015) and Whiting (2010) on veritism.

4550 accuracy (believing false propositions only can be consistent). As we can see,  
4551 Kolodny's objection is malleable.<sup>13</sup>

4552 Another possibility is that Kolodny and accuracy-first theorists have a  
4553 different understanding of what "ought" means. We can make a distinction  
4554 between normativity in the rule-following sense (as in: Relative to domain  
4555 D, A ought to X) and normativity in the reason-involving sense (as in: A has  
4556 a reason to X).<sup>14</sup> For example, the rules of etiquette require of agents to be  
4557 polite, but agents might lack a reason to be polite. By way of analogy with  
4558 the rules of etiquette, perhaps accuracy-first theorists are merely interested  
4559 in arguing that the rules of rationality require consistency. This would be  
4560 compatible with Kolodny's view—namely, that agents do not have a reason  
4561 to be consistent. Both views would then be compatible with each other.

4562 It is true that accuracy-first theorists see Consistency as a demand of rati-  
4563 onality. However, it is implausible that accuracy-first theorists are *merely*  
4564 concerned with normativity in the rule-following sense. Accuracy-first theo-  
4565 rists like Joyce tie norms of rationality to epistemic value, as in the following:

4566 THE NORM OF TRUTH. An epistemically rational agent must strive  
4567 to hold a system of full beliefs that strikes the best attainable overall  
4568 balance between the epistemic good of fully believing truths and  
4569 the epistemic evil of fully believing falsehoods (1998, 577).

4570 THE NORM OF GRADATIONAL ACCURACY. An epistemically rational  
4571 agent must evaluate partial beliefs on the basis of their gradational  
4572 accuracy, and she must strive to hold a system of partial beliefs that,  
4573 in her best judgment, is likely to have an overall level of gradational  
4574 accuracy at least as high as that of any alternative system she might  
4575 adopt (1998, 579).

4576 Satisfying the requirements of rationality is different from, say, satisfying  
4577 the requirements of etiquette. The former has a privileged relationship to  
4578 value. Epistemically rational agents want to optimize their overall balance of  
4579 epistemic value. Accordingly, it would be surprising that Joyce and others are  
4580 merely concerned with normativity in the rule-following sense. Specifically,

---

13 I thank a referee for inviting me to discuss this possibility.

14 See Parfit (2011, 144–148) on this distinction.

4581 it would be surprising that, while rationality has some sort of privileged  
4582 relationship to value, it is merely normative in the rule-following sense.<sup>15</sup>

4583 Under the assumption that Kolodny and accuracy-dominance theorists  
4584 agree upon a specific version of veritism and the meaning of “ought,” the nat-  
4585 ural reaction is to think that at least one of the above arguments is mistaken—  
4586 either the objection from truth-conduciveness is inconclusive, or accuracy-  
4587 dominance arguments fail. After all, how can there be no reason to be consis-  
4588 tent and reasons against being inconsistent? If there is something wrong with  
4589 being inconsistent, there must be something good with being consistent!

4590 However, this natural reaction presupposes that there is always a connection  
4591 between (i) reasons for being consistent (as in Normativity+) and (ii) reasons  
4592 against being inconsistent (as in Normativity–). Call this the Coextensivity  
4593 Thesis, as in the following:

4594 COEXTENSIVITY THESIS. Arguments in favour of Normativity–  
4595 count as arguments in favour of Normativity+ (and vice versa).

4596 Those who endorse the Coextensivity Thesis think that (i) and (ii) express the  
4597 same normative relation.

4598 If the Coextensivity Thesis were correct, then Kolodny’s objection from  
4599 truth-conduciveness would be inconclusive. Under the assumption that the  
4600 Coextensivity Thesis is correct, two kinds of considerations can vindicate  
4601 the view that one ought to be consistent—namely, reasons be consistent and  
4602 reasons against being inconsistent. Kolodny argues for the *absence* of reasons  
4603 in favour of being consistent. But if the Coextensivity Thesis is correct, *such*  
4604 *considerations are just half of the story*. We also need to consider whether there  
4605 are reasons against being inconsistent in the balance, since they count as  
4606 reasons for being consistent. Accuracy-dominance arguments entail that one  
4607 ought not to be inconsistent. So, even if Kolodny is right that there is no reason  
4608 to satisfy Consistency, this does not entail that it is false that one ought to be  
4609 consistent. Insofar as there are arguments against inconsistency (as suggested  
4610 by accuracy-dominance arguments), there is a reason to be consistent.

4611 However, if the Coextensivity Thesis is false, then accuracy-dominance  
4612 arguments are compatible with the objection from truth-conduciveness. Here  
4613 is why. Kolodny argues that there is no reason to be consistent: he denies that  
4614 one ought to be consistent, as in Normativity+. However, if the Coextensivity

---

15 I thank a referee for inviting me to clarify this possibility.

4615 Thesis is false, we can deny Normativity+ without denying Normativity—.  
 4616 In other words, even if it is false that one ought to be consistent, perhaps  
 4617 one ought not to be inconsistent. The same goes for accuracy-dominance  
 4618 arguments. According to such arguments, inconsistent combinations of beliefs  
 4619 are dominated. So, one ought not to be inconsistent. But if the Coextensivity  
 4620 Thesis is false, this does not entail that one ought to be consistent.

### 3.3 *Reasons to be Consistent and the Coextensivity Thesis*

4622 So, is the Coextensivity Thesis true? This depends on what “a reason to be  
 4623 consistent” means. Suppose, like Kolodny, that “a reason to be consistent”  
 4624 concerns each individual consistent option one has (see section 3.1. That is,  
 4625 suppose that “a reason to be consistent” means something like “a consideration  
 4626 that counts in favour of *each* individual consistent options one has.” For  
 4627 Kolodny, nothing can be said in favour of some consistent combinations of  
 4628 attitudes. So, under this interpretation of what “a reason to be consistent”  
 4629 means, we do not necessarily have a reason to be consistent.

4630 Relative to this interpretation of what “a reason to be consistent” means,  
 4631 the Coextensivity Thesis does not seem plausible. For reasons found in  
 4632 Snedegar (2018), we can make a distinction between reasons for Consistency  
 4633 (as in Normativity+) and reasons against inconsistency (as in Normativity—).  
 4634 The distinction comes from the following account of reasons for and reasons  
 4635 against endorsed by Snedegar:

4636 My view puts a strong condition on reasons for and a weak condi-  
 4637 tion on reasons against. For some objective to provide a reason  
 4638 for an option, that option has to do the best with respect to the ob-  
 4639 jective. For some objective to provide a reason against an option,  
 4640 that option only has to do worse than some alternative. (2018,  
 4641 737)

4642 Snedegar roughly argues that the problem with views that lump together  
 4643 reasons against and reasons for is that there can be good reasons not to  $\phi$ ,  
 4644 even if there are worse alternatives to  $\phi$ -ing.<sup>16</sup> For instance, suppose that I am  
 4645 trying to decide what to drink. I might have conclusive reason not to drink  
 4646 gin, but this does not entail that I have a reason to drink any beverage that

---

16 See Snedegar (2018) for more details.

4647 isn't gin. I should definitely not drink petrol, even if petrol isn't gin. This is  
4648 compatible with my having conclusive reason not to drink gin.

4649 Snedegar's observation sits well with accuracy-dominance arguments dis-  
4650 cussed in Section 2. Indeed, recall the options agents have in Table 1. Clearly,  
4651 there is conclusive reason not to go for the inconsistent option, since neither  
4652 believing nor disbelieving  $p$  is better than being inconsistent at every possible  
4653 world. However, this does not entail that there is a reason in favour of every  
4654 alternative to the inconsistent option. For instance, disbelieving  $p$  when  $p$  is  
4655 true (or believing  $p$  when  $p$  is false) is worse than being inconsistent. So, as  
4656 in the gin and petrol case, reasons against inconsistency are logically weaker  
4657 than reasons for Consistency.

4658 This suggests that accuracy-dominance arguments do not vindicate  
4659 Normativity+ on their own. Of course, when combined with the Coex-  
4660 tensivity Thesis, these arguments support Normativity+. But Kolodny's  
4661 interpretation of what "a reason to be consistent" means conflicts with the  
4662 Coextensivity Thesis. So, while accuracy-dominance arguments support  
4663 Normativity-, it is an open question whether they also support Normativity+.

4664 Here is a response to my argument on behalf of the accuracy-dominance  
4665 theorist. We can regroup the consistent options in Table 1 under a single  
4666 option. Call this the consistent option. With respect to the consistent option,  
4667 Snedegar's distinction does not apply. If there is conclusive reason not to go for  
4668 the inconsistent option, and the only option left is the "regrouped" consistent  
4669 option, then reasons against inconsistency favour the consistent option. So,  
4670 could there be a sense in which the Coextensivity Thesis is true?<sup>17</sup>

4671 My response to this objection goes as follows. This way of framing the prob-  
4672 lem cannot make sense of Kolodny's objection concerning some consistent  
4673 options. *There is something wrong with some consistent combinations of beliefs*  
4674 —some consistent combinations of beliefs are entirely wrong or improbable  
4675 on the evidence. Kolodny is right to point out that nothing can be said in  
4676 favour of these combinations of attitudes. The only way to make sense of  
4677 Kolodny's objection is *not* to regroup all the consistent options under a single  
4678 label, precisely because relevant normative distinctions can (and should) be  
4679 made between some consistent options.

