Vol. 17, No. 1, 2021

EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY

UDC 101 | ISSN 1845-8475

Vol. 17, No. 1, 2021

REGULAR ARTICLES

BOOK REVIEWS

EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDC 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
https://doi.org/10.31820/ejap

Publisher, Editorial office

University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia
Phone: +385 51 669 794
E-mail: eujap@ffri.uniri.hr

Web address: https://www.eujap.uniri.hr

Printed by Impress, Opatija (Croatia)
30 copies

TABLE OF CONTENTS

REGULAR ARTICLES

DISCUSSION NOTES

BOOK REVIEWS

EuJAP | Vol. 17 | No. 1 | 2021
UDC: 165.731:159.9.01
https://doi.org/10.31820/ejap.17.1.1

TRUE GRIT AND THE POSITIVITY OF FAITH

Finlay Malcolm[1] and Michael Scott[2]

[1]University of Hertfordshire; [2]University of Manchester

Original scientific article – Received: 02/10/2020 Accepted: 24/01/2021

ABSTRACT

Most contemporary accounts of the nature of faith explicitly defend what we call 'the positivity theory of faith' – the theory that faith must be accompanied by a favourable evaluative belief, or a desire towards the object of faith. This paper examines the different varieties of the positivity theory and the arguments used to support it. Whilst initially plausible, we find that the theory faces numerous problematic counterexamples, and show that weaker versions of the positivity theory are ultimately implausible. We discuss a distinct property of faith that we call 'true grit', such that faith requires one to be resilient toward the evidential, practical, and psychological challenges that it faces. We show how true grit is necessary for faith, and provides a simpler and less problematic explanation of the evidence used to support the positivity theory.

Keywords: *Propositional faith; objectual faith; desire; evaluative belief; positive attitude*

Introduction

Does faith require a positive attitude towards the object of faith? That is, does faith require that one desire or approve of the object of one's faith, or regard it as a good or desirable thing? Accounts of faith that endorse this position, which we will call *positivity theories*, are prevalent in recent literature in the field. A widely canvassed type of argument for positivity theory appeals to examples that appear to show that faith, in contrast with

Correspondence: f.malcolm@herts.ac.uk

belief or acceptance, must be accompanied by a positive attitude. For example,

1. Ava believes in ghosts.
2. Ava believes that Donald Trump will win a second term in 2020.

are attitudes that Ava could have even if she thought that ghosts are malevolent beings, or that Trump winning a second term would be a bad thing. In contrast,

3. Ava has faith in ghosts.
4. Ava has faith that Donald Trump will win a second term in 2020.

seem to require that Ava positively evaluate ghosts or Trump winning. Moreover, expressions of faith directed toward objects that the speaker does not consider favourably such as

5. I have faith in our impending demise.
6. I have faith that Donald Trump will destroy the world.

look like infelicitous or inapt things to say. The positivity theory is usually advanced with some version of these arguments. The theory is sometimes restricted to significant varieties of faith, such as religious or propositional faith, and there are differences in how the positive valency metaphor is cashed out, be it in terms of desires or evaluative beliefs. However, most recent accounts of faith support a version of positivity theory; no contemporary account, to our knowledge, rejects it.

We will review in section one the positivity thesis in its different forms and in section two the arguments put forward in its defence. We agree that faith and a positive evaluation of its content are closely associated but argue, partly on the basis of counterexamples set out in section three, that this is a contingent rather than a necessary relation. While there are some fallback positions available to the theory, which we will explore in section four, the proposed necessary connection between faith and a positive evaluation of its object or content should be rejected. Moreover, in section five, we argue that there are other widely acknowledged properties of faith that provide a simpler explanation for why faith often goes along with a positive attitude. Specifically, a property of faith we call *true grit*: its relationship with a disposition to resist epistemic, practical and psychological considerations to give up on the object or content of faith. True grit, we argue, does justice to both the examples and intuitions that motivate positivity theory without the requirement that faith be accompanied by a positive attitude.

1. What is Positivity Theory?

Positivity theories are commonly focussed on propositional faith, where faith is an attitude with a propositional content; for example, faith that God is good, that Brazil will win the World Cup, or that things will turn out well. According to Robert Audi:

> even if propositional faith is not reducible to a kind of belief, it is reducible to a complex of beliefs and attitudes, for example to some degree of belief that p and a positive attitude toward p's being the case. (2011, 79)

As Audi indicates here, positivity is seen as a way to distinguish faith from cognate propositional attitudes such as belief.[1] Positivity theory is also often advanced for objectual faith, or *faith-in S*, where faith has a non-propositional object such as a person, an institution, or political system. According to Audi:

> There is a further characteristic (already foreshadowed) of both propositional and…[objectual] faith. Both require a positive evaluative attitude toward their object. (2011, 67)

Others who take positive valency as essential for faith include William Alston (1996, 12), Lara Buchak (2014, 53), Daniel Howard-Snyder (2017, 56-57), Walter Kaufman (1958, 113) and John Schellenberg (2005, 133). Less boldly, Daniel McKaughan (2018, 198) says positivity is a characteristic of 'paradigm cases' of faith-that and faith-in. Alvin Plantinga, mainly concerned with Christian faith, claims that someone with Christian faith '(paradigmatically) finds the whole scheme of salvation enormously attractive, delightful, moving, a source of amazed wonderment' (2000, 292). Although 'paradigmatic' is open to interpretation, we take this to be the view that positivity is necessary for some broad but restricted (in some to-be-specified way) class of faith states.

[1] Whether propositional faith requires belief or, more modestly, acceptance, is a matter of contention. On a standard view, to accept a proposition is to use it as if it were true in one's theoretical and practical reasoning (Cohen 1992; Jackson forthcoming); one can choose to accept *p* even if one does not believe it to be true. The accounts of acceptance somewhat differ, however. Even though Alston's (1996) account of acceptance draws from Cohen (1992), he diverges from Cohen by maintaining that acceptance is 'not just on an "as if" basis… To accept [*p*] is to accept [*p*] as true' (18), rather than accepting *p* as if *p* were true. For the purposes of evaluating the positivity theory, whether faith requires belief or acceptance is not crucial: comparable arguments and examples about the differences between faith and belief can be constructed for faith and acceptance. For simplicity, therefore, we will take belief to be the cognitive constituent of faith.

Alston draws attention to two *different* ways in which faith is positive. He notes that merely believing *p* can be considered a ‘positive attitude’ towards *p* (Alston 1996, 12) but the positivity of faith is something different:

> It necessarily involves some pro-attitude toward its object. If S is said to have faith that democracy will eventually be firmly established everywhere, that implies not only that S believes that this will happen but that S looks on this prospect with favor. If S were strongly opposed to universal democracy, it would be somewhere between inapt and false to represent S as having *faith* that democracy will triumph. Whereas one can truly and unproblematically be said to *believe* that democracy will win out even if one views the prospect with horror. (Alston 1996, 12)[2]

Let’s call the positivity of belief *B-positivity*. In what way is B-positivity positive? What Alston has in mind, we take it, is that believing *p* to be true includes, among many other things, the disposition to use *p* in one’s reasoning and to endorse or assert *p* in various circumstances. Belief that *p* is thereby ‘positive’ because the believer is disposed to rely on and agree with it.[3] In contrast, disbelief goes along with the ‘negative’ dispositions to disagree with and reject *p*. However, as Alston makes clear, the kind of positive attitude that he is interested in is not B-positivity. Audi makes a similar point:

> If I have faith that God loves human beings, I have not just a cognitive attitude (the kind that, like belief, may be called true or false), but something more: a certain positive disposition toward the state of affairs being so, i.e. actually obtaining (toward the truth of the proposition, in another terminology). (Audi 2011, 54)

Call this second kind of positivity *F-positivity*. F-positivity and B-positivity are distinct properties. B-positivity towards *p* is not only compatible with a lack of F-positivity towards *p*, but also a negative evaluation of *p*. So, someone who believes that democracy will be universally established is B-positive towards that proposition but may be entirely neutral about that prospect or even, as Alston notes, ‘regard it with

[2] As noted in footnote 1, Alston’s view is that faith requires either belief or acceptance.

[3] In line with fn. 1 above, a similar positive attitude may characterise acceptance; acceptance that *p* similarly involves the disposition to use *p* in one’s reasoning and to endorse or assert *p* under various circumstances and so may similarly be understood to share B-positivity. Indeed, although Alston presents positivity as a characteristic of belief, he endorses an acceptance theory of faith (Alston 1996).

horror' (cf. also Howard-Snyder, 2019, 5). It is F-positivity that is being posited in the positivity theory and that is the focus of these arguments and of this paper.

What, then, is F-positivity? For Alston, it is 'some pro-attitude towards its object'. Here are some other proposals:

> This is an attitude of a kind that at least normally has motivational as well as cognitive elements. The point is (roughly) that faith that something will occur entails taking that to be a good thing. (Audi 2011, 67)

> A positive conative orientation toward the object of faith consists in being for its truth, favoring its being the case, wanting it to be so, giving its truth a positive evaluation, regarding it as good or desirable, and the like. (Howard-Snyder 2017, 48)

> some sort of…positive affective-evaluative attitude toward the person or content that is the object of one's faith…someone who has faith that *God exists* or that *God will be faithful to such and such promises* will care about whether the propositions in question are true, will want them to be the case, or will consider the truth of these propositions or the obtaining of these states of affairs to be good or desirable. (McKaughan 2018, 198)

And according to John Schellenberg

> it seems possible to develop examples of cases where one has faith that *p* without a desire that *p* be true. Accordingly, so as not to be misleading, I suggest that we avoid the notion of a pro-attitude and instead deploy the weaker notion of a *favourable evaluation* of the state of affairs reported by *p* (and, by extension, of the truth of *p*). This *is* entailed by faith that *p*. (2005, 133)

There are significant differences in these accounts of F-positivity. To see this, consider some different options for analysing the belief/desire constituents of F-positivity.[4] Suppose that *R* has faith that *p* (or faith in s). On a belief theory of F-positivity:

[4] This paper will follow these authors in working within the framework of Humean or belief-desire psychology that distinguishes between beliefs and desires as categories of mental state with distinct

> BEL: *R* believes that *p* is good or that it is desirable that *p* be true (or believes that *s* is good or desirable).

And according to the desire theory of F-positivity:

> DES: *R* desires that *p* or approves of *p* (or desires/approves of *s*).

Either one of these conditions could be understood as providing a complete analysis of positivity, as follows:

> X-BEL: Only BEL is true
> X-DES: Only DES is true

There are also two obvious ways of combining them:

> CON: Both BEL and DES are true
>
> DIS: Either BEL or DES are true

Alston is completely clear on where he stands, at least with respect to propositional faith. He supports a pure desire account, i.e. X-DES. Audi proposes that the positivity of faith only requires one to regard the object of faith as a good thing, which he seems to allow could be either an evaluative belief or a desire. Schellenberg clearly rejects X-DES but his preferred notion of 'favourable evaluation' appears to encompass *either* desire or belief. So, we take both authors to support DIS. In other work, Howard-Snyder (2013, 367) adopts a varied stance towards positivity, and so is likely also a supporter of DIS. Similarly, McKaughan also gives space to positive evaluations that could be interpreted as either desires or beliefs about *p* (i.e. as approving of *p* or believing that *p* is a good thing).

There is a reason for thinking, contrary to X-DES and CON, that F-positivity need not be a desire-like attitude. Adapting an example from John Schellenberg, imagine that Paul is a supporter of a political party and places his faith in its leadership. Following a leadership contest, not only does Paul's preferred candidate fail to win, the successful candidate is someone that Paul finds both personally repellent and morally

dispositional profiles. This is sometimes put in terms of direction of fit. Desires (and desire-like states such as wishes, hopes, plans and so on) have a world-to-mind direction of fit: the agent desires to bring the world into accordance with the content of the desire. Beliefs have a mind-to-world direction of fit: the content of the belief should fit with the way that the world is. Desires, unlike beliefs, are taken to motivate the agent to bring about action.

reprehensible. Despite his misgivings and resentment of the candidate's success, Paul recognises that the new leader is the best prospect for achieving the aspirations of the political party. Accordingly, Paul maintains his faith in the leadership, and loyally commits to campaign for it. As Schellenberg points out, faith can be positive by virtue of recognising that the object of faith is desirable without any favourable feelings towards it: 'something may intellectually be seen as desirable – as *worthy* of desire – without actually being desired, when relevant psychological obstacles are present' (2005, 133).

For these reasons, we take the positivity theorists to be committed to (at least) the more modest DIS. The availability of plausible counterexamples to X-DES and CON make DIS the more plausible position.[5]
To sum up, positivity theorists support a theory on the following lines:

> *Positivity Theory* (PT). Necessarily, if R has faith that *p* (or in *s*) then R desires that or approves of *p* (or desires or approves of *s*), or believes that *p* (or *s*) is good or desirable.

Additionally, some restrict PT to religious faith or to paradigm cases of faith.

Before proceeding, we need one further distinction. It is very widely held that faith motivates the agent with faith (e.g. Bishop 2007, 117; Howard-Snyder 2017, 56-57; Schellenberg 2005, 127-66; Swinburne 2001, 211). This theory, which we will call *faith internalism*,[6] in its simplest form says that

> Necessarily, if R has faith that *p* or faith in *s*, then R is (to some extent) motivated to act on that faith.

Now, faith internalism could dovetail with PT in the following way. Suppose that the motivation to act is explained by the presence of a desire-like, or evaluative state, in line with Humean psychology (see footnote 4). It follows from faith internalism that faith must be accompanied by a desire-like attitude. This affords a neat way of bringing together positivity theory and faith internalism: the desire towards or approval of the object of faith posited by PT could also motivate the agent. Audi (2011, 67, cf. also Howard-Snyder 2019, 3) appears to suggest this connection. Faith internalism and positivity theory are clearly distinct theories. However,

[5] Note that if the F-positivity of propositional faith is cashed out as an evaluative belief, it will involve *two* positive cognitive attitudes: belief in the propositional content (which is B-positive) and belief that that the content is good (which is F-positive).

[6] For an overview of a comparable current debate in metaethics see Björnsson et. al. (2015).

since the connections between them play a role in later discussion it is useful to make clear at this stage that these theories are independent. There are three main reasons for this.

First, PT can be satisfied by faith being accompanied by evaluative beliefs (i.e. BEL) about the object of faith rather than desires. So, positivity theory (assuming, again Humean psychology) is compatible with faith being motivationally inert. Second, the desire-like state that motivates the agent with faith does not have to be about the object of faith, as PT requires. Suppose that Jane has faith that Brazil will win the next World Cup. She enthusiastically supports the team but she is motivated by a desire to please her father (who is a big supporter of the Brazilian team) rather than a desire that the team wins. She may not be aware that this is the desire that motivates her. Her psychological state satisfies faith internalism – she supports the team – but not PT because her motivating desire is not directed towards the content of faith. Third, faith internalism does not require that the desire-like state that motivates the agent with faith is positive. Suppose, to take a minor variation on our example, that Jane is motivated by a fear of her father's displeasure (and he would be displeased if Brazil lost). Again, she may not be aware that this is the desire that is motivating her to support the team. Unlike PT, faith internalism is not picky on the kind of attitude that motivates the agent to act on her faith. The attitude does not have to be positive evaluation or approval – it could be fear, selfishness, vanity, etc. – provided that it disposes the agent to act on that faith.

It can be seen, therefore, that while an agent whose faith involves a positive and motivating evaluation of the object of faith will satisfy both PT and faith internalism, these two theories are independent.

2. Arguments for the Positivity of Faith

A useful initial classification among the arguments advanced for positivity theory is between those that exploit (a) examples of the kinds of attitudes that are appropriately regarded as faith, and (b) examples that contrast faith with related attitudes.

Examples of (a) are found in the writings of Walter Kaufmann, one of the first philosophers to draw attention to positivity: 'One can say: "I have faith I shall recover." One cannot say, without doing violence to language: "I have faith that I have cancer."' (1958, 113). Lara Buchak takes a similar approach. According to Buchak,

> in order for a proposition to be a potential object of faith [...] the individual must have a positive attitude towards the truth of the proposition. This can be seen by noting that while I can be said to have or lack faith that you will quit smoking, I can't appropriately be said to have or lack faith that you will continue smoking. (2014, 53)

Positivity theory is taken to be supported by (a) because genuine faith appears to go along with a positive evaluation of its object. Examples of (b) are particularly prominent in discussion of propositional faith and point up differences between faith that *p* and other propositional attitudes, in particular belief. Alston's example of the difference between faith and belief that democracy will triumph is a case in point. Faith that *p*, it seems, must have an extra 'positive' property not necessary for mere belief that *p*. As is clear from the quotations above, the arguments for positivity theory employ two different types of evidence. Some arguments (c) use examples of faith and related attitudes to bolster intuitions about the kind of thing that faith is, while others (d) appeal to considerations about linguistic felicity. Buchak and Kaufmann, for instance, emphasise the oddity of *saying* that someone has faith if they don't also have a positive attitude toward the object. Alston appeals to either (c) or (d) considerations: it is 'somewhere between inapt and false' to say that S has faith that democracy will triumph if S does not see that prospect favourably. The strategy of (d), we take it, is that if there is something linguistically amiss with representing someone as having faith without an associated positive attitude, that supports the conclusion that the positivity is built into our concept of faith.[7]

We think that (c) is a more compelling strategy than (d). First, as Malcolm and Scott (2017) point out, it is questionable not only whether judgements made by hearers about linguistic felicity offer reliable evidence for a philosophical theory about the nature of faith, but also whether hearers are expressing linguistic intuitions rather than theoretical presuppositions.[8] Second, and more directly, the assumption that the proposed statements about faith are infelicitous seems to us unpersuasive. Take Buchak's and Kafumann's claims that

7. I have faith that you will continue smoking

[7] The notion of linguistic felicity is not fully spelled out by proponents of these arguments. It appears to be determined by the evaluation, by competent speakers of a language, that a given sentence of that language is ill-formed or does not make sense.

[8] For a review of the many challenges in unpicking facts about meaning from the judgments of speakers about linguistic felicity see Novek (2018). An empirically informed investigation into talk of faith, of the kind conducted in experimental pragmatics, is an intriguing but as yet unexplored prospect.

8. I have faith that I have cancer

are linguistically infelicitous. We agree that these are unusual things to say (indeed, sufficiently unusual that a hearer might reasonably ask ‘don’t you mean *believe* rather than *have faith*?’) but that is because what they are saying is so unusual rather than because there is something wrong with the utterances. Even by the positivity theorist’s own lights these utterances could be true, provided that the speaker has a positive attitude towards the hearer’s continuing to smoke or the speaker’s having cancer. It might be morally or prudentially inappropriate to assert (7) or (8) but neither are linguistically infelicitous. Alston proposes that

9. S has faith that universal democracy will triumph but is strongly opposed to it

is inapt. But this does not seem to involve any linguistic mistake even if (assuming Alston is right and faith necessarily involves a pro-attitude) S cannot have the combination of attitudes described by (9). Alston would presumably wish to maintain that we can understand (9) to argue that it is saying something untrue. Indeed, this may be his point: (9) is inapt not in the sense that its meaning is unclear or that it deploys a misuse of language but that it is obviously untrue. For these reasons, we take (c) to offer the most promising way of arguing for PT.[9]

3. Faith without Positivity

Having considered what positivity is and the arguments for the positivity theory, is PT true? We believe that faith often goes along with a positive attitude but that the connection is contingent rather than necessary. That is, we support the more modest theory:

> (PT*): Faith is usually but contingently accompanied by positive attitudes towards its object or content.

We will flesh out the details of this contingent relationship later in the paper; our focus here is on whether the necessary connection between faith and positive attitudes is defensible. In this section we will set out several counterexamples to PT; in the following we will look at two ways of revising PT to accommodate these counterexamples. Space considerations

[9] The case for PT is sometimes advanced alongside one for faith internalism (Audi 2011, 67; Howard-Snyder 2019, 3). However, as we have seen, these are independent theories; we will return to the connections between them in section four.

limit the range and number of counterexamples we can give, and we have found in discussion that interlocutors vary in how intuitively appealing they find the given putative examples of faith. Our aim, therefore, is to raise doubts for the reader about the simplicity or necessity of a relation between faith and positivity and to motivate consideration of an alternative account.

A commonplace observation about faith is that it is frequently accompanied by misgivings, either about the object or content of faith. Faith can be difficult to maintain and is often talked about as something that individuals struggle with. This aspect of faith has been a focus of discussion in recent papers in the field (not least by some of the supporters of positivity theory), with most attention being given to how faith is maintained in the face of doubt (Pojman 1986; Schellenberg 2005; Howard-Snyder 2013; McKaughan 2013, 2018). We find similar considerations raised in historical and theological treatments of religious faith. According to one recent survey of religious faith, "[d]oubt emerged as inevitable, as concomitant to faith, occasionally a virtue, more often as a struggle, an ailment to be overcome" (Andrews 2016, 2). However, the kinds of misgivings that go along with faith clearly extend beyond doubts about the truth of one's faith.[10] Religious faith may be clouded by despair, torment, anger, feelings of abandonment, sadness and dark nights of the soul. As McKaughan (2018) has demonstrated, such feelings were widely felt and documented by Mother Theresa. For instance, in her personal diaries from around 1961 she writes,

> Since [19]49 or [19]50 this terrible sense of loss—this untold darkness—this loneliness this continual longing for God—which gives me that pain deep down in my heart—Darkness is such that I really do not see—neither with my mind nor with my reason—the place of God in my soul is blank—There is no God in me—when the pain of longing is so great—I just long & long for God—and then it is that I feel—He does not want me—He is not there. (Kolodiejchuk 2007, 349)

More generally, it seems, during a crisis of faith, that negative feelings about the content or object of faith can come to the fore while positive feelings and judgments can go into abeyance, even if only for brief periods. The insistence on positivity as a *necessary* condition for faith seems at odds with this view about crises of faith. Moreover, crises of faith are not only restricted to religious cases. For example, one can continue to have faith in a person who engages in frustrating and self-destructive behaviour, even

[10] For some recent empirical data see Dura-Vila (2016).

though this behaviour may cause one to doubt the merits of one's faith and to feel anger towards and disappointment in that person.

The problem presented for PT by crises of faith is straightforward: faith in crisis can become detached, if only briefly, from positive evaluations or beliefs about the object or content of that faith. We do not, however, regard these cases as losses of faith. Indeed, faith is often seen as helping one to get through such crises. This does not, of course, establish that there is no connection between faith and positive evaluations. It does show, however, that PT as it stands is untenable. Faith is not indefatigably positive: it may endure even when positive attitudes about the object or content of faith are in abeyance.

If objectual and propositional faith is possible without a positive attitude when faith is in crisis, can they also come apart in less challenging circumstances? Consider the following example.

> [A] Ellis is travelling to a conference in Shanghai where he is due to deliver a presentation. Ellis does not know the country and his flight schedule leaves him little time to get from the airport to the conference venue. But his old friend Thomas, who works in China, has agreed to pick him up at the airport to drive him to the event. Ellis, who regards Thomas not only a good friend but also a conscientious person, has faith in Thomas to be there to collect him on time and get him to the conference (as well as propositional faith that Thomas will do these things). Over the course of the flight, however, Ellis spots a major problem with the presentation, one that he cannot clear up in the time he has. If only, Ellis thinks, Thomas could slack off on this occasion and be a little late and I could miss the presentation slot and save the embarrassment of a poor presentation. He retains his faith with respect to Thomas collecting him and getting him to the venue on time but he neither believes these would be good things, nor does he desire them.

If the positivity theory is right, Ellis should have lost his faith in Thomas over the course of the flight. But this does not seem right. In key respects Ellis' attitudes and dispositions are unchanged. He has not undergone loss of confidence in Thomas: he stills expects Thomas to be there on time, he has not made any alternative plans so still relies on Thomas to be there on time. It is simply that, with respect to some things that Ellis has faith in Thomas to do, he has changed his evaluation of their merit: he doesn't positively evaluate Thomas' timeliness in this context. Ellis' hope that

Thomas be late is perhaps unfair to Thomas since Ellis' predicament is his own fault, but it seems possible. Ellis might even say things like: 'While I have full faith in Thomas to pick me up from the airport and get me to the conference on time, I hope he doesn't'. We would be puzzled if Ellis said: 'Since looking again at my presentation, I've totally lost faith in Thomas getting me to the venue on time'. More generally, faith can persevere even though one's positive attitudes about its content or object do not. Contrary to PT, it seems possible to have faith in s x-ing (or faith that s *x*s) while lacking a positive attitude toward the object or content.

Consider two further, connected examples:

> [B] Silvia has faith that the Biblical miracle stories are true. However, she has always been troubled by the story of the Miracle at Cana. She does not understand why Jesus would have transformed water into wine; it seems to her a pointless exercise. Moreover, she disapproves of drunkenness and the encouragement thereof. She retains her faith that Jesus transformed water into wine at Cana, along with her faith in the other miracle stories, despite neither believing that it was a good thing to do nor approving of it.
>
> [C] Ryan has faith that the teachings of his church are based on the word of God. However, he finds some of these teachings a struggle, in particular those related to the sinfulness of homosexuality. This is because Ryan is coming to terms with the fact that he is gay. Ryan has faith that the church's teachings on homosexuality are true but he does not look on them with any favour; he certainly does not desire them to be true, nor, given his own experiences, does he understand how God could will them. Nevertheless, his faith holds.

Silvia, insofar as she has a view of the content of her faith, considers the miracle ethically dubious. Ryan struggles with his faith and is unable to look positively on some aspects of it. Both are examples of faith without DIS. The cases are connected because they trade on the fact that faith is often directed towards a body of propositions to which agents are committed, rather than just one. Religious faith may encompass numerous propositions, sometimes codified in creeds, commitment to which are considered important to membership in a religious tradition. Political systems, particularly those associated with revolutionary movements, provide another example. In such cases, the agent may not view all of the requisite propositions with the same favourable attitude; some may be seen either neutrally but taken on trust as among the requisite commitments of the political position. Faith that *p*, that *q*, that *r*, etc., therefore seems possible without a positive attitude towards all of the propositions in question.

Consider one more example:

> [D] Martha has faith that those people that God does not save will go to Hell and that such decisions are predestined. She does not claim to understand how this arrangement can be just or good; her faith is such that she eschews questions about its merits or thinking through its ethical implications. Nor does she desire it to be true; indeed, her feelings towards this are closer to dread, not least in case she should be among those who are not saved.

Martha's faith is not in crisis; she has not had her positive attitudes challenged or upset. Indeed, Martha might never have seen predestination in a positive light but as an incomprehensible mystery, or as an inescapable fact of religious reality about which she makes no evaluative judgement. Her faith is manifested by her resolute conviction that this is part of a religious reality, and that this conviction manifests in her life and thinking, rather than her approving of it. Moreover, her lack of a positive stance might be deliberate: she intends to refrain from forming an evaluative appraisal of predestination because she thinks it at best a pointless or at worst inappropriately presumptive attitude on her part.

In some of these cases – notably Silvia and Ryan – it seems possible that their attitudes towards the object or content of their faith is conflicted. For example, perhaps Silvia has a negative attitude to

10. Jesus transformed water into wine.

while *also* having other positive attitudes towards the proposition. According to PT, faith that *p* requires desire or approval of *p*, or a belief that *p* is desirable or good; PT does not say that faith that *p* is incompatible with non-positive attitudes towards *p*. So, the examples can be brought into line with PT by assuming that positive attitudes are also in play.

We agree that individuals can be conflicted in this way, most clearly in the case of desires. Silvia might approve of Jesus' miraculous intervention at Cana while also disapproving of what he did. We also agree that if this is how things are with Silvia, then PT is consistent with the example. However, PT claims a necessary connection between faith and positive attitudes about its object or content. So, for the conflict defence to work, Silvia – along with individuals in any number of similar cases – *must* be conflicted. This, it seems to us, is less plausible than PT* without independent argument. Silvia may be conflicted, and this may be the most probable explanation of her attitude, but it does not seem necessary that she approve of (10) alongside her misgivings about it.

PT therefore runs into a number of problem cases. It seems that faith need not be accompanied by positive attitudes either when undergoing a crisis of faith or in much more mundane circumstances (such as [A]); positive attitudes need not extend to all contents or objects of one's faith (as in [B] and [C]); faith also seems possible in cases where a positive attitude may not even be seen by the agent as appropriate (as in [D]).

4. Modest Positivity Theory

We have indicated our preference for PT* which posits a contingent connection between faith and positive attitudes. But is there a more modest version of PT that is compatible with the counterexamples but retains the necessary link between faith and positivity? There seem to us two main options.

First, positivity theorists could attempt to find a more attenuated necessary connection between faith and positive attitudes towards its content or object. The onus is on the positivity theorist to flesh out the details of this connection. But since nobody has yet attempted this modification, and with a view to being constructive, here is a proposal of refined positivity theory:

> (RPT) Necessarily, if R has faith that *p* (or in *s*) then R desires that *p* (or approves of *s*) or believes that *p* is good (or that *s* is good), or some relevantly connected faith judgement is accompanied by a desire or approval of its content or object or belief that its content or object is good.

This formulation is modelled on attempts to refine motivational internalism in metaethics, which are perhaps an object lesson in the difficulties of finding plausible attenuated necessary connections (see Björklund et al. 2012). The central idea behind RPT is that while there may be cases of individuals with faith that *p* (or in *s*) without a corresponding positive attitude towards *p* (or *s*), they must have a positive attitude to some proposition or object closely related to *p* (or *s*). For example, Ellis may not think positively about Thomas' getting him to the venue on time, but he presumably does regard Thomas' timeliness or at least more generally

Thomas' organisational abilities favourably. Is RPT, or something like it, defensible?

An initial problem for RPT is to specify what makes faith relevantly connected to a positive attitude. Suppose that Ellis has a high regard for Thomas' abilities as a cook and eagerly anticipates the culinary feast that Thomas will lay on for him. Ellis therefore appears to have a positive attitude that is connected to his faith in Thomas' timeliness, since both are concerned with his prospective meeting with Thomas. However, this is presumably not a *relevant* connection since his feelings about Thomas' cooking should not have a bearing on whether he has faith in Thomas' timeliness. Supporters of RPT will need to specify more closely the 'relevant connection' to avoid these problems. Second, RPT comes at a significant cost to plausibility. Once the positivity theorist has conceded that the arguments for PT considered in section one are unsuccessful, and that there are counterexamples to these theories, why continue to maintain that there is a *distant* necessary connection rather than conceding that there is no necessary connection at all? Third, there is an alternative to RPT compatible with the counterexamples that preserves the intuition that faith and positive attitudes are closely related: that there is a regular but contingent connection between faith and a positive attitude towards its content or object, i.e. PT*. This would account for the fact that we expect faith to be positive and that it usually is, while allowing that under certain circumstances it is not. Notably, the exploration of causal links between faith and positive attitudes has been the focus of a growing body of empirical investigation (Ögtem-Young 2018, Pargament 2010, and Pargament and Cummings 2010). We will say more about philosophical accounts of faith that make this relationship plausible in the following section.

It is useful to consider a specific way of developing an answer to the first objection, i.e. specifying the relevant connection needed for RPT. An appealing way of doing this is to posit a relationship between propositional and objectual faith.[11] Take the example of Ellis. While he may not have a positive attitude towards the proposition that Thomas will get him to the venue on time, it is plausible that his propositional faith is based on faith *in* Thomas (or in Thomas' reliability), and that he thinks favourably of Thomas (or of Thomas' reliability). Similarly, while Martha may not have a positive attitude toward the proposition that the afterlife is predestined, she may have faith in the authority that inspired her propositional faith (the church, the Bible, God, etc.) and have a favourable attitude towards it. This suggests the following approach to defending refined positivity theory: in

[11] We would like to thank an anonymous referee for this journal for suggesting this example.

cases where agents with propositional faith appear to lack a positive attitude towards the proposition in question, the propositional faith is based on an objectual faith with a content that is viewed positively.

The proposed ways of expanding the examples of Ellis and Martha to include objectual faith are plausible: their propositional faith may be inspired by objectual faith and, moreover, they may have a positive attitude associated with their objectual faith that they lack towards the proposition. This is, however, compatible with our preferred theory PT*, i.e. that faith is usually but only contingently associated with positive attitudes. A defence of PT or RPT will need to posit a necessary connection between propositional and objectual faith, that is, (non-positive) propositional faith is not merely causally related to (positive) objectual faith, but the former *requires* the latter. This is certainly not self-evident. If we look more broadly at the literature on faith, there is little, if any support for the view that there is a necessary relation between objectual and propositional faith. One potentially sympathetic voice in favour of the dependency of propositional faith on objectual faith is William Alston: ‘It seems plausible that wherever it is clearly appropriate to attribute “faith that,” there is a “faith in” in the background’ (1996, 13). However, Alston appears to take the connection between propositional and objectual faith to be causal rather than necessary. Moreover, he offers this remark as an intuition about propositional faith rather than a substantive theory that he aims to defend. Making progress with the proposed defence of positivity theory, therefore, will require new arguments for dependency relations between propositional and objectual faith, which are yet to be forthcoming.