4680 At best, this reply shows that, under a different interpretation of what "a  
4681 reason to be consistent" means, the Coextensivity Thesis is true. But Kolodny's  
4682 argument still succeeds relative to another interpretation of this expression.

---

17 I thank a referee for inviting me to discuss this objection.

4683 When Kolodny discusses the normativity of Consistency, he discusses the nor-  
 4684 mativity of the individual consistent options one has, including the ones that  
 4685 are entirely wrong or improbable on the evidence. The accuracy-dominance  
 4686 theorist can claim that one ought to be consistent, but that is simply because  
 4687 the expression “one ought to be consistent” here refers to something logically  
 4688 weaker than what Kolodny has in mind.<sup>18</sup>

### 3.4 *An Escape Route for the Accuracy-Dominance Theorist?*

4690 The accuracy-dominance theorist could then offer the following objection.  
 4691 Suppose there is an accuracy-dominance argument against one’s attitudes.  
 4692 Accordingly, one can identify at least one collection of attitudes that veritisti-  
 4693 cally dominates one’s current state. If agents can identify at least one set of  
 4694 attitudes that is better than their current state, then they have a reason to take  
 4695 the dominating set of attitudes, which will be consistent. Doesn’t this support  
 4696 the view according to which one ought to be consistent? If agents ought to  
 4697 take dominating combinations of beliefs, and such combinations of beliefs  
 4698 are consistent, then this seems to entail that agents ought to be consistent.<sup>19</sup>

4699 This objection carries weight depending on what accuracy-dominance  
 4700 arguments prove. Let me explain.

4701 Suppose the contender is right. Then, accuracy-dominance vindications  
 4702 are akin to the Truth Vindication, the Knowledge Vindication or the Reasons  
 4703 Vindication discussed in Section 1. If one has inconsistent combinations of  
 4704 beliefs (say, one believes  $p$  and also believes  $\neg p$ ), the Truth Vindication says  
 4705 that agents ought to maintain the true one (and abandon the false one), the  
 4706 Knowledge Vindication says that agents are only permitted to maintain the  
 4707 known one, and the Reasons Vindication says that agents are only permitted  
 4708 to maintain the reasonable one (and ought to abandon the *unreasonable* one).  
 4709 In any case, satisfying such norms means that agents will cease entertaining  
 4710 inconsistent combinations of beliefs.

4711 The contender makes a similar point. If one has inconsistent combinations  
 4712 of beliefs, one should go for the option dominating inconsistent combinations  
 4713 of beliefs. But if that is right, the accuracy-dominance argument merely entails

---

18 My response might not convince some readers. In any case, we can draw a lesson from this discussion. We have learned that the expression “a reason to be consistent” is ambiguous. Some readings of this expression are a problem for Kolodny’s argument, and other readings of this expression conflict with vindicating Normativity+.

19 I thank a referee for bringing this objection to my attention.

4714 that agents ought (or have reasons) to have *some* combinations of beliefs, not  
4715 *any* consistent combination of beliefs. In other words, the argument leaves  
4716 out some consistent combinations of beliefs.

4717 This brings us back to the discussion in Section 1. What do we expect from  
4718 a good vindication of the normativity of Consistency? For many philosophers,  
4719 a good vindication of Consistency should cover all the possible consistent  
4720 combinations of beliefs. If the contender is right, then accuracy-dominance  
4721 arguments can explain the significance of some consistent combinations of  
4722 beliefs—namely, the dominating ones. But this is not what we were looking for.  
4723 The explanation should apply to *all* the consistent combination of beliefs. To be  
4724 clear: Some philosophers might not be interested in this specific interpretation  
4725 of the “Why-Be-Consistent” debate. It should be clear that, with respect to  
4726 other understandings of the question, the contender is right.

#### 4722 **4 Conclusion and Implications in the Debate on the** 4728 **Normativity of Structural Rationality**

4729 This paper supports the view that there are two theses concerning the nor-  
4730 mativity of Consistency: Normativity+ and Normativity-. While accuracy-  
4731 dominance arguments support Normativity-, they might not necessarily sup-  
4732 port Normativity+. This is so, because the Coextensivity Thesis might be false.  
4733 In fact, one way to reconcile Kolodny’s objection from truth-conduciveness  
4734 with accuracy-dominance arguments is to deny the Coextensivity Thesis.


4735 These clarifications concerning Normativity+ and Normativity- allow us  
4736 to rethink the debate on the normativity of structural rationality. Indeed, a  
4737 popular strategy for arguing against the normativity of structural rationality  
4738 is to point out that there is no reason to satisfy some specific rational require-  
4739 ments (such as Consistency). Kolodny’s objection from truth-conduciveness  
4740 is a good illustration of such arguments. These arguments are compelling if  
4741 we focus on Normativity+. But this might be a mistake. Perhaps that, when it  
4742 comes to formal requirements like Consistency, the only view we should try  
4743 to vindicate is Normativity-.

4744 The argument of this paper allows us to make sense of some pre-  
4745 theoretically correct assumptions structural requirements of epistemic  
4746 rationality such as Consistency. Plausibly, there is something wrong,  
4747 suboptimal or disvaluable with inconsistent combinations of beliefs. The  
4748 mistake might have been to try to explain this assumption in terms of *an*

4749 *obligation to be consistent*. But if I am right, we might only be able to explain  
 4750 this assumption in terms of *an obligation not to be inconsistent*. Hence,  
 4751 requirements like Consistency might merely be normative in a weak sense.

4752 The good news is that we can now make sense of such a possibility. If the  
 4753 Coextensivity Thesis is false, it makes perfect sense to say that one ought not  
 4754 to be inconsistent without also saying that one ought to be consistent. There  
 4755 might not be something good with being structurally rational, but it seems  
 4756 patently clear that there is something bad with being structurally irrational.\*

Marc-Kevin Daoust

 0000-0002-6710-1944

Harvard University

mk.daoust@live.ca

## 4761 References

- 4762 DE BONA, Glauber and STAFFEL, Julia. 2018. "Why Be (Approximately) Coherent?"  
 4763 *Analysis* 78(3): 405–415, doi:10.1093/analys/anx159.
- 4764 BROOME, John A. 2013. *Rationality Through Reasoning*. Hoboken, New Jersey: John  
 4765 Wiley and Sons, Inc., doi:10.1002/9781118609088.
- 4766 COATES, Allen. 2012. "Rational Epistemic Akrasia." *American Philosophical Quarterly*  
 4767 49(2): 113–124.
- 4768 DAOUST, Marc-Kevin. 2020. "The Explanatory Role of Consistency Requirements."  
 4769 *Synthese* 197(10): 4551–4569, doi:10.1007/s11229-018-01942-8.
- 4770 DORST, Kevin. 2019. "Lockeans Maximize Expected Accuracy." *Mind* 128(509):  
 4771 175–211, doi:10.1093/mind/fzx028.
- 4772 EASWARAN, Kenny. 2016. "Dr. Truthlove or: How I Learned to Stop Worrying and  
 4773 Love Bayesian Probabilities." *Noûs* 50(4): 816–853, doi:10.1111/nous.12099.
- 4774 EASWARAN, Kenny and FITELSON, Branden. 2015. "Accuracy, Coherence, and Ev-  
 4775 idence." in *Oxford Studies in Epistemology*, volume V, edited by Tamar Szabó  
 4776 GENDLER and John HAWTHORNE, pp. 61–96. Oxford: Oxford University Press,  
 4777 doi:10.1093/acprof:oso/9780198722762.003.0003.
- 4778 FITELSON, Branden. 2016. "Coherence." Unpublished manuscript, [http://fitelson.org](http://fitelson.org/coherence/coherence_duke.pdf)  
 4779 /coherence/coherence\_duke.pdf.
- 4780 GOLDMAN, Alvin I. 2015. "Reliabilism, Veritism, and Epistemic Consequentialism."  
 4781 *Episteme* 12(2): 131–143, doi:10.1017/epi.2015.25.

---

\* This research was supported by the Social Sciences and Humanities Research Council (grant #756-2019-0133). Thanks to Samuel Dishaw, Branden Fitelson, Daniel Laurier and Justin Snedegar for helpful comments.