There is another fallback position available to positivity theory. This is to revise the account of the positive valence of faith whereby the positivity of faith is effectively guaranteed by faith internalism. As we saw in section two, if faith internalism (and Humean psychology) is true, then if R has faith that *p* (or in *s*), R will have some desire-like state that will dispose her to in some way act on that faith. Now, if faith that *p* (or in *s*) has this effect on R’s plans and objectives, *p* and *s* make a difference to R. They make a difference for R because if R did not have faith that *p* or in *s* (or had faith in some other proposition or object) R would be differently motivated. To this extent, *p* or *s* matter to R. Accordingly faith internalism goes along with the following theory:

> IPT (Internalist positivity theory): Necessarily, if R has faith that *p* or in *s* then *p* or *s* matter to R.

This, of course, is only tenuously a ‘positivity’ theory: it falls far short of the requirements of PT. The object or content of faith matters to R only in the sense that they make a difference to what she is motivated to do. As we saw in section two, this requires neither that R's motivating attitude is about the content or object of faith, nor that the attitude that motivates R is positive in the sense given by PT. Nevertheless, IPT does preserve a necessary connection between faith and some notion of positivity.

To see the potential appeal of IPT, it is useful to consider an argument from Howard-Snyder about the connection between faith and what we care about:

> one cannot have faith that something is so without at least some tendency to feel disappointment upon learning that it's not so. That's because one can have faith that something is so only if one cares that it is so; and one can care that something is so only if one has some tendency to feel disappointment upon learning that it's not so. (Howard-Snyder 2013, 360)

Now, the connection between caring and a disposition to feel disappointment, offered here as a priori, seems to us misplaced. In some of the examples we considered in section three (notably Ellis and at least some examples of those with crises of faith) the connection looks doubtful. Here are two more examples. Suppose my son enters the sack jumping race at a local fete. I have faith that he is going to win. However, I learn that the prize for second place – a chocolate dinosaur – would please him far more than the Snakes & Ladders game reserved for the winner (a copy of which he already has). He comes in second. Am I disappointed? Not even a little bit. This isn't because I didn't care that he would win while he was in the race or lacked faith that he would win. Rather, when he came in second, I ceased to have those attitudes and felt delighted (and maybe a little relieved) that he secured the prize he would prefer. Second, suppose I support a minor English football team at the low end of the National League. I recognise their many weaknesses but nevertheless have faith that they will manage to stay in the league and avoid relegation. It turns out that this does not happen. Instead, through a serious of extraordinary victories they secure a place in the higher English Football League. Am I disposed to be disappointed that they didn't stay in the same league? Clearly not. I am delighted they did even better than I had faith that they would.

Our point in drawing attention to Howard-Snyder's argument, however, is not for the connection he makes between faith and disappointment but the one between faith and caring. This connection is intuitively plausible but can be secured by IPT rather than PT: agents with faith care about the content or object of their faith because it matters to them. It matters because the content or object of their faith makes a difference to their motivations and what they plan to do. IPT, therefore, preserves an intuitive connection between faith and what the agent cares about.

IPT offers a viable fallback position by conceding the problematic aspects of PT, i.e., that the agent with faith requires a positive attitude towards the object or content of faith. As such, IPT escapes the counterexamples considered in section three that target this feature of PT. However, the 'positivity' of faith proposed by IPT derives from its impact on the motivational profile of agents that have faith. As such, IPT glosses faith internalism. Even if IPT is true, therefore, PT* is still needed to account for the connection between faith and one's approvals or positive evaluative judgements of the object or content of faith.

5. Faith and True Grit

We have argued that PT is false and instead endorsed (a) PT*, which posits a contingent connection between faith and positive attitudes, and (b) IPT, which connects faith and what matters to the agent with faith, in a way that follows from faith internalism. In this concluding section we will focus on some widely recognised, necessary properties of faith – that we will call *true grit* – that explain why faith should usually go along with positive attitudes, that is, why we would expect PT* to be true. We will also argue that the examples considered in section two that purportedly lent support to PT can be explained by true grit rather than necessarily being accompanied by a positive attitude.[12]

As a starting point, it seems platitudinous that faith is not fickle. Someone who has faith that *p* or faith in *s* does not give up on *s* or reject *p* on the least reason to do so. Even if it is not acted on, a disinclination to give up on the object of faith seems one of the minimal necessary requirements for either objectual or propositional faith. Unsurprisingly, this idea shows up in most theories of faith, albeit under various guises: Bishop (2007), for instance, talks of 'commitment', Buchak (2017) 'steadfastness', Howard-Snyder (2013) 'resilience', Kvanvig (2013) 'retention', Malcolm and Scott (2017) 'resistance', Matheson (2018) 'grit', and McKaughan (2018) 'perseverance'. One cannot have faith without in some way and to some extent sticking with the object of faith. Can we specify this disposition in a less metaphorical way, while preserving its platitudinousness?

[12] To the best of our knowledge, the expression *grit*, and the psychological literature associated with it (Duckworth et al. 2007), was first connected with the resilience of faith in a workshop presentation by Malcolm (2017). The first published work to make the connection was Matheson (2018).

It is useful to consider a comparable literature in the social sciences where the notions of grit and, in particular, resilience have been explored (for recent overviews, see Bourbeau 2018; Jacelon 1997; Luthar et al. 2000; and Stewart and Yuen 2011), as well as where the connections between faith and resilience have been to the fore (see Pargament 2010; Pargament and Cummings 2010; Ögtem-Young 2018). This literature has admitted several characterisations of resilience, four of which look particularly relevant: the ability of an agent to 'bounce back' from a setback (Block and Thomas 1955), to 'adapt' to challenging circumstances by changing various attitudes and behaviours (Zautra et al. 2010), to 'persist' or exhibit 'staying power' (Masten et al. 1990), and to 'resist' the challenges presented by adverse circumstances (Rutter 2006). How might these be applied to psychological attitudes such as faith? Suppose an agent S has some attitude A under circumstances C that challenge or provide a reason for her to not have that attitude. The four characterisations of resilience suggest four corresponding ways in which S might be resilient with respect to A in C:

i. *Bouncing Back*: S is disposed to regain A after its loss as a result of C.
ii. *Adaptation*: S is disposed to modify her thinking and other attitudes to retain A in response to C that would otherwise cause her not to have A.
iii. *Persistence*: S is disposed to persist in exhibiting A in C.
iv. *Resistance*: S is disposed to resist, at least to some extent, factors that would lead her to cease having A.

Faith could exhibit resilience in any of these ways. Suppose, for instance, I have faith in a friend's honesty, but I am presented with compelling evidence that he has acted dishonestly. Even if I lose my faith for a while, I may be (i) disposed to regain it later, or I may be (ii) inclined to change other attitudes – such as my views about the credibility of the source of the challenging evidence – to preserve my faith, or I may be (iii) disposed to continue to voice my conviction that my friend is honest in the face of this contrary evidence, or I may be simply (iv) disposed to be unpersuaded by that evidence (at least up to a point).

There is much more to say about the merits of these analyses of resilience but since our focus is on faith, it is (iv) – the resistance analysis – that seems to us the most promising. There are two reasons for this. First, (iv) is the least demanding of the four analyses: anybody who rebounds, adapts or persists in their attitudes in C thereby satisfies (iv) by resisting factors that would undermine A. Indeed, (i), (ii) and (iii) can each be seen as ways of resisting C: by adaptation, resistance or rebounding. Second, (i), (ii) and

(iii) each face counterexamples. Someone might lose their faith without any inclination to regain it: 'I used to have faith in democracy as the best political system but after the political turbulence of last couple of years I've given up on the idea'. Similarly, while (ii) and (iii) – adaptation and persistence – may often characterise faith, neither one seems necessary. I may maintain my faith in a sports team but substantially change the manner in which I voice and act on that faith after my team undergoes a crushing defeat. I thereby adapt but do not persist in my earlier behaviours. In contrast, my faith may exhibit a degree of persistence but not be adaptable. I may be disposed to continue enthusiastically to support my team after various defeats, but not disposed to adapt my behaviour and attitudes to preserve my faith in the event of a major loss. So, in general, (iv) is successful because it does not restrict the ways in which S may be disposed to resist C with respect to her faith. The concept of resistance – discussed in the social sciences but hitherto not explicitly considered in philosophy – provides us with a helpfully minimal analysis of the kind of resilience that faith is widely taken to exhibit.

Can we say more about the challenging circumstances C? Discussion has tended to focus in particular on counterevidence to the truth of the propositional content of faith (Howard-Snyder 2013, 367-68; Buchak 2017; Matheson 2018; McKaughan 2018), less so on non-epistemic factors (though see examples from Howard-Snyder 2017 for non-epistemic examples of resilience).[13] But faith goes along with resisting practical and psychological challenges. Consider, for instance, the demands of having faith in a society in which public expressions of faith are liable to be met with persecution and mistreatment. Sustaining faith in such a context incurs a practical cost: it is difficult to do and carries with it significant risks. This is, of course, a somewhat extreme case; faith does not have to be so resilient that it persists even under these circumstances. However, faith must be able to withstand *some* practical costs: one does not have faith in someone if one is disposed to defame them in exchange for an Oreo Bar. Additionally, one can have faith in someone who behaves in an exasperating and emotionally wearing manner. Again, faith need not require a heroic degree of determination and steeliness. But it does need to exhibit some degree of resistance to psychological pressures. One does not have faith that democracy is good if one is disposed to change one's mind about it because of the tiresome and provocative behaviour of one democratically elected leader. In general, faith disposes the faithful agent

[13] In the social psychology literature on 'grit', Duckworth et al. (2007) point to non-epistemic factors, but when grit is addressed in recent analytic philosophy, the focus is clearly on resisting epistemic reasons (Morton and Paul 2019).

to withstand practical, emotional and psychological costs as well as contrary evidence.[14]

Since we are looking for a minimal, widely acceptable, and necessary property of faith, we will remain pluralists about the specification of the attitude of faith itself. For example, on a doxastic theory of propositional faith, faith disposes the agent to resist pressures (evidential, practical or psychological) to disbelieve the propositional content of faith; on a nondoxastic theory (e.g. Alston 1996) the attitude in question may be acceptance. On a trust theory of objectual faith (e.g. McKaughan 2016), the attitude will be a disposition to resist evidential, practical or psychological factors to break one's trust with the object of faith; on the theory that objectual faith is a goal-directed attitude (e.g. Kvanvig 2013), the agent will resist evidential, practical and emotional pressures to give up on that objective. We will remain neutral on these contentious areas of debate.

Drawing the elements of this theory together, we propose that it is necessary that faith disposes the agent to resist, to some extent, giving up on the object or content of faith (be it trust of or allegiance to the object, or a belief or support of the proposition, etc.) in response to epistemic, psychological or practical pressures to do so. For convenience we will call this property of faith *true grit.*[15] Faith is *true* in the sense that it exhibits an allegiance or attachment to the object or content of faith (which may be characterised differently as belief, acceptance, trust, commitment, etc.); it is *gritty* in the sense that it is a disposition to resist (to some extent) challenging circumstances that would undermine that allegiance or attachment.

True grit is a distinct property from the positivity of faith. Someone with true grit is undeterred in their commitments by evidential, practical and psychological factors but they are not thereby invariably positive in their attitudes about the object or content of their faith. One may exhibit true grit without approving of or having a positive evaluative belief about the object of faith. On the other hand, a positive attitude toward *p* (or about *s*) and true grit commonly go along with each other, as PT* predicts. The most

[14] Although it is not central to our argument, the notion of contrary evidence needs more careful handling than it is sometimes given. If I have access to incontrovertible evidence that *p* is true I will be unmoved by counterevidence to this belief. However, we usually take faith to be characterised by resistance to counterevidence that is not counterbalanced by the evidential resources at the agent's disposal. (For more discussion of the idea that faith 'goes beyond the evidence', see Buchak 2012; Malcolm 2020).

[15] We use this well-worn expression not with the aim of making a connection with established theories of grit in the social science, or existing accounts of faith that focus on grit as a salient property, but simply as a familiar expression to cover the minimal properties of faith that we are positing.

straightforward way in which someone might have true grit towards an object or proposition is to desire or positively evaluate that object or the truth of that proposition, where these desires and beliefs have a causal role in sustaining one's resistance to circumstances C. For example, the resilience needed for faith that God will save us will be strengthened by the desire that God will save us or the judgement that this is a valuable thing. Positive attitudes towards the object or content of faith bolster one's resistance reasons to give up on that object or disbelieve that content. The true grit of faith, therefore, fits well with PT*.

Can true grit also explain the examples used to support PT? Let us take three. First, why do we find cases of faith like (11) but not like (12)?

11. Peter has faith that Franz will give up smoking.
12. Peter has faith that Franz will continue smoking.

Buchak proposes that the absence of a positive attitude towards Franz continuing to smoke explains the difference. But this isn't convincing. Suppose that Peter wishes Franz ill and believes that Franz's death would be a good thing; suppose he also believes that Franz's continuing to smoke raises the chances of this happening. Even with the requisite positive desires and beliefs in place, that Franz will continue to smoke still looks like an odd thing for Peter to have faith about. True grit does better: it is the peculiarity of someone having an attitude of true grit towards Franz continuing to smoke that accounts for why (12) seems an odd candidate for faith. The circumstances in which (12) might be true are ones in which Peter persists, for example, in maintaining that Franz will continue to smoke despite evidence that he has given up. That is, where Peter exhibits the true grit he needs for faith.

Why do we find instances of faith like (13) but not (14)?

13. I have faith that I will give up smoking.
14. I have faith that I will continue smoking.

Not, it seems, because of anything to do with positivity. There is nothing unusual about a desire to smoke nor, unfortunately, about a desire to continue to smoke. So, the presence or absence of a positive attitude does not explain why (14) is an odd case. True grit does. To commit to give up smoking, for many, requires resolve in the face of a variety of pressures: putting aside the evidence of past failures to stop, determination to give up despite the nagging need to smoke; practical avoidance of circumstances in which one will be tempted to change one's mind. In contrast (14), except under unusual circumstances, is not a suitable subject of true grit. Indeed, it is the reverse: typically, someone doesn't need any resilience to believe that they will continue to smoke since they just need to *give in* to it.

Why do we find instances like (15) but not (16)?

15. Peter has faith that he will survive cancer.
16. Peter has faith that he will die of cancer.

According to the positivity theory, it is because the latter, unlike the former, is not something about which an agent usually has a positive attitude. Equally, however, the proposition that one will die of cancer is not usually something that people take a gritty attitude towards. For example, we do not usually find someone determined to uphold the judgement that they have cancer in the face of evidence that the diagnosis should be overturned. In contrast, we do find agents that have cancer with a gritty attitude towards their survival. For example, someone may be disposed to persist with this attitude when confronted with increasingly negative prognoses and the practical and emotional challenges that come with the worsening condition.

Now, it could be objected that we often find people who commit to gloomy assessments, of which Peter's judgement in (16) is an extreme case, and are gritty in maintaining those assessments. They are pessimists. Doesn't this show that we still need to appeal to positivity to explain why such pessimistic commitments do not count as faith? Not so. First, there is nothing about pessimistic judgements that requires they should be gritty. For example, someone who favours pessimistic beliefs or assumptions simply because they think those beliefs are true or those assumptions prudent, does not thereby hold to those beliefs or assumptions grittily. The kind of pessimistic judgement that satisfies true grit is less commonplace. The gritty pessimist would be disposed to, for example, disregard plausible contrary evidence to their beliefs, persist with the beliefs in the face of emotional and practical pressures to adopt less gloomy judgements, and so on. Second, PT does not exclude individuals from having faith in pessimistic beliefs: PT is a constraint on the kinds of attitudes required for faith, not on their content. For example, someone may form a pessimistic judgement about the future but also desire or in some way to approve of that outcome. Notably, for the reasons already given for the connection between true grit and positivity, we should expect that someone who is gritty in their pessimistic judgements will regard them positively. Neither true grit nor PT, therefore, exclude the possibility of faith in pessimistic judgments.

Another purported advantage of the positivity theory is that a positive attitude distinguishes merely believing something from having faith in it. A difference between belief in ghosts and faith in ghosts, or between belief and faith that Trump will win a second term in 2024, is that the faith attitudes must be accompanied by a positive view of their objects. Here too, the true grit theory provides a simpler explanation of the contrasting cases without needing to appeal to positivity. For example, to have faith in ghosts or faith that Trump will win a second term in 2024 requires true grit – that is to resist a variety of countervailing considerations – whereas belief in these matters can be surrendered merely on the basis of evidence that it is not true.

These considerations suggest that even if we put aside the objections to PT, the appeal to positivity as an explanation of the examples used to support PT may be dispensable in favour of one that appeals to true grit. Moreover, in some examples, such as (11) and (12), true grit appears to provide a better explanation of the intuitive difference than positivity.

REFERENCES

Alston, William P. 1996. 'Belief, Acceptance, and Religious Faith'. In *Faith, Freedom, and Rationality*, edited by J. Jordan and D. Howard-Snyder, 3–27. Lanham: Rowman & Littlefield.

Andrews, Frances. 2016. 'Introduction'. In *Doubting Christianity: The Church and Doubt*, edited by Frances Andrews, Charlotte Methuen, and Andrew Spicer. Studies in Church History 52. Cambridge: Cambridge University Press.

Audi, Robert. 2011. *Rationality and Religious Commitment*. New York: Oxford University Press.

Bishop, John. 2007. *Believing by Faith: An Essay in the Epistemology and Ethics of Religious Belief*. Oxford: Clarendon Press.

Björklund, Fredrik, Gunnar Bjornsson, John Eriksson, Ragnar Francén Olinder, and Caj Strandberg. 2012. 'Recent Work on Motivational Internalism'. *Analysis* 72 (1): 124–37. https://doi.org/10.1093/analys/anr118.

Björnsson, Gunnar, Caj Strandberg, Ragnar Francén Olinder, John Eriksson, and Fredrik Björklund, eds. 2015. *Motivational Internalism*. Oxford Moral Theory. New York: Oxford University Press.

Block, Jack, and Hobart Thomas. 1955. 'Is Satisfaction with Self a Measure of Adjustment?' *The Journal of Abnormal and Social Psychology* 51 (2): 254–59. https://doi.org/10.1037/h0048246.

Bourbeau, Philippe. 2018. 'A Genealogy of Resilience'. *International Political Sociology* 12 (1): 19–35. https://doi.org/10.1093/ips/olx026.

Buchak, Lara. 2012. 'Can It Be Rational to Have Faith?' In *Probability in the Philosophy of Religion*, edited by Jake Chandler and Victoria S. Harrison, 225–48. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199604760.003.0012.

———. 2014. 'Rational Faith and Justified Belief'. In *Religious Faith and Intellectual Virtue*, edited by Laura Frances Callahan and Timothy O'Connor, 49–73. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672158.003.0003.

———. 2017. 'Faith and Steadfastness in the Face of Counter-Evidence'. *International Journal for Philosophy of Religion* 81 (1–2): 113–33. https://doi.org/10.1007/s11153-016-9609-7.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. 'Grit: Perseverance and Passion for Long-Term Goals.' *Journal of Personality and Social Psychology* 92 (6): 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087.

Durà-Vilà, Glòria. 2016. *Sadness, Depression, and the Dark Night of the Soul*. London; Philadelphia: Jessica Kingsley Publishers.

Faigin, Carol Ann, and Kenneth I. Pargament. 2011. 'Strengthened by the Spirit: Religion, Spirituality, and Resilience through Adulthood and Aging'. In *Resilience in Aging: Concepts, Research, and Outcomes*, edited by Barbara Resnick, Lisa P. Gwyther, and Karen A. Roberto, 163–80. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-0232-0_11.

Howard-Snyder, Daniel. 2013. 'Propositional Faith: What It Is and What It Is Not'. *American Philosophical Quarterly* 50 (4): 357–72.

———. 2017. 'Markan Faith'. *International Journal for Philosophy of Religion* 81 (1): 31–60. https://doi.org/10.1007/s11153-016-9601-2.

———. 2019. 'Can Fictionalists Have Faith? It All Depends'. *Religious Studies* 55 (4): 447–68. https://doi.org/10.1017/S0034412518000161.

Jacelon, Cynthia S. 1997. 'The Trait and Process of Resilience'. *Journal of Advanced Nursing* 25 (1): 123–29. https://doi.org/10.1046/j.1365-2648.1997.1997025123.x.

Jackson, Elizabeth. 2020. 'Belief, Faith, and Hope: On the Rationality of Long-Term Commitment'. *Mind*, no. fzaa023. https://doi.org/10.1093/mind/fzaa023.

Kaufmann, Walter Arnold. 1958. *Critique of Religión and Philosophy*. Princeton: Princeton University Press.

Kolodiejchuk, Brian. 2007. *Mother Teresa - Come Be My Light: The Revealing Private Writings of the Nobel Peace Prize Winner*. New York: Doubleday.

Kvanvig, Jonathan L. 2013. 'Affective Theism and People of Faith'. *Midwest Studies in Philosophy* 37 (1): 109–28. https://doi.org/10.1111/misp.12003.

Luthar, Suniya S., Dante Cicchetti, and Bronwyn Becker. 2000. 'The Construct of Resilience: A Critical Evaluation and Guidelines for Future Work'. *Child Development* 71 (3): 543–62.

Malcolm, Finlay. 2017. 'Perseverance and the Value of Faith: Lessons from the Psychology of Grit'. In . University of Manchester.

———. 2020. 'The Moral and Evidential Requirements of Faith'. *European Journal for Philosophy of Religion* 12 (1): 117–42. https://doi.org/10.24204/ejpr.v0i0.2658.

Malcolm, Finlay, and Michael Scott. 2017. 'Faith, Belief and Fictionalism'. *Pacific Philosophical Quarterly* 98 (S1): 257–74. https://doi.org/10.1111/papq.12169.

Masten, Ann S., Karin M. Best, and Norman Garmezy. 1990. 'Resilience and Development: Contributions from the Study of Children Who Overcome Adversity'. *Development and Psychopathology* 2 (4): 425–44. https://doi.org/10.1017/S0954579400005812.

Matheson, Jonathan. 2018. 'Gritty Faith'. American Catholic Philosophical Quarterly. 2018. https://doi.org/10.5840/acpq201858152.

Mckaughan, Daniel J. 2013. 'Authentic Faith and Acknowledged Risk: Dissolving the Problem of Faith and Reason'. *Religious Studies* 49 (1): 101–24. https://doi.org/10.1017/S0034412512000200.

McKaughan, Daniel J. 2016. 'Action-Centered Faith, Doubt, and Rationality'. *Journal of Philosophical Research* 41 (9999): 71–90. https://doi.org/10.5840/jpr20165364.

———. 2018. 'Faith through the Dark of Night: What Perseverance amidst Doubt Can Teach Us about the Nature and Value of Religious Faith'. Faith and Philosophy. 2018. https://doi.org/10.5840/faithphil2018327101.

Morton, Jennifer M., and Sarah K. Paul. 2018. 'Grit'. *Ethics* 129 (2): 175–203. https://doi.org/10.1086/700029.

Noveck, Ira. 2018. *Experimental Pragmatics: The Making of a Cognitive Science*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316027073.

Ögtem-Young, Özlem. 2018. 'Faith Resilience: Everyday Experiences'. *Societies* 8 (1): 10. https://doi.org/10.3390/soc8010010.

Pargament, Kenneth I., and Jeremy Cummings. 2010. 'Anchored by Faith: Religion as a Resilience Factor'. In *Handbook of Adult*

Resilience, edited by J. W. Reich, A. J. Zautra, and J. S. Hal, 193–210. New York, NY, US: The Guilford Press.

Plantinga, Alvin. 2000. *Warranted Christian Belief*. New York: Oxford University Press.

Pojman, Louis. 1986. 'Faith without Belief?' *Faith and Philosophy* 3 (2): 157–76. https://doi.org/10.5840/faithphil19863213.

Rutter, Michael. 2006. 'Implications of Resilience Concepts for Scientific Understanding'. *Annals of the New York Academy of Sciences* 1094: 1–12. https://doi.org/10.1196/annals.1376.002.

Schellenberg, J. L. 2005. *Prolegomena to a Philosophy of Religion*. Ithaca, N.Y.: Cornell University Press.

Stewart, Donna E., and Tracy Yuen. 2011. 'A Systematic Review of Resilience in the Physically Ill'. *Psychosomatics* 52 (3): 199–209. https://doi.org/10.1016/j.psym.2011.01.036.

Swinburne, Richard. 2001. 'Plantinga on Warrant'. *Religious Studies* 37 (2): 203–14. https://doi.org/10.1017/S0034412501005601.

Zautra, Alex J., John Stuart Hall, and Kate E. Murray. 2010. 'Resilience: A New Definition of Health for People and Communities'. In *Handbook of Adult Resilience*, edited by A. Reich, J. Zautra, and J. S. Hall, 3–29. New York, NY, US: The Guilford Press.

EuJAP | Vol. 17 | No. 1 | 2021
UDC: 111:130.31
https://doi.org/10.31820/ejap.17.1.2

PURE POWERS ARE NOT POWERFUL QUALITIES

Joaquim Giannotti[1]

[1]University of Birmingham

Original scientific article – Received: 11/11/2020 Accepted: 27/01/2021

ABSTRACT

There is no consensus on the most adequate conception of the fundamental properties of our world. The pure powers view and the identity theory of powerful qualities claim to be promising alternatives to categoricalism, the view that all fundamental properties essentially contribute to the qualitative make-up of things that have them. The pure powers view holds that fundamental properties essentially empower things that have them with a distinctive causal profile. On the identity theory, fundamental properties are dispositional as well as qualitative, or powerful qualities. Despite the manifest difference, Taylor (2018) argues that pure powers and powerful qualities collapse into the same ontology. If this collapse objection were sound, the debate between the pure powers view and the identity theory of powerful qualities would be illusory: these views could claim the same advantages and would suffer the same problems. Here I defend an ontologically robust distinction between pure powers and powerful qualities. To accomplish this aim, I show that the collapse between pure powers and powerful qualities can be resisted. I conclude by drawing some positive implications of this result.

Keywords: *Pure powers; powerful qualities; dispositionalism; collapse objection; dispositional essentialism*

Correspondence: philosophy@joaquimgiannotti.com

1. The Qualitative and the Dispositional

Fundamental properties are an elite minority that suffices to characterise all things completely and form a minimal basis on which all non-fundamental properties supervene (Lewis 1983, 1986, 2009). It is typically claimed that physics is in the business of discovering the fundamental properties of our world. Properties such as *charge*, *mass*, and *spin* are often invoked as plausible candidates. Yet there is no consensus on the most adequate conceptions of the fundamental properties of our world. The disagreement runs deep and covers several longstanding metaphysical questions (cf. Armstrong 2005). Here my focus is on whether the nature of fundamental properties is *qualitative/categorical* or *dispositional.*

Three monist views offer an answer to this question. *Categoricalism* holds that all fundamental properties are essentially and purely qualitative, or pure qualities (e.g., Lewis 1986; Armstrong 1997). The *pure powers view* holds that all fundamental properties are essentially and purely dispositional, or *pure powers* (e.g., Mumford 2004; Bird 2007a). The *identity theory of powerful qualities* holds that all fundamental properties have a dual nature: they are essentially both dispositional *and* qualitative, or *powerful qualities* (e.g. Martin 1993, 2008; Heil 2003, 2012). This view is committed to a distinctive three-fold identity claim: a fundamental property's dispositionality is identical with its qualitativity, and each of these is identical with the property itself (e.g., Heil 2003, 111; Taylor 2018, 1424).[1]

To elucidate these positions, we need two clarifications: one concerns what it is for a property to be a certain way, the other regards the notions of dispositionality and qualitativity.

For the sake of the discussion, let us assume that to say that a property P is essentially such-and-such means that it is true in virtue of P's nature that P is such-and-such, or that P's nature grounds that P is such-and-such. As is now standard, if P is essentially such-and-such, then necessarily P is such-and-such, but the converse does not hold.

Now let us clarify dispositionality and qualitativity. Dispositionality is a matter of what a thing is disposed to do in various possible circumstances by virtue of having certain properties. Call these *dispositional properties*. We can think of dispositional properties as those that *cannot* be specified

[1] In the literature, we can find also *mixed views*. These hold that some fundamental properties are essentially qualitative, and others are essentially dispositional (e.g., Ellis 2001, 2002, 2012; Ellies and Lierse 1994). In what follows, I restrict my attention to monist views of fundamental properties.

independently of any causal roles. These can be regarded as descriptions or ways of conceptualizing or specifications that typically refer to the manifestation of distinctive effects in distinctive circumstances. Accordingly, a property such as that of *having a determinate charge* is plausibly dispositional: it cannot be specified independently of the causal role of producing an electromagnetic force that electrons, say, play or possess. An *essentially* dispositional property, or power, is one for which it is true in virtue of its nature that it cannot be specified independently from any causal roles.

Some philosophers take qualitativity to be a matter of how a thing is in virtue of possessing some actual or occurrent properties (e.g., Strawson 2008, 278; Heil 2010, 70; Heil 2012, 59). However, on this understanding, every actual dispositional property would be qualitative. Such a result is unsavoury for those who wish to preserve the mutual exclusivity of dispositional and qualitative properties (e.g., Armstrong 2005; Bird 2007). Therefore, we need to opt for a different characterisation. Since my target in this paper is the powerful qualities view, the qualitative should be characterised in a way which permits one to coherently hold that a property is dispositional as well as qualitative. Consequently, I will not follow those who take the qualitative to be the non-dispositional. For example, I will part ways with Alexander Bird, who claims that a qualitative property requires us "to deny that it is necessarily dispositional" (2007a, 66–67).

Typically, examples of qualitative properties include shape and colour properties (*being spherical*, *being scarlet*), structural properties (*having a determinate crystalline structure*), geometrical properties (*having an angle of a determinate measure*), and spatio-temporal properties (*having a determinate location*). Two things group these properties: one is that they contribute to the make-up of objects that instantiate them, the other is that they *can* be specified independently of any causal roles. An instance of the property of *having a tetrahedral molecular structure* is plausibly qualitative for it contributes to the make-up of a bearer, say a diamond, and its characterisation does not force us to invoke any causal role. However, dispositional properties also contribute to the make-up of their bearers. For instance, the property of *having a determinate charge* is a part of the make-up of an electron. To draw the distinction, we should privilege the fact that qualitative properties can be specified independently from any causal roles. Accordingly, I submit that an essentially qualitative property, or quality, is one for which it is true in virtue of its nature that it can be specified independently from any causal role. Some ambiguity is nonetheless inescapable. For this reason, examples of qualities are contentious. I will mention properties such as that of *having a certain quantity of charge* and *having a certain quantity of matter* as plausible candidates of fundamental qualities (e.g., Giannotti 2019). These two quantitative properties seem to be exhaustively specified without any reference to causal roles and are more plausible fundamental properties than colours or geometrical shapes.

On this characterisation of the qualitative–dispositional distinction, some properties can be both dispositional and qualitative in the sense that they can be characterised (overtly or covertly) in terms of the causal *as well as* non-causal roles they play. It is one of the aims of this paper to clarify this view.

Two remarks on this characterisation of the qualitative–dispositional distinction are needed. First, it is not meant to be a reductive analysis of dispositionality and qualitativity. For the purposes of this paper, a general sense of these notions will suffice.

Second, this characterisation is not the only game in town. For example, it is orthodoxy amongst categoricalists and dispositionalists to define powers and qualities in mutually exclusive terms. But one is not forced to do so. As I explained, here we need to adopt a different conception of the qualitative and the dispositional.[2]

2. The Collapse Objection

This paper concerns the pure powers view and the identity theory of powerful qualities. These views are manifestly distinct. It is one thing to claim that the nature of all fundamental properties is purely dispositional, however, it is another to claim that it is dispositional as well as qualitative. It seems that only the identity theory is *prima facie* committed to the view that fundamental properties are essentially dispositional as well as qualitative.

Contrary to the appearances, it has been recently argued that there is no real, ontologically robust distinction between the pure powers view and the identity theory of powerful qualities. Henry Taylor (2018) argues that they collapse into the same view. Call this the *collapse objection*.

If the collapse objection were sound, then the pure powers view and the identity theory could claim the same advantages and would suffer the same problems. The debate between these views would be illusory. Since each position claims to be preferable over other options in the debate about

[2] See Ingthorsson (2013) for an overview of various ways in which the qualitative–dispositional distinction is spelled out in the literature.

fundamental properties, it is crucial to assess Taylor's collapse objection. My aim in this paper is to show that the collapse does not obtain: pure powers and powerful qualities share some relevant features and yet do not coincide. To do so, I defend an ontological demarcation between pure powers view and the identity theory.

Here is the plan. In the remainder of this section, I lay out a few assumptions that are needed for delineating the scope of this paper. In Section 3, I articulate the notion of a *part of a property,* which Taylor (2018) invokes to characterise the pure powers view and the identity theory. As it will become clear in due course, a suitable interpretation of this notion serves the purposes of this paper well. In Sections 4 and 5, I illustrate the pure powers view and the identity theory, respectively. In Section 6, I formulate the collapse objection in a more precise way. In Section 7, I show how to resist it by expanding on a strategy that I hinted at elsewhere (Giannotti 2019). I conclude in Section 8 by identifying some theoretical advantages of this result.

To begin with, let us acknowledge that Taylor's objection is not meant to undermine the prospects of dispositionalism *tout court*. Rather it is meant to show that two prominent dispositionalist approaches—namely, the pure powers view and the identity theory—fail to be ontologically distinct. Taylor offers his *compound view*, which I will outline in Section 4, as a positive alternative that preserves the dispositionalist spirit while escaping the collapse objection. If the argument in Section 7 is correct, we are not forced to embrace the compound view. This is good news for both the pure powers theorist and the identity theorist.

Second, it is not my aim to defend the correctness of either the pure powers view or the identity theory. It is one thing to show that the collapse between these views can be escaped. It is another thing to show that either of them is true. Here my focus is on the former task.

Third, the pure powers view and the identity theory can come in a variety of flavours. There are various ways of understanding the claim that powers are pure (e.g., Taylor 2018). Likewise, there are several ways of interpreting the claim that a property's dispositionality and its qualitativity are identical (I discuss some of these in Giannotti 2019). Furthermore, it is possible to articulate accounts that renounce the identity claim and yet share *some* similarities with powerful qualities (e.g., Tugby 2012; Taylor 2018; Giannotti 2019; Williams 2019). The discussion of the collapse objection is restricted to the versions of pure powers and powerful qualities that I present in what follows. However, I shall neglect the question of whether these versions are the best ones on the market.

Bearing these remarks in mind, I will turn to illustrate Taylor's notion of a part of a property.

3. "Parts" of Properties

Taylor (2018) invokes the notion of a 'part' of a property to illustrate some of the claims that are attributed to both the pure powers view and the identity theory. For example, in describing the pure powers view, he says that "there is no part of a property's nature that is non-powerful" (2018, 1433). Unfortunately, Taylor does not elucidate what it is for a property to have parts. Let us assume that he makes no category mistakes. By doing so, we can articulate and evaluate a more fine-grained version of the collapse objection. We can also reformulate both the pure powers view and the identity theory in a way that illuminates the structure of fundamental properties as portrayed by these views. These advantages suggest that the notion of a part of a property is serviceable in casting a light on the "metaphysical workings" of pure powers and powerful qualities. However, my aim is *not* to offer a complete metaphysics of parts of properties. Some readers will find the choice of sticking with parts questionable and potentially confusing. The following remarks will address some initial reservations.

First, Taylor's notion of a part of a property is not mereological. The claim that a property has parts should not be understood as the claim that a property is made of more basic elements—parts—that constitute it. Therefore, talk of parts should not be construed as implying that fundamental properties are bundles or aggregates of parts. This mereological interpretation would threaten the claim that pure powers are fundamental ontological entities: arguably, if pure powers are composed of parts, these parts are more fundamental than the pure powers themselves.

There is another compelling reason for avoiding the mereological interpretation. If parts are themselves purely dispositional or purely qualitative properties, and if these are themselves made of parts, then a regress of parts emerges: the parts that make fundamental properties have further parts, which in turn have further parts, and so on *ad infinitum*. A somewhat similar problem occurs if we take parts as entailing the existence of other properties with parts: the latter would bring into existence further properties with parts, which in turn would bring into existence further properties with parts, and so on, *ad infinitum* again. These problematic consequences give us reasons for favouring a different approach.

A more promising way of thinking of parts is to take them as *features* or *aspects* of fundamental properties that ground and explicate dispositional roles and qualitative features, where these roles and features do not entail the existence of other properties with parts. It is one of the aims of this paper to clarify this idea. Since the appeal to parts should preserve the view that pure powers are fundamental, parts are better regarded as being, in some sense, dependent upon the properties of which they are parts. Elsewhere, I proposed that parts can be thought of as aspects that are ontologically dependent upon properties (Giannotti 2019). In this paper, I wish to maintain a more flexible stance. We do not need to decide which relation better captures the link between parts and properties. Nor does the reader have to accept that parts of properties are the sort of aspects that I have in mind in Giannotti (2019).

In the literature, three views that adopt a similar conception of parts of properties are worthy of mention. In chronological order, the first one is the *two-sided* version of powerful qualities (Martin and Heil 1999). On this view, properties have dispositional and qualitative *sides* or aspects, which can be abstracted from the unitary property itself. For example, the property of *having a determinate charge* is two-sided in the sense that it can be thought of as a quality or power. We can regard it as the quality of *having a specific quantity of charge* or the *disposition to produce an electromagnetic field*. Another view is Tugby's (2012) qualitative dispositional essentialism. On this view, properties have a qualitative nature that grounds the dispositional aspect. The qualitative aspect is an inherent nature that grounds the causal roles associated with a property. Qualitative and dispositional aspects stand in a grounding relationship. Since it is the qualitative aspect of a property P that governs the causal roles that things play by virtue of instantiating P, Tugby's view is closer to the categoricalist camp than the dispositional one. The third view, as anticipated, is the dual-aspect account I put forward in Giannotti (2019). On this view, fundamental properties have dispositional and qualitative aspects that supervene on the property of which they are aspects and play distinct theoretical roles.[3]

The cited views are a few examples of how we can think of the relation between parts of properties—or aspects—and properties in non-mereological terms. They represent evidence of the plausibility of the idea

[3] Another philosopher who thinks that some properties have parts in a non-mereological sense is David Armstrong (1997). Structural universals have more basic universals as non-mereological constituents. Note, however, that if parts of properties are to be understood *à la* Armstrong, then these are presumably less fundamental then their parts. This conception is therefore inadequate for making sense of fundamental properties in terms of parts.

that properties can have parts in a non-mereological sense, which is suitable to various ontological interpretations.

Equipped with parts of properties, let us move onto the discussion of pure powers

4. Pure Powers

The pure powers view holds that all fundamental properties are essentially *purely dispositional*, or pure powers (e.g., Mumford 2004; Bird 2007a, 2016). In this section, my aim is to illustrate Taylor's conception of pure powers, which I will adopt for the sake of the discussion.

At first glance, the purity claim seems to convey the idea that the dispositional nature of a pure power exhausts its being. As Taylor puts it, to say that a power is pure is to say that "the whole nature of a property is powerful: all of it is powerful and there is no part of a property's nature that is non-powerful" (2018, 1433). Crucially, this interpretation of the purity claim in terms of *complete powerfulness* is the one that Taylor endorses. He thinks of complete powerfulness as "the most natural way to interpret" the purity of pure powers (2018, 1433). By appealing to the notion of a part of a property, we can formulate the purity claim in terms of **Complete Powerfulness**, where Greek letters denote parts of properties:

> **Complete Powerfulness.** For every part α of a fundamental property, α is dispositional.

This formulation captures the view that a pure power is completely powerful: since powers are essentially dispositional, it is in virtue of its nature that a pure power has only dispositional parts. Of course, we must supplement **Complete Powerfulness** with a characterisation of the notion of a *dispositional part*. Here is my preferred one.

> **Dispositional Part**. A part α of a property P is dispositional if and only if there is a causal role or cluster of causal roles in virtue of α that is played or possessed by every object that has P.

The proposed characterisation regiments the idea that if a property has some dispositional parts, then an object instantiating it plays some causal roles by virtue of these parts. Differently put, the dispositional parts of a property are those that ground the causal roles of a bearer of that property. Tugby (2012) defends a similar characterisation: the dispositional aspects of a property are the causal roles that a bearer plays by virtue of instantiating that property. Here is an example to illustrate. Let us consider the property of *having a determinate mass* and the causal role of producing a gravitational force. Under **Dispositional Part**, if there is a part of the property of *having a determinate mass* such that any massive object plays the causal role of producing a gravitational force in virtue of it, then this part is dispositional. It is important to note that **Dispositional Part** is not meant to elucidate the notion of dispositionality. Instead, it expresses the relation between dispositional parts and causal roles. If we interpret the pure powers view in terms of **Complete Powerfulness**, this position holds that every part of the fundamental properties grounds the possession of some causal roles.

It is also important to stress that Taylor does not defend the claim that **Complete Powerfulness** is the *only* plausible interpretation of the purity claim. He makes a different claim, namely that **Complete Powerfulness** is the most natural interpretation of the claim that a power is purely dispositional. For the sake of the discussion, I will grant this point.

In addition to **Complete Powerfulness**, the version of pure powers view under scrutiny endorses two other claims. Let us call them **Actuality** and **Non-Armstrongianism**.

Actuality captures the idea that pure powers are actual, here-and-now properties of their bearers. As Taylor puts it, pure powers "are real, actual features of objects" (2018, 1431). We can formulate this claim as follows:

> **Actuality**. Every fundamental property is an actual and real property of its bearers.

According to **Actuality**, if the property of *having a determinate charge* is a fundamental power, then it is also an actual and real property of its bearers.

Now let us consider **Non-Armstrongianism**. The pure powers theorist denies that fundamental properties are qualitative in *the sense of* being Armstrongian qualities, namely not essentially dispositional (e.g., Armstrong 1997). Here is one way to formulate this claim:

> **Non-Armstrongianism**. Fundamental properties are not Armstrongian qualities.

Non-Armstrongianism allows us to distinguish the pure powers views (as thought of *à la* Taylor) from views like qualitative dispositional essentialism (Tugby 2012). On both views, certain parts of a property P ground the causal profile of something that instantiates P. Therefore, on both positions, the dispositional profile associated with P is necessary. However, the necessity flows from two incompatible sources. On the pure powers view, the dispositional profile is grounded in some dispositional parts; by contrast, on qualitative dispositional essentialism, it is grounded in some qualitative aspects.

To sum up, the version of the pure powers view that faces the collapse objection endorses three distinctive claims: **Complete Powerfulness**, **Actuality**, and **Non-Armstrongianism**. Now let us turn to illustrate the identity theory of powerful qualities.

5. Powerful Qualities

The identity theory holds that all fundamental properties have a dual nature: they are at once dispositional and qualitative, or powerful qualities (Martin 2008, 64). As Martin and Heil put it, "in virtue of possessing a property [powerful quality], an object possesses both a particular dispositionality and a particular qualitative character" (1999, 45–46).[4]

On this view, a fundamental property is essentially such that it can be characterised in terms of the causal roles played by things that instantiate it *and* the qualitative features these things have in virtue of it. With 'qualitative features', I have in mind features that can be described or conceptualized or specified without involving, overtly or covertly, any reference to manifestations of distinctive effects in characteristic circumstances.

Qualitative features are typically associated with qualities. Structural, geometrical, and mathematical features would be paradigmatic examples of qualitative features. If the property of *having a determinate charge* were a powerful quality, it would ground some causal roles, such as that of producing an electromagnetic field, and some qualitative features which something instantiating this property can be said to have. For example, it seems that by virtue of instantiating the property of *having a determinate charge*, objects can *also* be specified qualitatively in terms of having a certain quantity of charge (which can be measured in coulombs). To give another example, consider the property of *having a determinate spin*. If it were a powerful quality, it would ground some causal roles a bearer plays

[4] This formulation would make Tugby's (2012) qualitative dispositional essentialism a version of powerful qualities.

and some of its qualitative features. For example, the causal role of producing a certain magnetic moment and the qualitative feature of having a specific quantity that can have only values that are multiples of ħ/2, where ħ is the reduced Planck constant.

We can appeal to Taylor's parts of properties to define the notion of a powerful quality in precise terms as follows.

> **Powerful Quality**. A property P is a powerful quality if and only if P essentially has some dispositional parts and some qualitative parts.

The notion of a dispositional part is the same that has been introduced in Section 2. Now we need to characterise that of a qualitative part. A promising formulation, which captures the idea of qualitativity, is the following one.

> **Qualitative Part.** A part α of a property P is qualitative if and only if there is a qualitative feature in virtue of α that is possessed by every object that has P.

Put differently, a qualitative part of P is one which grounds the existence of a qualitative feature of a bearer of P which is neither overtly nor covertly dispositional. Suppose once again that the property of *having a determinate charge* is a powerful quality. If there is a part of this property in virtue of which a bearer has a feature that does not involve, either overtly or covertly, any manifestations of distinctive effects in characteristic circumstances, then this part is qualitative. To use the previous example, a qualitative part of the property of *having a determinate charge* is that which grounds a mathematical, non-dispositional feature of an electron, such as the possession of a certain quantity of charge that can be measured in coulombs. The powerful qualities theorist would maintain that the property of *having a determinate charge* has some dispositional parts in addition to this qualitative part. These ground the causal roles that the electron plays by virtue of *having a determinate charge*, such as that of producing an electromagnetic force.[5] Putting these pieces together, we get that it is true in virtue of its nature that a powerful quality has some parts that ground some causal roles and some others that ground some qualitative features.

[5] Some caution is needed. We should not take **Qualitative Part** to be a definition of a quality. Otherwise, this definition would be problematically circular. Rather **Qualitative Part** simply captures the relation between the qualitativity of a bearer and some parts of a powerful quality which is instantiated by such a bearer.

Someone could worry about inferring a robust distinction between dispositional and qualitative parts from a distinction between dispositional roles and qualitative features. Surely, it is one thing to describe a charged object in quantitative or mathematical terms, but it is another thing to describe it in causal or dispositional terms. However, such a distinction does not guarantee that such descriptions pick out different parts of a property of *having a determinate charge*. Instead of succumbing to this objection, identity theorists embrace the possibility that the same property can be at once dispositional as well as qualitative. To put it in terms of parts, identity theorists champion the idea qualitative and dispositional parts are, in a sense that I shall explain below, identical.

The identity theory endorses a distinctive three-fold identity claim between a property's dispositionality, its qualitativity, and the property itself (e.g., Heil 2003, 2012; Martin 2008; Taylor 2013). Heil formulates it as follows:

> If P is an intrinsic property of a concrete object, P is simultaneously dispositional and qualitative; P's dispositionality and qualitativity are not aspects or properties of P; P's dispositionality, Pd, *is* P's qualitativity, Pq, and each of these *is* P: Pd = Pq = P. (Heil 2003, 111)

Difficulties in understanding this identity claim obfuscate the merits of powerful qualities. Here I do not wish to defend its correctness (for a discussion about some plausible interpretations, see Giannotti 2019). The clause "P's dispositionality and qualitativity are not aspects or properties of P" is meant to rule out the idea that powerful qualities are conjunctive properties made of purely dispositional and purely qualitative properties. Recall that on the proposed characterisation, parts are non-mereological aspects of properties that are dispositional and qualitative in virtue of playing the theoretical roles of grounding causal and qualitative features, respectively. Therefore, we should not think of dispositional and qualitative parts as purely dispositional and purely qualitative properties. Nor are these parts such that they bring into existence further purely dispositional or qualitative properties.

Now let us reformulate the identity claim in terms of parts of properties as follows.

> **Identity.** For every fundamental property P, (1) P has at least one dispositional part and P has at least one qualitative part, (2) every dispositional part of P is numerically identical with a qualitative part of P and *vice versa*, and (3) no part of P is a proper part.

It is plainly obvious that **Identity** is different from Heil's formulation. However, the difference is not metaphysically deep: **Identity** preserves the original three-fold claim.

Clause (1) reformulates the claim that the properties are both dispositional and qualitative in terms of parts of properties.

Clause (2) does the same for the identity claim between a property's qualitativity and its dispositionality. Under the adoption of **Dispositional Part** and **Qualitative Part**, clause (2) says that every part of a property that grounds some causal roles is a part that also grounds some qualitative features. Take a particle that instantiates the property of *having a determinate mass*. Under the assumption that this is a fundamental property, (2) implies that the part of this property that grounds the particle's causal role of producing a gravitational field is identical with the part that grounds one of the qualitative features of the particle, such as that of having a certain quantity of matter measurable in kilograms.

Clause (3) is a reformulation of the identity claim between a property's dispositionality and qualitativity and the property itself. The proposed interpretation borrows the proper/improper distinction from mereology to recover the idea that the dispositionality and qualitativity of a property are identical to the property itself. No proper part is identical with the object of which it is a part, but improper parts always are. It might be useful to acknowledge that views that endorse something like (3) already appeared in the literature. For example, Locke (2012) and Smith (2016) discuss versions of 'austere quidditism' and 'moderately austere quidditism' that take fundamental properties *to be identical with* and individuated by their qualitative suchness, which is an *aspect* of fundamental property (these views, however, differ with respect to the thinness of the qualitative aspect of fundamental properties). Smith (2016, 251–253) compares moderately austere quidditism with the identity theory explicitly. Smith's moderately austere quidditism holds that:

> […] the property and its qualitative nature are identical ($P = P_Q$), but the property and its dispositionality are plausibly distinct ($P \neq P_D$) despite the fact that, as a matter of metaphysical necessity, an object instantiates P (and hence P_Q) if an only if it instantiates P_D. (Smith 2016, 252)

By contrast, as **Identity** states, the identity theory holds that a property and its dispositional parts are identical (that is, $P = P_D$). I will return to Smith's view in the final section.

Having outlined the pure powers view and the identity theory, we can now discuss the collapse objection.

6. The Collapse Argument

As Taylor notes, the identity theorist must embrace **Complete Powerfulness** (2018, 1434). Otherwise, a powerful quality would have some non-dispositional parts which would falsify **Identity**. Thus both the pure powers theorist and the identity theorist take fundamental properties to be completely powerful.

Like the pure powers theorist, the identity theorist also takes powerful qualities to be actual properties of their bearers. For example, Heil regards powerful qualities *as* qualities because they are "here and now, actual, not merely potential, features of objects, of which they are qualities" (2012, 59). This claim expresses a commitment to **Actuality**.

Lastly, on the identity theory, we must deny that fundamental powerful qualities are qualitative in the Armstrongian sense—namely, essentially non-dispositional. Otherwise, **Identity** could not hold. Thus the identity theorist embraces **Non-Armstrongianism**.

As I explained in Section 3, the pure power theorist holds **Complete Powerfulness**, **Actuality**, and **Non-Armstrongianism**. The pieces of the collapse objection are now put together. As Taylor puts it:

> […] The two views share the same commitments concerning the ontology of properties: both accept that properties are powers, both accept that they are 'qualities' in the same ways, and both accept the same interpretation of the 'purity' claims. (Taylor 2018, 1435)

In the remainder of the paper, I will show how to resist the collapse. I will explain that even if the pure powers view and the identity theory endorse **Complete Powerfulness**, **Actuality**, and **Non-Armstrongianism**, these views do not coincide *because* pure powers and powerful qualities are essentially distinct: not everything that is true in virtue of the nature of a pure power is also true in virtue of the nature of a powerful quality.

Before proceeding any further, it is worth stressing that the collapse objection targets *only* views of pure powers and powerful qualities, which endorse **Complete Powerfulness**, **Actuality**, and **Non-Armstrongianism**. A

straightforward way to escape the collapse would be to adopt a view of fundamental properties which renounces one of these claims.
The rejection of **Actuality** seems to be the most problematic option though. It would imply that fundamental properties are not actual features of their bearers. Such a view should strike us as implausible.

The rejection of **Non-Armstrongianism** implies that fundamental properties *are* essentially non-dispositional qualities. If we follow this approach, both the pure powers view and the identity theory must be abandoned. Neither the pure powers theorist nor the identity theorist can contemplate this decision.

Something similar can be said for the option of giving up **Complete Powerfulness**. Because not all their parts would be dispositional, this solution would imply that fundamental properties are neither pure powers nor powerful qualities—at least as standardly construed. But nor would they automatically be Armstrongian qualities. The denial of **Complete Powerfulness** is, in fact, compatible with views that take fundamental properties to have *both* dispositional and qualitative, non-dispositional parts.[6] These views, which I shall not discuss here, demand the acceptance of a new kind of properties. Surely, this will be a fair cost for some. However, before we pay it by the coin of ontology, it is worth exploring whether we can resist the collapse objection without abandoning **Complete Powerfulness**.

7. Escaping the Collapse

Elsewhere, I suggested that the identity theorist can argue that it is a "dual nature" (Martin and Heil 1999, 46; Martin 2008, 45; see also Giannotti 2019) that makes a powerful quality dispositional and qualitative; in contrast, a pure power has a powerful but not qualitative nature. Call this the *distinct nature strategy*.

Lamentably, Giannotti (2019) merely gestures toward the distinct nature strategy without offering a clear articulation. Since there I take the qualitative to be a matter of the actual contribution to the make-up of bearers, the lack of elucidation is problematic: it leads us to misleadingly think that pure powers are not qualitative in the sense of being actual (e.g., Taylor forthcoming). However, since I acknowledge that pure powers are

[6] For example, Taylor (2018, 1438–1439) offers a compound view of properties that have dispositional and qualitative, non-dispositional parts. Giannotti (2019) and Williams (2019) put forward similar views. For a critical discussion of Giannotti's dual-aspect account, see Taylor (forthcoming).

actual, the manoeuvre is meant to take a different shape. In this section, I will endeavour to follow the distinct nature strategy through, thereby showing that it is indeed a promising option for resisting the collapse objection. The upshot of this strategy is that both the pure powers view and the identity theory endorse **Complete Powerfulness**, **Actuality**, and **Non-Armstrongianism** and yet pure powers and powerful qualities have different natures—that is, they are essentially distinct.

The distinct nature strategy aims to establish the soundness of the following argument:

(1) If pure powers and powerful qualities are essentially distinct, then they are ontologically distinct kinds of properties.
(2) If pure powers and powerful qualities are ontologically distinct kinds of properties, then the pure powers view and the identity theory do not amount to the same view.
(3) Pure powers and powerful qualities are essentially distinct.

Therefore:

(4) The pure powers view and the identity theory do not amount to the same view.

If sound, this argument establishes that pure powers and powerful qualities *do not* "share the same commitments concerning the ontology of properties" (Taylor 2018, 1435). The conclusion (4) is the denial of the conclusion of the collapse argument, namely that pure powers and powerful qualities "are not distinct" (Taylor 2018, 1438). Importantly, (4) is compatible with the possibility that the pure powers view and the identity theory share *some* commitments. That is, (4) is consistent with Taylor's claim that the two views share **Complete Powerfulness**, **Actuality**, and **Non-Armstrongianism**.

The success of the distinct nature strategy hangs on premise (3). Recall that I adopted a conception of essentiality according to which the claim that a property P is essentially such-and-such means that it is true in virtue of P's nature that P is such-and-such (Section 1). Accordingly, (3) states that pure powers and powerful qualities differ with respect to what is true in virtue of their own nature.

Both the pure powers view and the identity theory endorse **Complete Powerfulness**, which describes or characterises the nature of fundamental properties. Since both views endorse it, we should expect that **Complete Powerfulness** entails the same view about what it is essential to fundamental properties. If there is a difference between pure powers and powerful qualities in what is essential to them, then premise (3) is true—namely, it is true that pure powers and powerful qualities have different natures. The proposed framework of parts of properties is extremely serviceable for establishing this claim.

Let us start by observing that **Complete Powerfulness** entails the following characterisation of the nature of pure powers and powerful qualities:

> **Essential Dispositionality**. For every essential part α of a fundamental property, α is dispositional.

This should be uncontroversial: if all parts of a power are dispositional, so are its essential parts. **Essential Dispositionality** says that every essential part of a fundamental property grounds some causal roles that bearers of such a property play. Both the pure powers theorist and the identity theorist happily accept **Essential Dispositionality**. According to the pure powers theorist, a power's nature is exhausted in its powerfulness (e.g., Bird 2007a, 100). The identity theorist also embraces **Essential Dispositionality**. A powerful quality does not comprise any non-dispositional parts; otherwise, **Identity** would be false (e.g., Heil 2003, 111). So far, so good. However, there is a crucial difference: *only* the identity theory is committed to **Identity**, namely the claim that the dispositional and qualitative parts of a property are identical. And the conjunction of **Identity** and **Essential Dispositionality** entails another claim:

> **Essential Qualitativity.** For every essential part α of a fundamental property, α is qualitative.

This claim is true of powerful qualities. It states that every essential part of a fundamental powerful quality grounds some qualitative features that bearers of such a powerful quality has. **Essential Dispositionality** and **Essential Qualitativity** taken together capture the spirit of the identity theory nicely (e.g., Martin and Heil 1999, 45–46): by virtue of the essential parts of a fundamental powerful quality P, every object that instantiates P has a particular dispositionality and a particular qualitativity. Note that the identity theorist cannot separate **Essential Dispositionality** and **Essential Qualitativity**. By embracing **Identity**, **Complete Powerfulness** gives the identity theorist a two-for-one deal: **Identity** and **Complete Powerfulness** entail both **Essential Dispositionality** and **Essential Qualitativity**.

The pure powers theorist *does not embrace* **Identity**. But it is only under the assumption of **Identity** that **Complete Powerfulness** entails *both* **Essential Dispositionality** and **Essential Qualitativity.** Therefore, on the pure powers view, **Complete Powerfulness** *does not entail* **Essential Qualitativity**. This seems quite right: the pure powers view denies that the essence of fundamental powers is to ground qualitative, non-causal features of bearers.

It appears, therefore, that **Essential Dispositionality** and **Essential Qualitativity** are both true of the identity theory. And **Essential Dispositionality** is true of the pure powers view. But **Essential Qualitativity** is not true of the pure powers view. Nor does it capture the nature of pure powers. Premise (3) of the distinct nature argument is consequently true: pure powers and powerful qualities are essentially distinct. The collapse is escaped. We can draw an ontological demarcation between the pure powers view and the identity theory on the grounds of their difference with respect to the truth of **Essential Qualitativity**.

Someone might worry that the distinct nature argument jeopardizes the robustness of the qualitative–dispositional distinction. However, the identity theorist would embrace this result. If the identity theory is true, then the difference between dispositional and qualitative parts does not demarcate a real distinction among properties. The identity theorist would stress that there is no incoherence in claiming that the same part of a property can ground some causal roles *and* some qualitative features. Yet the proposed characterisation of dispositional and qualitative parts does not automatically establish that this is indeed the case.

Here I do not wish to establish the correctness of the identity theory. So, I will just hint at a possible strategy to show that the qualitative and dispositional parts of a property are indeed identical. Distinctness of roles, the identity theorist could argue, does not reflect distinctness of parts. Just as the same person can play the role of a parent and of a Judge of the Supreme Court, so could the same part play both the role of grounding some causal roles and of grounding qualitative features. I leave the task of fleshing out this argument and the discussion of potential objections to a separate work.

For now, let us focus on the implications of the distinct nature strategy. If someone does not adopt **Identity**, then **Essential Dispositionality** and **Essential Qualitativity** do not entail each other. Therefore, someone could endorse one thesis while rejecting the other. For instance, the pure powers theorist can accept **Essential Dispositionality** while denying **Essential Qualitativity**. Presumably, the categoricalist embraces **Essential Qualitativity** while rejecting **Essential Dispositionality** (e.g., Lewis 1986; Armstrong 1997). For example, the categoricalist would say that by virtue of parts that belong to the property of *having a determinate charge*, an electron has some qualitative feature such as that of having a certain quantity of charge that can be measured in coulombs. In contrast, the electron's dispositionality would obtain in virtue of something distinct from the property itself. For example, it could hold in virtue of some laws of nature (e.g., Armstrong 1997).

The above considerations vindicate the claim that pure powers and powerful qualities are essentially distinct. Only the identity theory appears to be committed to *both* **Essential Dispositionality** and **Essential Qualitativity**. In contrast, the pure powers view appears to be committed *only* to **Essential Dispositionality.** Such a difference between the pure powers view and the identity theory strongly suggests that pure powers and powerful qualities have different natures: every part of a powerful quality grounds both causal roles and qualitative features; by contrast, every part of a pure power grounds causal roles only.

Before we move on, however, it is important to emphasise that the failure of the collapse between pure powers and powerful qualities does not imply that Taylor's (2018) considerations are wholly incorrect. He is right in thinking that the pure powers view and the identity theory share *some commitments* about the ontology of fundamental properties. The mistake, if I am right, is to infer the collapse between these views from these shared commitments.

Against the distinct nature strategy, someone might argue that the denial of **Essential Qualitativity** undermines the **Actuality** of pure powers (Taylor forthcoming raises a similar objection against Giannotti's dual-aspect account). Therefore, this approach would render the pure powers view implausible.

Here is one way to spell out this objection: if the nature of a pure power does not contribute to the qualitativity of an object that instantiates it, then nothing secures the reality of such a pure power. For instance, an opponent could argue that if *having a determinate charge* were not to contribute to an electron's qualitative make-up, then nothing would ground the actuality and reality of this property. The threat would extend to every putative fundamental power.

An easy way out would be to embrace **Essential Qualitativity**. Accordingly, it would be part of the nature of fundamental pure powers to ground some qualitative features had by a bearer, thereby securing the reality of such powers. Unfortunately, this option opens the door to the collapse. If **Essential Qualitativity** were true of the pure powers view, then the distinct nature strategy would fail; the pure powers view would indeed coincide with the identity theory.

The previous objection barks but does not bite. Surely, the pure powers theorist must safeguard **Actuality**. However, it is a mistake to think that **Essential Qualitativity** is the *only* way to do so. Recall that **Essential Qualitativity** is a claim about the qualitative features that are grounded by some parts of a property. Of course, if one conceives of the qualitative features of an object as a matter of its actual make-up, then **Essential Qualitativity** ensures the actuality of the relevant properties. However, on the proposed characterisation, **Actuality** and **Essential Qualitativity** come apart. The former is a claim about the possession of powers; the latter concerns their nature. The possibility of holding **Actuality** without **Essential Qualitativity** is good news for the pure powers theorist who wishes to escape the collapse objection once and for all.

One promising option to ground **Actuality**, which escapes the collapse objection, is to argue that it is the *possession* of a power by a bearer that makes it actual and real. Differently put, what grounds the reality and actuality of a pure power is the fact that it is possessed by some object. This approach captures what George Molnar says by claiming that powers are actual and real *in the sense* that "having a power is … having an actual property" (2003, 99). Thus **Actuality** should not be confused with a claim about the nature of pure powers.

Now I turn to conclude by pointing out some positive implications of the distinct nature strategy for both the pure powers view and the identity theory.

8. Divide et Impera

It goes without saying that by resisting the collapse objection, we make the dispute between the pure powers view and the identity theory substantive again. However, both positions also enjoy less obvious merits and drawbacks that concern their opposite stance on **Essential Qualitativity**. In what follows, I will point out an issue that the pure powers theorist can *prima facie* escape by rejecting **Essential Qualitativity**. To level the playing field, I will then consider an objection against pure powers that the identity theorist can *prima facie* avoid by embracing **Essential Qualitativity**. I do not aim to adjudicate which views handle these objections better. Nor is my aim to establish that these issues fatally wound either position. Rather, my purpose is to show that the ontological distinction between pure powers and powerful qualities has important consequences regarding the advantages of either doctrine.

The pure powers view is often introduced as a form of anti-quidditism (e.g., Bird 2007, 70–79). Since there is no univocal understanding of what quidditism is, the opposition can be construed in a variety of ways. Here I shall not attempt to reconstruct the debate surrounding this notion. Others have already done so meticulously (e.g., Locke 2012, Smith 2016, Wang 2016). Instead, let us consider again moderately austere quidditism (Section 5)—according to which fundamental properties are individuated by a *qualitative* nature (Smith 2016, 250). This view allows for causally indistinguishable possible worlds that differ just by a permutation or replacement of fundamental properties. For example, there can be possible worlds where the causal roles played by our worldly *charge* and *mass* are the same and yet the properties that play such roles are swapped.[7] Pure powers theorists argue that we should block possibilities of this sort.