- 4782 GRECO, Daniel. 2014. "A Puzzle about Epistemic Akrasia." *Philosophical Studies* 167(2):  
4783 201–219, doi:10.1007/s11098-012-0085-3.
- 4784 HOROWITZ, Sophie. 2014. "Epistemic Akrasia." *Noûs* 48(4): 718–744, doi:10.1111/no  
4785 us.12026.
- 4786 JOYCE, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of*  
4787 *Science* 65(4): 575–603, doi:10.1086/392661.
- 4788 KIESEWETTER, Benjamin. 2016. "You Ought to  $\phi$  Only If You May Believe That You  
4789 Ought to  $\phi$ ." *The Philosophical Quarterly* 66(265): 760–782, doi:10.1093/pq/pqw012.
- 4790 —. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press, doi:10.1093/  
4791 oso/9780198754282.001.0001.
- 4792 KOLODNY, Niko. 2005. "Why Be Rational?" *Mind* 114(455): 509–563, doi:10.1093/mi  
4793 nd/fzi509.
- 4794 —. 2007a. "State or Process Requirements?" *Mind* 116(462): 371–385, doi:10.1093/mi  
4795 nd/fzm371.
- 4796 —. 2007b. "How Does Coherence Matter?" *Proceedings of the Aristotelian Society* 107:  
4797 229–263, doi:10.1111/j.1467-9264.2007.00220.x.
- 4798 LASONEN-AARNIO, Maria. 2020. "Enkrasia or Evidentialism? Learning to Love Mis-  
4799 match." *Philosophical Studies* 177(3): 597–632, doi:10.1007/s11098-018-1196-2.
- 4800 LEITGEB, Hannes and PETTIGREW, Richard. 2010. "An Objective Justification of  
4801 Bayesianism I: Measuring Inaccuracy." *Philosophy of Science* 77(2): 201–235, doi:10  
4802 .1086/651317.
- 4803 LITTLEJOHN, Clayton. 2018. "Stop Making Sense? On a Puzzle About Rationality." *Philosophy and Phenomenological Research* 96(2): 257–272, doi:10.1111/phpr.122  
4804 71.
- 4805
- 4806 PARFIT, Derek. 2011. *On What Matters, Volume One*. Oxford: Oxford University Press.  
4807 Edited and introduced by Samuel Scheffler, doi:10.1093/acprof:osobl/9780199572  
4808 809.001.0001.
- 4809 PETTIGREW, Richard. 2013. "Accuracy and Evidence." *Dialectica* 67(4): 579–596, doi:10  
4810 .1111/1746-8361.12043.
- 4811 —. 2016a. *Accuracy and the Laws of Credence*. Oxford: Oxford University Press, doi:10  
4812 .1093/acprof:oso/9780198732716.001.0001.
- 4813 —. 2016b. "Jamesian Epistemology Formalised: An Explication of "The Will to Believe." *Episteme* 13(3): 253–268, doi:10.1017/epi.2015.44.
- 4814
- 4815 SNEDEGAR, Justin. 2018. "Reasons For and Reasons Against." *Philosophical Studies*  
4816 175(3): 725–743, doi:10.1007/s11098-017-0889-2.
- 4817 STEINBERGER, Florian. 2019. "Accuracy and Epistemic Conservatism." *Analysis* 79(4):  
4818 658–669, doi:10.1093/analys/any094.
- 4819 TITELBAUM, Michael G. 2015. "Rationality's Fixed Point (or: In Defense of Right  
4820 Reason)." in *Oxford Studies in Epistemology*, volume V, edited by Tamar Szabó  
4821 GENDLER and John HAWTHORNE, pp. 253–294. Oxford: Oxford University Press,  
4822 doi:10.1093/acprof:oso/9780198722762.003.0009.

- 4823 WAY, Jonathan. 2010. "The Normativity of Rationality." *Philosophy Compass* 5(12):  
4824 1057–1068, doi:10.1111/j.1747-9991.2010.00357.x.
- 4825 —. 2013. "Intentions, Akrasia, and Mere Permissibility." *Organon F* 20(4): 588–611,  
4826 <http://www.klemens.sav.sk/fiusav/doc/organon/2013/4/588-611.pdf>.
- 4827 WHITING, Daniel. 2010. "Should I Believe the Truth?" *Dialectica* 64(2): 213–224, doi:10  
4828 .1111/j.1746-8361.2009.01204.x.
- 4829 WORSNIP, Alex. 2018a. "What is (In)coherence?" in *Oxford Studies in Metaethics*, vol-  
4830 ume XIII, edited by Russ SHAFER-LANDAU, pp. 184–206. Oxford: Oxford University  
4831 Press, doi:10.1093/oso/9780198823841.003.0009.
- 4832 —. 2018b. "The Conflict of Evidence and Coherence." *Philosophy and Phenomenologi-  
4833 cal Research* 96(1): 3–44, doi:10.1111/phpr.12246.

PROOF

# Review of Soames (2018)

FRASER MACBRIDE

4835 SOAMES, Scott. 2018. *The Analytic Tradition in Philosophy, Volume 2: A New*  
4836 *Vision*, Princeton: Princeton University Press.

4837 *A New Vision* is the sequel to Soames' *The Analytic Tradition in Philosophy,*  
4838 *Volume I: The Founding Giants* (Princeton UP, 2014). *Founding Giants* covered  
4839 Frege, Moore and Russell. *New Vision* covers Wittgenstein's *Tractatus*,  
4840 the rise of logical empiricism and its downfall, the advances in logic due to  
4841 Gödel, Tarski, Church and Turing, Tarski's theory of truth, and contrasting approaches  
4842 to ethics and meta-ethics in the 1930s. Soames describes his goal as  
4843 being to identify major insights and achievements, distinguishing them from  
4844 major errors or disappointments. His declared focus is explication and evaluation  
4845 of arguments in the texts of Wittgenstein, Carnap *et al.* Thereby Soames  
4846 conceives of himself as "arguing with the greats" rather than historians of  
4847 analytic philosophy. He thereby seeks to avoid the perils of antiquarianism  
4848 which besets history of philosophy when it is bowed down by too much attention  
4849 to historical-textual detail, whilst his engagement with the secondary  
4850 literature is sparse.

4851 I do believe that it is possible to do insightful history of philosophy by  
4852 interrogating dead philosophers as though they were walking amongst us—  
4853 possible because it's actually been done. Exemplars of this kind of work  
4854 are Jonathan Bennett's *Kant's Analytic* (1966) and *Kant's Dialectic* (1974),  
4855 volumes which have stood the test of time, proving fruitful for philosophers  
4856 and historians of philosophy alike. But I don't think that there's a simple  
4857 equation which determines that more history, more textual detail means less  
4858 philosophy—because sometimes more of that is just what's needed to channel  
4859 the philosophy of our forebears. It's because Soames hasn't done enough to get  
4860 the history and the texts right that I think he quite often gets their philosophy  
4861 wrong.

4862 Soames' story in *New Vision* is, as he says, a "complicated" one—  
4863 understandably so because his aim is to engage directly with the arguments of the  
4864 greats and they gave a lot of arguments. As a consequence, *New Vision* might

4865 better be characterised as a collection of interrogative episodes rather than as  
 4866 an extended dialogue. To provide an impression of the whole, I'm going to  
 4867 evaluate one such episode in which Soames attempts to strike up an argument  
 4868 with Wittgenstein.

4869 In *New Vision* Soames takes Wittgenstein to task for what he describes as  
 4870 “among the darkest and most implausible aspects of the *Tractatus*”, Wittgen-  
 4871 stein’s metaphysics of simples and atomic facts configured from them, ideas  
 4872 which Soames does not consider to have had much interest or influence any-  
 4873 way (Soames 2018, 23). Where does Soames think Wittgenstein went wrong?  
 4874 To be blunt: because Wittgenstein had the ill-fortune to come before Kripke.  
 4875 Soames credits Kripke with the land mark discovery that metaphysical and  
 4876 epistemic modalities needn’t march in step but have the potential to diverge,  
 4877 so propositions might be necessary whilst being *a posteriori* and *a priori*  
 4878 though contingent. For Soames this discovery was one of the most remark-  
 4879 able achievements of analytic philosophy in the 20<sup>th</sup> century. But coming  
 4880 before Kripke, Soames claims, Wittgenstein mistakenly identified necessarily  
 4881 true propositions with propositions knowable *a priori*. According to Soames  
 4882 it’s this very mistake, “the notorious tractarian collapse of the modalities”,  
 4883 that led Wittgenstein down the false path to his misbegotten metaphysics of  
 4884 simples and atomic facts (Soames 2018, 14).

4885 Wittgenstein famously advanced his atomism by arguing that if there were  
 4886 only complexes all the way down, “then whether a proposition had sense  
 4887 would depend on whether another proposition was true” (2.0211). This would  
 4888 be an intolerable consequence because, Wittgenstein continued, “[i]t would  
 4889 be impossible to form a picture of the world (true or false)” (2.0212). Since  
 4890 it is possible for us to form a true or false picture of the world, Wittgenstein  
 4891 concluded that the analysis of complexes must terminate in absolute simples.  
 4892 Soames reconstructs Wittgenstein’s argument along the following lines.

4893 Suppose  $S_1$  is a statement affirming the existence of a complex designated  
 4894 by the logically proper name “ $O$ ”. In order for  $S_1$  to “have sense”, by which  
 4895 Wittgenstein means be true or false,  $S_1$ ’s constituent expressions, including  
 4896 “ $O$ ”, must have meaning. In order for “ $O$ ” to have meaning,  $O$  must exist.  
 4897 Because  $O$  is a complex,  $O$  exists if and only if its parts ( $a$ ,  $b$ ,  $c$ ) are arranged a  
 4898 certain way. Let  $S_2$  be the statement whose constituent expressions include  
 4899 logically proper names for  $O$ ’s parts and which says that  $O$ ’s parts are so  
 4900 arranged. Then whether  $S_1$  has sense depends upon whether  $S_2$  is true. But  
 4901 in order for  $S_2$  to have sense its constituent expressions must have meaning  
 4902 too, which they do only if  $O$ ’s parts exist. Since  $O$ ’s parts are complexes too,

4903 they exist if and only if their parts are arranged a certain way. Hence whether  
 4904  $S_2$  has sense depends upon whether another statement  $S_3$  which says that the  
 4905 parts of  $O$ 's parts are so arranged is true, and so on without end. Represent  
 4906 this chain of meaning-truth dependencies as an unending sequence:

4907  $(S) (S_1 \rightarrow S_2), (S_2 \rightarrow S_3), (S_3 \rightarrow S_4), \dots$

4908 Now the key interpretative question is why does Wittgenstein take this regress  
 4909 of one sentence's meaningfulness presupposing the truth of another to be  
 4910 vicious? For Soames it's vital to appreciate that this regress presupposes a  
 4911 chain of necessary connections whereby the existence of a complex is analysed  
 4912 in terms of the existence and arrangement of its parts: necessarily  $O$  exists if  
 4913 and only if  $O$ 's parts exist and they're arranged a certain way, necessarily  $O$ 's  
 4914 parts exist iff the parts of  $O$ 's parts exist and they're arranged a certain way,  
 4915 and so on without end. We can represent this chain as an unending sequence  
 4916 of necessary conditionals:

4917  $(S_{\square}) \square(S_1 \rightarrow S_2), \square(S_2 \rightarrow S_3), \square(S_3 \rightarrow S_4), \dots$

4918 According to Soames, we have seen, Wittgenstein presupposes that necessity  
 4919 and *a priori* knowability coincide. Hence, for Soames' Wittgenstein,  $(S_{\square})$  is  
 4920 equivalent to another non-terminating sequence of *a priori* knowable condi-  
 4921 tional:

4922  $(S_{\text{apriori}}) a \text{ priori knowable } (S_1 \rightarrow S_2), a \text{ priori knowable } (S_2 \rightarrow S_3), a \text{ priori know-}$   
 4923  $\text{able } (S_3 \rightarrow S_4), \dots$

4924 Soames now reasons that if there were no simples "it would follow that  
 4925 *knowing* that [" $O$ "] means what it does" and hence knowing the meaning of  
 4926 the sentences in which " $O$ " occurs, "would require *knowing* the proposition  
 4927 that  $a$ ,  $b$  and  $c$  are composed in the right way" (p. 13). But the same reasoning  
 4928 can be repeated for its parts: "*knowing* that they exist and that propositions  
 4929 about them are meaningful, and have the senses that they do, would require  
 4930 *knowing* the existence of still further objects, as well as the meaningfulness  
 4931 of still further names for those objects and so on without end" (pp. 13–14).  
 4932 Soames concludes: "Thus, if there were no metaphysically simple objects,  
 4933 then one couldn't *know* the meaning of any sentence or perhaps whether it  
 4934 even had a meaning" (p. 14).

4935 Soames' reconstruction of Wittgenstein's argument isn't plausible. Even  
 4936 supposing that  $(S_{\square})$  and  $(S_{\text{apriori}})$  are equivalent it doesn't follow that this

4937 imposes a requirement upon what must be actually known by a speaker  
 4938 who grasps “*O*”. A proposition’s being knowable (*a priori* or otherwise) is  
 4939 quite different from its being known – possibility doesn’t entail actuality.  
 4940 So even if it is *a priori* knowable that  $S_1 \rightarrow S_2$ , it doesn’t follow that anyone  
 4941 actually knows this, much less that a speaker has to actually know  $S_2$  in order  
 4942 to actually know  $S_1$ . Soames supposes that ( $S_{\text{apriori}}$ ) imposes an unending,  
 4943 therefore unsatisfiable set of necessary conditions upon actually knowing  
 4944 that *O* exists. But because ( $S_{\text{apriori}}$ ) covers only the weaker modality of what  
 4945 is knowable, it remains open that a speaker might know  $S_1$  and not know  $S_2$   
 4946 even if  $S_1 \rightarrow S_2$  is *a priori* knowable.

4947 The upshot is that Soames fails to explain how the *a priori* knowability of  
 4948  $S_1 \rightarrow S_2$  etc. imposes a requirement upon what must be known by someone  
 4949 who understands “*O*”. All that Soames establishes is that if there is complexity  
 4950 all the way down, then there is an indefinite potential for unpacking *O*’s complex-  
 4951 ity, a potential that can be realised by actually coming to know *a priori*  
 4952  $S_1 \rightarrow S_2$ ,  $S_2 \rightarrow S_3$  etc. This might be a surprising view to hold. But since Soames  
 4953 hasn’t shown that speakers would have to actually exhaust (*per impossibile*)  
 4954 the potential for unpacking *O*’s complexity in order to grasp “*O*”’s meaning,  
 4955 Soames sheds no light upon Wittgenstein’s claim that if there was complex-  
 4956 ity all the way down, it would be impossible to say something about *O* (or  
 4957 any other object). So it’s hard to see that Soames succeeds in striking up a  
 4958 conversation with Wittgenstein rather than talking past him.

4959 Where Soames has gone adrift is failing to factor in Wittgenstein’s own  
 4960 insistence that a non-terminating sequence of meaning-truth dependencies  
 4961 would make it impossible to “form”, or more literally “draw up” [“entwerfen”],  
 4962 “a picture of the world (true or false)” (2.0212). By “picture of the world (true or  
 4963 false)”, Wittgenstein doesn’t simply mean “bearer of truth or falsity” but points  
 4964 us further into the interior of the *Tractatus* where a more demanding notion  
 4965 of a proposition and what it is to grasp a proposition awaits us – Wittgenstein’s  
 4966 picture theory. It’s because a non-terminating sequence of meaning-truth  
 4967 dependencies is incompatible with the possibility of a proposition in this  
 4968 more demanding sense that Wittgenstein concludes that there cannot be  
 4969 complexity all the way down (as I argue in 2018, 188–190).

4970 Let me elaborate briefly upon this alternative interpretation. When we read  
 4971 further into the *Tractatus* we find that a proposition is a complete picture  
 4972 of reality in the sense that when a speaker understands a proposition, they  
 4973 have an exact knowledge of how objects must be arranged for that statement  
 4974 to be true or false and which arrangements of them are thereby left open.

4975 And this is information a speaker can uptake with effortless facility: “The  
 4976 proposition is a picture of reality, for I know the state of affairs presented by  
 4977 it, if I understand the proposition. And I understand the proposition, without  
 4978 its sense having been explained to me” (4.021). So a speaker must already  
 4979 actually know everything she/he needs to know to understand how things  
 4980 must be arranged for a proposition to be true even if the proposition isn’t  
 4981 one she/he has heard before. But a speaker couldn’t have knowledge of what  
 4982 it takes for a proposition to be true (or false) and what is thereby left open  
 4983 if she/he had *per impossibile* to check and see whether a non-terminating  
 4984 sequence of meaning-truth dependencies was satisfied for every expression of  
 4985 their language. A speaker wouldn’t be in a position to know straightaway that  
 4986 the expressions of their language were meaningful but only have a supertask  
 4987 ahead of them. As finite agents, speakers could never confirm that more than  
 4988 an initial segment of the sequence was satisfied, so never be in a position  
 4989 to exercise the consummate facility with language with which Wittgenstein  
 4990 credits speakers.

4991 By contrast to Soames’ account, this interpretation has the merit of making  
 4992 immediate contact with what speakers are required to know to understand a  
 4993 language and it makes sense of Wittgenstein’s argument at 2.0211-2.0212 in  
 4994 the wider context of Wittgenstein’s commitment to the picture theory. It’s a  
 4995 further consequence of this interpretation that what Soames describes as the  
 4996 “notorious Tractarian collapse of the modalities” plays no significant role in  
 4997 Wittgenstein’s argument – Soames’ original mistake was to read the *Tractatus*  
 4998 through “Kripke goggles.”

4999 I have concentrated upon one interrogative episode of *New Vision* to give a  
 5000 representative impression, but I might have taken issue with other episodes  
 5001 where, it seems to me, Soames’ arguments falter for lack of engagement with  
 5002 the historical texts. Consider, for example, his dismissal of the Tractarian  
 5003 conception of a proposition as a propositional sign in its projective relation to  
 5004 the world in favour of his own cognitive act type theory. Or his criticism of  
 5005 the *Aufbau* that Carnap failed to realise that statements expressed in purely  
 5006 logical vocabulary have no empirical content when, Soames has forgotten,  
 5007 “ $\exists x \exists y (x \neq y)$ ” consists of purely logical vocabulary but remains verifiable or  
 5008 falsifiable depending on how many things there are.