The rejection of **Essential Qualitativity** secures this result. Fundamental pure powers lack qualitative parts. So, they also lack qualitatively quidditistic aspects which could individuate them. Pure powers are individuated by their essential causal roles which are, on the proposed framework, grounded in their dispositional parts. By contrast, the identity theory faces an odd consequence. Under the assumption that the quidditistic nature of a property is a matter of its qualitativity, powerful qualities would turn out to have quidditistic parts because of **Essential Qualitativity**. Perhaps shockingly, the identity theory would emerge as a form of quiddistim. In fairness, it is worth noting that the identity theory would block worrisome scenarios where causal roles are swapped (the identity between qualitative and dispositional parts would prevent the swapping). However, the worry that there is something odd about this upshot remains: typically, quidditistic features are supposed to be non-dispositional (Smith 2016, 252). Perhaps the identity theorist could insist that the felt sense of oddity is a remnant of an ill-conceived attachment to the idea that dispositionality and qualitativity are mutually exclusive. Be that as it may, facing the odd consequence is a drawback that the pure powers view easily avoids.

The rejection of **Essential Qualitativity** is not without problems, however. For example, E. J. Lowe argues that an ontology of nothing but pure powers cannot fix the identity of fundamental properties (2010, 12–14; 2012, 217–228). Lowe's objection is as follows. What individuates, and therefore identifies, pure powers is their causal roles. But from the

[7] Not all form of moderately austere quidditism entail this possibility. For example, Smith's (2016) non-recombinatorial version imposes necessary connections between qualitative natures and their causal roles, thereby blocking swapped scenarios akin to the one illustrated here.

viewpoint of an ontology of nothing but pure powers, the causal roles involve other pure powers. In turn, these will be individuated by some causal roles which involve further pure powers. And on it goes. The problem, Lowe claims, is that pure powers cannot get their identity fixed if they owe it to other ones. According to Lowe, the metaphysics of pure powers lack the resources for accommodating suitable *individuators* of properties, which ought to be qualitative features. Of course, Lowe's objection as well as his claim about qualitative individuators can be challenged. But suppose, for the sake of argument, that Lowe is right. Does the identity theory suffer the same problem?

It does not seem so. By embracing **Essential Qualitativity**, the identity theorist can locate Lowe's individuators in the qualitative features that bearers of fundamental powerful qualities have by virtue of instantiating them. These qualitative features do not involve the acquisition of pure powers. Therefore, the inadmissible regress of identity is blocked from the get-go. At least in principle, the identity theory does have the resources for meeting Lowe's challenge. Since the pure powers theorist cannot pursue the same strategy, the identity theorist can claim an advantage.

A lot more could be said about the previous objections and how to address them. But my goal is not to adjudicate a winner between pure powers and powerful qualities. The lesson here is that there are objections that target only one view but not the other. An explanation of this fact, if I am right, lies in the different commitments about the metaphysics of fundamental properties, particularly concerning **Identity** and **Essential Qualitativity**, that these views endorse. The discrepancy is beneficial. If one of the above objections were to be lethal for one view, the other could still be available. Overall, the possibility of retaining a robust ontological distinction between pure powers and powerful qualities is advantageous for the advocates of both theories.

Acknowledgements

Thanks to María Pía Méndez Mateluna, Rani Anjum, Lorenzo Azzano, Donatella Donati, Toby Friend, Giacomo Giannini, Stephen Mumford, Andrea Raimondi, Matthew Tugby, Barbara Vetter, Lisa Vogt, Neil E. Williams, a set of anonymous reviewers for helpful discussion on various parts of this manuscript, and to the audience of the New Foundations of Dispositionalism conference in Exeter.

REFERENCES

Anjum, Rani Lill, and Stephen Mumford. 2011. *Getting Causes from Powers*. Oxford: Oxford University Press.

Armstrong, David M. (1978). *Universals and Scientific Realism: A Theory of Universals Volume 2*. Cambridge: Cambridge University Press.

Armstrong, David M. 1989. *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.

Armstrong, David M. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.

Armstrong, David. M. 2005. 'Four Disputes about Properties'. *Synthese* 144: 309–320. https://doi.org/10.1007/s11229-005-5852-7

Bird, Alexander. 2007a. *Nature's Metaphysics*. New York: Oxford University Press.

Bird, Alexander. 2007b. 'The Regress of Pure Powers?'. *The Philosophical Quarterly* 57(229): 513–534. https://doi.org/10.1111/j.1467-9213.2007.507.x

Bird, A. 2016. 'Overpowering: How the Powers Ontology Has Overreached Itself'. *Mind* 125(498): 341–383. https://doi.org/10.1093/mind/fzv207

Ellis, Brian. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.

Ellis, Brian. 2002. *The Philosophy of Nature*. Chesham: Acumen.

Ellis, Brian. 2012. 'The Categorical Dimensions of the Causal Powers'. In *Properties, Powers and Structures*, edited by Alexander Bird, Brian Ellis and Howard Sankey, 11–27. New York: Routledge.

Ellis, Brian and Caroline Lierse. 1994. 'Dispositional Essentialism'. *Australasian Journal of Philosophy* 72(1): 27–45. https://doi.org/10.1080/00048409412345861

Giannotti, Joaquim. 2019. 'The Identity Theory of Powers Revised'. *Erkenntnis*: 1–19. https://doi.org/10.1007/s10670-019-00122-5

Heil, John. 2003. *From an Ontological Point of View*. Oxford: Oxford University Press.

Heil, John. 2010. 'Powerful Qualities'. In *The Metaphysics of Powers: Their Grounding and their Manifestations*, edited by Anna Marmodoro, 58–73. New York: Routledge.

Heil, John. 2012. *The Universe as We Find It*. Oxford: Oxford University Press.

Lewis. David K. 1983. 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy* 61(4): 343–377. https://doi.org/10.1080/00048408312341131

Lewis, David K. 1986. *On the Plurality of Worlds*. Oxford: Blackwell

Lewis, David K. 2009. 'Ramseyan humility'. In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 203–222. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/9780262012560.003.0009

Locke, Dustin. 2012. 'Quidditism Without Quiddities'. *Philosophical Studies* 160: 345–363. https://doi.org/10.1007/s11098-011-9722-5

Lowe, E. Jonathan. 2010. 'On the Individuation of Powers'. In A. In *The Metaphysics of Powers: Their Grounding and their Manifestations*, edited by Anna Marmodoro, 8–26. New York: Routledge. https://doi.org/10.4324/9780203851289

Lowe, E. Jonathan. 2012. 'Asymmetrical Dependence in Individuation'. In *Metaphysical Grounding: Understanding the Structure of Reality,* edited by Fabrice Correia and Benjamin Schnieder, 214–234. Cambridge: Cambridge University Press. https://doi.org/10.1093/analys/anu046

Ingthorsson, Rögnvaldur. 2013. Properties: Qualities, Powers, or Both? *Dialéctica* 67(1): 55–80. https://doi.org/10.1111/1746-8361.12011

Martin, Charles B. 1993. The Need of Ontology: Some Choices. *Philosophy: The Journal of the Royal Institute of Philosophy* 68(266): 502–522. https://doi.org/10.1017/s0031819100041863

Martin, Charles B. 2008. *The Mind in Nature*. Oxford: Oxford University Press.

Martin, Charles B., and Heil, John. 1999. The Ontological Turn. *Midwest Studies in Philosophy* 23(1): 34–60. https://doi.org/10.1111/1475-4975.00003

Molnar, George. 2003. *Powers*. New York: Oxford University Press.

Mumford, Stephen. 2004. *Laws in Nature*. London: Routledge.

Schaffer, Jonathan. 2005. 'Quiddistic Knowledge'. *Philosophical Studies* 123: 1–32. https://doi.org/10.1007/s11098-004-5221-2

Smith, Deborah. 2016. 'Quid Quidditism Est?'. *Erkenntnis* 81(2): 237–257. https://doi.org/10.1007/s10670-015-9737-y

Strawson, Galen. 2008. The Identity of the Categorical and the Dispositional. *Analysis* 68(4): 271–282. https://doi.org/10.1093/analys/68.4.271

Taylor, Henry 2018. Powerful Qualities and Pure Powers. *Philosophical Studies* 175(6): 1423–1440. https://doi.org/10.1007/s11098-017-0918-1

Taylor, Henry. Forthcoming. Powerful Problems for Powerful Qualities. *Erkenntnis*. https://doi.org/10.1007/s10670-019-00199-y

Tugby, Matthew. 2012. Rescuing Dispositionalism from the Ultimate Problem: Reply to Barker and Smart. *Analysis* 72 (4): 723–731. https://doi.org/10.1093/analys/ans112

Yates, David 2012. 'The Essence of Dispositional Essentialism'. *Philosophy and Phenomenological Research* 87(1): 93–28. https://doi.org/10.1111/j.1933-1592.2011.00568.x

Wang, Jennifer 2016. 'The Nature of Properties: Causal Essentialism and Quidditism'. *Philosophy Compass* 11 (3): 168–176. https://doi.org/10.1111/phc3.12307

Williams, Neil E. 2019. *The Powers Metaphysics*. New York: Oxford University Press.

EuJAP | Vol. 17 | No. 1 | 2021
UDC: 179:616-036.8
https://doi.org/10.31820/ejap.17.1.4

ACTS THAT KILL AND ACTS THAT DO NOT — A PHILOSOPHICAL ANALYSIS OF THE DEAD DONOR RULE

Cheng-Chih Tsai[1]

[1] Center for Holistic Education,
MacKay Medical College

Original scientific article – Received: 12/10/2020 Accepted: 15/05/2021

ABSTRACT

In response to recent debates on the need to abandon the Dead Donor Rule (DDR) to facilitate vital-organ transplantation, I claim that, through a detailed philosophical analysis of the Uniform Determination of Death Act (UDDA) and the DDR, some acts that seem to violate DDR in fact do not, thus DDR can be upheld. The paper consists of two parts. First, standard apparatuses of the philosophy of language, such as sense, referent, truth condition, and definite description are employed to show that there exists an internally consistent and coherent interpretation of UDDA which resolves the Reduction Problem and the Ambiguity Problem that allegedly threaten the UDDA framework, and as a corollary, the practice of Donation after the Circulatory Determination of Death (DCDD) does not violate DDR. Second, an interpretation of the DDR, termed 'No Hastening Death Rule' (NHDR), is formulated so that, given that autonomy and non-maleficence principles are observed, the waiting time for organ procurement can be further shortened without DDR being violated.

Keywords: *DDR; UDDA; DCDD; NHDR; vital organ; causation*

Correspondence: cctsai@mmc.edu.tw

Introduction

In the practices of vital-organ transplantation, while doctors typically want to procure a vital-organ as early as possible, the Dead Donor Rule (DDR) requires them to wait till the donor is dead, for otherwise the procurement would constitute, presumably, an act of killing. For some authors (cf. Veatch 2008), this amounts to the impossibility of a lawful vital-organ transplantation. Commenting on this situation, Robert Truog maintains that current practices in organ procurement do cause the death of the person if death is understood in the 'scientific way',[1] and claims that while the long-term solution to this problem should be to reframe the ethics of vital-organ donation in terms of the principle of respect for autonomy and the principle of non-maleficence rather than the DDR, the short term solution is to "conceptualize current approaches to defining death as socially acceptable 'legal fictions'"[2] (Truog 2015, 1885). Walter Sinnott-Armstrong and Franklin Miller (2013), on the other hand, offer a more radical solution by showing that there is nothing wrong with killing *per se*, hence DDR can be safely dropped.

The present paper proposes an alternative way out. By resorting to standard apparatuses in the philosophy of language and putting DDR and other relevant regulations or practices, such as UDDA and DCDD, under scrutiny,[3] I show that there is an interpretation of UDDA that captures a certain aspect of our intuition about the death of a person,[4] and, with respect to which, current practices of DCDD do not violate DDR. Hence there is neither a short-term need to regard UDDA as merely creating legal fictions nor a long-term need to abandon DDR. More specifically, I discuss two conceptual problems which, with DDR upheld, seem to threaten the present definition of UDDA and current practices of organ transplantation,[5] and resolve them in philosophical terms. Then I go one step further to formulate NHDR (No Hastening Death Rule), a version of DDR, which I claim to capture the spirit of donor protection better than the DDR taken at face value, and with NHDR, the waiting time for an organ procurement can be further shortened.

[1] The scientific standard mentioned in Truog (2015), attributed to Bernat, defines death as "the permanent cessation of functioning of the organism as a whole" (Truog 2015, 1892).

[2] By regarding a current definition of death as merely a 'legal fiction', one is reluctant to accept that the definition has captured the true notion of death.

[3] These are the abbreviations for the 'Uniform Determination of Death Act', and 'Donation after the Circulatory Determination of Death' respectively.

[4] Shewmon (2004; 2010) nicely demonstrates the intrinsic difficulties in obtaining a uniform definition of death.

[5] See Veatch (2008; 2010) for example. Note that many authors have proposed that DDR should be revised or dropped, see, for instance, Miller, Truog, and Brock (2010) and Sinnott-Armstrong and Miller (2013).

1. Upholding DDR in the Face of UDDA

By the Dead Donor Rule (DDR), I mean the following:

> **DDR** A vital organ *H* of a person *A* can be procured for donation at time *t* only if *A* is already dead at *t*.

Here, I should clarify what I mean by a 'vital organ' first. A 'vital organ' can mean a *type* of organs the removal of which would generally lead to the death of the owner. Heart, for example, is a vital organ in this sense, while appendix is not. However, a 'vital organ' can also denote a *specific* organ of a person the removal of which would lead to the person's death. Although heart is a vital organ in the first sense, a specific heart might not be *vital* in the second sense if its owner will be blown to pieces by a bomb a split second later or if its owner is currently receiving a new heart through a heart transplantation—after all, the removal of the (old) heart would not in any way hasten his death. To avoid further confusion, I shall refer to a vital organ in the first sense by a 'vital-organ', and reserve the term 'vital organ' for a vital organ in the second sense, and throughout Section 1, we shall only understand the 'vital organ' in DDR in the first sense, namely, as "vital-organ". In other words, in Section 1, we are concerned with

> [DDR$_1$] A vital-organ *H* of a person *A* can be procured for donation at time *t* only if *A* is already dead at *t*.

In general, interpreting DDR in the first sense would yield us a rule which is more strict, because, according to it, insofar as an organ is a vital-organ, it can only be procured after the owner is dead, even if the organ is actually not *vital* for that person.[6] In this section, issues about the UDDA-DDR framework and the practices of vital-organ procurement in the US will be formulated as two conceptual problems, which will then be settled by linguistic and philosophical means.

1.1 The Reduction Problem—How can Death Amount to Brain Death?

In Truog (2007), it says that in 2005, Dr. Sanjay Gupta, a neurosurgeon and Senior Medical Correspondence for CNN, told Larry King, "Well, you know, a dead person really means that the heart is no longer beating [...] people do draw a distinction between brain dead and dead" (Truog 2007,

[6] Certainly, in rare cases, a non-vital-organ can happen to be *vital* for a particular person as well, whose procurement will be blocked by the DDR in the second sense but allowed by the DDR in the first sense. But, for simplicity, we shall ignore such cases in this paper.

274), which amounts to publicly disagreeing with current medical and legal criteria of death. Indeed, some authors, such as D. Alan Shewmon and Robert D. Truog, have tried and succeeded in convincing key figures in medical ethics, including some members of the President's Council on Bioethics,[7] to the extent that the Council admits that "[…] it would be difficult to deny that the body of a patient with total brain failure can still be alive, at least in some cases" (Miller and Truog 2011, 72).

In this subsection, we discuss whether it is justifiable to define the death of a person in terms of the 'death' of one of his or her organs. Now, the death statement 'John is dead' clearly cannot be defined by 'The brain of John is dead'. The subject of the former sentence, namely 'John', is a proper name referring to a person, while that of the latter, namely 'the brain of John', is a definite description denoting an organ, yet they share the common predicate 'dead'. Apparently, we cannot define the death of something in terms of the death of some other thing if we do not know the extension of 'death' in the first place.

A natural solution to the above problem is to stress that the phrase 'brain-death' itself by no means suggests that the death of a person is characterized by the 'death' of his brain; it only suggests that the death of a person is to be determined by some condition of his brain. For example, using '*b*-death' instead of 'death' for the brain condition in question is a way out. To avoid future confusion, let me introduce new symbols to stand for some predicates that concern us in this paper.

- 'John is D' stands for 'John is dead'.
- 'John is D_b' stands for 'John is brain-dead'.
- 'The brain is ${}_bD$' stands for 'the brain is *b*-dead'.

The brain-dead definition of death can then be summarized as follow:

> John is D if John is D_b, and John is D_b if the brain of John is ${}_bD$. (*)

While the subject of a sentence of the form 'John is D' is still a proper name that refers to a person, and the criterion for the person's death is still expressed in terms of a sentence whose subject is a definite description, namely 'the brain of John', which denotes the brain of the person, the latter sentence is no longer a statement about the death of a "person", and the predicate is ${}_bD$ instead of D or D_b.

[7] See *President's Council on Bioethics: Controversies in the Determination of Death* 2008.

Note that regardless of what the content of $_bD$ truly is, the presence of a definite description in (*) alone generates a semantic issue concerning personal death. As 'John is *D*' is about John, while 'the brain of John is $_bD$' is about John's brain, it is unlikely that the two sentences can be synonymous.[8] Intuitively, if John loses his entire brain, then we would say that John is dead. Nevertheless, the $_bD$ statement about John's brain now becomes either truth-valueless (analogous to the claim that 'the King of France is bald' makes no statement when there is simply no King of France at present) or false (if Russell's theory of definite description is to be adopted).[9] So, the practice of defining the death of a person in terms of certain property of some part of the person's body seems problematic. Following the spirit of a Strawson/Wolfram framework[10] which regards 'the King of France is bald' as making no statement, one can claim that if John has lost his brain then a sentence token 'John is *D*' is truth-valueless. On the other hand, according to the Russellian framework, if John has no brain then he cannot be *D*, which is even more absurd. Imagine trying to complete a sentence that begins with "Pew is blind if and only if the eyes of Pew are …", while soon reckoning that it is possible that Blind Pew simply has no eyes.

Elbourne claims that for certain sentences in which definite descriptions are embedded under propositional attitude verbs and conditionals, the Fregean analysis of definite descriptions is superior to the Russellian analysis. For example, "Hans wants the ghost in his attic to be quiet tonight" (Elbourne 2010, 8) does not entail that Hans wants that there exists exactly one ghost in his attic[11] (the Russellian way). Rather, it *presupposes* the existence of exactly one ghost in his attic (the Fregean way). Similarly, "If the ghost in his attic is quiet tonight, Hans will hold a party" is not to be rephrased as "If there is exactly one ghost in his attic and it is quiet tonight, Hans will hold a party" (Elbourne 2010, 2). Again, it *presupposes* the existence of exactly one ghost in his attic.

This framework helps us to better analyze the problem that I raised two paragraphs back. Analogous to Elbourne's analysis,[12] the sentence 'John is *D* if and only if the brain of John is $_bD$' is *not* to be translated as 'John is

[8] In the sense that the truth condition of one is governed by the other.

[9] According to Russell's theory, for the *b*-death statement to be true, John has to have a brain to start with. More specifically, *the F* is *Q* if and only if (i) there is an *x* such that *Fx*, (ii) for all *y*, if *Fy* then *x*=*y*, and (iii) *Qx*.

[10] See Wolfram (1989).

[11] For simplicity, here I assume that 'Hans wants *A* and *B*' implies 'Hans wants *A*'.

[12] Note that Elbourne's analysis is primarily for embedded statements, but we find an analogous phenomenon here.

D if and only if there is exactly one brain of John and it is ${}_{b}D$'. John's having a brain is no longer a necessary condition for his death. Rather, (*) only *presupposes* the existence of a brain of John and when John has lost his brain, (*) is no longer applicable. What can we say about John's death if at the instance of his death he does not have a brain?[13]

Recall that the Fregean account of sense and reference tells us that the sense of 'John' in 'John walks' determines the referent [John]$_w$ of 'John' when the sentence is tokened in world *w*, and the token is true provided that [John]$_w$ lies in [walks]$_w$, the set of all things that walk. Would such a mechanism work for a death statement of the form 'John is dead' as well? At the most abstract level, it would still work, but in practice it does not. If John is blown into pieces by a bomb, then there simply is no entity left in the world that can be said to be the 'referent' of 'John', but we can still claim that he is dead. Presumably, outlining the truth condition of a death statement without assuming that 'John' refers is a better approach. According to this approach, we only need to resort to something denoted by the definite description 'the brain of John' and see if it lies in the extension of the predicate ${}_{b}D$. More specifically, the no-brain (or no-body) problem mentioned earlier can be resolved by 1) taking the definite description 'the brain of John' as *presupposing* the existence of a unique referent rather than *asserting* its existence, and 2) in case John has lost his brain (or his entire body), we simply stipulate that he is dead because his brain no longer exists—in other words, 'the brain of John is ${}_{b}D$' vacuously holds. So, in contrast to the Russellian account, the brain of John is ${}_{b}D$ if and only if John has exactly one brain and the brain meets the criteria associated with ${}_{b}D$ or John no longer has a brain. Specifically, when John has lost his brain, the sentence 'John is *D*' is *false*, *truth-valueless*, and *true* according to the Russellian account, the Fregean account, and the present account, respectively.

In sum, despite that the subject of 'John is dead' is a proper name for a person, the truth condition of the statement can be described by another sentence whose subject is a definite description denoting one specific organ of John. Here the definite description itself is to be interpreted more in the Fregean than in the Russellian way. However, when the definite description fails to denote, the death statement will still have a definite truth value, rather than remains undecided.

[13] Strictly speaking, when John has gotten two brains, (*) is inapplicable also.

1.2 The Ambiguity Problem—Do We Have Two Distinct Notions of Death?

In 1981, The Uniform Determination of Death Act (UDDA) was approved as a model state law for the United States. It states that an individual who has sustained either (1) irreversible cessation of circulatory and respiratory functions, or (2) irreversible cessation of all functions of the entire brain, including the brain stem, is dead. Furthermore, a determination of death must be made in accordance with accepted medical standards.

A general concern now arises.

(i) The *death* + *death* = *life* Problem

If UDDA is understood as a definition, which defines D as the disjunction of D_h and D_b, where D_h is a short-hand for the state of a person who meets (1), while D_b is a short-hand for the state of a person who meets (2), then we immediately encounter the alleged *death* + *death* = *life* problem. To be more precise. Let A be one that is D_h already but not yet D_b, and B be one that is D_b already but not yet D_h. According to UDDA, they are both dead, but given that the transplantation option is available, can we not make use of *A*'s brain and *B*'s heart-lung system and build a living being from two dead persons?

This is not as problematic as it sounds. Imagine that a certain creature, Two-Eye say, is composed of a left eye and a right eye. A Two-Eye is impaired if at least one of its eyes is broken. Then if recombination is possible, we can surely expect to get a non-impaired Two-Eye from a pair of impaired Two-Eyes. After all, the problem is with personal identity rather than with life-and-death.

If neither the brain nor the heart-lung system is essential[14]—that is, they are replaceable—then the resulting living individual should neither be A nor be B, as they are both dead already, and dead people are not expected to come back to life. But as the Two-Eye case demonstrates, there is nothing odd here at all.

If the brain is the essential part of a person, however, then the resulting living individual should of course be A. But, isn't A already dead by UDDA? How come he/she comes back to life after the transplantations? Shouldn't this prove that UDDA is problematic? Not really! The point is that if clause (1) is to be considered as a sufficient condition for death, and

[14] Here by 'essential', I mean essential of for personal identity, not essential for life.

it involves irreversibility in its terms, then to announce the death of *A* before the transplantation, we should have thought of the possibility of transplantation. Given that after the transplantation, *A* is apparently alive, we should realize that the prior announcement of the death had been premature—*A*'s circulatory and respiratory functions were not yet "irreversibly-ceased" in the first place. So, there is no *death + death = life* problem for us to worry about here.

Alternatively, we can regard UDDA as merely listing two criteria of death, which together characterize death, rather than regarding each of clause (1) and clause (2) as semantically capturing the essence of death. For example, if we follow the idea that what is essential for a person is his/her brain, and what is essential for a person's life is his/her brain function, then UDDA amounts to characterizing the *brain* condition of a person through two criteria which are pragmatically, rather than semantically, related to the underlying condition of the brain.

For brevity and clarity, I will introduce the following abbreviations. If clause (1) of UDDA is met, I shall say that the person's brain is ${}_{b}D_{1}$, and if clause (2) of UDDA is met, I shall say that the person's brain is ${}_{b}D_{2}$. UDDA recognizes the irreversible cessation of all functions of the entire brain, i.e. ${}_{b}D_{2}$, as a criterion for death, which allows doctors to procure vital-organs from D_b (brain-dead) victims without violating the Dead Donor Rule. However, with the acceptance of UDDA, Donation after the Circulatory Determination of Death (DCDD) can be a protocol for vital-organ donation as well: the life support equipment for severely brain damaged patients are removed until the patients meet the traditional circulatory and respiratory criteria for death and then the organs are removed.

Now, the listing of these two distinct criteria for death in UDDA makes the concept of death sound ambiguous and compromising. In particular, on the face of it, one can be dead without her brain being ${}_{b}D_{2}$—meeting the criterion of ${}_{b}D_{1}$ suffices—which seemingly contradicts the guiding principle that *D* (death) is characterized by D_b (brain-death), which in turn is the main driving force of UDDA's coming to being.

More specifically, we can imagine that[15] (a) while John's brain is still functioning, some foreign creature rips his heart and lung out from his chest in an instant, and, Dr. Who, who is obsessed with brain research, happily takes this chance and pronounces John dead based on the fact that John's circulatory and respiratory functions have irreversibly ceased, and immediately procures John's brain for research. Intuitively we would

[15] See Lizza (2011) for a treatment of this problem through the example of decapitation.

expect that as John's brain can still work for a split second after his heart and lung are ripped away, he is not dead yet and Dr. Who is doing harm to a living person rather than merely manipulating the corpus of a dead man. Analogously, we can imagine that (b) John's brain has met the $_{b}D_{2}$ dead criterion but his heart is still beating strongly for unknown reasons. Some people may find it difficult, as Dr. Gupta did, to suppress the intuition that John is not dead yet,[16] despite that John has met the criteria of death prescribed by UDDA.

A natural reaction to the above objections would be to reformulate UDDA. However, UDDA is the outcome of the collective wisdom of many individuals, and has been in use for several decades. So, despite that there are issues that need to be dealt with more carefully—especially the conceptual ones like the ambiguity problem I have just mentioned—insofar as these issues can be adequately explained in linguistic/philosophical terms, our priority should be in keeping it rather than altering it. Now, as we have seen in the previous subsection, while a death statement concerns a person, its truth condition only resorts to some particular organ of the person. So, there can be no ambiguity problem in UDDA at all. Clause (1) and clause (2) collectively characterize the condition of death for an individual—in our terms, John is D iff John's brain is $_{b}D$, and John's brain is $_{b}D$ iff John's brain is either $_{b}D_{1}$ or $_{b}D_{2}$—and so long as one of the clauses holds, the person is dead. In other words, 'John is still alive' amounts to the conjunction of two clauses.

With the help of these two notions of $_{b}D_{1}$ and $_{b}D_{2}$, we can observe that, our daily uses of the term 'brain-death' can actually mean two different things. It can mean either that John's brain is $_{b}D$ or that John's brain is $_{b}D_{2}$. However, please bear in mind that in this paper the term 'brain-death' is reserved for the first reading only.

Now, recall that scenario (a) seems more disturbing than scenario (b). So far as scenario (b) is concerned, nowadays many surgeons have been practicing the procurement of a beating heart from a person whose brain meets the $_{b}D_{2}$ criterion, without feeling that the donor is still alive. It appears that the concept of the death of a person can be a constructed idea that we can gradually adapt to. Can one's uneasiness towards scenario (a) be similarly resolved in the future? The answer is probably 'no'.

[16] Shewmon, Truog, and a minority of the President's Council would be reluctant to accept that John is already dead as well. See for example, Brugger (2013).

In defining death, many authors have attempted to resort to higher brain death,[17] or the so called 'cerebral death', instead of the whole brain death. It reflects the fact that John Locke's psychological continuity account of personal identity is not something that easily fades. On the one hand, higher brain death proponents maintain that the whole brain death, or the ${}_{b}D_{2}$ criterion, is not necessary for the death of a person. On the other hand, they may suspect that the irreversible cessation of circulatory and respiratory functions, or the ${}_{b}D_{1}$ criterion, is not sufficient for the death of a person, as the higher brain may still be functional. The latter is precisely the concern that my scenario (a) tries to raise: is it possible that a person who is still conscious, hence alive, be mistakenly pronounced dead according to clause (1) of UDDA?

A brain's being ${}_{b}D$ is the disjunction of its being ${}_{b}D_{1}$ and its being ${}_{b}D_{2}$, thus the application condition of ${}_{b}D_{2}$ is, in practice, more strict than that for ${}_{b}D$—otherwise we would not need the other criterion, namely, ${}_{b}D_{1}$, in the first place. So, a brain can indeed be ${}_{b}D$ without being ${}_{b}D_{2}$, as the higher brain death proponents would have maintained. However, given that we only have two criteria of death, can the other criterion, namely ${}_{b}D_{1}$, truly capture the higher brain death so that the person in scenario (a) would not be mistakenly pronounced dead?

Note that the ${}_{b}D_{1}$ criterion, like the ${}_{b}D_{2}$ criterion, serves as an indication that the brain of the person is ${}_{b}D$.[18] However, the ${}_{b}D_{2}$ criterion normally takes longer to meet than the criterion for the ${}_{b}D_{1}$ criterion. In practice, the ${}_{b}D_{2}$ test is usually done when the heart-lung system is still working (with the help of an artificial life support system, if needed) so it involves a long period of waiting time before checking for brain activities for a second time, while the ${}_{b}D_{1}$ test is usually done when the heart-lung failure is imminent and it involves a waiting time of only several minutes. The hidden consensus here is that if the circulatory and respiratory functions have ceased for that amount of time, the brain would have been in the state of ${}_{b}D$ even though we have not gone through the usual ${}_{b}D_{2}$ test procedure for it.[19]

In sum, the cessation of circulatory-respiratory functions can indeed be seen as a sign that indicates that the brain of the individual in question is ${}_{b}D$ already. While we may not have the means to directly assess the condition of the brain, we may still pronounce the patient dead according

[17] See DeGrazia (2005) and McMahan (2002).

[18] In Bernat's words, "the circulatory criterion is valid only because it leads to the brain criterion" (Bernat 2013, 28).

[19] I do not claim that this notion of brain-death and the 'higher-brain-death' amount to the same thing. But it is certainly plausible.

to the ${}_{b}D_{1}$ criterion, because without the help of an artificial life support, the cessation of the circulatory-respiratory function is a sure sign that the ‘key’ brain functions would cease in several minutes if nothing is to be done about it. So, the list of two criteria of death in UDDA itself does not make the notion of death ambiguous. A truth condition may come with two criteria, and insofar as the criteria collectively shape the right concept, the listing of two criteria causes no harm.

Now, back to scenario (a). Taking the above into account, what can we say about it? In (a), the heart and lung of John are ripped away in an instant, so, *on the face of it*, his circulatory and respiratory functions have irreversibly ceased, thus he has met the first clause of UDDA and can be pronounced dead, which seems to contradict our intuition that he is not dead yet. However, remember that in this particular scenario, we *have not* gone through the waiting time of several minutes as required by accepted medical standards. So, we cannot say that the ${}_{b}D_{1}$ has been met or that the brain is already ${}_{b}D$. As the story has already suggested, John might still have a split second of consciousness left after his heart and lung were ripped away, thus the instant ripping away of John’s heart/lung does not entail his death right away—the usual several minutes of waiting is still needed for us to pronounce that the brain of John is ${}_{b}D_{1}$.

This indeed safeguards human life. Recall that all the issues concerning DDR, UDDA and DCDD etc. arise because of the possibility of vital-organ transplantation. Now, even though John’s original heart-lung system has been ripped away, there is still the possibility that, with the most advanced medical technology, a new heart-lung system can be transplanted into John’s chest and begin to function in less than a minute’s time, just before John’s brain is forever damaged. In other words, ‘the circulatory and respiratory functions of the heart-lung system of John’ *cannot* be said to be irreversibly lost at the time of ripping, because that definite description ‘the heart-lung system’ denotes a system of John that is in his chest (or somewhere nearby), regardless of whether it is the original one or a replacement. When we say that the president of America has always been male, we by no means mean that Joe Biden has always been a male. Rather, we mean that each president of America has been male to date. Analogously, to say that the heart-lung system of John has irreversibly ceased functioning, we should ensure that no possible replacement heart-lung system, such as an artificial heart-lung system, or a new heart-lung system with a transplanted heart, can succeed as ‘the heart-lung system of

John'[20] and function properly, and this necessitates the waiting of several minutes before the pronouncement of death. So scenario (a) amounts to a premature judgement of death. John was not dead yet immediately after the ripping away of his heart-lung system—his brain was neither ${}_bD_2$ nor ${}_bD_1$ yet, even though his *original* biological heart-lung system had indeed irreversibly ceased to function.