Fraser MacBride

 0000-0003-2374-044

University of Manchester

fraser.macbride@manchester.ac.uk

5009  
5010  
5011  
5012

5013

## References

- 5014 BENNETT, Jonathan. 1966. *Kant's Analytic*. Cambridge: Cambridge University Press.  
5015 Reissued as Bennett (2016a), doi:10.1017/cbo9780511554506.
- 5016 —. 1974. *Kant's Dialectic*. Cambridge: Cambridge University Press. Reissued as Bennett  
5017 (2016b).
- 5018 —. 2016a. *Kant's Analytic*. Philosophy Classics. Cambridge: Cambridge University  
5019 Press. Reissue of Bennett (1966), doi:10.1017/cbo9781316492901.
- 5020 —. 2016b. *Kant's Dialectic*. Philosophy Classics. Cambridge: Cambridge University  
5021 Press. Reissue of Bennett (1974), doi:10.1017/cbo9781316492949.
- 5022 MACBRIDE, Fraser. 2018. *On the Genealogy of Universals: The Metaphysical Origins of*  
5023 *Analytic Philosophy*. Oxford: Oxford University Press, doi:10.1093/oso/97801988  
5024 11251.001.0001.
- 5025 SOAMES, Scott. 2018. *The Analytic Tradition in Philosophy, Volume 2: A New Vision*.  
5026 Princeton, New Jersey: Princeton University Press, doi:10.1007/s11098-019-01267-  
5027 3.



# Review of Oppy (2018)

MARIO SCHÄRLI

5028

5029 OPPY, Graham, ed. 2018. *Ontological Arguments*. Cambridge: Cambridge  
5030 University Press.

5031 A shadow of criticism has followed ontological arguments for almost a thou-  
5032 sand years and counting. Irrespectively, the arguments continue to intrigue  
5033 philosophical thought, and no decline is in sight. In particular, modal versions  
5034 of the argument formulated by Hartshorne, Lewis, Plantinga and Gödel in  
5035 the 1960's and 70's have helped to dispel the widely held suspicion that a  
5036 simple logical blunder lies behind ontological arguments. As a result, recent  
5037 discussions have shifted from assessing the argument's validity towards its  
5038 soundness and dialectical efficacy. This requires engaging with the philosoph-  
5039 ical issues inevitably raised by the argument, such as questions about the  
5040 nature of concepts and arguments, existence and possibility. These have since  
5041 stood at the forefront of the debate.

5042 The concerns united by reference to ontological arguments form the subject  
5043 matter of a recently published volume edited by Graham Oppy, himself one  
5044 of the most prolific authors on the topic in the past 25 years. His informative  
5045 introductory essay underlines important differences between the arguments  
5046 commonly called "ontological." Oppy suggests abandoning the search for  
5047 unity suggested by the description "the ontological argument." Instead, the  
5048 commonalities should be viewed genealogically: "What is distinctive of onto-  
5049 logical arguments is that their formulation has the right kind of connection  
5050 to Anselm's argument" (p. 11). Hence, fruitful engagement with and criticism  
5051 of ontological arguments proceeds by cases.

5052 This sets the tone for the volume's first group of articles which are devoted to  
5053 defenders and critics of the ontological argument, namely: Anselm, Aquinas,  
5054 Descartes, Leibniz, Kant, Hegel, Gödel, Lewis, Plantinga, and Tichý. A second  
5055 group of three essays dealing with overarching systematic issues surrounding  
5056 the preceding arguments complements the volume. Here we find treatments of  
5057 the relation between conceivability and possibility, the "fallacy" of begging the  
5058 question, and the relation between existence, characterization and modality.

5059 Overall, the volume provides readers with informative up-to-date discussions  
 5060 of ontological arguments of scholarly value by senior researchers in the field.  
 5061 (With the notable exception of M. Inwood’s article on Hegel which lacks  
 5062 engagement with the literature on the subject.) At the same time, the essays  
 5063 are written in an accessible manner, rendering the volume suitable as an  
 5064 accompaniment to graduate-level courses on the subject. Due to limitations  
 5065 in space, I will refrain from summarizing and discussing all the contributions.  
 5066 For that purpose, Oppy’s introduction (pp. 2–5) is well suited. Instead, I will  
 5067 focus on three contributions I found particularly worth discussing.

5068 The majority of ontological arguments treated in the volume—Anselm’s,  
 5069 Leibniz’s, Gödel’s, Plantinga’s—are shown to be deductive in nature by their  
 5070 interpreters. A noticeable rift opens up between them and Descartes’ argu-  
 5071 ment, according to Lawrence Nolan’s interpretation. His article represents an  
 5072 important scholarly contribution because it virtually reverses the standard  
 5073 deductive reading, and plausibly so.<sup>1</sup> Developing a suggestion hinted at by  
 5074 M. Gueroult and J. Barnes, Nolan interprets Descartes’ so-called ontological  
 5075 argument as “the report of an intuition in the sense of a non-discursive, self-  
 5076 validating, intellectual apprehension” (p. 54). The aims Descartes pursues  
 5077 with the argument are persuasive rather than argumentative: all he points  
 5078 to serves the purpose of getting the meditator to have the relevant intuitive  
 5079 insight.

5080 A strength of Nolan’s reading is that it allows us to make good sense of  
 5081 passages (e.g. *Med.* V, AT VII 68–69 and *Princ.* I., §15), where Descartes clearly  
 5082 glosses the cognition of God as an intuitive insight; these have always been  
 5083 difficult to accommodate within deductive interpretations of the argument.  
 5084 Moreover, Nolan convincingly shows that his reading coheres with Descartes’  
 5085 skepticism towards a formal-deductive understanding of reasoning voiced in  
 5086 the *Rules* as well as the other philosophical doctrines he adheres to (pp. 57–65).  
 5087 However, the intuitive reading of the argument has to confront the follow-  
 5088 ing difficulty: what about the passages where Descartes overtly argues in a  
 5089 deductive manner?

5090 Nolan uses two principal interpretive moves to provide a coherent picture  
 5091 in these cases (pp. 54, 66–71). First, he convincingly shows that the overtly  
 5092 argumentative passages, commonly taken to be Descartes’ argument, are best  
 5093 read as rebuttals of possible criticisms. They allow the meditator’s intuition  
 5094 not to be distracted by an unjustified conception, e.g. by understanding the

---

1 Cf. also his earlier “The Ontological Argument as an Exercise in Cartesian Therapy” (2005).

5095 distinction between essence and existence as a real rather than merely rational  
5096 distinction. Second, he argues that, for historical reasons, Descartes aimed  
5097 to present his philosophy in a manner adjusted to the scholarly discourse of  
5098 his day, which put great emphasis on the syllogistic demonstrability of God's  
5099 existence.

5100 While the latter may be correct as a matter of historical fact, Nolan's line of  
5101 interpretation may be bolstered by a more penetrating understanding of the  
5102 relation between intuition and deduction. It is Descartes' view that deduction  
5103 is necessary in case one does not have clear and distinct, intuitive insight at  
5104 one's disposal (p. 61). Although this legitimizes ascribing priority to intuitive  
5105 over deductive insight, it does not imply a merely historical explanation of the  
5106 occurrence of deduction. Rather, one might—in line with Descartes' general  
5107 manner of proceeding in the *Meditationes*—explain the deductive arguments  
5108 as necessary steps towards intuitive insights. It helps to take into consideration  
5109 what the condition for distinctly perceiving a given content is: being able to  
5110 tell it apart from others (*Princ. I.*, §45). If that is the case, then the arguments  
5111 delivered to fend off criticisms are not merely negative or persuasive, but  
5112 positively contribute to the distinctness of the meditator's perception and  
5113 thus to its intuitiveness. This should not be understood as a criticism, but as  
5114 additional support for Nolan's reading. In my view, Nolan's essay represents a  
5115 lasting contribution to our understanding of Descartes. Moreover, it points to a  
5116 version of the ontological argument that might merit systematic development  
5117 in the light of recent advancements in the epistemology of intuition.<sup>2</sup>

5118 Other than authors who defend the ontological argument, the volume  
5119 features some of its most important critics in Aquinas, Kant, and Lewis.  
5120 Among these, L. Pasternack's perceptive and well-informed article on Kant  
5121 is one of the best discussions currently available. It sets the record straight  
5122 on the nature of Kant's case against the ontological argument. Contrary to  
5123 popular wisdom, the latter extends well beyond the familiar line "existence is  
5124 not a real predicate". Pasternack distinguishes two main strands of criticism  
5125 within Kant's argumentation in the *Critique of Pure Reason*, the first of which  
5126 targets an analytic, the second a synthetic reading of the judgment "God  
5127 exists" (p. 102). Kant argues that an ontological argument insisting on the  
5128 analytic reading of the statement is dialectically flawed, i.e. does not add up

---

2 First and foremost: Chudnoff (2013).

5129 to an argument at all (pp. 104, 106)<sup>3</sup>, while a synthetic reading rests on the  
5130 thesis that existence is a “real predicate” which Kant disputes (pp. 106–115).

5131 The soundness of the second part of Kant’s criticism rests on an argument  
5132 against existence being a real predicate, which Pasternack deems inconclusive.  
5133 His rendering is as follows: if existence were a property of objects, then a  
5134 concept specifying all the properties of the object, but lacking existence as a  
5135 mark, would fail to fully articulate the object in question, leading to a “mis-  
5136 match” between concepts and their objects. This argument is unconvincing  
5137 for two related reasons. First, it leaves open why this mismatch should be  
5138 deemed problematic according to Kant, which needs to be established for the  
5139 argument to be sound. Pasternack appears to agree on this point, which leads  
5140 to the second weakness of the argument: it is susceptible to the “obvious re-  
5141 buttal” Pasternack puts forward. Basically, it consists in making the mismatch  
5142 disappear by allowing existence to be a mark of concepts (p. 114).

5143 However, a more convincing reading of Kant’s argument is possible. Imme-  
5144 diately after the passage Pasternack quotes in support of his reading, Kant  
5145 writes: “Even if I think in a thing every reality except one, then the missing  
5146 reality does not get added when I say the thing exists, but it exists encum-  
5147 bered with just the same defect as I have thought in it; otherwise something  
5148 other than what I thought would exist” (A600/B628). It emerges from this  
5149 sentence that the alleged “mismatch”, i.e. a concept’s not fully capturing all  
5150 the properties its instances exhibit, is not what is at issue, at least as far as  
5151 Kant perceives matters. On the contrary, his point concerns instantiation, or  
5152 the relation between concepts and objects, in general. This addresses the first  
5153 weakness of Kant’s case as interpreted by Pasternack. But what is Kant’s point  
5154 then?

5155 Kant argues that a concept’s instantiation does not correspond to any *addi-*  
5156 *tion* of properties to it; rather, a concept’s instantiation amounts to the object’s  
5157 having *just* the properties the concept specifies. A plausible way of construing  
5158 this claim is: A concept’s content consists in the conditions an object has to  
5159 meet in order to count as an instance of it. Kant can be understood as showing  
5160 that this view cannot be upheld if one understands existence as a property. The  
5161 reasoning can be understood as follows. If existence were a property of objects  
5162 and being an instance of a concept is to exist, then a concept’s instances would

---

3 It is, of course, a common criticism of ontological arguments that they are question-begging; e.g. Aquinas raises a similar point according to B. Leftow’s reconstruction (pp. 47, 49, 51) and P. van Inwagen discusses the issue concerning the modal ontological argument in his contribution (pp. 238–249).

5163 consequently have to bear the property “existence”. If “falling under a concept”  
 5164 consists in an object’s *conforming* to the conditions set by the concept and  
 5165 existence is one of these conditions, then existence would have to be a mark  
 5166 of the concept. But this would render some existence-judgments analytic—a  
 5167 view Kant takes himself to have refuted at this point in the discussion. If one  
 5168 grants this, it follows that existence is not a mark of any concept. But if it still  
 5169 holds that instances of concepts exist, then a concept’s instantiation consists  
 5170 in an object’s conforming to the conditions set *and* exhibiting the property  
 5171 “existence” *additionally*. As the latter is not part of a concept’s content, this  
 5172 content’s identity therefore cannot consist in a specification of what it takes  
 5173 to be its instance, no matter how completely or incompletely it captures an  
 5174 object’s properties.

5175 Kant therefore does not argue that a mismatch between concepts and  
 5176 objects is problematic as such, but that a specification of a concept’s content  
 5177 in terms of conditions instances have to meet is impossible given that one  
 5178 accepts the following three theses:

- 5179 (1) a concept is individuated by the conditions on objects to count as in-  
 5180 stances of it;
- 5181 (2) existence is a property of objects;
- 5182 (3) existence-judgments are synthetic.

5183 Kant’s point therefore is: the view that existence is a property is indicative  
 5184 of a misunderstanding of what concepts and “falling under a concept” are.  
 5185 Pasternack’s rebuttal misses the mark in relation to this issue, for accepting  
 5186 existence as property and conceptual content *either* leads to the implausible  
 5187 view that all existence-judgments are analytic *or*, if they remain synthetic,  
 5188 precludes a conception of concepts as specifying the conditions of what it is  
 5189 to fall under them.

5190 Alongside “the usual suspects”, Pavel Tichý’s work on the ontological ar-  
 5191 gument makes an unexpected appearance in the volume. As is convincingly  
 5192 shown by G. Oddie’s essay, Tichý offers one of the most penetrating and re-  
 5193 vealing interpretations of Anselm’s *Proslogion* III, i.e. the passage serving as  
 5194 inspiration for what is known as “the modal ontological argument”. Tichý  
 5195 delivers a logically valid reconstruction of Anselm’s argument as well as an  
 5196 unfamiliar axiological criticism of its soundness.

5197 The reading rests on Tichý’s ontology, fundamental to which is the dis-  
 5198 tinction between “two entirely different *types* of entity” (p. 199): *individuals*

(such as Donald Trump) and *offices* (such as “the President of the U.S.A”).  
 Intuitively, offices are either occupied by an individual or not, where occupancy is to be understood as a property of the office rather than the individual. Formally, offices are partial functions mapping world-time pairs to individuals which are undefined when the office goes unoccupied. What an office is—its essence—is given as a set of conditions called *requisites* which have to be borne by occupants to count as such (p. 203).

Within this framework, the modal ontological argument aims to derive the necessary occupation of “the divine office” (p. 205), which Tichý interprets as “*that individual office such that no individual office is greater than it*” (p. 206). Anselm’s formula, thus understood, singles out a second-order office, that is, an office occupied by a first-order office rather than an individual. Glossing over the details of the reconstruction, the *Proslogion* III argument derives necessary existence as a requisite of the greatest office, yielding the conclusion that the divine office is necessarily occupied. This yields a “valid” argument according to Oddie (p. 209).


Compared to the standard modal ontological argument known from the writings of Harthshorne and Plantinga, Tichý’s interpretation of the argument has one key advantage. The standard version treats existence in all possible worlds as an essential property of God and derives God’s existence from His/Her possible existence plus S5. The argument is often criticized for begging the question because the premise that God’s existence is possible cannot be substantiated in a non-circular fashion. G. Priest’s offers one way of putting the difficulty (p. 265).<sup>4</sup> According to the premises of the argument, the following two entailments hold: (1) God’s actual existence follows from His/Her possible existence; (2) God’s actual existence entails His/Her possible existence. “God exists” and “possibly, God exists” are therefore equivalent according to the argument’s premises. Hence, presupposing the possibility of God’s existence is question-begging insofar as it is equivalent to presupposing God’s existence. By contrast, Tichý’s reconstruction derives the necessary occupancy of the divine office via an axiological premise, namely: necessarily occupied offices are always greater than ones which are not (p. 208). The truth of this premise can be assessed independently, and hence God’s necessary existence gets established in a more satisfactory way.

<sup>4</sup> Ways of stating and resolving the difficulty are discussed in the articles of J. Spencer, J. Rasmussen, P. van Inwagen in the volume. Rasmussen tries to make progress on the issue by providing an independent argument for God’s possibility turning on the modal properties of value (pp. 183–185). I find his argument unconvincing, but due to limitations in space, I cannot give my reasons here.

5233 However, this premise also renders the argument unsound according to  
 5234 Tichý. The claim that necessarily occupied offices are always greater than  
 5235 ones which are not is subject to counterexamples, one of which is: the office  
 5236 “*the discoverer of the incompleteness of arithmetic*” is contingently occupied,  
 5237 whereas the office “*the pick of the morally most depraved*”, where “pick” refers  
 5238 to a choice function to be applied in case of a tie, is necessarily occupied. Yet,  
 5239 the former is plausibly “greater” than the latter (p. 212). Therefore, Oppy con-  
 5240 cludes with Tichý, the argument rests on an implausible axiology of existence.  
 5241 Attempts at weakening the relevant requisite (e.g. either being God or else  
 5242 the pick of the morally best) will, while ending up necessarily occupied, fail  
 5243 to prove the existence of God at all world-times, for God is not merely the  
 5244 relatively morally best being, but the absolutely best (pp. 212–213).

5245 Oddie’s simultaneously fascinating and accessible discussion of Tichý’s  
 5246 reconstruction will hopefully lead to the recognition of what strikes me as the  
 5247 most convincing version of a *Proslogion* III-style modal ontological argument.  
 5248 Further discussion may delve deeper into the axiological questions raised by  
 5249 Tichý. As is always the case with criticisms resting on counterexamples, they  
 5250 may show *that*, but not explain *why*, some thesis is false. What principled  
 5251 reason against Anselm’s axiology can be given?\*

5252 Mario Schärli

5253  0000-0002-9380-4555

5254 Université de Fribourg

5255 mario.schaerli@unifr.ch

## 5256 References

- 5257 CHUDNOFF, Elijah. 2013. *Intuition*. Oxford: Oxford University Press, doi:10.1093/acpr  
 5258 of:oso/9780199683000.001.0001.
- 5259 NOLAN, Lawrence. 2005. “The Ontological Argument as an Exercise in Cartesian  
 5260 Therapy.” *Canadian Journal of Philosophy* 35(4): 521–562, doi:10.1080/00455091.2  
 5261 005.10716601.

---

\* Research going into this article has been funded by the Swiss National Science Foundation, Grant No. P1FRP1\_188051. I am grateful to Anna Katsman and Jelscha Schmid for critical remarks on a draft version of this article.