Another alleged problem related to the ambiguity problem is the following.

(ii) The *Reversing the Irreversible* Problem

While the Denver case of successful heart transplantation was hailed as a great medical achievement, Veatch draws our attention to the fact that the procurement of hearts from DCDD patients for organ donation seems to involve reversing the irreversible (Veatch 2008). Imagine that a critically ill patient John has chosen to forgo life-sustaining treatment and donate his heart. After the withdrawal of life-sustaining equipment, his heart stops and after several minutes of waiting, he is pronounced dead because his circulatory and respiratory functions are regarded as irreversibly lost, and his heart is procured and transplanted into the chest of another patient, Smith say, who was on ECMO and has been waiting for a new heart for some time. Smith lives well after the transplant, which implies that the new heart is beating well in his body. Now, according to Veatch's insight, a moment ago it was declared that the circulatory and respiratory functions of John's heart-lung system were irreversibly lost, and now the heart is beating again in another person's body, does this act of transplantation not amount to reversing the irreversible?

Veatch's point sounds convincing, and according to this view, the Denver doctors were guilty of procuring vital organs before the donor was dead, thus had violated the homicide law by killing the donor for his or her heart. But, as this act saved the lives of the organ recipients, we may choose to just muddle through. Or, alternatively, we can see the death of the donor as just a legal fiction: the donor is not *really* dead yet, but based on UDDA and current medical criteria for ${}_bD_1$, the donor is 'dead' already, even if the donor's heart actually beats nicely in another person's chest later. Possible solutions to Veatch's challenge, other than muddling through or regarding UDDA as creating legal fictions, include 1) deleting the first criterion of UDDA or disallowing the procurement of hearts from patients who seemingly have (but in fact have not) met the DCDD criterion hereafter,

[20] Note that here the *de dicto*, or small scope, reading of the definite description is the intended interpretation, otherwise vital organ transplantation would not be possible in the first place.

because these patients are not dead yet, and procuring their vital-organs violates the DDR; 2) not altering UDDA but simply dropping the DDR, thus allowing the procurement of vital-organs from patients who seemingly have (but in fact have not) met the DCDD criterion; 3) replacing the 'irreversibility' requirement of UDDA by 'permanence' so that, insofar as the procurement is performed, the person's circulatory and respiratory functions have ceased permanently, hence he is dead regardless of whether the organ is reversible.

We will not consider options 1) and 2) as they involve either dropping UDDA or abandoning DDR, and the goal of the paper is to show that they can be held without inconsistency. We will not accept 3) either, because it literally alters UDDA—namely, by replacing 'irreversible' by 'permanent'. Nevertheless, in Section 2, I will return to 3) and see it as a failed attempt to shortening the waiting time for a death pronouncement. In the meantime, I would only stress that with the help of a careful linguistic analysis of the predication of 'irreversible', we can show that there is no 'reversing the irreversible' involved in DCDD donation in the first place.

The following example prepares us for this point. Imagine that an alien creature needs two functional hearts, an *L*-heart and an *R*-heart, for it to be alive. So both hearts are vital for such creatures. Now, the function of an *L*-heart will be irreversibly lost after it stops beating for 4 minutes and the function of an *R*-heart will be irreversibly lost after it stops beating for 2 minutes. Furthermore, an *R*-heart will stop beating after its corresponding *L*-heart has stopped beating for 2 minutes, and similarly, an *L*-heart will stop beating after its corresponding *R*-heart has stopped beating for 2 minutes. So a creature can be pronounced dead after its *L*-heart has stopped beating for 4 minutes or after its *R*-heart has stopped beating for 2 minutes. Moreover, while an *L*-heart is transplantable, an *R*-heart is not. Now, suppose Alice is such a creature, and the function of her *L*-heart has stopped beating for 4 minutes. Can we transplant Alice's *L*-heart into Betty's body without violating the DDR? Yes, because Alice is already dead by the criteria set for these creatures, yet the transplanted *L*-heart still has a chance of beating again inside Betty's body. If the transplantation is successful, and Alice's original *L*-heart is now beating again in Betty's body, are we not reversing the irreversible? Surely not. What is irreversibly lost is the function of the *L*-heart *in Alice's body*—the stopping of the *L*-heart for 4 minutes has implied that the corresponding *R*-heart has stopped for 2 minutes, which in turn implies that its corresponding *L*-heart, *in Alice's body*, will not be functioning again. It says nothing about the function of the *L*-heart in Betty's body.

The fact that **the *L*-heart of Alice** is irreversible at the time of procurement and the fact that **the *L*-heart of Betty** is beating afterward[21] do not contradict each other, even though the former is spatial-temporally continuous with the latter (they are virtually the same heart). In conclusion, there is no ‘reversing the irreversible’ problem at all.

Now, back to our case concerning DCDD. The irreversibility of John’s circulatory and respiratory functions is, after all, a property of John rather than of a particular heart/lung system, while the beating again of the donated heart is merely a property of the heart, and these facts do not contradict each other. John’s failing to have his circulatory-respiratory functions re-established on site reveals the fact that he is *brain-dead* in the sense that his brain is *b*-dead, but that does not imply that his *former* heart-lung system cannot be functional in another person’s body.

To sum up this section, the criteria of death listed in UDDA help materialize the truth condition of a death statement, by drawing our attention to the conditions of some suitable organ of an individual. When it comes to the pinning down of the semantics of a death statement, all we need is a way of finding out whether some portion of a body can be considered as the brain of the individual, and whether the circulatory and respiratory functions of a heart-lung system or all functions of the entire brain are irreversibly lost. Finally, *brain-death* says more about the **non-existence of a functioning brain** than about the **existence of a non-functioning brain**.

2. Reinterpreting the Dead Donor Rule—The No Hastening Death Rule

In this section, we explore the possibility of understanding the ‘vital organ’ in DDR in the second sense, and show that with this interpretation, the waiting time for procurement can be further shortened without the rule being violated. In particular, we shall adopt Shewmon’s insight concerning the interpretation of DDR, develop it into a more workable version, and show that certain seemingly hasty procurements of vital-organs are not killing acts at all.

[21] Note that (the function of) Betty’s *L*-heart cannot be said to be irreversible before the transplantation. The *L*-heart may be not functioning before the transplantation, but with the transplantation option, *it* is not irreversible yet.

2.1 Causing Death as Hastening Death

In the Conclusion section of Shewmon (2004), we find:

> Regarding organ transplantation, the important and truly meaningful question is not ‘When is the patient dead?’ but rather ‘When can organs X, Y, Z … be removed without *causing* or *hastening death* or harming the patient in anyway?’ (Shewmon 2004, 297, Emphasis added)

More explicitly, in Shewmon, and Shewmon (2004, 110), we see:

> […] This approach to heart/lung retrieval does not *cause* or *hasten death*, because once circulation has effectively ceased due to the effect of progressive hypoxia on the heart, the dying or decaying process continues just the same regardless whether the nonbeating heart and nonfunctioning lungs remain physically in the circulationless body or not. (Shewmon, and Shewmon 2004, 110, emphasis added)

Note that in these passages, Shewmon wrote as if ‘causing death’ and ‘hastening death’ amount to roughly the same thing. However, we will see that, depending on how we conceive of causation, while the two notions can indeed be interchangeable if a specific but-for styled account of causation restricted to the causation of death is adopted, they can also mean radically different things according to other accounts, such as according to a version of NESS[22] which regards the death of *A* as some *event* incorporating all details of the way the death comes about.

For simplicity of treatment, I shall define the vitality of an organ in terms of ‘causing death’ first, and then identify ‘causing death’ with ‘hastening death’ as Shewmon seems to have suggested, only after a particular notion of death causing is subscribed later.

Recall that in this section we shall understand the ‘vital organ’ in DDR in the second sense. As a consequence, by Contraposition and some other elementary logical rules, DDR can be rephrased as

> [DDR$_2$] If *A* is not already dead at *t*, then an organ *H* of a person *A* can be procured for donation at time *t* only when *H* is not vital then.

[22] NESS is a short for Necessary Element in a Set of conditions Sufficient for the effect. See, for instance, Moore (2009) and references therein for how it works.

Now, defining vitality of an organ in terms of ‘causing death’, we have

> [Vitality] An organ *H* of *A* is *vital* at *t* if the procurement of *H* from *A* at *t* would cause *A*’s death,

As the DDR is clearly a rule concerning the living rather than the dead, and for a dead person no organ is vital, the antecedent of [DDR$_2$] can be dropped, and then [DDR$_2$] and [Vitality] can be combined into a single rule.

> [*] An organ *H* of a person *A* can be procured for donation at time *t* only if the procurement of *H* would not cause *A*’s death.

This is a decent rule. However, what exactly do we mean by ‘causing of a death’ here and what are the causal relata in question?[23] While this paper is no place for us to review a full range of accounts of causation and give the causation in question a particular theory-laden interpretation, we can at least consider two standard accounts of causation, namely the counterfactual account of David Lewis and the NESS account of Richard W. Wright, and see whether they are up to the job of characterizing the causation in [*].

Recall that the counterfactual account of causation faces the challenge of pre-emption. Take the famous Suzy and Billy throwing rock scenario for example. Suzy and Billy both threw a rock at a bottle, Suzy’s rock hit the bottle first and broke it. Intuition seems to suggest that Suzy’s throwing the rock is the cause of the breakage of the bottle. Nonetheless, according to the counterfactual theory of causation, had Suzy not thrown the rock, the bottle would have been broken by Billy’s rock, so Suzy’s throwing of the rock is not the cause. This is counter-intuitive.

The NESS account solves the problem by reckoning that while Suzy’ throwing a rock is not a but-for cause, it is indeed a necessary element of a set of conditions sufficient for the breakage of the bottle. So Suzy’s throwing is a NESS cause of the breakage. Nevertheless, according to this account Billy’s throwing a rock is a NESS cause as well. This is counter-intuitive too.

To solve the counter-intuitive conclusions mentioned above, the counterfactual theorist and NESS theorist often resort to the fact that by examining the way the bottle was broken to pieces one can establish that the underlying causation at work is Suzy’s rock breaking the bottle rather

[23] See Hall and Paul (2004) and the references therein.

than Billy's rock breaking the bottle. In other words, by fine-graining the effect, so that 'the breakage of the bottle' contains more details about how it is broken, both the but-for test and NESS test remain plausible accounts of causation.

Hereafter, I will apply a time-frame analysis to (1) the rock-throwing case, (2) a famous hypothetical of McLaughlin and (3) the organ procurement case, which is the primary concern of this paper, and show that there is a better strategy dealing with the causation of death than indefinitely fine-graining the effect.

We start with an analysis of a but-for statement. In saying that but for the procurement of *H*, *A* would not have died, we surely do not mean that but for the procurement of *H*, *A* would live forever. Rather, we seem to have a time frame such that at some time $\tilde{t}>t$, both the procurement of *H* and the death of *A* has happened, at t and t' respectively, with $t'>t$, and had the procurement of *H* not occurred, the death of *A* would occur at some other time $t''> \tilde{t}$. The problem here is that this reference time $\tilde{t}$ seems arbitrary—being the time the discourse takes place.

Applied to the rock-throwing case, the time frame can be illustrated as in Figure 1.

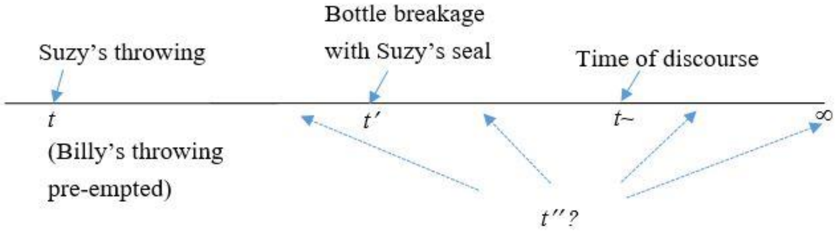

Figure 1. The Rock-Throwing Case

We have the following candidates for a but-for account based on the location of t''—the time of the breakage of the bottle had Suzy not thrown the rock.

1) [But-for 1] $t'' = \infty$ But for Suzy's throw, the bottle would never be broken
2) [But-for 2] $t'' > \tilde{t}$ But for Suzy's throw, the bottle would still be unbroken at the time of discourse.

3) [But-for 3] $t'' > t'$ But for Suzy's throw, the bottle would still be unbroken at t'.
4) [Fine-grained] The location of t'' is *irrelevant*, even $t'' < t'$ is acceptable. What matters is Suzy's signature/seal in the breakage.

Now, as the primary concern of this paper is not a general account of causation, we will be content with applying the above framework only to the causation of death. In that case, option 1) and 2) are to be ruled out right away as, first, even if the procurement had not been carried out, the death of patient would have been bound to happen at some later time, and, second, an organ donation case can be reviewed at any time and there is no apparent reason why the discourse time should play a role, despite that a but-for statement usually, on the face of it, takes the form of "… would not have happened".

Option 4) is more subtle and it seems to capture many, if not most, people's intuition. As long as a high-speed camera captures the detail of the breakage of the bottle and reveals that it's Suzy's rock that is involved in the physical process of the breakage of the bottle, then some would think it's Suzy's rock throwing that caused the breakage, even if had Billy's throwing not been pre-empted by Suzy's throwing, Billy's rock would have broken the bottle at a time t'' earlier than t'.

I would not try to challenge this intuition here, but would rather draw the reader's attention to the following scenario adopted from a famous hypothetical of McLaughlin, which is discussed in McLaughlin (1925) and Moore (2009). A man, Dan say, was to travel to the desert with a bottle of water. Before he set off, one rival of his, Bob say, added poison to the water, intending to kill him, while an hour later, another rival of his, Sue say, without knowing what Bob had done, emptied the bottle, also intending to kill Dan. Dan died of thirst in the desert in the end. Now, what was the cause of his death?

According to a coarse-grained but-for test, both rivals' acts *aren't* but-for causes for Dan's death, while according to a coarse-grained NESS test, both rivals' acts *are* NESS causes. Yet, according to a finer-grained but-for test, Sue's act is the cause of Dan's death by thirst, because but for Sue's act, Dan would not have died of thirst. On the other hand, according a finer-grained NESS test, Bob's act isn't the NESS cause of the death of Dan, because adding poison cannot be said to be a necessary element of a set of conditions sufficient for Dan's death by thirst. The scenario can be summed up in the following way (see Figure 2).

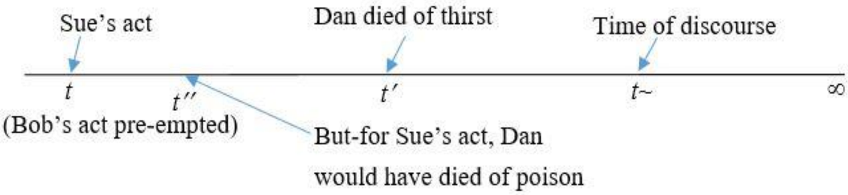

Figure 2. The Water Keg Case

Note that while, indeed, but for Sue's act, Dan would not have died of thirst, and Sue's act alone is a necessary element of a set of conditions sufficient for Dan's dying of thirst, we can consider the following twist of the story before asserting that Sue's act is the cause of Dan's death. Suppose, that Sue was not a rival of Dan, and she knowingly emptied the bottle to avoid Dan's being poisoned by Bob. Having no clean water to refill the bottle, Sue had done her best to *save/prolong* the life of Dan. It is simply ridiculous to say that her act is the cause of Dan's death.

Now, many theorists assume that actual causation is a factual causation[24], but, if that is true, then the fact that we are reluctant to deem Sue's act a cause of Dan's death after learning the mindset of Sue together with the fact that the twist of the story does not affect the underlying physical facts represented in the picture should prompt us to have a second thought about embracing 4).

Finally, back to our original context, I propose that, instead of embracing 4), we adopt 3), and the procurement of *H* *causes*, or more straightforwardly *hastens*, the death of *A* only when $t'' > t'$, as shown in Figure 3.

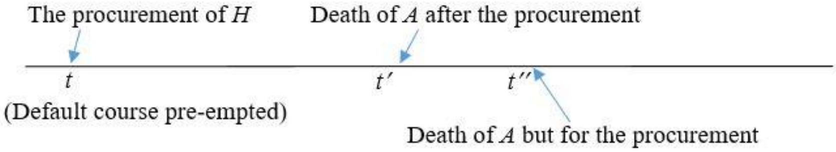
The procurement of H
Death of A after the procurement
t
t'
t''
(Default course pre-empted)
Death of A but for the procurement

Figure 3. A Death-Hastening Procurement

[24] See Moore (2009) for the repeated emphasis on this.

As a consequence of this account, if $t'' < t'$, then even if the death of *A* does bear the signature of procurement, we *cannot* say that the procurement of *H causes* the death of *A*, after all, it *prolongs* life rather than *hastens* death. We then arrive at a new version of DDR, which can be term NHDR (No Hastening Death Rule)

> **NHDR** If a person *A* is to donate his/her organ *H* then the procurement of *H* should not hasten *A*'s death.

Note that as vitality is understood as hastening death through procurement, NHDR amounts to the slogan: 'No Vital Organ Procurement!'. This partly explains why people, such as Veatch, are tempted to suspect that DDR cannot be consistently held in the practices of vital-organ transplantation. However, as we have repeatedly stressed, vital-organs are not necessarily vital organs, and a vital-organ *H* can by all means be non-vital for *A* yet becomes a vital organ for *B* after the transplantation. There is no problem with the slogan.

Before we look more closely at how the NHDR scheme works in the practice of organ procurements, we need to digress for a while to discuss an issue relevant to the re-formulation of DDR, so as to be better prepared for the analysis. Recall that we mentioned earlier that replacing 'irreversibility' by 'permanence' in the definition of death can shorten the waiting time for vital-organ procurements so that an organ can be procured well before it is damaged (Bernat 2013). However, this amounts to either changing UDDA or violating the DDR, because permanence does not imply irreversibility. After all, irreversibility is a modal property, which involves a set of possible worlds, but permanence only concerns the actual world. Irreversibility is a state of an entity which is characterized by its possible behaviors at various possible worlds, but permanence is not. To say that something has irreversibly lost some feature that it once exhibited, we need only to look at its current state and then, by consulting past statistics and predictions by experts, assert the irreversibility. But to say that the lost is permanent, we are talking about a four-dimensional continuum which constitutes the world line of the individual, therefore we can pass judgement without resorting to past statistics or future predictions about people in similar conditions. We simply need to check the whole continuum of an individual and find out whether the feature indeed never reappears. A patient whose heart has stopped but has not yet met the irreversibility criteria of UDDA—for example, the required several minutes' waiting time has not yet elapsed—may actually be 'permanently-dead' because no one attempted to resuscitate him. Therefore, permanence does not entail irreversibility.

On the other hand, it is imaginable that a patient who has been pronounced dead based on the irreversibility requirement of UDDA can be brought back to life by a miraculous divine action. Thus irreversibility[25] does not entail permanence either.

The advantage of the move to replace 'irreversibility' by 'permanence' in UDDA is of course that no doctor would be accused of procuring the heart from a heart stopping donor whose heart has not met the permanence requirement of death, because the doctor's act of procurement itself would guarantee that no heart would ever be beating again in the chest of the donor. But this move is in practice unacceptable, because it will allow an ER staff who is reluctant to perform CPR to a heart stopping patient to defend himself/herself by saying that "the heart-beat monitor be my witness, at the time of the patient's arrival, his heart has stopped permanently".

However, sticking to the irreversibility requirement of UDDA would, as Bernat (2013) stresses, allegedly increase the waiting time before procurement, because biological irreversibility generally comes much later than the irreversibility judged by current medical technology. Furthermore, modern medicine has made ECMO a standard equipment in major hospitals, thus theoretically a heart stopping patient cannot be declared dead before ECMO has been tried. But such waiting is in most cases unnecessary, a waste of resource, and even harmful to the patient and her family. What can we say about this? I think, as John Lizza has elaborated in Lizza (2005), irreversibility needs qualification. We have, to name just a few, logical irreversibility, metaphysical irreversibility, physical irreversibility, biological irreversibility, technological irreversibility, situational irreversibility (imagine you have a heart attack in the middle of the Sahara desert), and societal irreversibility (imagine you have signed a DNR[26]) etc. How is the irreversibility in UDDA to be understood?

Biological irreversibility seems to be a nice candidate. However, taking into account the conjecture that life on earth starts as a result of a coincidental lightning strike to a suitable earth environment,[27] there is always a chance that a dramatic event would bring a heart stopping patient back to life. Therefore, biological irreversibility is an unrealistic, even

[25] People usually presuppose some practical constraints on reversibility. For example, if the story of Jesus' raising Lazarus is true, we would still regard Lazarus' state as dead before the raising, because under "normal" conditions a person in that state has no chance to be brought back to life.

[26] A shorthand for 'Do Not Resuscitate'.

[27] For instance, see Hess, Piazolo, and Harvey (2021).

vacuous, notion to be considered as the underlying interpretation for irreversibility involved in the UDDA. In contrast, a notion of irreversibility based on a social-norm which takes biological, technological, situational, and legal considerations all into account can turn out to be more realistic. For a general account about how social norms can play a significant role in the ethics of killing, see Tsai (2017).

Now, back to the main concern of this section. I claim that without taking the move to replace irreversibility by permanence, NHDR itself allows us to shorten the waiting time for the procurement of a heart that has stopped beating—especially when we have had the consent from a donor who very much liked to donate his or her heart, and we can make sure that the anesthesia will be properly administrated during the operation (so that autonomy and non-maleficence that Troug (2015), cares about will be safeguarded) without violating DDR. The details are as follows.

2.2 Alternative Ways of Dying

Recall that in contrast to *hasty* DCDD (the procurement of a stopping heart without waiting long enough to ensure irreversibility), treatment-withdrawal with DNR has become an acceptable practice in many societies today. In other words, a dying patient can ask for the withdrawal of the life sustaining equipment and dying as a consequence, and no member of the medical staff would be accused of killing the patient by shutting down the life-sustaining system. On the other hand, the procurement of a heart that has stopped beating without waiting for several minutes to make sure that the heart of the donor has met the irreversibility criterion is disallowed as it violates [$\mathrm{DDR}_{\mathrm{I}}$].

However, if NHDR is adopted instead, then a hasty DCDD does not always violate DDR. To decide whether such an act of procurement violates DDR, we are to see whether the act hastens death, by comparing the times of death associated with the procurement and the non-procurement (the default) respectively. In Figure 4, compare the following value-time diagrams of a case of treatment-withdraw with DNR and a case of hasty DCDD.

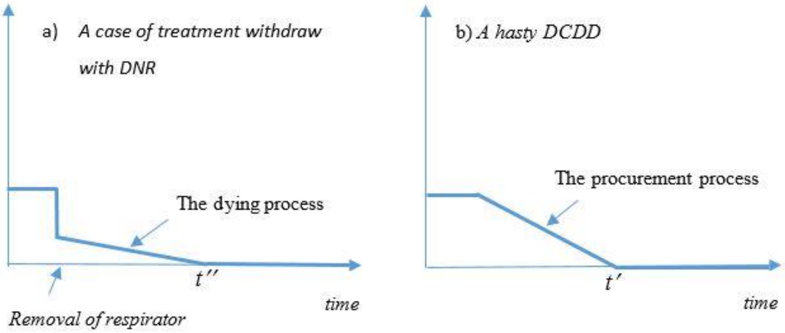

Figure 4. Two Ways of Dying

In case that autonomy and non-maleficence are both guaranteed, there is no reason why a) is allowed while b) is disallowed. Judged from the graphs, a) and b) are both processes from life to death. They are simply two ways of dying. And so far as death time is concerned, the procurement does not hasten death insofar as $t' \geq t''$, so it does not violate NHDR. After all, in the case of a hasty DCDD for a patient with DNR, while it is hasty in the sense that the donor is not dead yet, so long as $t' \geq t''$, it does *not hasten death*, and thus what has been procured is not a *vital* organ and NHDR has not been violated.

If nowadays we can accept, unlike some decades ago, that removing a life support device does not always constitute an act of killing, we should accept that a hasty DCDD does not necessarily constitute an act of killing as well. When one has decided to be let die and does not mind which course her dying process will take, death by treatment-withdrawal with DNR and death by DCDD really make little difference.

The analysis scheme above is new but its conclusion—namely, hasty DCDD is not always wrong—is by no means new, as it has long been observed in Shewmon (2004) and Shewmon and Shewmon (2004). Nevertheless, the analytic scheme of this paper indeed grants us an easy and a principle-based way to explain why certain seemingly unacceptable acts are actually acceptable, as the following imaginary scenario demonstrates.

> A criminal has jumped from the top of a 101-story building to seek death. He will be dead in a few seconds. Before he hits the ground, he is offered a final chance to payback to society. With his consent, a cushion will be provided to delay the death, and

> a hi-tech ultra-fast snatcher can procure his heart from his chest a split second before he eventually hits the ground, and the heart can then be used to save someone's life soon after. If he agrees to the proposal, do we commit homicide by procuring his heart right before he crashes? Is his heart really a vital organ then? Will DDR be violated?

The answers to the last three questions are clearly, I suppose, all 'no' as suggested by Figure 5.

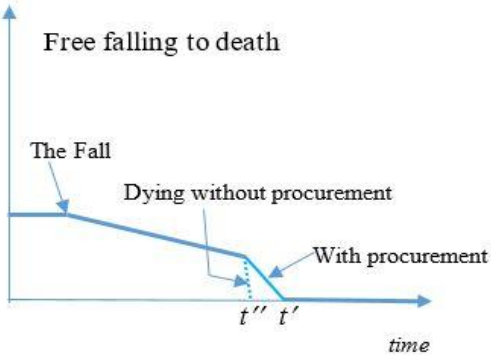

Figure 5. A Non-Killing Procurement

3. Conclusion

Let me sum up what we have achieved in this paper, by reviewing some of the claims in Marquis (2010) which maintains that DCDD donors are not dead. There Marquis claims that DCDD proponents often 'appeal to permanence' or 'appeal to a norm' to show that DCDD donors are dead, but the two appeals both fail. In Section 1, I have stressed that irreversibility and permanence are different things, so we should not substitute permanence for irreversibility and appeal to permanence. Similarly, substituting norm for irreversibility and then appealing to norm won't work either, as criteria in UDDA are clearly biological in nature. Therefore, we agree with Marquis that the two appeals he addresses in his paper indeed fail. However, that does not imply that DCDD donors are not dead. It only shows that proponents of DCDD often appeal to wrong items. In Section 1, I have, without appealing to either permanence or norm, shown that so long as UDDA are properly understood and obeyed, DCDD donors are dead already within the scheme. Furthermore, in Section 2, I have shown that so long as autonomy and non-maleficence principles are observed, some hasty DCDDs—i.e. procurements done before DCDD donors are dead—can be compatible with a newly interpreted DDR too.[28]

Acknowledgments

I would like to thank three anonymous reviewers of this journal for insightful criticisms and constructive suggestions that helped me turn this paper into a solid and consistent piece of work. Thanks are also due to other reviewers that have previously commented on earlier versions of this paper. A special thank is due to Yi, Fu-Yun, for being the sole audience of a draft presentation upon which the current paper was developed, and her comments were very much appreciated.

The work was supported in part by the Ministry of Science and Technology, Taiwan [grant numbers MOST 102-2410-H-715-001-MY3 and MOST 107-2410-H-715 -004 -MY3]. The author is obliged.

REFERENCES

Bernat, James. L. 2013. 'On Noncongruence between the Concept and Determination of Death'. *Hastings Center Report* 43 (6): 25-33.

Brugger, E. Christian. 2013. 'D. Alan Shewmon and the PCBE's White Paper on Brain Death: Are Brain-Dead Patients Dead?' *Journal of Medicine and Philosophy* 38 (2): 205–218.

DeGrazia, David. 2005. *Human Identity and Bioethics*. Cambridge: Cambridge University Press.

Elbourne, Paul. 2010. 'The Existence Entailments of Definite Descriptions'. *Linguistics and Philosophy* 33 (1): 1-10.

Hess, Benjamin L., Sandra Piazolo, and Jason Harvey. 2021. 'Lightning Strikes as a Major Facilitator of Prebiotic Phosphorus Reduction on Early Earth'. *Nature Communications* 12 (1): 1535. https://doi.org/10.1038/s41467-021-21849-2

Joffe, Ari. 2018. 'DCDD Donors Are Not Dead'. *Hastings Center Report* 48 (6): S29-S36.

Lizza, John P. 2005. 'Potentiality, Irreversibility, and Death'. *Journal of Medicine and Philosophy* 30 (1): 45-64.

[28] Ari Joffe also claims that DCDD donors are not dead in Joffe (2018), in which four reasons that death is not merely permanent are given and three common objections to this idea are refuted. Again, I agree with him that substituting permanence for irreversibility is a bad idea. However, DCDD donors can be dead on the irreversibility criteria alone (Section 1), and my new interpretation of DDR can even allow some hasty DCDDs (Section 2).

Lizza, John P. 2011. 'Where's Waldo? The "Decapitation Gambit" and the Definition of Death'. *Journal of Medical Ethics* 37 (12): 743-746

Marquis, Don. 2010. 'Are DCD Donors Dead?' *Hastings Center Report* 40 (3): 24-31.

McLaughlin, James Angell. 1925. 'Proximate Cause'. *Harvard Law Review* 39 (2): 149-199.

McMahan, Jeff. 2002. *The Ethics of Killings: Problems at the Margins of Life*. Oxford: Oxford University Press.

Miller, Franklin G., Robert D. Truog, and Dan W Brock. 2010. 'The Dead Donor Rule: Can it Withstand Critical Scrutiny?' *Journal of Medicine and Philosophy* 35 (3): 299-312.

Miller, Franklin G., and Robert D. Truog. 2011. *Death, Dying, and Organ Transplantation: Reconstructing Medical Ethics at the End of Life*. New York: Oxford University Press.

Moore, Michael S. 2009. *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. New York: Oxford University Press.

Paul, Laurie Ann, and Edward J. Hall. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.

Shewmon, D. Alan. 2004. 'The Dead Donor Rule: Lessons from Linguistics'. *Kennedy Institute of Ethics Journal* 14 (3): 277-300.

Shewmon, D. Alan. 2010. 'Constructing the Death Elephant: A Synthetic Paradigm Shift for the Definition, Criteria, and Tests for Death'. *Journal of Medicine and Philosophy* 35 (3): 256-98.

Shewmon, D. Alan, and Elisabeth Seitz Shewmon. 2004. 'The Semiotics of Death and its Medical Implications'. In, *Brain Death and Disorders of Consciousness*. Advances in Experimental Medicine and Biology, Vol. 550, edited by Calixto Machado and D. Alan Shewmon, 89-114. New York: Kluwer Academic/Plenum Publishers.

Sinnott-Armstrong, Walter, and Franklin G. Miller. 2013. 'What Makes Killing Wrong?' *Journal of Medical Ethics* 39 (1): 3-7.

Truog, Robert D. 2007. 'Brain Death - Too Flawed to Endure, Too Ingrained to Abandon'. *Journal of Law, Medicine and Ethics* 35 (2): 273-281.

Truog, Robert D. 2015. 'Defining Death: Getting It Wrong for All the Right Reasons'. *Texas Law Review* 93 (7): 1885-1914.

Tsai, Cheng-Chih. 2017. 'Killing, a Conceptual Analysis'. *Ethical Perspectives* 24 (3): 467-499.

Veatch, Robert M. 2008. 'Donating Hearts after Cardiac Death—Reversing the Irreversible'. *New England Journal of Medicine* 369: 672-673.

Veatch, Robert M. 2010. 'Transplanting Hearts after Death Measured by Cardiac Criteria: the Challenge to the Dead Donor Rule'. *Journal of Medicine and Philosophy* 35 (6): 313-329.

Wolfram, Sybil. 1989. *Philosophical Logic, An Introduction*. London and New York: Routledge.

EuJAP | Vol. 17 | No. 1 | 2021
UDC: 115:161.2
https://doi.org/10.31820/ejap.17.1.7

IS THERE CHANGE ON THE B-THEORY OF TIME?

Luca Banfi[1]

[1] University College Dublin

Review article – Received: 02/06/2020 Accepted: 01/07/2021

ABSTRACT

The purpose of this paper is to explore the connection between change and the B-theory of time, sometimes also called the Scientific view of time, according to which reality is a four-dimensional spacetime manifold, where past, present and future things equally exist, and the present time and non-present times are metaphysically the same. I argue in favour of a novel response to the much-vexed question of whether there is change on the B-theory or not. In fact, B-theorists are often said to hold a 'static' view of time. But this far from being innocent label: if the B-theory of time presents a model of temporal reality that is static, then there is no change on the B-theory. From this, one can reasonably think as follows: of course, there is change, so the B-theory must be false. What I plan to do in this paper is to argue that in some sense there is change on the B-theory, but in some other sense, there is no change on the B-theory. To do so, I present three instances of change: Existential Change, namely the view that things change with respect to their existence over time; Qualitative Change, the view that things change with respect to how they are over time; Propositional Change, namely the view that things (i.e. propositions) change with respect to truth value over time. I argue that while there is a reading of these three instances of change that is true on the B-theory, and so there is change on the B-theory in this sense, there is a B-theoretical reading of each of them that is not true on the B-theory, and therefore there is no change on the B-theory in this other sense.

Keywords: *Change; B-theory of time; existence; properties; propositions*

Correspondence: banfi.luca@ymail.com

Introduction

The purpose of this paper is to explore the connection between *change* and the *B-theory of time*, sometimes also called the *Scientific view of time*, according to which reality is a four-dimensional spacetime manifold, where past, present and future things equally exist, and the present time and non-present times are metaphysically the same. I argue in favour of a novel response to the much-vexed question of whether there is change on the B-theory or not.[1]

In fact, B-theorists are often said to hold a 'static' view of time. But this far from being an innocent label: if the B-theory of time presents a model of temporal reality that is static, then there is no change on the B-theory. From this, one can reasonably think as follow: of course there is change, so the B-theory must be false. What I plan to do in this paper is to argue that *in some sense* there is change on the B-theory, but *in some other sense*, there is *no* change on the B-theory. To do so, I present three instances of change: Existential Change, namely the view that things change with respect to their existence over time; Qualitative Change, the view that things change with respect to how they are over time; Propositional Change, namely the view that things (i.e. propositions) change with respect to truth value over time. I argue that while there is a reading of these three instances of change that is true on the B-theory, and so there is change on the B-theory *in this sense*, there is a reading of each of them that is not true on the B-theory, and therefore there is no change on the B-theory *in this other sense*.

1. Three Instances of Change

Bubbles, chemical reactions, flowers, butterflies, human beings (and so on) exist, but do not exist forever. More generally, many things change with respect to existence. Hence, the following counts as an instance of change:

> EXISTENTIAL CHANGE**:** things change with respect to existence over time.[2]

[1] For a classic discussion see McTaggart (1927), Prior (1968) and Williams (1951); for a contemporary discussion see Sider (2011, Ch. 11) and Williamson (2013, ch. 8).

[2] Existential Change may take different forms. Some believe it to be true, since they believe that things both begin and cease to exist, such as Lowe (2003, 2006, 2009), Prior (1968), and Zimmerman (2008); others, think of it to be true because things begin to exist, but then do not cease to do so (Correia and Rosenkranz 2018).

It is important to be clear about the meaning of ‘exist’ in Existential Change: ‘exist’ means here the same as ‘being’ or ‘being something’ or ‘being identical to something’ in the most unrestricted sense. What I am assuming here is the standard meaning of *existence* assumed by most contemporary metaphysicians.[3] So, for a cat to exist is for it to be something, for a car to exist is for it to be something, and so on. Nothing more or less.

For the sake of convenience, it is useful to introduce a more formal way of expressing Existential Change, and the following instances of change. To do so, let’s appeal to the language of *free tense logic*,[4] the language that implements the language of free logic with the so-called *tense operators* such as the past tense operator ‘It was the case that’ or ‘It is the case at some past time that’ (symbolised as ‘P’) and the future tense operator ‘It will be the case that’ or ‘It is the case at some future time that’ (symbolised as ‘F’). From these, one can further define the operator ‘It is sometimes the case that’ or ‘It is the case at some time’ (symbolised as ‘S’, where ‘$\mathrm{S}\varphi$’ is defined as ‘[5]$\mathrm{P}\varphi \lor \varphi \lor \mathrm{F}\varphi$’), and the operator ‘It is always the case that’ or ‘It is the case at all times’ (symbolised as ‘A’, where ‘$\mathrm{A}\varphi$’ is defined as ‘$\mathrm{P}\varphi \land \varphi \land \mathrm{F}\varphi$’).

[3] See van Inwagen (2009) for a detailed discussion of the view.

[4] I say *free* tense logic instead of *quantificational* tense logic, since the latter arguably entails the falsity of Existential Change (see, for instance, Bacon (2013) and Williamson (2013)). I will not rehearse all the details of the argument here, but a brief discussion may help. Given quantificational tense logic, every instance of ‘$\exists x\ a = x$’ is proved to be true (where ‘a’ stands for any singular term in that language, ‘$\exists x\ a = x$’ says that there is something identical to a). Given the temporal analogue of the rule of necessitation, one can infer from the true φ that it is always the case that φ, such that one can infer ‘$\mathrm{A}(\exists x\ a = x)$’ (which says that it is always the case that there is something identical to a) from ‘$\exists x\ a = x$’. Now, by the rule of generalization, that allows one to infer from the true φ the true $\forall x\varphi$, one can infer ‘$\forall y\ \mathrm{A}(\exists x\ y = x)$’ (which says that everything always exists) from ‘$\mathrm{A}(\exists x\ a = x)$’. So, given quantificational tense logic, one can prove that everything always exists, thus contradicting Existential Change.
If one adopts a *free* tense logic, one can deny that every instance of ‘$\exists x\ a = x$’ is proved to be true in the first place. In fact, given a free tense logic, one can reject the axiom of universal instantiation of quantificational tense logic, namely ‘$\forall x\varphi x \rightarrow \varphi \mathrm{a}$’, from which ‘$\exists x\ a = x$’ is inferred, and accept the weaker axiom of free universal instantiation, such as ‘$\forall y(\forall x\varphi x \rightarrow \varphi y)$’ (which says that for every way everything in the domain is, there is something we can name in the domain that is in that way). Therefore, those who accept a free tense logic have the logical resources to accept that some instances of ‘$\exists xa = x$’ are false, when ‘a’ fails to denote a member of the domain of quantification, and also to resist the above argument for the inconsistency between quantificational tense logic and Existential Change.

[5] In fact, there the list of tense operators I present is not exhaustive, as there are further tense operators such as ‘It has always been the case that’ (symbolised as ‘H’) and ‘It is always going to be the case that’ (symbolised as ‘G’). One can also make use of the so-called metric tense operators, operators of the form ‘It was the case *n* units of time ago’.

So, one can regiment Existential Change as:

> EXISTENTIAL CHANGE: Sometimes, something is not always something. (Formally: ‘$S(\exists x \neg A(\exists y\, y = x))$’)

There is not just change in what exists over time, however. Tim was a kid, and he is an adult, Lisa is seated, even though she was standing, and Barack Obama was the US President, but he is not anymore. Therefore, a further instance of change is the following:

> QUALITATIVE CHANGE: things change with respect to how they are over time.[6]

If we understand the predicates ‘is adult’, ‘is seated’ and ‘is the US President’ as expressing *properties* or *qualities*, Qualitative Change amounts to the view that things change with respect to properties or qualities over time: Tim does not always bear the property of being an adult, Lisa does not always bear the property of being seated, and Barack Obama does not always bear the property of being the US President.

Qualitative Change is naturally linked with the view that things *persist* through time. Take Lisa: it is not just that Lisa, for example, changes as she is seated, but she was standing, but also that Lisa exists and she is seated, and *existed* and she was standing; analogously, Lisa will exist and will be standing too.

More specifically, for Lisa to change with respect to her properties, she must remain in existence through time. That’s the dynamic phenomenon philosophers call *persistence*:

> PERSISTENCE: things persist through time.

For Lisa to change with respect to her properties is not for Lisa to begin to exist when she gains the property of being seated, and for her to cease to exist when she loses the property of being seated and gains the property of being standing. As a matter of fact, Lisa exists both when she is seated and when she was and will be standing. It is one and the same thing that both changes and persists through time: it is one and the same Lisa that is seated and exists, and did exist and was standing.

[6] Those who believe in Qualitative Change include, for example, Hinchliff (1996) and Prior (1968, 78–9).

So, Qualitative Change and Persistence count as a single instance of change (or so I claim): call this instance of change *Qualitative Change plus Persistence*. One can express Qualitative Change plus Persistence more formally as follows, where F stands for some property:

> QUALITATIVE CHANGE PLUS PERSISTENCE: Sometimes, some *x* bears some F but *x* does not always bear F. (Formally: ‘S(∃*x*∃F (F*x* ∧ ¬AFx))’)

The last instance of change that I wish to consider is change in what is the case, or what is true, over time. One can derive such instance of change from either Existential Change or Qualitative Change. Consider Lisa: since Lisa exists, it the case that Lisa exists. Moreover, since Lisa is seated, it is the case that Lisa is seated. But Lisa does not always exist and is not always seated. Therefore, it is not always the case that Lisa exists or that Lisa is seated. By taking talks about “being the case” as equivalent to “being true”, where the primary bearers of truth and falsehood are *propositions*, then one can say that the propositions *that Lisa exists* and *that Lisa is seated* are true, but not always so. As a matter of fact, they change with respect to truth value. Here is the third instance of change:

> PROPOSITIONAL CHANGE: things (i.e., propositions) change with respect to truth value over time.[7]

Very much as with Existential Change and Qualitative Change plus Persistence, one can express Propositional Change in more formal terms as follows, where ‘*p*’ stands for a propositional variable:

> PROPOSITIONAL CHANGE: Sometimes, there is some true proposition that is not always true. (Formally: ‘S(∃p (p ∧ ¬A(p))’

This concludes the presentation of three entirely plausible instances of change.

In this paper, my purpose is to argue that the B-theory of time, which I will introduce in more detail in the following section (Section 2), is consistent with certain readings of such instances of change; thus, *in this sense*, there is change on the B-theory. Nevertheless, I argue that there is a reading of each of them that is false on the B-theory, and therefore that *in this other sense*, there is no change on the B-theory. The plan for the remaining part

[7] The view is extensively defended in Borgaard (2012) and Cappelen and Hawthorne (2009) among others.

of the paper is as follows. In Section 2, I say more about how to understand the B-theory; in Section 3 I discuss the connection between Existential Change and the B-theory; in Section 4, I discuss the connection between Qualitative Change plus Persistence and the B-theory; in Section 5, I discuss the connection between Propositional Change and the B-theory.

2. The B-theory of time

B-theorists, those who defend the *B-theory of time*,[8] typically hold that reality consists of a four-dimensional block universe, the spacetime of relativistic physics[9] (in virtue of which it is sometimes also called the *Scientific view of time*), where past, present and future things equally exist, and the present time and non-present times are metaphysically the same. On this view, for the present time to be present does not designate anything of metaphysical significance, as 'present' is an indexical expression that refers to the time of utterance of such expression.

Consider dinosaurs, for example. B-theorists hold that dinosaurs exist very much as you and me; and the same goes for times: B-theorists think that the time at which dinosaurs are located exists very much as this time, the time at which we are located. In fact, B-theorists think of time as very similar to space: very much as all places, and things located at such places, equally exist, all times, and things located at such times, equally exist. To this, B-theorists add that as for some place to be the place that is *here* does not designate anything of metaphysical importance, for some time to be the time that is *present* does not designate anything of metaphysical importance: 'here' and 'present' are merely indexical expressions that refer, respectively, to the place and time at which they are uttered.

A further B-theoretic commitment on which I want to focus is how B-theorists usually interpret tenses and tense operators, as this will help to introduce the B-theoretical readings of Existential Change, Qualitative Change plus Persistence and Propositional Change. B-theorists think of tense operators as being fully reducible; to describe how reality ultimately looks like, B-theorists do not make use of any tense operator. This idea is captured by what Sider says in following quotation:

> [B-theorists] do not admit tense operators into their fundamental ideology, since they can describe temporal reality

[8] Some supporters of the B-theory include Deng (2013), Dyke (2002), Leininger (2021), Mozersky (2015) and Sider (2001; 2011) and Williams (1951).

[9] As characterized by the pioneering research of Einstein (1952) and Minkowski (1952).

> without them—by quantifying over past and future entities and predicating features of them relative to times. (Sider 2011, 241)

Now we have all the ingredients we need to proceed with the discussion. Let's begin by exploring the connection between Existential Change and the B-theory.

3. Existential Change and the B-theory

Existential Change is the view that things change with respect to their existence, and B-theorists might happily grant that Existential Change is true on the B-theory. Consider the more formal version of Existential Change, namely 'Sometimes, something is not always something'. Now, there is a reading of Existential Change that is true on the B-theory, and to see that we must be clear about what it is for something to be such that it is *sometimes* something, or that it *sometimes* exists. As stated in the previous section, tense operators such as 'Sometimes' or, equivalently, 'It is sometime the case that' are fully reducible on the B-theory.[10] In fact, according to B-theorists, tense operators are fully reducible to quantifiers over past, present and future times, as Sider makes clear in the passage quoted in the previous section: for something to be such that it sometimes exists is for it to be such that there is a time at which it exists. In other words, B-theorists reduce expressions of the form 'sometimes, x exists' to expressions of the form 'x exists at some time t'. Accordingly, Existential Change reduces to the following:

> EC-1: For some times t and t_1, some x is such that x exists at t but x does not exist at t_1

And EC-1 is true on the B-theory. Consider dinosaurs, for example: it is true that dinosaurs exist at some times but not at others.

However, contemporary research on the topic suggests that expressions of the form 'x exists at t' are inherently ambiguous (Correia and Rosenkranz 2019; Deasy 2019; Markosian 2014): on one reading, they are equivalent to expressions of the form 'x is located at t'; whereas on another reading, they are equivalent to expressions of the form 'at t, x is something'. Thus, on the first reading, 'x exists at t' is understood in *locational* terms, such

[10] Defenders of Existential Change like Prior (1968) or Crisp (2007) accept the partial (but not full) reducibility of tense operators to quantifiers over times, where times are intended as maximal, consistent, and sometimes-true propositions. However, since times are defined as "sometimes true" propositions, tense operators do not fully reduce to quantifiers over times.

that to say that something exists at a time is to make a claim about where things are located *in* time; on the second reading, '*x* exists at *t*' is understood in *perspectival* terms, such that to say that something exists at a time t is to make a claim about what there is relative to (i.e. from the perspective of) *t*.

If expressions of the form '*x* exists at *t*' are equivalent to expressions of the form '*x* is located at *t*', then Existential Change is true on the B-theory, since one can think of EC-1 as equivalent to the following:

> EC-2: For some times t and t_1, some x is such that x is located at t but x is not located at t_1

EC-2 is true on the B-theory.

However, one can read Existential Change as a thesis about what there is in time, rather than about where things are located in time. On this reading, Existential Change becomes a thesis about there being change in what there is over time. And *on this reading*, Existential Change is false on the B-theory. Or so I argue.

In order to develop this argument, I wish to consider the modal analogue of the B-theory, namely *Modal Realism*, on which actual things exist just as possible things do, and the actual world and non-actual worlds are metaphysically the same, notably defended by Lewis (1986). Modal Realists understand expressions of the form '*x* exists at world *w*' very much as B-theorists understand expressions of the form '*x* exists at time *t*'. Accordingly, we can disambiguate between two readings of expressions of the form '*x* exists at world *w*', as either equivalent to '*x* is located at *w*' or 'at *w*, *x* is something'.

Now, consider the following quotation from Lewis (1986):

> The phrase 'at W' which appears within the scope of the quantifier, […] works mainly by restricting the domains of quantifiers in its scope, in much the same way that the restricting modifier 'in Australia' does. […] [However] I do not suppose that they must restrict all quantifiers in their scope, without exception. […] 'At some small worlds, there is a natural number too big to measure any class of individuals' can be true even if the large number that makes it true is no part of the small world. (Lewis 1986, 6)

In the first half of the quotation, Lewis is referring to the first reading of '*x* exists at world *w*', where *w*, a world, is where some *x* is located. In the second half of the quotation, Lewis speaks of the second reading of '*x* exists at world *w*', and says that sentences such as 'At *w*, there is a natural number too big to measure any class of individuals' can be true, even if such natural number is not located at *w*. In other words, irrespective of the location of such natural number, it is true of it that it is something, or it exists, given Modal Realism. So, Lewis is here suggesting that when attached to claims about what there is, irrespective of the location, in the modal space, the phrase 'at *w*' is irrelevant: that there is a natural number too big to measure any class of individuals is true even at worlds at which it is not located.

Let's apply the understanding of expressions of the form 'at *t*, *x* is something' proposed by Lewis to the temporal case. To do so, consider the sentence 'at *t*, there is a dinosaur': 'There is a dinosaur' is true on the B-theory, even if there are no dinosaurs located at *t*. So, when attached to claims about what there is, irrespective of the location, in time, the phrase 'at *t*' seems to be irrelevant on the B-theory too: 'There is a dinosaur' is true on the B-theory, even at times at which dinosaurs are not located.

In light of that, one can argue that there is a reading of Existential Change that is false on the B-theory as there is a reading of *Permanentism*, the view that 'Everything always exists', namely the negation of Existential Change, which is true on the B-theory. To see that, let's reduce Permanentism to the view 'Everything exists at every time',[11] and let's disambiguate between two versions of Permanentism in accordance with the disambiguation of the expression '*x* exists at *t*':

P1: Everything is located at every time
P2: At every time, everything is something

P1 appears to be false:[12] it is false that dinosaurs are located at every time, as they are not located at this time. However, 'Everything is something' is (trivially) true at every time: it is true at this time that some dinosaur exists, for example, even though there are no dinosaurs located at this time. So, there is a reading of Permanentism, namely P2, that is true on the B-theory.

[11] Given the reducibility of tense operators proposed by B-theorists, for something to be always something is for it to exist at every time.

[12] Unless one accepts some less-standard view on which, for example, there are just eternally existent atoms. On this view, P1 turns out to be true.

However, when '*x* exists at *t*' is read at 'at *t*, *x* is something', namely when expressions of the form '*x* exists at *t*' are read as making claims about what there is, irrespective of the location, in time, P2 is equivalent to Permanentism. In very much the same way, one can read Existential Change as making a claim about what there is, irrespective of the location, in time:

> EC-3: for some times t and t_1, at t, x is something and at
> t_1 x is nothing

However, the B-theory is inconsistent with *this* reading of Existential Change, since on the B-theory there is no change in what there is, irrespective of the location, in time.

As we have seen, there is a sense in which Existential Change is true on the B-theory, and so in this sense there is change on the B-theory; however, there is also a reading of Existential Change that is inconsistent with the B-theory, such that in this other sense, there is no change on the B-theory. We can make a very similar claim with respect to Qualitative Change plus Persistence, as I argue in the following section.

4. Qualitative Change plus Persistence and the B-theory

As we needed to be clear about what is for something to sometimes exist on the B-theory, in order to explore its connection with Existential Change, we now have to be clear about what is for something *to sometimes bear some property* on the B-theory, in order to explore the connection between the B-theory and Qualitative Change plus Persistence (from now on, simply QCP).[13]

Think again of the above quotation from Sider (2011, 241). Sider remarks that B-theorists can describe things over time by "predicating features of them *relative to times* (italic mine)", such that for something to sometimes bear some property is for it to bear some property at some time. Therefore, B-theorists think of expressions of the form 'sometimes, *x* is F' as reducing to '*x* bears F at some time *t*'. However, different B-theorists understand expressions of the form '*x* bears F at some time *t*' in different ways. In this section, I explore different B-theoretical interpretations of such expression, and I explore their connection with QCP.

[13] For a discussion of how the B-theory connects with what I call Qualitative Change plus Persistence see Cameron (2015, 152-159) and Wasserman (2006).

To begin with, consider what Lewis says in the following quotation:

> Let us say that something *persists* iff, somehow or other, it exists at various times; this is the neutral word. Something *perdures* iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it *endures* iff it persists by being wholly present at more than one time. (Lewis 1986, 202)

Accordingly, *Perdurance* is the view on which things persist by perduring and *Endurance* is the view that things persist by enduring. More recently, philosophers have introduced a further notion of persistence called *Exdurance* (Hawley 2001; Sider 1996), on which things persist by *exduring*, namely by having different temporal counterparts at different times. Depending on whether one endorses Perdurance, Endurance or Exdurance (plus the B-theory) one delivers a different interpretation of '*x* bears F at *t*'. My plan in what follows is to discuss each option in order.

4.1 B-theoretic Perdurance

Perdurance as defended, for example, by Heller (1984), Lewis (1986) and Quine (1950) is the view that things persist by having different temporal parts at different times. Think of Lisa again. Lisa persists by having different temporal parts at different times. Then, on Perdurance, for Lisa to change from being standing to being seated, is for Lisa to have a temporal part that is standing and a later temporal part that is seated.

For the sake of a better understanding of Perdurance, we must get a better grip on what temporal parts are. Temporal parts are usefully characterized by analogy with spatial parts: as things have spatial parts, such as Lisa has a spatial part, such as her arm, and another spatial part, such as her leg, Lisa also has temporal parts, such as one that is standing and another that is seated.

Both spatial and temporal parts can be understood as spatially and temporally extended: as Lisa's arm is extended through space, Lisa's temporal part that is standing can be taken to be temporally extended too. However, as Lewis (and many others) defines the notion of temporal parts, temporal parts exist *at times*, and since times are instantaneous objects, temporal parts are naturally understood as instantaneous too: thus, *x* is an instantaneous temporal part of *y* if and only if (Sider 2001, 59) *x* is part of *y*, *x* overlaps, or shares, any part of *y*, and *x* exists only at a single time. In what follows, for the sake of simplicity, when I speak of 'temporal parts' I mean 'instantaneous temporal parts'.

Thus, perdurantists naturally read expressions of the form 'x bears F at t' as equivalent to 'one of x's temporal parts is F and is located at t'. Accordingly, there is a reading of QCP that is true on the B-theory, namely:

> QCP-1: for some times t and t_1, there is some x, y and z such that y, one of x's temporal parts, is F and is located at t, and z, another of x's temporal parts, is not-F and is located at t_1

So, perdurantists accept the truth of QCP by treating possession of properties as possession of properties relative to times; and then, by analysing possession of properties relative to times as possession of properties by temporal parts. In this sense, there is change on B-theory, as there is change given Perdurance.

However, there is also a reading of QCP that is ultimately inconsistent with Perdurance. Let me expand on that. On this alternative reading, it is Lisa the thing that is seated *simpliciter* and it is Lisa who changes with respect to that property. On the contrary, defenders of B-theoretic Perdurance think that it is not Lisa, but one of her temporal parts – call it T-Lisa – which is seated *simpliciter*. However, as an instantaneous object, T-Lisa does not persist, and therefore does not change with respect to her being seated, as there is no other time at which it is located and bears the property of being standing. So, T-Lisa does not change with respect to the property of being seated: T-Lisa is always seated.

The result is that B-theoretic Perdurantists reduce QCP to the eternal possession (*simpliciter*) of properties by temporal parts. Moreover, on the assumption that there is a sense in which B-theorists accept Permanentism––the view that everything always exists—B-theorists accept that in some sense it is true that temporal parts *always* exist. So, B-theoretic Perdurantists reduce QCP to the eternal possession of properties by eternally existent temporal parts. But one can read QCP as being about Lisa's changing with respect to her properties while persisting through time, while the B-theoretic Perdurantist's explanation of such a phenomenon bottoms out in permanent facts about the eternal properties of (eternal) temporal parts. B-theoretic Perdurance is thus inconsistent with *this* reading of QCP.

In the next section I show how a similar argument applies to B-theoretic Exdurance too.

4.2 B-theoretic Exdurance

Distinct from Perdurance, Exdurance as defended by, for instance, Hawley (2001) and Sider (1996), among others, is the view that ordinary things are not temporally extended things, but instantaneous temporal parts, or *stages*, and for something to persist is for it to have different temporal counterparts at different times. Think of Lisa: Lisa persists by having different temporal counterparts at different times. Then, on Exdurance, for Lisa to change from being standing to being seated, is for Lisa, an instantaneous stage, to be seated, and to have a temporal counterpart that is standing.

One of the main novelties of Exdurance is the introduction of the notion of *temporal counterparts*: to get a better sense about what a temporal counterpart is, an analogy with the modal case is instructive. David Lewis' notion of *modal counterpart* (Lewis 1968) is probably the best place for that: the point here is to get a sense of how, for example, Lisa modally persists. For Lisa to modally persist is for Lisa to have various modal counterparts at various worlds, where for something to be a modal counterpart of Lisa is to *resemble* Lisa in all her *relevant features* (Lewis 1968, 114; Sider 2001, 111–2). For example, 'Lisa is seated but might be standing' is true because Lisa modally persists by having a modal counterpart at some world, which resembles Lisa in all her relevant features, and it is standing. The same applies to the temporal case. For Lisa to persist is for Lisa to have different temporal counterparts at various times, which resemble Lisa in all her relevant features. Then, 'Lisa is seated but was standing' is true because the Lisa that is seated has a temporal counterpart at some time, which resembles Lisa in all her relevant features, and it is standing.

Thus, Exdurantists naturally read expressions of the form 'x is F at t' as 'x is F and is located at t', by treating the variable 'x' as taking in only instantaneous stages. In other words, according to Exdurantists, the name 'Lisa' does not refer to an object that exists at different times, but to an instantaneous stage. Accordingly, there is a reading of QCP that is true on the B-theory, namely:

> QCP-2: for some times t and t_1, there is some x such that x is F and is located at t, and there is some y, one of x's temporal counterparts, such that y is not-F and is located at t_1

So, exdurantists accept QCP by treating possession of properties as possession of properties relative to times; and then, by analysing possession of properties relative to times as possession of properties by instantaneous stages. In this sense, there is change on the B-theory, as there is change given Exdurance.

Still, there is an alternative reading of QCP that is ultimately inconsistent with Exdurance. On this reading of QCP, it is Lisa who is seated, and it is *that* thing that persists through time: distinct from defenders of Perdurance, defenders of Exdurance accept that. However, on this reading of QCP, ‘Lisa’ refers to a temporally extended object and not to an instantaneous thing, as it does according to Exdurance. In light of that, there is a sense in which things do not persist given Exdurance, and that’s the sense in which things persist given this reading of QCP: Lisa exists, but did exist and will exist too. As an instantaneous thing, however, Lisa does not persist in *this* sense on Exdurance: instantaneous things are, by definition, things that do not exist at multiple times, and in this sense, it is false that Lisa did exist and will exist too given Exdurance.

So, in that sense, Lisa does not persist on Exdurance; but if Lisa does not persist in this sense, Lisa does not change too, as there is no other time at which it exists and is, for example, standing. If she does not change and persist in this sense, namely the sense in which things change and persist given this reading of QCP do, Exdurance is inconsistent with *this* reading of QCP. Thus, there is no change on the B-theory, as there is no change on Exdurance.

There are further elements that make one worry about the consistency of QCP, on this reading, and Exdurance. First, while given this reading of QCP, the Lisa that exists and is seated is one and the same with the Lisa that did exist and was standing, given Exdurance, the Lisa that exists and is seated is not one and the same with her earlier temporal counterpart that is standing. As a matter of fact, given Exdurance, the two are not identical, but resemble each other with respect to their relevant features. Such resemblance-relation, however, is *deliberately* context sensitive, as it is the notion of “relevant features”. As a matter of fact, we may deliberately refer to one set of features S in one context according to which the Lisa that is seated and the Lisa that is standing resemble each other, and to another set of features S* in another context according to which the two do not resemble each other.

As a consequence, there is no fixed set of relevant features according to which temporal counterparts resemble each other: such set varies from situation to situation and the choice of such set is entirely arbitrary. However, on Exdurance things persist by being related via a relation of resemblance in relevant features; so, Persistence becomes a deliberately context sensitive phenomenon as well. On the contrary, given the reading of QCP under consideration, Persistence is not deliberately context sensitive: it is one and the same thing, namely Lisa, that changes and persists over time.

Moreover, given Exdurance, we have a series of instantaneous stages that persist by resembling each other with respect to some relevant features. In other words, we have a series of instantaneous stages lined up in time, related to one another by a relation of resemblance in all the relevant features. In such a series, we have the Lisa that is seated and the Lisa that is standing. But *who* is the persisting Lisa? *This* Lisa, namely the Lisa that is seated, or *that* Lisa, the Lisa that is standing? The choice is entirely *arbitrary*. [14] On the contrary, given the reading of QCP under consideration, there is no choice to be made: there's only one Lisa, and *that*'s the persisting thing, and that's the Lisa that is seated.

Exdurance is thus inconsistent with *this* reading of QCP, and in this sense there is no change on Exdurance, and then on the B-theory. Let's now move to the final view I wish to discuss, namely B-theoretic Endurance.

4.3 B-theoretic Endurance

Endurance, as defined by Lewis, is the view that things persist by being wholly present at different times. Defining Endurance in these terms raises several difficulties.[15] It is not my aim here to try to fix some of such difficulties. What I plan to do, instead, is to look at a couple of ways in which self-described endurantists characterize the view and expand on their connection with QCP.

4.3.1 B-theoretic Relationalism

Let's begin with the view I call *Relationalism* as defended by Mellor (1998) and Mozersky (2015), among others, according to which expressions of the form 'x is F at t' are interpreted as 'x is F-at-t', namely the view on which things have different time-indexed properties. Time-indexed properties are properties such as being-a-kid-at-t or being-red-at-t_1 and so on. Think of Lisa: for Lisa to change from being standing to being seated, given Relationalism, is for Lisa to be standing-at-t_1 and to be seated-at-t.

[14] Note that the question is not "who is the Lisa that is seated?", as the Lisa that is seated is plausibly taken to be the Lisa that exists at the time of utterance of the sentence 'Lisa is seated'.

[15] See Sider (2001, 63–68) for a discussion of such problems.

Granted the relationalist's reading of expressions of the form '*x* is F at *t*', there is a reading of QCP that is true on the B-theory, namely:

> QCP-3: for two times t and t_1, there is some x such that x is F-at-t and not-F-at-t_1

So, Relationalists accept QCP by treating possession of properties as possession of properties relative to times; and then, by treating properties as time-indexed properties, where the index corresponds to time relative to which the relevant property is said to be possessed in the first place.[16] In this sense, there is change on the B-theory, as there is change on Relationalism.

While there is a reading of QCP that is true given Relationalism, there is a further reading of QCP that is inconsistent with Relationalism: on this reading, Lisa changes with respect to her being seated, and being seated is a *temporary property*. I claim that what are temporary properties given this reading of QCP become *eternal properties* given Relationalism. To see that, consider the property of being seated: given QCP, being seated is a temporary property, where a temporary property is a property that is sometimes but not always possessed. As a matter of fact, Lisa is seated, but not always seated. In other words, Lisa does not bear any indexed property such as the property of being-seated-at-*t*, as she is simply seated, and not always so.

On the contrary, B-theoretic Relationalists think that for Lisa to be such that she was standing and is seated reduces to her bearing the properties of being-standing-at-t_1 and being-seated-at-*t*, where t_1 is earlier than *t*. The problem is then that Lisa never changes with respect to being-seated-at-*t* and being-standing-at-t_1. Consider being-seated-at-*t*: being-seated-at-*t* is an eternal property, where for a property to be eternal given the B-theory is for it to be such that if something bears it, it always bears it.[17] As a matter of fact, it is true at every time that Lisa bears the property of being-seated-at-*t*. If so, however, Lisa never changes with respect to this property: Lisa is always seated-at-*t*.

Given the reading of QCP under consideration, however, Lisa changes with respect to the property of being seated, which is a temporary, rather than eternal, property, that is a property that Lisa has, but not always. Even

[16] The view is notably criticized in Lewis (1986, 204), with the so called *temporary intrinsics objection*.

[17]It is important to notice that unlike the B-theory, on the view on which there is change in what exists, irrespective of the location, in time, a property that something always bears, namely an eternal property, would be a property that something bears *whenever it exists*.

if B-theoretic Relationalists attempt to explain this reading of QCP, they do so by reducing temporary properties onto eternal ones, such as time-indexed properties. Since, on *this* reading QCP, properties such as being seated are temporary, rather than eternal, B-theoretic Relationalism is ultimately inconsistent with it.

Before concluding this section, I wish to consider a slightly different version of Relationalism as defended, for example, by van Inwagen (1990):[18] on this view, expressions of the form '*x* is F at *t*' are interpreted as '*x* is-F-at *t*', namely the view on which things bear different relations with different times. Think of Lisa again: for Lisa to change from being standing to being seated, given this version of Relationalism, is for Lisa to be-standing-at t_1 and to be-seated-at *t*.