PROOF



# Review of Antonelli (2018)

HAMID TAIEB

5263 ANTONELLI, Mauro. 2018. *Vittorio Benussi in the History of Psychology: New*  
 5264 *Ideas of a Century Ago*, Studies in the History of Philosophy of Mind 21,  
 5265 Cham: Springer.

5266 The history of phenomenology has not been a peaceful and autonomous pro-  
 5267 cess taking place independently of any competitors. On the contrary, from  
 5268 the very beginning of their inquiries, phenomenologists had to struggle with  
 5269 several rival explanatory schemes in psychology. The most important among  
 5270 them were physiological psychology (of various sorts) and psychoanalysis.  
 5271 Both of these scientific projects tried to minimize the importance of conscious-  
 5272 ness in the explanation of the mind, the first by treating consciousness as  
 5273 some sort of epiphenomenal outcome of brain and other nervous processes,  
 5274 the second by describing it as a blind domain, driven by underlying mental  
 5275 acts to which consciousness itself has no access. Interestingly, however, phe-  
 5276 nomenology did not ignore these two competing explanatory schemes; on  
 5277 the contrary, it entered into manifold discussion with them, trying to estab-  
 5278 lish more and more precisely the “division of scientific labour” among these  
 5279 three approaches. Evidence of this engagement is plentiful. With respect to  
 5280 physiological psychology, the discussion goes as far back as Franz Brentano,  
 5281 who tried to combine his “descriptive psychology”, also called “descriptive  
 5282 phenomenology”, with “genetic psychology”, that is, physiological psychology;  
 5283 and it has had a long and complex history, up to the most recent papers pub-  
 5284 lished in the journal *Phenomenology and the Cognitive Sciences*. With respect  
 5285 to psychoanalysis, phenomenologists such as Merleau-Ponty and Ricœur  
 5286 engaged in detail with the thought of Freud (who, by the way, had been a  
 5287 student of Brentano); there have also been more recent attempts to combine  
 5288 these two traditions, for example by Lohmar and Brudzinska (2012). However,  
 5289 as shown by Mauro Antonelli, the first ecumenical hero in this history, who  
 5290 combined in a harmonious way all three disciplines—that is, phenomenology,  
 5291 physiological psychology, and psychoanalysis—was Vittorio Benussi.

5292 In reading Antonelli's book, one comes to realize that Benussi, who is described as an "*Einzelgänger*" (p. 238), is a figure as important as he is unknown.  
 5293 Antonelli very nicely combines detailed analysis of Benussi's philosophy of  
 5294 mind with description of the historical and scientific background in which  
 5295 Benussi developed his work. Benussi's life was rich, but also "tragic", as  
 5296 Antonelli emphasizes. Born in Trieste in 1878, Benussi moved to Graz at the age  
 5297 of 18, where he studied with and was influenced by Meinong, and through  
 5298 him by Brentano, with whom Meinong had studied. In Graz, Benussi did  
 5299 not have a permanent academic position: he was a temporary assistant in  
 5300 Meinong's psychology laboratory and worked at the university library to earn  
 5301 enough money to live; but with access to Meinong's laboratory, becoming  
 5302 even its "*de facto* director" (p. 112), he developed his own research agenda.  
 5303 After Trieste was absorbed by Italy following the First World War, he became  
 5304 an Italian citizen, and as a result he lost his position as a librarian in Graz,  
 5305 and was forced to move to Padua. He then fell into a deep depression, despite  
 5306 being hired as a professor at the University of Padua soon after arriving in  
 5307 the city. He committed suicide in 1927 at the age of forty-nine by drinking  
 5308 cyanide, just as in a dream years earlier.

5310 After a short but useful introduction (ch. 1), which explains the *raison d'être*  
 5311 for a monograph on Benussi, Antonelli presents the state of the art in  
 5312 psychology in the German-speaking world at the end of the 19th century  
 5313 and provides a brief overview of Brentanian and Meinongian philosophy and  
 5314 psychology (ch. 2). Following these helpful chapters of contextualization, and  
 5315 a biographical sketch of Benussi (ch. 3), Antonelli enters into the details of  
 5316 Benussi's work and impressive research program. Benussi is mostly known  
 5317 for having developed a theory of *Gestalt*. He was a member of the so-called  
 5318 "Graz School" of *Gestalt* theory, which was opposed to the "Berlin School" of  
 5319 Wolfgang Köhler and his associates. *Gestalten* are, roughly speaking, complex  
 5320 but unitary entities based on a series of elements, to which, however, they  
 5321 are not reducible; for example, a melody is a *Gestalt*, which is based on but  
 5322 not reducible to the series of sounds that compose it. Benussi emphasized the  
 5323 importance of subjective activity in the production of *Gestalten*, whereas the  
 5324 Berlin School had an objectivist account of them (see ch. 4.3 and 4.6, which  
 5325 present in detail Benussi's views, including his evolution on the topic, due in  
 5326 part to objections from the Berlin Gestaltists). However, as clearly shown by  
 5327 Antonelli, Benussi's research extended far beyond Gestalt theory; among the  
 5328 topics on which he worked were the classification of mental acts, the distinction  
 5329 between intentional content and object, sensory illusions, judgments and

5330 “assumptions” (or “pseudo-judgments”), the theory of “productive presenta-  
5331 tions” (which explains, among others things, the constitution of *Gestalten*), the  
5332 relation between emotions and cognition, and time-consciousness (ch. 4.4);  
5333 beyond these rather classical themes of Brentanian and Meinongian psychol-  
5334 ogy (ch. 4.2), but also of Würzburgian *Denkpsychologie*, another source of  
5335 inspiration for him, Benussi worked on testimony, including lie detection  
5336 (ch. 4.7), unconscious mental phenomena, including their relation to dreams  
5337 (ch. 4.5 and 5.4), and the influence of the body on emotions (ch. 5.2), as well  
5338 as mental analysis (ch. 5.1) and hypnosis (ch. 5.3), these themes being mostly  
5339 develop in his later, Padua period, perhaps due to the fact that he had no labo-  
5340 ratory allowing him to continue his work on sensation and *Gestalt* (p. 261). On  
5341 all these themes, the reader will find original and highly interesting develop-  
5342 ments, due first to Benussi’s careful experimentations and analyses, founded  
5343 on methodological reflections about psychology and its relation to philosophy  
5344 (ch. 4.1), and second to Antonelli’s clear and detailed reconstruction, made  
5345 possible by an impressive knowledge of Benussi’s work, including his *Nach-*  
5346 *lass* (which is presented at the end of the volume, along with a bibliography  
5347 and a list of the lecture courses that Benussi delivered at the universities of  
5348 Graz and Padua), and by a rare sense of synthesis. The Conclusion (ch. 6)  
5349 shows that Benussi’s work could be applied to draw interesting connections  
5350 between phenomenology and enactivism on the one hand, and contemporary  
5351 neurosciences, biology, and pragmatics on the other.

5352 Obviously, it is impossible in this review to address all of the topics listed  
5353 above. I would like to focus on one aspect of Benussi’s work, namely, his  
5354 account of emotions, which will also be the occasion to discuss some cru-  
5355 cial methodological points that he defends about psychology. In the Brenta-  
5356 nian tradition, an important psychological thesis, which is not based on any  
5357 empirical-inductive generalization, but is meant to be an *a priori* truth, is that  
5358 no emotion can take place without an underlying presentation: emotions are  
5359 all *about* something, or have an object, and this object is provided to them by a  
5360 presentation on which they, thus, depend. Interestingly, this thesis is attacked  
5361 by Benussi, who holds explicitly that such a view is a mere philosophical  
5362 *speculation* (pp. 277–278). His position is based on specific empirical findings,  
5363 as he wanted psychology to rely on experience, and thus adopted a “theoret-  
5364 ical minimalism” (pp. 145–147, Antonelli quoting an expression from Sadi  
5365 Marhaba); in this respect, according to Antonelli, Benussi’s approach is to be  
5366 placed somewhere between the philosophical phenomenology of Husserl and  
5367 the experimental phenomenology of Stumpf (p. 320).

5368 What then was Benussi's empirical ground for his thesis of the non-  
5369 intentionality of emotions? He applied his "analytic" method in psychology,  
5370 the idea being that the mental life is a "harmonious coordination of  
5371 autonomous elementary functions" (as Benussi puts it) that one can  
5372 "disarticulate", pretty much on the model of vivisection (p. 262). One of  
5373 the tools that Benussi used for performing these vivisections was hypnosis.  
5374 Now, one state to which he was able to lead the persons on whom he was  
5375 testing his hypotheses was that of "basic sleep", a state in which, supposedly,  
5376 subjects had their "conscious intellectual life" interrupted while being still  
5377 able to have some specific feelings. Once put in these states, the subjects  
5378 were suggestible, and Benussi would invite them to have specific emotions,  
5379 such as hate. When they came back to consciousness, they were asked to  
5380 report what they experienced. Now, according to their testimonies, they did  
5381 indeed experience specific emotions such as hate, but given the absence of  
5382 intellectual awareness these emotions were deprived of any object (p. 278). In  
5383 fact, the test subjects reported a series of "kinaesthetic and muscle sensations",  
5384 which Benussi apparently took to be constitutive of emotions. All this was  
5385 proof, for Benussi, that intentionality is not necessary to emotions, and  
5386 thus that the philosophical thesis that emotions are based on an underlying  
5387 presentation is speculative. Note that Benussi defended the view that  
5388 emotions might be intimately linked with an "organic-visceral sensitivity"  
5389 (as Antonelli puts it, p. 315), to the extent that they might be generated by  
5390 viscera and other organs, including the lungs (pp. 303-304); as such, they  
5391 would be the product of a "physiological unconscious" (p. 316). Benussi  
5392 was thus connecting the mind closely to the body, and through it to the  
5393 evolution of the species; in this, as Antonelli emphasizes, Benussi anticipated  
5394 various contemporary theories, notably those of Antonio Damasio and Jaak  
5395 Panksepp, and evolutionism more broadly.