I am persuaded to think that this version of Relationalism is inconsistent with the reading of QCP under consideration too. As a matter of fact, very much as time-indexed properties, relations to times always hold: Lisa always bears the relation of being-standing-at with t_1 and the relation of being-seated-at with *t*. On this reading of QCP, being seated, for example, is a temporary property, that becomes a permanent relation that Lisa bears with some time given this version of Relationalism. Therefore, also this version of Relationalism is inconsistent with *this* reading of QCP.

4.3.2 B-theoretic Adverbialism

Adverbialism, the view defended by Haslanger (1989), Johnston and Forbes (1987) and Miller and Braddon-Mitchell (2007), among others, is the view according to which expressions of the form '*x* is F at *t*' reduce to '*x* is-at-*t* F', where the instantiation-relation between properties and their bearers is time-indexed. Then, for Lisa to change from being standing to being seated is for Lisa to be-at-t_1 standing and to be-at-*t* seated.

Granted the adverbialist's understanding of expressions of the form '*x* is F at *t*', there is a reading of QCP that is true one the B-theory:

> QCP-4: for some times *t* and t_1, there is some *x* such that
> *x* is-at-*t* F and *x* is-not-at-t_1 F

So, Adverbialists accept QCP by treating possession of properties as possession of properties relative to times; and then, by treating the instantiation-relation as time-indexed, where the index corresponds to time

[18] Thanks to an anonymous reviewer for this journal for pressing me to consider this version of Relationalism too.

relative to which the relevant property is said to be possessed in the first place. In this sense, there is change on the B-theory, as there is change on Adverbialism.[19]

I think that there is an argument like the one I raised against B-theoretic Relationalism to show that there is a reading of QCP that is inconsistent with B-theoretic Adverbialism: on this reading of QCP, Lisa changes with respect to her being seated, as she is temporarily seated.[20] What I believe to be problematic is that there is a reading of QCP on which *temporary ways* of bearing properties are transformed into *eternal ways* given B-theoretic Adverbialism. If we say that for something to bear a property temporarily is for it to be such that it bears some property but not always, then on this reading of QCP, Lisa, for instance, temporarily bears the property of being seated, as she is seated but not always so. In other words, Lisa does not bear-at-*t* some property, but she simply bears the property of being seated. However, very much as time-indexed properties are always had, the time-indexed instantiation-relation always holds, since if something bears-at-*t* some property *F*, it always bears-at-*t F*. As a matter of fact, bearing-at-*t* is an eternal way of bearing properties: things do not change with respect to their bearing certain properties if they bear-at-times properties. Given our example, for Lisa to be-at-*t* seated is for Lisa to always be-at-*t* seated, as it is the case at every time that Lisa is-at-*t* seated: Lisa does not change with respect to her being-at-*t* seated.

On the contrary, on the reading of QCP under consideration, Lisa changes with respect to her being seated, as Lisa temporarily bears the property of being seated. Even if B-theoretic Adverbialists attempt to explain this reading of QCP, they do so by reducing the temporary instantiation of properties to an eternal one, such as the time-indexed instantiation of properties. Since on *this* reading of QCP, the instantiation of properties is temporary, rather than eternal, B-theoretic Adverbialism is ultimately inconsistent with it.

This concludes the discussion of how the B-theory connects with QCP. In the following, and last, section, I plan to say more about Propositional Change and the B-theory of time.

[19] For a famous objection against Adverbialism see Lewis (2002), according to whom Adverbialism lands us in a version of Bradley's regress.

[20]For some objections to Adverbialism see Lewis (2002).

5. Propositional Change and the B-theory

In Section 1 of this paper, I argued that one can derive Propositional Change from either Existential Change or Qualitative Change plus Persistence: since Lisa exists, but not always, the proposition *that Lisa exists* is true, but not always; analogously, since Lisa is seated, but not always, the proposition *that Lisa is seated* is true, but not always.

From this, one can infer that since there is a reading of Existential Change and Qualitative Change plus Persistence that is inconsistent with the B-theory, there is a reading of Propositional Change that is inconsistent with the B-theory. Let's expand on that.

To begin with, let's consider B-theoretical views on which expressions of the form 'sometimes, p is true' are interpreted as 'p is true at t', namely on which propositions have truth-value relative to times. Here are two versions of the B-theoretic proposal: on one conception, analogous to the modal case where propositions are *properties of worlds*, propositions are considered as *properties of instants* (Lewis 1979); on another conception, propositions are *functions from instants to truth values* (Sider 2001, 20–1). On both views, Propositional Change is true because the following is true:

> PC-1: for some times t and t_1, there is some p such that p is true at t, but p is not true at t_1

More precisely, on the view that propositions are properties of instants, to say that the proposition *that Lisa exists* is true relative to a certain time t is just to say that t possesses the property of being a time at which Lisa exists. Hence, on this view, to say that a certain proposition changes in truth value over time is just to say that the property F of times identified with that proposition is possessed by some but not all times.

On the view on which propositions are functions from times to truth values, the truth of *that Lisa exists* depends on the instant of time we plug into the function. Hence, to say that propositions change in truth value is to say that the function f identified with a proposition delivers *truth* for some but not all times as inputs. In both views, Propositional Change turns out to be true, and therefore there is change in this sense on the B-theory.

In doing so, both views preserve the truth of Propositional Change by interpreting what is for something to be sometimes true in terms of truth-relative to times. Doing that, however, make them inconsistent with an alternative reading of Propositional Change: on this reading of Propositional Change, propositions are not true relative to times, very much as one can read Existential Change as the view on which things do not exist relative to times, and one can read Qualitative Change plus Persistence as the view on which things do not have properties relative to times. On this reading, propositions have truth values *simpliciter*.

So, on this reading of Propositional Change, some proposition, such as *that Lisa exists*, is true *simpliciter*, but not always. On this reading of Propositional Change, propositions do not change with respect to truth value if they always have the truth value they have. In fact, on this reading of Propositional Change, propositions that are true relative to times, are always true if true: if it is true at *t* that Lisa is seated, it is always true at *t* that Lisa is seated, as it is true at every time that it is true at *t* that Lisa is seated. Thus, on this reading of Propositional Change, propositions that are true relative to times always have the truth value they have. Thus, treating truth as relative to times lead to a reduction of Propositional Change, such as PC1, which is ultimately inconsistent with the reading of Propositional Change under consideration.

Conclusion

The result of this paper is that, granted different understandings of what it is for things to change, we end up having different responses to the question as to whether there is change on the B-theory. By considering three instances of change, such as Existential Change, Qualitative Change plus Persistence and Propositional Change, I argued that we can read those theses such that they are all true on the B-theory. In this sense, there is change on the B-theory. However, I claimed there are alternative readings of each of them that are false on the B-theory: so, in this other sense, there is no change on the B-theory.

Acknowledgments

I would like to thank Daniel Esmonde Deasy and Antoine Athanassiadis for many helpful conversations and comments on the material of this paper.

REFERENCES

Bacon, Andrew. 2013. ‘Quantificational Logic and Empty Names’. *Philosophers’ Imprint* 13 (24): 1–21. http://hdl.handle.net/2027/spo.3521354.0013.024.

Cameron, Ross. 2015. *The Moving Spotlight*. Oxford: Oxford University Press.

Cappelen, Herman, and John Hawthorne. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Crisp, Thomas M. 2007. 'Presentism and the Grounding Objection'. *Noûs* 41 (1): 90–109. https://doi.org/10.1111/j.1468-0068.2007.00639.x.

Correia, Fabrice, and Sven Rosenkranz. 2018. *Nothing to Come*. Berlin: Springer.

Correia, Fabrice, and Sven Rosenkranz. 2020. 'Temporal Existence and Temporal Location'. *Philosophical Studies* 177 (7): 1999–2011. https://doi.org/10.1007/s11098-019-01295-z

Deasy, Daniel. 2019. 'Characterising Theories of Time and Modality'. *Analytic Philosophy* 60 (3): 283–305. https://doi.org/10.1111/phib.12147.

Deng, Natalja. 2013. 'Fine's McTaggart, Temporal Passage, and the A Versus B-debate'. *Ratio* 26 (1): 19–34. https://doi.org/10.1111/j.1467-9329.2012.00526.x.

Dyke, Heather. 2002. 'McTaggart and the truth about time'. In *Time, reality, and axperience*, edited by Craig Callender, 137–152. Cambridge: Cambridge University Press.

Einstein, Albert. 1952. 'On the Electrodynamics of Moving Bodies'. In *The Principle of Relativity*, edited by Arnold Sommerfeld, 35–65. New York: Dover Publishing.

Haslanger, Sally. 1989. 'Endurance and Temporary Intrinsics'. *Analysis* 49 (3): 119–125. https://doi.org/10.2307/3328113.

Hawley, Katherine. 2001. *How Things Persist*. Oxford: Oxford University Press.

Heller, Mark. 1984. 'Temporal Parts of Four Dimensional Objects'. *Philosophical Studies* 46 (3): 323–334. https://doi.org/10.1007/bf00372910.

Hinchliff, Mark. 1996. 'The Puzzle of Change'. *Philosophical Perspectives* 10: 119–136. https://doi.org/10.2307/2216239.

Johnston, Mark and Graeme Forbes. 1987. 'Is There a Problem About Persistence?'. *Aristotelian Society Supplementary Volume* 61 (1): 107–156. https://doi.org/10.1093/aristoteliansupp/61.1.107.

Leininger, Lisa. 2021. 'Temporal B-coming: Passage Without Presentness'. *Australasian Journal of Philosophy* 99 (1): 130–147. https://doi.org/10.1080/00048402.2020.1744673.

Lewis, David K. 1968. 'Counterpart Theory and Quantified Modal Logic'. *Journal of Philosophy* 65 (5): 113–126. https://doi.org/10.2307/2024555.

Lewis, David K. 1979. 'Attitudes De Dicto and De Se'. *The Philosophical Review* 88 (4): 513–543. https://doi.org/10.2307/2184843.

Lewis, David K. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, David K. 2002. 'Tensing the Copula'. *Mind* 111 (441): 1–13. https://doi.org/10.1093/mind/111.441.1.

Lowe, Jonathan E. 2003. 'Substantial Change and Spatiotemporal Coincidence'. *Ratio* 16 (2): 140–160. https://doi.org/10.1111/1467-9329.00212.

Lowe, Jonathan E. 2006. 'How Real is Substantial Change?'. *Monist* 89 (3) 275–293. https://doi.org/10.5840/monist200689312.

Lowe, Jonathan E. 2009. 'Serious Endurantism and the Strong Unity of Human Persons'. In *Unity and Time in Metaphysics*, edited by Ludger Honnefelder, Benedikt Schick and Edmund Runggaldier*,* 67–82. Berlin: Walter de Gruyter.

Markosian, Ned. 2014. Time. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed November 17, 2020. https://plato.stanford.edu/entries/time/#TheBThe

McTaggart, John. M. E. 1927. *The Nature of Existence: Volume II*. Cambridge: Cambridge University Press.

Mellor, Hugh. 1998. *Real Time II*. London: Routledge.

Miller, Kristie, and David Braddon-Mitchell. 2007. 'There Is No Simpliciter Simpliciter'. *Philosophical Studies* 136 (2): 249–278. https://doi.org/10.1007/s11098-007-9074-3.

Minkowski, Hermann. 1952. 'Space and time', In *The Principle of Relativity*, edited by Arnold Sommerfeld, 73–91. New York: Dover Publishing.

Mozersky, Joshua. 2015. *Time, Language and Ontology*. Oxford: Oxford University Press.

Prior, Arthur. 1968. *Papers on Time and Tense*. Oxford: Oxford University Press.

Quine, Willard V. 1950. 'Identity, Ostension, and Hypostasis'. *The Journal of Philosophy* 47 (22): 621–632. https://doi.org/10.2307/2021795.

Sider, Ted. 1996. 'All the World's a Stage'. *Australasian Journal of Philosophy* 74 (3): 433–453. https://doi.org/10.1080/00048409612347421.

Sider, Ted. 2001. *Four-Dimensionalism*. Oxford: Oxford University Press.

Sider, Ted. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

van Inwagen, Peter. 1990. ‘Four-Dimensional Objects’. *Noûs* 24 (2): 245–55.
https://doi.org/10.2307/2215526.

van Inwagen, Peter. 2009. ‘Being, Existence, and Ontological Commitment’. In *Metametaphysics*, edited by David Chalmers, David Manley and Ryan Wasserman, 472–506. Oxford: Oxford University Press.

Wasserman, Ryan. 2006. ‘The Problem of Change’. *Philosophy Compass* 1(1): 48–57.
https://doi.org/10.1111/j.1747-9991.2006.00012.x.

Williams, Donald C. 1951. ‘The Myth of Passage’. *The Journal of Philosophy* 48 (15): 457–472.
https://doi.org/10.2307/2021694.

Williamson, Timothy. 2002. ‘Necessary Existents’. *Royal Institute of Philosophy Supplement* 51: 233–251.
https://doi.org/10.1017/S1358246100008158.

Williamson, Timothy. 2013. *Modal Logic as Metaphysics*. Oxford: Oxford University Press.

Zimmerman, Dean W. (2008). ‘The Privileged Present: Defending an “A-theory” of Time’. In *Contemporary Debates in Metaphysics*, edited by Ted Sider, John Hawthorne and Dean. W. Zimmerman, 211–225. Oxford: Blackwell.

EuJAP | Vol. 17 | No. 1 | 2021
UDC: 130.1:159.9
https://doi.org/10.31820/ejap.17.1.3

AGAINST PHENOMENAL BONDING

S Siddharth[1]

[1]National Institute of Advanced Studies
(A recognized research centre of University of
Mysore), Benglauru, India

Discussion – Received: 20/02/2021 Accepted: 24/03/2021

ABSTRACT

Panpsychism, the view that phenomenal consciousness is possessed by all fundamental physical entities, faces an important challenge in the form of the combination problem: how do experiences of microphysical entities combine or give rise to the experiences of macrophysical entities such as human beings? An especially troubling aspect of the combination problem is the subject-summing argument, according to which the combination of subjects is not possible. In response to this argument, Goff (2016) and Miller (2017) have proposed the phenomenal bonding relation, using which they seek to explain the composition of subjects. In this paper, I discuss the merits of the phenomenal bonding solution and argue that it fails to respond satisfactorily to the subject-summing argument.

Keywords: *Panpsychism; combination problem; subject-summing; phenomenal bonding; constitutive panpsychism*

Introduction

Panpsychism, the view that phenomenal consciousness or *experientiality* is possessed by all fundamental physical entities, faces an important challenge in the form of the combination problem: how do experiences of microphysical entities combine or give rise to the experiences of

Correspondence: siddharth.nias@gmail.com

macrophysical entities such as human beings? (Chalmers 2016a)[1] An especially troubling aspect of the combination problem is the *subject-summing argument*, according to which the combination of subjects is not possible. In response to this argument, Goff (2016) and Miller (2017) have proposed the *phenomenal bonding* relation, with which they seek to explain the composition of subjects. In this paper, I argue that the phenomenal bonding solution does not work. I begin by introducing the combination problem and the subject-summing argument in §1, followed by an evaluation of Goff's proposal in §2. Goff, even while proposing his solution, admits that we do not have a positive conception of the phenomenal bonding relation; Miller, however, argues that we do have such a conception. In §3, I argue against Miller's attempt at forming a positive conception. The upshot of this discussion is that a panpsychist's best bet is in pursuing non-constitutive approaches in response to the combination problem.

1. The Combination Problem

The combination problem facing panpsychism is the question of explaining how the experiences of macrophysical entities, such as human beings, emerge from the experiences of microphysical entities. The challenge in providing an acceptable answer to this question is that the combination of experiences seems unintelligible—experiences just do not seem to be the kind of things that can combine. The most famous articulation of the combination problem is by William James, who says,

> Take a hundred of them [feelings], shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it. (James 1890, 160, original emphasis)

[1] Also see Seager (1995), Goff (2006, 2009), and Coleman (2012, 2014) for more on the combination problem.

Here, James argues that a combination of 'feelings' is unintelligible, for each feeling is 'windowless'—the content of one feeling cannot seep into another, or be shared with another. Given this, if there were a 101st feeling emerging from a group of hundred feelings, such an emergence would be a 'totally new fact'—a case of *brute* emergence.

While James' argument talks of 'feelings' or experiences, it is the subjective component of experiences that has emerged as the most significant challenge—how do microphysical entities *qua* subjects (hereafter, *microsubjects*) combine to form other subjects such as macrophysical entities *qua* subjects (hereafter, *macrosubjects*)? (Chalmers 2016a) The combination of subjects, as Coleman (2012) notes, seems unintelligible and thus impossible, due to certain intuitions about the nature of subjects.[2] First, subjects seem to be *ontological unities*, or entities that, in the words of Galen Strawson (2009), are "fundamentally unified, utterly indivisible as the particular concrete phenomenon it is, simply in being, indeed, a total experiential field" (377-78). Such a unified subject experiencing a complex experience cannot be broken down into and understood in terms of multiple subjects, each experiencing one aspect or 'part' of the complex experience. In other words, a subject understood as an ontological unity cannot be broken down into 'parts'.[3] How can a macrosubject, then, be composed of microsubjects?

Closely related to the unity of a subject is its *privacy*[4]—a subject's experience is private to that subject, and it seems unintelligible how another subject could access the same token experiential content as the first subject. One could, perhaps, imagine a situation where two subjects experience identical experiential content. For example, consider a future where we have developed advanced scientific equipment that allows us to invoke specific experiences in a subject. Using this equipment, a scientist can bring about identical experience as of eating an apple in two friends. Such a situation would be a case where there are two tokens of the same experiential content (the experience as of eating an apple), each experienced by a distinct subject, and not a case where two distinct subjects

[2] Coleman (2012) himself does not use the terms that I use here—*ontological unity* and *privacy*—but makes the same point. For example, he is alluding to both unity and privacy when he says, "…our notion of a subject, is precisely the notion of a discrete, essentially *inviolable* sphere of conscious-experiential goings-on. My mind is separate from your mind, is separate from her mind, and so on. None of us has, nor can have, access to the consciousness of another, to *what it is like* for them" (Coleman 2012, 145, emphasis in original).

[3] See Barnett (2008) for more on the intuition that subjects are mereological *simples* i.e. without proper parts.

[4] The term *privacy* is borrowed from Roelofs (2019).

experience the same token experiential content. This is what is meant by the privacy of subjects—two subjects cannot experience the same token experiential content. If the experiential content of one subject cannot be experienced by another, how can the experiential content of microsubjects constitute the experience of a macrosubject? Thus, we see that the ontological unity and privacy of subjects seem to render the combination of subjects impossible.

This problem facing panpsychism has come to be known as the *no-summing-of-subjects argument* (Goff 2016) or simply the *subject-summing argument* (Chalmers 2016a). Goff articulates the argument as follows:

1. Conceptual Isolation of Subjects—For any group of subjects, instantiating certain conscious states, it is conceivable that just those subjects with those conscious states exist in the absence of any further subject.
2. Transparency Conceivability Principle—For any proposition P, if (A) P involves only quantifiers, connectives, and predicates expressing transparent concepts, and (B) P is conceivably true upon ideal reflection, then P is meta-physically possibly true.
3. Phenomenal transparency—Phenomenal concepts are transparent.
4. Metaphysical Isolation of Subjects—For any group of subjects, instantiating certain conscious states, it is possible that just those subjects with those states exist in the absence of any further subject (from 1, 2, and 3).
5. For any group of subjects, those subjects with those conscious states cannot account for the existence of a further subject (from 4).
6. Therefore, panpsychism is false (from 5) (Goff 2016, 291-92)

Premise 1 states that one can conceive of *n* subjects and their experiences without the existence of a further, *n+1th* subject. This, as noted above, is underpinned by the intuition that subjects are ontological unities. Premise 2 states that if any proposition that involves transparent concepts is conceivably true, it is also possibly true. Further, our concepts of experiential phenomena, including of subjects are transparent concepts, according to premise 3. Thus, from 1, 2 and 3, it follows that it is possible that n subjects exist without the sum of these n subjects—a further, n+1th subject—existing. If this were the case, it follows that panpsychism is false, for the existence of microsubjects cannot explain the emergence of macrosubjects (such as human subjects), leading to an explanatory gap. Faced with this explanatory gap, panpsychism loses its attraction as an alternative to physicalism and dualism.

A panpsychist could argue that the relation between microsubjects and macrosubjects is not one of composition but something else, such as ontological emergence. Chalmers (2016b) refers to panpsychist positions that propose that macroexperiences are composed of microexperiences as *constitutive panpsychism*, and those that do not as *non-constitutive panpsychism*. For the purpose of this paper, I ignore non-constitutive views, and deal only with the combination problem facing constitutive versions. I hence reserve the term 'panpsychism' for its constitutive version, unless otherwise specified.

2. The Phenomenal Bonding Response

In response to the subject-summing argument, Goff (2016) proposes the *phenomenal bonding* relation. He concedes that the mere existence of n subjects and their experiential content in themselves does not necessitate the presence of an $n+1^{th}$ subject. However, Goff argues that it is possible for the n subjects to enter into a relation—"be involved in some state of affairs" (Goff 2016, 292)—which necessitates the existence of a composite macrosubject. He calls this relation the phenomenal bonding (PB) relation. A collection of bricks in themselves do not compose a wall but do so only when they are related in a particular manner—spatially arranged in certain ways. Goff argues that similarly, a collection of subjects in themselves do not compose a further subject, but do so only when they are related by the PB relation. If we were to accept the phenomenal bonding relation, a panpsychist can respond to the subject-summing argument by arguing that premise 4 does not lead us to conclusions 5 and 6, for subjects which share the phenomenal bonding relation can account for a further subject of experience.

Goff himself admits that we have no positive conception of the PB relation. However, he contends that it is understandable why we have no conception of a relation between subjects, for we have neither perceptual nor introspective access to subjects barring our own. Despite having no positive conception of the PB relation, Goff thinks that there is no reason to deny that such a relation between subjects is possible; just as panpsychists have to identify some phenomenal property with the physical property 'charge' in a brute manner, the PB relation too will have to be identified with some physical relation (Goff 2016).

It is here that Goff's proposal faces a problem. The thrust of the subject summing argument is not that the subject-combination relation cannot be identified with some physical relation in a brute manner—we could, if we had good reasons to believe that subject combination is possible. Rather, as Coleman (2012) notes, it is that the notion of a composite subject itself seems incoherent, and thus impossible, on account of the ontological unity and privacy of subjects.[5,6] Given this, the subject summing argument ought to be understood as the problem of the unintelligibility, incoherence and thus, impossibility of relations such as the PB relation. By simply defining and stipulating the PB relation in terms of the role we want it to play, without either an argument for how subject combination is possible in the first place or a positive conception of the relation, Goff is assuming what ought to be argued for, and thus begging the question.

One can adopt such a method of defining relations in a brute manner to defend almost any unintelligible relation. For example, consider the example of 'volume' in Euclidean space. When there are two perfect cubes of 1-unit volume each conjoined together at one of their surfaces with no overlap of volume, the total volume of the newly formed cuboid would be 2 units. If one were to follow Goff's method, one can simply define a new relation called 'volume-contraction' such that when the two cubes are conjoined, the total volume would not be 2 units, but only 1.5 units. Further, it could be argued that while such a relation is unintelligible to us, this is so only because volume contraction is a brute fact about the world. One can immediately see that positing such a volume-contraction relation is wrong. Without a further positive characterisation of the volume-contraction relation, it is unintelligible to us how the total volume of two cubes with conjoined surface can be 1.5 units instead of 2 units. By proposing the volume-contraction relation as a brute posit, we would be assuming what ought to be explained (that such volume contraction is possible). Similarly, by simply defining the phenomenal bonding relation such that it fulfils the role of subject composition, Goff is assuming what ought to be explained in the first place.

3. Positive Conception of Phenomenal Bonding

Proponents of the PB relation can avoid begging the question if they are able to provide a positive conception of the relation. This is what Miller

[5] This distinction between two versions of the subject combination problem is made more clearly by Shani and Williams (2021). In the first version, similar to Goff's (2016) articulation, it is argued that no arrangement of subjects necessitates a composite subject, and hence, subject composition is impossible. According to the second, similar to Coleman's (2012) articulation, it is argued that the notion of a composite subject itself is unintelligible and incoherent, and hence, subject composition is impossible. Shani and Williams argue that the second is the stronger and more difficult challenge facing panpsychists.

[6] I would like to thank an anonymous reviewer for pressing this point.

(2017) attempts. He identifies three conditions a relation ought to fulfil to qualify as the phenomenal bonding relation:

- It must be a phenomenal relation i.e., there should be a *what-it-is-like* feel associated with it.
- Its relata should be subjects *qua* subjects.
- It must necessitate further subjects distinct from the subjects it holds between (Miller 2017).

Miller further identifies *co-consciousness* as the relation that fulfils these conditions and fits the role required of the PB relation. By co-consciousness, Miller refers to the "the relation in virtue of which conscious experiences have a conjoint phenomenology or a conjoint what-it-is-like-ness" (Miller 2017, 548). For example, when one looks at a bird while listening to it chirp, the auditory quality of the bird's chirp and the visual quality of its appearance are experienced together as a unified experience. The relation that unifies these two qualities to produce the conjoint phenomenology of our experience is what Miller refers to as co-consciousness.

Miller contends that the co-consciousness relation is known to us through our own experiences, for it feels some way for us to experience the qualities in a unified manner. That is, there is a phenomenal quality associated with the co-consciousness relation. It thus fulfils the first condition to fit the role of the PB relation. The second condition facing co-consciousness is that it ought to hold between subjects *qua* subjects. In the example given earlier, the co-consciousness relation holds between two qualities that are experienced by the *same* subject. Can we conceive of a similar co-consciousness relation that holds between two subjects instead of qualities? While Goff argued that we cannot conceive of any relations between subjects *qua* subjects because we have epistemic access through introspection only to one subject—our own—Miller contends that this limitation can be overcome. He proposes that one could form a positive conception of inter-subject co-consciousness through *analogical extension*.

Analogical extension, according to Miller, is a method of concept formation wherein we start with a case where we have a clear conception (hereafter, the *prototype* scenario), and use this conception to form a concept in another scenario that is not wholly similar to the first (hereafter, the *target* scenario). Some examples of analogical extension he gives are:

- We form a concept of the molecule as a physical object using visual representations of macrophysical entities that we have, though we do not have visual representations of molecules.
- We form a concept of the relation 'earlier than' as it applies to vast tracts of time (e.g., on cosmic scale) though we only experience events across much smaller periods (like a few second, days, months etc.).
- We form a concept of similarity of phenomenal states across subjects, though we only experience our own phenomenal states and conceive of them as being similar to each other.

In all these examples, we use the concept from a known scenario to form a concept in a different scenario. Miller contends that we can use this method to form a positive conception of the *inter-subject* co-consciousness relation based on our concept of *intra-subject* co-consciousness.

However, this approach does not work for the following reasons. First, consider the examples cited by Miller. It is important to note that in each of these examples, the relata in the prototype and target scenarios are of the same type. In the case of the 'earlier than' relation, the relata are events-in-time in both scenarios. In the case of phenomenal similarity, the relata are qualities-experienced-by-a-subject. In the case of molecules as physical objects, the relata are objects-in-space in both scenarios. In contrast to these three examples, in the case of co-consciousness, the relata in the prototype and target scenarios are *not* of the same type. The relata of the intra-subject co-consciousness relation—the prototype—are qualities experienced by a subject. On the other hand, in the case of the inter-subject co-consciousness relation—the target—the relata are *subjects qua subjects* and not qualities experienced by a subject (same or different subjects). This is as per Miller's own criteria that any relation has to meet to qualify as the phenomenal bonding relation (the second criterion listed above). Thus, unlike the examples used by Miller to outline analogical extension, the type of relata in the prototype and target scenarios are different in the case of the co-consciousness relation. For this reason, analogical extension cannot help us form a positive conception of co-consciousness between subjects *qua* subjects.

Even if we were to ignore this drawback, there is another problem in using analogical extension to form a positive conception of inter-subject co-consciousness. It seems that if one were allowed to use analogical extension to form a conception of inter-subject co-consciousness, one could use analogical extension to form positive conceptions of relations which we know are definitely not acceptable. Consider the example from earlier, of volume contraction of two cubes in Euclidean three-dimensional space, occupying 1-unit volume each, conjoined together with one overlapping surface and no overlapping portion of volume. Everyone would accept that such a volume-contraction relation is inconceivable. However, it seems that one can use analogical extension (of the sort required for inter-subject co-consciousness) to form a positive conception of the volume-contraction relation too. One could argue thus:

> Start with the following prototype scenario: volume contraction relation in cases where two cubes, each individually occupying 1-unit volume, overlap not just along a surface but in part of their volumes as well. In this case, the volume contraction relation—the relation between the cubes on account of which the total volume occupied by them together is less than 2 units—is intelligible and we have a positive conception of such a relation. Now, we can use the positive conception of volume contraction in volume-overlapping cases as the prototype scenario and form a positive conception of volume contraction in the scenario where there is overlap only along a surface (and no overlap of volume).

Would such a proposal be acceptable? Can we claim to have a positive conception of the volume-contraction relation based on this argument? Clearly, we cannot. The lesson here is that analogical extension works only in some cases. How do we know that co-consciousness relation is not like volume-contraction (where analogical extension does not work) but like phenomenal similarity (where analogical extension does work)? In the cases of inter-subject co-consciousness and volume-contraction, it is not just that we do not have a positive conception of these relations, but that we also have *a priori* reasons to believe that the relation in question leads to contradictions. For example, given our conception of Euclidean space, cubes and volumes, it is *a priori* true that the volume of non-overlapping cubes conjoined along a surface is just the sum of the volumes of the two cubes. Positing volume contraction without changing any of our initial conceptions (of what Euclidean space, cubes or volumes are) leads to a contradiction. Similarly, given ontological unity and privacy of subjects, positing co-consciousness relation between two subjects leads to contradictions—if inter-subject co-consciousness and composite subjects were possible, privacy and ontological unity of subjects would be false. In contrast, we have no *a priori* reason to believe that phenomenal similarity between qualities experienced by different subjects leads to any contradiction. Hence, we can use analogical extension to form a positive conception of this relation based on phenomenal similarity between qualities experienced by the same subject. Similarly, we have no *a priori* reason to believe that the 'earlier than' relation, when applied to vast tracts of time, leads to contradictions. Hence, we can use analogical extension to form a positive conception of this relation based on the known prototype.

To summarise, Miller's proposal to form a positive conception of the inter-subject co-consciousness relation through analogical extension does not work for two reasons. First, co-consciousness as known to us is a relation that holds between qualities and not between subjects *qua* subjects. On the other hand, the relation we want to form a positive conception of (inter-subject co-consciousness) is required to hold between subjects *qua* subjects. Second, the kind of analogical extension that is required from a positive conception of inter-subject co-consciousness can be used to form a positive conception of relations that we know are definitely not possible (such as the volume-contraction relation). This serves as a *reductio ad absurdum* against Miller's argument.

Miller's proposal is now in the same boat as Goff's—both fail to provide a positive conception of the phenomenal bonding relation. Without a positive conception, the phenomenal bonding solution simply assumes that composite subjects are possible, while the possibility of composite subjects is what the subject-summing argument questions in the first place.

4. Conclusion

The phenomenal bonding solution to the combination problem does not work, for we have no positive conception of such a relation, while we have good reasons to believe that such a relation is not possible. Goff's argument for the phenomenal bonding relation in the absence of a positive conception is not acceptable; neither is Miller's attempt at motivating a positive conception of the relation. In the absence of such a conception, proponents of this approach are guilty of begging the question against the subject-summing argument.

Where does this leave panpsychism? While it is only one approach to constitutive panpsychism that has been refuted here, it is likely that the challenge posed here would equally apply to any solution that seeks to explain combination of subjects—in the absence of a positive conception of the subject-composition relation, the solution would be guilty of assuming what ought to be argued for. Thus, in response to the combination problem, a panpsychist would be better off pursuing a non-constitutive ontology.

Acknowledgments

This research was carried out as part of my doctoral work, and forms a part of the thesis I submitted to the University of Mysore for award of the Ph.D. degree. I would like to thank Dr. Indrani Bhattacharjee and an anonymous referee for helpful comments on earlier drafts of this paper.

REFERENCES

Barnett, David. 2008. 'The Simplicity Intuition and Its Hidden Influence on Philosophy of Mind'. *Nous* 42 (2): 308–35. https://doi.org/10.1111/j.1468-0068.2008.00682.x

Chalmers, David J. 2016a. 'The Combination Problem for Panpsychism'. In *Panpsychism: Contemporary Perspectives*, edited by G. Brüntrup and L. Jaskolla. Oxford: Oxford University Press.