5396 These considerations about emotion are particularly interesting, as Ben-  
5397 nussi's views anticipate various contemporary hypotheses and debates. They  
5398 also seem to develop an account of emotions very much like that of William  
5399 James, for whom emotions are feelings of bodily processes. Now, in contem-  
5400 porary philosophy, the Jamesian account of emotions has been challenged in  
5401 favour of a model which defends the intentionality of emotions. (For a good  
5402 overview on contemporary theories of emotions, see [Scarantino and de Sousa](#)  
5403 [2018](#).) It would be interesting to compare Benussi's views on emotions with  
5404 those of contemporary philosophers, which Antonelli does not do, despite  
5405 his general willingness to make such comparisons with more recent thinkers.

5406 Independently of this, however, a question that is raised by this theoretical  
5407 conflict about emotions is that of the delimitation of the scope of Benussi's  
5408 research. Benussi criticizes speculative approaches to the philosophy of mind  
5409 and praises empirical inquiries. However, the people with whom Antonelli  
5410 compares him—not just Husserl, but also Brentano and Stumpf—all agree  
5411 on one important point: they admit *a priori* truths in philosophy of mind,  
5412 and they are very careful — especially Husserl—to distinguish this “eidetic  
5413 phenomenology”, which is about the nature or essence of mental acts and  
5414 states, from empirical psychology, which is devoted to the study of the mental  
5415 life of a determinate natural species (e.g. human beings). Benussi's attraction  
5416 to empirical research might have led him to neglect this distinction too much.  
5417 Indeed, the distinction does not play a major role in Antonelli's book. Keeping  
5418 this distinction in mind, however, leads to a more accurate determination of  
5419 the scope of one's psychological research, since it allows one to distinguish  
5420 in one's inquiries between what belongs to a mental phenomenon as such,  
5421 and what belongs to it insofar as it is implemented in a certain kind of living  
5422 being. This might have important consequences for the way one describes  
5423 and understands a given phenomenon. As regards emotions, couldn't one say  
5424 that the feelings Benussi is pointing to are not themselves the emotion of, say,  
5425 hate, but merely some bodily impressions that human beings contingently  
5426 co-experience while feeling hate? In that case, what Benussi's subjects are  
5427 reporting are these feelings, which they confuse with hate properly speaking  
5428 simply because they are concomitant, while hate as such, by its very nature  
5429 or essence, has another structure, being object-directed.

5430 Such interrogations can be extended to all dimensions of psychology, and  
5431 were in fact extended in this way by Husserl and others. As Antonelli shows,  
5432 Benussi developed, in parallel to Husserl, a genetic phenomenology which  
5433 studies how the subject passively and unconsciously constitutes the identity  
5434 of perceptual objects despite constant perceptual variations, organizes the  
5435 perceptual field, produces *Gestalten*, etc. But here too, Husserl pointed out  
5436 the possibility of an *a priori* knowledge, since these processes have their own  
5437 essential rules, which are independent of being instantiated in this or that  
5438 natural species (see e.g. Husserl's *Passive Synthesis*, Hua 9, 121.34–123.28, and  
5439 Elmar Holenstein's (1972, 22–25) study on association of ideas in Husserl).  
5440 In sum, a question that remains open when reading Antonelli's book, in  
5441 the discussion of emotions and elsewhere, is whether Benussi's criticism  
5442 of “speculative” philosophy goes too strongly in the opposite direction, by  
5443 blurring an important distinction found among other phenomenologists of

5444 his time. And behind this question is the more fundamental one of whether  
 5445 it is legitimate to accept something like a “philosophical psychology” which  
 5446 supposedly has its own proper task that is distinct from that of empirical  
 5447 psychology. Perhaps Benussi underestimated the importance of this issue.

5448 But these reflections should not distract us from the most important point:  
 5449 Antonelli’s book is a fascinating, well-informed, and admirably clear study  
 5450 which should be read by everyone interested in the history of psychology and  
 5451 phenomenology. It also extends the canon in the philosophy of mind by reha-  
 5452 bilitating an unduly neglected figure who managed to combine, long before  
 5453 others, the theoretical insights of phenomenology, physiological psychology,  
 5454 and psychoanalysis. There is no doubt that Vittorio Benussi’s theoretical  
 5455 project remains highly relevant.

5456

Hamid Taieb

5457

 0000-0003-3066-0387

5458

University of Hamburg

5459

[hamid.taieb@hu-berlin.de](mailto:hamid.taieb@hu-berlin.de)

## 5460 References

- 5461 HOLENSTEIN, Elmar. 1972. *Phänomenologie der Assoziation. Zu Struktur und Funktion*  
 5462 *eines Grundprinzips der passiven Genesis bei E. Husserl*. Phaenomenologica n. 44.  
 5463 Den Haag: Martinus Nijhoff Publishers.
- 5464 LOHMAR, Dieter and BRUDZINSKA, Jagna, eds. 2012. *Founding Psychoanalysis Phe-*  
 5465 *nomenologically: Phenomenological Theory of Subjectivity and the Psychoanalytic*  
 5466 *Experience*. Phaenomenologica n. 199. Dordrecht: Springer, doi:[10.1007/978-94-](https://doi.org/10.1007/978-94-007-1848-7)  
 5467 [007-1848-7](https://doi.org/10.1007/978-94-007-1848-7).
- 5468 SCARANTINO, Andrea and DE SOUSA, Ronald B. 2018. “Emotion.” in *The Stanford*  
 5469 *Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab,  
 5470 Center for the Study of Language and Information, [https://plato.stanford.edu/arc](https://plato.stanford.edu/archives/win2018/entries/emotion/)  
 5471 [hives/win2018/entries/emotion/](https://plato.stanford.edu/archives/win2018/entries/emotion/).

5472

Published by *Philosophie.ch*

5473

Verein philosophie.ch

5474

Fabrikgässli 1

5475

2502 Biel/Bienne

5476

Switzerland

5477

[dialectica@philosophie.ch](mailto:dialectica@philosophie.ch)

5478

<https://dialectica.philosophie.ch/>

5479

ISSN 1746-8361

5480

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

5481

5482

*Dialectica* is supported by the [Swiss Academy of Humanities and Social Sciences](https://www.snf.ch/).

5483

5484

Abstracting and Indexing Services

5485

The journal is indexed by Arts and Humanities Citation Index, Current Contents, Current Mathematical Publications, Dietrich's Index

5486

Philosophicus, IBZ — Internationale Bibliographie der Geistes- und

5487

Sozialwissenschaftlichen Zeitschriftenliteratur, Internationale Bibliographie der Rezensionen Geistes- und Sozialwissenschaftlicher Literatur, Linguistics

5488

and Language Behavior Abstracts, Mathematical Reviews, MathSciNet,

5489

Periodicals Contents Index, Philosopher's Index, Repertoire Bibliographique de la Philosophie, Russian Academy of Sciences Bibliographies.

5490

5491

5492

# Contents

5493

5494	PHILIPP BLUM, <i>Editorial: Dialectica Goes Open Access</i> . . . . .	1
5495	VINCENT CONITZER, <i>The Personalized A-Theory of Time and Perspective</i> . . .	3
5496	NADINE ELZEIN, <i>Determinism, “Ought” Implies “Can” and Moral Obligation</i>	33
5497	LUKE ROELOFS, <i>Consciousness, Revelation, and Confusion: Are Constitutive</i>	
5498	<i>Panpsychists Hoist by their Own Petard?</i> . . . . .	63
5499	DONNCHADH O’CONNELL, <i>In Defence of Facts: Grounding, Essential Prop-</i>	
5500	<i>erties and the Unity Problem</i> . . . . .	97
5501	HSENG XIONG JOSHUA GOH, <i>Strevens’s Counterexample to Lewis’s “Causa-</i>	
5502	<i>tion as Influence”, and Degrees of Causation</i> . . . . .	127
5503	MARC-KEVIN DAoust, <i>Consistency, Obligations, and Accuracy-Dominance</i>	
5504	<i>Vindications</i> . . . . .	141
5505	FRASER MACBRIDE, <i>Review of Soames (2018)</i> . . . . .	159
5506	MARIO SCHÄRLI, <i>Review of Oppy (2018)</i> . . . . .	165
5507	HAMID TAIEB, <i>Review of Antonelli (2018)</i> . . . . .	173

PROOF