———. 2016b. 'Panpsychism and Panprotopsychism'. In *Panpsychism: Contemporary Perspectives*, edited by G. Brüntrup and L. Jaskolla, 19–47. Oxford: Oxford University Press.

Coleman, Sam. 2012. 'Mental Chemistry: Combination for Panpsychists'. *Dialectica* 66 (1): 137–66. https://doi.org/10.1111/j.1746-8361.2012.01293.x.

———. 2014. 'The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence'. *Erkenntnis* 79: 19–44. https://doi.org/10.1007/s10670-013-9431-x.

Goff, Philip. 2006. 'Experiences Don't Sum'. *Journal of Consciousness Studies* 13 (10–11): 53–61.

———. 2009. 'Why Panpsychism Doesn't Help Us Explain Consciousness'. *Dialectica* 63 (3): 289–311. https://doi.org/10.1111/j.1746-8361.2009.01196.x

———. 2016. 'The Phenomenal Bonding Solution to the Combination Problem'. In *Panpsychism: Contemporary Perspectives*, edited by G. Brüntrup and L. Jaskolla, 283–302. Oxford: Oxford University Press.

James, William. 1890. *The Principles of Psychology*. New York: Henry Holt & Company.

Miller, Gregory. 2017. 'Forming a Positive Concept of the Phenomenal Bonding Relation for Constitutive Panpsychism'. *Dialectica* 71 (4): 541–62. https://doi.org/10.1111/1746-8361.12207

Roelofs, Luke. 2019. *Combining Minds: How to Think About Composite Subjectivity*. New York: Oxford University Press.

Seager, William. 1995. 'Consciousness, Information and Panpsychism'. *Journal of Consciousness Studies* 2 (3): 272–88.

Shani, Itay, and Heath Williams. 2021. ‘The Incoherence Challenge for Subject Combination: An Analytic Assessment’. Unpublished manuscript, last modified February 18, 2021.

Strawson, Galen. 2009. *Selves: An Essay in Revisionary Metaphysics*. Oxford University Press.

EuJAP | Vol. 17 | No. 1 | 2021
UDC: 17.034:159.947.5
https://doi.org/10.31820/ejap.17.1.9

MOTIVATIONAL INTERNALISM AND THE SECOND-ORDER DESIRE EXPLANATION

Xiao Zhang[1]

[1] Birmingham, West Midlands, U.K.

Discussion – Received: 23/04/2020 Accepted: 20/07/2021

ABSTRACT

Both motivational internalism and externalism need to explain why sometimes moral judgments tend to motivate us. In this paper, I argue that Dreier' second-order desire model cannot be a plausible externalist alternative to explain the connection between moral judgments and motivation. I explain that the relevant second-order desire is merely a constitutive requirement of rationality because that desire makes a set of desires more unified and coherent. As a rational agent with the relevant second-order desire is disposed towards coherence, she will have some motivation to act in accordance with her moral judgments. Dreier's second-order desire model thus collapses into a form of internalism and cannot be a plausible externalist option to explain the connection between moral judgments and motivation.

Keywords: *Motivational internalism; externalism; second-order desire; practical rationality*

1. Smith's Internalist Challenge

In metaethics, motivational internalism is roughly the view according to which there is a necessary connection between moral judgments and motivation (Blackburn 1998; Gibbard 1990, 2003; Smith 1994, 1996a, 1996b, 1997). Externalism, in contrast, maintains that this connection is at best a contingent one (Brink 1989, 45-49, 1997; Copp 1995, 1997;

Correspondence: XXZ619@alumni.bham.ac.uk

Lillehammer 1997; Shafer-Landau 2003, 145-147; Sayre-McCord 1997; Svavarsdóttir 1999; Zangwill 2003, 2008). Yet, even if externalism were true, its defenders would still need to explain why at least our moral judgments usually tend to motivate us.

Michael Smith, in his ground-breaking work on this topic, criticizes externalism with the famous fetishism argument. The fetishism argument begins from an ordinary observation that is normally accepted by both internalists and externalists. Suppose that I am engaged in a discussion with a fundraiser for a local charity that aims to improve the situation of homeless people. Let us further imagine that, initially, I have no intention to donate any money to the charity because I think that some homeless people should seek employment instead of relying on charities' help. During the conversion, however, the fundraiser tries to persuade me that the majority of homeless people cannot work for different personal reasons. And, even if some of them could really work, they cannot always successfully secure jobs sufficiently quickly. The fundraiser further explains that her charity raises money not only to provide basic necessities for homeless people, but also to run political campaigns that hopefully can resolve the issues faced by the homeless. Now, if I am convinced by the fundraiser, I will begin to believe that it is morally right for me to give at least some money to the charity. Usually, we can also expect that I will thereby come to have some motivation to actually do so.

The previous case illustrates how, when you change your moral judgment about whether you should give some money to a local charity, your corresponding motivation to make the donation also tends to change accordingly. This phenomenon is so common that we can conclude that 'a change in motivation follows reliably in the wake of a change in moral judgment' (Smith 1994, 71). Both internalists and externalists then face the burden of having to explain why this is the case.

As internalists generally believe that there exists a necessary connection between moral judgments and motivation, it will be easier for them to explain the previous phenomenon. Internalists have already introduced different forms of internalism that can explain the reliable connection between moral judgments and motivation. For example, Smith puts forward a form of conditional internalism which suggests that practical rationality is a condition that must be satisfied in order for there to be a reliable connection between moral judgments and motivation. Here, we can see Smith's (1994, 61) own formulation of internalism:

> *The Practicality Requirement:* [Necessarily], if an agent judges that it is right for her to φ in circumstances C, then either she is motivated to φ in C or she is practically irrational.

Although externalists deny that there is that kind of an internal connection between moral judgments and motivation, they still need to explain why at least sometimes moral judgments tend to motivate us. As externalists claim that moral judgments at most motivate contingently, they would see to provide an explanation of why we generally tend to be motivated to act in accordance with our moral judgments from something else. At this point, Smith has assumed that the externalists would have to explain the connection between moral judgments and motivation by relying on a certain additional desire, namely the *de dicto* desire to do whatever is right (Smith 1994, 74; Smith 1997, 112).

Yet, according to Smith (1994, 75; 1997, 113), the previous externalist explanation that relies on the *de dicto* desire to do whatever is right is counterintuitive. To see this, let us imagine that an agent judges that it is right to help her friends and family and also has a corresponding desire to help her friends and family. On the externalist account, the agent's desire to help her friends and family in this case derives from her non-derivate desire—the *de dicto* desire to do whatever is right. This means that the agent desires to help her friends and family because she desires to do whatever is right, and helping her friends and family just happens to be the right thing. However, the externalist account of how an agent should be motivated does not seem to fit our ordinary understanding of good people's psychology. We would normally expect that a morally good person cares non-derivatively about the well-being of her friends and family rather than the abstract property of moral rightness. Therefore, Smith (1994, 75) argues that if an agent were motivated by the *de dicto* desire to do whatever is right, she would have a moral fetish.

2. Dreier's Second-order Desire Model

Externalists have tried to avoid Smith's fetishism objection by attempting to explain the reliable connection between moral judgments and motivation in ways that do not rely on the *de dicto* desire to do whatever is right (Copp 1995,1997; Cuneo 1999; Dreier 2000; Lillehammer 1997). In order to pursue this externalist strategy successfully, externalists will need to explain the recognized reliable connection between moral judgments and motivation in a way that is both compatible with externalism and able to avoid the fetishism objection. In this section, I focus on James Dreier's (2000) second-order desire model.

In order to explain what such a second-order desire is, Dreier begins from a maieutic end. A maieutic end is an end that is 'achieved through the process of coming to have other ends' (Schmitz 1994, 228; cf. Dreier 2000, 630). Suppose that you want to have a rewarding career, and, because of this, you want to pursue a career in medicine. Pursuing a career in medicine necessarily requires adopting other ends, such as the goal of relieving the patients' suffering and the goal of saving their lives. Effectively, the end of having a rewarding career in this case is also an end to have other ends in professional life, all of which make the career you end up choosing rewarding. Here, the end of having a career in medicine is a maieutic end because it can only be pursued through having other ends.

The previous discussion suggests that having a maieutic end requires having some other ends. In this way, a maieutic end resembles a second-order desire the having of which also requires having first-order desires. Dreier thinks that we should be able to explain the reliable connection between moral judgments and motivation by assuming that a most ordinary agent has the second-order desire to desire to do what she judges to be right. This enables us to formulate the following view:

> *The Second-order Desire Model:* Take an agent who has a second-order desire to desire to do what she judges to be right. If that agent judges that it is right for her to φ in circumstances C, then her relevant second-order desire will produce a first-order desire to φ in her, given that this desire is a desire she desires to have.

In response to Smith's objection, Dreier provides three reasons why he thinks that Smith is wrong (Dreier 2000, 636-637). Dreier first argues that nobody in the debate should complain about the relevant second-order desire itself because we should expect that an ordinary moral agent will have that desire. Imagine an agent who is not sure about what the right-making features of an action are. Suppose that the agent is then asked: if someday you are able to figure out what the right-making features of the action are, would you hope to be motivated by those right-making features? As Dreier puts it, we would certainly expect the agent to say 'yes'—to confirm that she would desire herself to be motivated by the right-making features of an action in the future. If the agent instead hoped that she would not be motivated by those right-making features in the future whatever they are, she would not seem to count as a good moral agent.

Secondly, Dreier also considers whether the relevant second-order desire would play too much of a role in the previous account of moral motivation, which he believes to be the most important concern behind the internalist objections. If that were the case, internalists could argue that the relevant second-order desire is merely another kind of the *de dicto* desire to do whatever is right. Yet, according to Dreier, the relevant second-order desire plays only a limited role in the account—a role that is not objectionable. Once the second-order desire in question produces the relevant first-order desire in the wake of a change in one's moral judgments, the relevant second-order desire does not need to maintain the first-order desire after that. Consequently, the relevant second-order desire plays, according to Dreier, only a very limited causal role in explaining how an agent becomes motivated to act in accordance with her moral judgments.

Dreier's own illustration of this second point is the following (Dreier 2000, 636-637). Let us imagine that David judges that it is right to stop using chimps in medical research. In this case, the relevant second-order desire in him would generate a first-order desire to stop doing so in the way described above. After this point, David's first-order desire can play a motivating role by itself, and it can even produce other first-order desires. For example, that first-order desire to end using chimps in medical research can generate a new first-order desire to use other substitutes or a first-order desire to stop other researchers who continue to use chimps in their medical research. That said, all of David's first-order desires in this case are *de re* desires that are not derivative of any other first-order desires and so they cannot be accused of being fetishistic.

The third and last point Dreier makes can be seen as a further development of his second claim. Dreier claims that the resulting first-order desires are not conditional on rightness. To see why this would be the case, Dreier (2000, 637) invites us to compare the following two formulas that both try to describe David's relevant first-order desire in the previous example:

1. David desires that David does *x*
2. David desires that David does *x* so long as *x* is right

According to Drier, the first formula describes David's first-order desire in the previous case correctly, whereas the second formula appears to misunderstand David's first-order desires. Arguably, if David comes to have the first-order desire to end using chimps in medical research, this first-order desire will thereafter exist and function without being influenced by the judgment that it is wrong to use chimps in medical research. This view coheres with Dreier's second point according to which the relevant second-order desire plays only a limited causal role in an agent's process of acquiring motivation.

3. The Second-Order Desire Model Collapses into A Form of Internalism

In the previous section, I examined whether the second-order desire model, as a version of externalism, is able to explain the reliable connection between moral judgments and motivation in a non-fetishistic way. At least for the purpose of this paper and on the basis of Dreier's responses, I am willing to grant that perhaps it can. Rather, what I want to challenge next is whether this account is compatible with externalism itself. In the rest of this section, I will, however, try to argue that the relevant second-order desire required by Dreier's model is a constitutive requirement of rationality itself. This means that the fundamental problem of the second-order desire model is that it collapses into a form of internalism and so the response cannot be available to externalists.

Let us then begin from of what practical rationality is generally thought to consists.[1] According to Michael Smith himself, in order to be fully rational, an agent has to meet four requirements: she should have no false beliefs, she should have all the relevant true beliefs, she should have a systematically justifiable set of desires and she should not suffer from any physical or psychological disturbances (Smith 1994, 156-161; 1995, 112-116; 1996a, 160; 2002, 311-315).

To see why the relevant second-order desire to have the first-order desires that match one's moral judgments would be required by the previous

[1] Here, it is worthwhile to consider an objection to Smith's concept of practical rationality, which was first stated by Alex Miller (2003, 221), and echoed by Roskies (2003, 53) and Strandberg (2013, 29-31). On Miller's view, Smith tends to think that, when the additional condition—being practically rational—has not been met, something blocks the normal way in which moral judgments give rise to motivation. This entails that, when the relevant condition has been met, the cases in which moral judgments fail to motivate cannot exist. This entailment thereby leads to the concern that Smith formulates his view merely by precluding all the situations where the counterexamples could be put forward. All that is left of internalism is thus the claim that internalism is true except when internalism is not true. Actually, it seems that the condition in which there is no connection between moral judgments and motivation has been given an insubstantial characterization. Because of this, the resulting forms of conditional internalism become trivial.
Yet, Smith's conditional internalism will not be trivially true simply because it formulates the condition which it then uses to deal with counterexamples. In this section, I will discuss that in order to count as fully rational, an agent has to satisfy four requirements. This description of the requirements for being fully rational provides an informative, substantial characterization of the condition. In Section 4.2, I will provide an independent, substantial explanation of the condition in which moral judgments must lead to motivation. Since moral agents in the counterexamples fail to satisfy the proposed internalist condition, it is understandable that they remain unmotivated by their relevant moral judgments.

constitutive requirements of rationality, we need to focus on the third requirement of rationality—that of having a systematically justifiable set of desires. By this third requirement, Smith means that a rational agent must have coherent and unified sets of desires. This is to say that a rational agent's desires do not first of all tend to conflict with each other—they do not pull the agent towards different directions at the same time. Additionally, the desires in the set support each other: they are in harmony with each other.

We can then consider in more detail how we should understand what it is for a set of desires to be coherent and unified. For example, if I feel cold, I may come to have a desire to turn up the heating and to put on more clothes. I might also come to have a desire not to open the window as doing so would bring even more cold air into the room. I might even have a higher-order desire to desire to take measures to keep the room warm. In addition, if there are other people in the room, I might continue to desire that those people also both desire and do as I do. My desires in this case are what Smith calls a systematically justifiable set of desires. It is evident that my desires aim at the same direction and they support one another rather than contradict with each other. Because of this, having such a set of coherent and unified desires should be thought of as rational—the desires in the set will finally lead to achieving what you most care about.

Of course, sometimes there will unavoidably be situations in which you will have different first-order desires that are not very coherent or unified, and sometimes those desires may even contradict each other. For example, you may have a set of desires concerning which methods of transport you would like to use for travelling. This set can include a desire to take a bus to work, a train when travelling to other cities nearby, and a desire to fly when you go abroad. At least initially, could the previous set of desires be made more coherent and unified?

At this point, Smith argues that a fully rational agent's disposition towards coherence and unity will under some circumstances change her desires (Smith 1994, 159-161; 1997, 94). The rational disposition towards coherence and unity can, for example, produce general desires that will support the more specific desires and also these new general desires will in some cases destroy some of the previous first-order desires that do not fit them. Smith would then ask you to consider whether the previous specific desires would be more systematically justifiable if a more general desire which could justify and explain those specific desires were added to your psychological make-up. For example, you could add a general desire—a desire to choose the most affordable and convenient means to go where you want to go—to your set of desires. This general desire could justify the previous set of desires by explaining why you would not want to travel to a faraway country by bus given that it is obvious that traveling by plane to another country is often more convenient and more economical. With the new added general desire, the relevant set of desires will be more systematically justifiable and thus more unified and also rationally preferable.

Analogously, we can argue that the second-order desire to desire to do what you judge to be right would be required by rationality, exactly in the same way as the general desire to travel in the most economical and convenient way is required in the case above. Consider, for example, an agent who has various moral desires, desires to treat her friends well, to keep her promises, to not cause physical harm to anyone and so on. These first-order desires are all distinct from one another because they are all related to different kinds of behaviour. However, a second-order desire to desire to do what one judges to be right would in this case justify and explain why the agent has the previous desires to do all the different things that she judges to be right. It could then be argued that the desiderative set also becomes more rationally preferable as a consequence of having that second-order desire.[2]

If, as I have just argued, the relevant second-order desire discussed by Dreier is required by the fundamental constitutive requirements of rationality—coherence and unity, having a second-order desire to do what one judges to be right (that will produce a first-order desire) is a matter of fulfilling a constitutive requirement of rationality. This means that a rational agent, who satisfies the constitutive requirements of rationality, would have the second-order desire to desire to do what she judges to be right. Assuming that the agent must also have the desires that she desires to have (given her constitutive disposition towards coherence), Dreier's second-order desire model thus entails that a rational agent will necessarily have at least some motivation to act in accordance with her moral judgments. If the agent did not, she would be less coherent and less rational as well.

[2] Someone might worry that the argument in this section does not show that rationality requires an agent to have the relevant second-order desire. Rather, the argument only shows that rationality requires an agent to have the relevant second-order desire if she has already had various moral desires. This worry seems to treat various moral desires as a premise of my argument. But that is not the reason why I employ various moral desires in my example. Those various moral desires are employed merely to show that the relevant second-order desire can make an agent's various moral desires more rationally preferable when she conducts moral deliberation. It would be too demanding to assume that an agent cannot have different moral desires when conducting moral deliberation.

This consequence furthermore means that Dreier's second-order desire model actually entails a form of conditional internalism the acceptance of which creates a commitment to a necessary connection between moral judgments and motivation. As a result, it seems that what Dreier proposes on the basis of the relevant second-order desires cannot be an entirely new externalist solution to the problem of moral motivation. Rather, the second-order desire model has actually collapsed into a form of internalism.

4. Responses to Objections

I have argued that the second-order desire model collapses into a form of conditional internalism. Before concluding, I shall consider two objections which the externalists might put forward to my argument. The first objection claims that the second-order desire model does not collapse into a form of internalism. The second objection further claims that, even if the second-order desire model collapses into another proposal, that proposal cannot be a form of internalism. In response, in the rest of this section, I will argue that both potential objections are implausible.

4.1 A Response to Sayre-McCord's Objection

The externalists may refuse to accept my argument that the second-order desire model collapses into a form of internalism. They could, for example, challenge the claim that a more general back-ground desire can make a given set of desires more coherent, unified and therefore also more rational (Sayre-McCord 1997, 75). If, arguably, a more general desire cannot make a given set of desires more coherent and unified, then an agent who comes to have such desires cannot be thought of as more rational. Furthermore, it could also be claimed that the relevant second-order desire of Dreier's model cannot contribute to making an ordinary moral agent more rational either. If this were right, we would have no reason to believe that the second-order desire model collapses into a form of internalism as I have suggested.

To illustrate this concern, we can consider Geoffrey Sayre-McCord's case of choosing an ice cream (Sayre-McCord 1997, 75). If we suppose that Smith's view is true, then, if I have a desire for coffee ice cream, my set of desires could be argued to exhibit more coherence and unity if a more general unconditional desire for ice cream were added to my current desiderative profile. My set of desires could be claimed to be more coherent and unified because the newly added general desire would be able to explain why I desire to enjoy coffee ice cream.

In this situation, eating coffee ice cream will satisfy both my desire for coffee ice cream and my general unconditional desire for ice cream. Sayre-McCord then objects that it is not plausible to think that satisfying the previous two desires would make me any more rational than how rational I am with merely my original desire (Sayre-McCord 1997, 76). So, he thinks that adding more desires, including more general desires, to a desiderative profile cannot itself enhance an agent's rationality as Smith suggests.

It seems that the crucial dispute between Sayre-McCord and Smith is over whether adding a more general desire to an agent's desiderative profile can make the agent more rational. I think that Sayre-McCord is right in claiming that merely satisfying more desires cannot itself make an agent more rational. Yet, the number of satisfied desires is not what Smith's view of rationality is based on. The key point of his view is that sometimes adding a more general desire to an existing set of desires can make the set more coherent and unified. This is the real reason why Smith would think that adding a more general desire can in the previous case make my desire set more rationally preferable.

In the previous case, it is supposed that I initially have a desire to have coffee ice cream. Usually, my desire to have coffee ice cream will move me to get it when it is available. Despite this, if I only had this one desire, I would presumably often ask myself: why do I choose to have coffee ice cream rather than other flavours or even other kinds of dessert (Smith 1997, 94)? The desire to have coffee ice cream itself does not seem to be able to answer this question. Yet, if a general, unconditional desire to eat what I enjoy eating, for example, were added to my desire set, this more general desire would be able to explain my specific desire to have coffee ice cream. The desire to eat coffee ice cream would no longer appear to be arbitrary, but rather it would be well-supported by the more general desire. In this way, my desire set has turned out to be more coherent, unified and thus more rationally preferable.

4.2 A Response to Bromwich's Objection

The externalists may continue to reject my argument by presenting Bromwich's (2010, 344; 2011, 75) challenge which claims that Smith's conditional internalism fails to capture the necessary connection between moral judgments and motivation. According to Bromwich, once practical rationality is inserted between moral judgments and motivation, it becomes unclear whether the motivation is still internal or built in to those moral judgments. Factors external to moral judgments—such as practical rationality—are now necessary for moral judgments to cause motivation.

Arguably, it is not an agent's moral judgments that produce motivation, but rather it is an agent's disposition towards coherence required by practical rationality that produces motivation to act accordingly (Bromwich 2011, 75; Svavarsdóttir 1999, 165). As a result, the necessary connection that is supposed to exist between moral judgments and motivation actually exists between an agent's disposition towards coherence and her motivation. If Bromwich's objection were right, then Smith's conditional internalism would not be regarded as an internalist view, and this furthermore means that even if the second-order desire model does collapse, it does not collapse into a form of internalism.

In order to see why Bromwich's challenge is implausible to accept, let us first consider why Smith thinks that moral judgments produce motivation essentially. Smith begins by analysing the concepts employed in moral judgments (in a similar way as we could try to analyse other concepts). Take the concept of 'a bachelor' for illustration. When I think that Mark is a bachelor, what I am thinking of is that Mark is a male and unmarried. This is because the concept of bachelorhood can be reductively analysed in terms of being male and unmarriedness. Similarly, Smith claims that moral concepts can be reductively understood to be about reasons for actions (Smith 1994, 62). When an action is judged to be right or wrong, a part of this thought is always that there are at least some reasons either to perform or refrain from doing the action.

The content of an agent's moral judgments, that is, the content of the thought that there are reasons for actions can be investigated further. Smith's (1994, 151-152) proposal is that, when an agent believes that there are reasons for her to carry out a certain action, she essentially believes that her fully rational version would want her to do that action in the actual situation she is in. So, for example, an agent's judgment that it is right to help innocent people is a judgment about what she has reasons to do. And, the content of this judgment, according to Smith, is that her fully rational version would want the agent to help the innocent in the situation she is in.

At this point, based on the content of her moral judgments, the agent has two options: either she will desire to help those innocent people to get rid of the plight or she will lack that desire. As I have discussed in Section 3, because practical rationality can be thought to consist at least in part of a disposition to have coherent mental states, a practically rational agent is disposed towards coherence. It is then plausible to suggest that a desire to help innocent people coheres better with the belief according to which the agent's fully rational version would want her to help the innocent. This means that, when an agent is practically rational, she will desire to act in accordance with her moral judgments, or so Smith argues.

The above discussion shows that an agent's disposition towards coherence plays an important role in Smith's explanation of how moral judgments produce motivation. But it would be implausible to thus claim that there actually exists a necessary connection between an agent's disposition towards coherence and her motivation as Bromwich does. Bromwich's challenge claims that it is an agent's disposition towards coherence that produces motivation. In contrast, Smith's own explanation suggests that it is an agent's moral judgment that produces corresponding motivation. On Smith's view, an agent's disposition towards coherence does not cause motivation, but rather it causes her motivation to cohere with the content of her moral judgment. The problem with Bromwich's challenge is that it mistakes the causal role that a moral judgment plays for the causal role that an agent's disposition towards coherence plays within Smith's conditional internalism. Given that Bromwich's challenge is based on a misunderstanding of Smith's view, we should still believe that there exists a reliable connection between moral judgments and motivation within Smith's internalist framework.

5. Conclusion

In this paper, I have argued that Dreier's second-order desire model collapsed into a form of internalism and thus cannot be available as an externalist option. We normally assume that a rational agent who has made a genuine moral judgment about what the right thing to do is has at least some motivation to perform that action, otherwise, she would be thought of as less coherent. I explained how rationality can be used to account for previous intuitions. It turns out that rationality itself requires that a rational agent has a second-order desire to desire what she judges to be right so that this desire makes the agent's set of desires more unified and coherent. Furthermore, because a rational agent who has the relevant second-order desire is disposed towards coherence, she will have some motivation to act in accordance with her moral judgments. As a consequence, Dreier's second-order desire model collapses into a form of internalism that is conditional on rationality. This also means that Dreier's second-order desire model cannot be used as an externalist alternative to explain the reliable connection between moral judgments and motivation.

Acknowledgments

I would like to thank two anonymous reviewers for this journal for their extremely helpful comments.

REFERENCES

Blackburn, Simon. 1998. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon Press.

Brink, David. 1989. *Moral Realism and the Foundation of Ethics*. Cambridge: Cambridge University Press.

———. 1997. 'Moral Motivation'. *Ethics* 108 (1): 4-32. https://doi.org/10.1086/233786.

Bromwich, Danielle. 2010. 'Clearing Conceptual Space for Cognitivist Motivational Internalism'. *Philosophical Studies* 148 (3): 343-367. https://doi.org/10.1007/s11098-008-9331-0.

———. 2011. 'How Not to Argue for Motivational Internalism'. In *New Waves in Ethics*, edited by Thom Brooks, 64-87. London: Palgrave Macmillan.

Copp, David. 1995. 'Moral Obligation and Moral Motivation'. *Canadian Journal of Philosophy* Supplementary Volume 21: 187-219. https://doi.org/10.1080/00455091.1995.10717438.

———. 1997. 'Belief, Reason, and Motivation: Michael Smith's "The Moral Problem"'. *Ethics* 108 (1): 33-54. https://doi.org/10.1086/233787.

Cuneo, Terence. 1999. 'An Externalist Solution to the "Moral Problem"'. *Philosophy and Phenomenological Research* 59 (2): 359-380. https://doi.org/10.2307/2653676.

Dreier, James. 2000. 'Dispositions and Fetishes: Externalist Models of Moral Motivation'. *Philosophy and Phenomenological Research* 61 (3): 619-638. https://doi.org/10.2307/2653615.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Clarendon Press.

———. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Lillehammer, Hallvard. 1997. 'Smith on Moral Fetishism'. *Analysis* 57 (3): 187-195. https://doi.org/10.1093/analys/57.3.187. https://doi.org/10.1093/analys/57.3.187.

Miller, Alexander. 2003. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.

Roskies, Adina. 2003. 'Are Ethical Judgments Intrinsically Motivational? Lessons from "Acquired Sociopathy"'. *Philosophical Psychology* 16 (1): 51-66. https://doi.org/10.1080/0951508032000067743.

Sayre-McCord, Geoffrey. 1997. 'The Metaethical Problem'. *Ethics* 108 (1): 55-83. https://doi.org/10.1086/233788.

Schmidtz, David. 1994. 'Choosing Ends'. *Ethics* 104 (2): 226-251. https://doi.org/10.1086/293599.

Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell Publishing.

———. 1995. 'Internal Reasons'. *Philosophy and Phenomenological Research* 55 (1): 109-131. https://doi.org/10.2307/2108311.

———. 1996a. 'Normative Reasons and Full Rationality: Reply to Swanton'. *Analysis* 56 (3): 160-168. https://doi.org/10.1093/analys/56.3.160.

———. 1996b. 'The Argument for Internalism: Reply to Miller'. *Analysis* 56 (3): 175-184. https://doi.org/10.1093/analys/56.3.175.

———. 1997. 'In Defense of "The Moral Problem": A Reply to Brink, Copp, and Sayre-McCord'. *Ethics* 108 (1): 84-119. https://doi.org/10.1086/233789.

———. 2002. 'Evaluation, Uncertainty and Motivation'. *Ethical Theory and Moral Practice* 5 (3): 305-320. https://doi.org/10.1023/A:1019675327207.

Strandberg, Caj. 2013. 'An Internalist Dilemma—and an Externalist Solution'. *Journal of Moral Philosophy* 10 (1): 25-51. https://doi.org/10.1163/174552412x625754.

Svavarsdóttir, Sigrún. 1999. 'Moral Cognitivism and Motivation'. *The Philosophical Review* 108 (2): 161-219. https://doi.org/10.2307/2998300.

Zangwill, Nick. 2003. 'Externalist Moral Motivation'. *American Philosophical Quarterly* 40 (2): 143-154.

———. 2008. 'The Indifference Argument'. *Philosophical Studies* 138 (1): 91-124. https://doi.org/10.1007/s11098-006-9000-0.

BOOK REVIEW

Maria Paola Feretti
THE PUBLIC PERSPECTIVE: PUBLIC JUSTIFICATION AND THE ETHICS OF BELIEF
Rowman & Littlefield, 2018
ISBN-10 1786608723
ISBN-13: 978-1786608727
Hardback, $126.00
e-Book, $38.00

IVA MARTINIĆ
University of Rijeka, Faculty of Humanities and Social Sciences

In her book, *The public perspective: public justification and the ethics of belief,* Maria Paola Ferretti discusses in an interesting and original way the question of how moral and political rules can be made justifiable to all individuals living in pluralistic societies, where each person has a potentially different notion of the good life. This is a fundamental question in a free and pluralist society. Ferretti adheres to the idea that government activities must be justified to all citizens for public purposes, so that those who are subject to them can freely assent. This refers to the philosophical concept of public justification. Ferretti contributes to the debate by supporting the idea that public justification is only conceivable if people agree on a shared ethics of belief. Through this concept, she refers to a collection of epistemic and moral principles that lead to the reshaping of the beliefs that form our public worldview. Ferretti claims that Locke's concept of the ethics of belief is firmly founded in the liberal tradition and it might be revitalized to address important aspects of contemporary liberalism.

The book is divided into six chapters. After the introduction, Ferretti launches a debate in chapter 2, Public Reasoning and Agreement, by contrasting two prominent models: John Rawls's and Gerald Gaus's, to examine the link between justification and agreement in liberal political theory. She moves to chapter 3, The Ethics of Belief and the Liberal Tradition, where she advocates John Locke's ethics of belief as a theory that may be useful in reducing conflict in situations where it cannot be eliminated and disagreement should be accepted rather than solved. In

Correspondence: ivamartinic@gmail.com

chapter 4, Having Reasons and Giving Reasons, Ferretti proposes that, rather than focusing on people as reasonable, we should focus on reasonable beliefs. She illustrates the difference between beliefs that are apt for public justification and beliefs that are not. In chapter 5, Facing Disagreement, she discovers points of agreement between people who hold opposing views but live in the same community rather than in separate communities. She examines and assesses her ideas in the real world in chapter 6, Equal Freedom, where she explains that her concept of equal freedom restricts the types of justifications that can be used to justify proposals for public norms. When determining whether a proposal is justified, we must consider if it is compatible with others's equal freedom, or whether it respects them as moral agents. Ferretti finishes with chapter 7, Liberal Multiculturalism, and explores the idea of respect for people as free, which involves respect for the fact that people exercise their freedom in groups.

Ferretti introduces public justification as a debate that takes place on multiple levels, including epistemology, metaethics, institutional design, and tolerance. She tells her readers that not all sides of the argument will be considered, and that many questions will have to be overlooked. Ferretti focuses on the relevance of free moral agency and the idea that people do not reach the same reasonable conclusions while exercising free moral agency.

Ferretti begins the topic in Chapter 2 with a focus on the link between public reason, justification, and agreement in liberal political theory, with John Rawls's consensual model and Gerald Gaus's convergence model being discussed as two contrasting approaches to the use of public reason. The agreement on principles of public order, according to both Rawls and Gaus, must be guided by reasons that are recognized as such from the evaluative point of view of each citizen.

Ferretti criticizes Rawls's shared agreement. According to this conception, public justification is based on reasons that can be expected to be shared by reasonable people when entering public debates. Ferretti's objection is that it is not clear how we will create room for a new consideration that could indicate to us that generally accepted premisses are wrong to uphold the principles of justice, if we have to reason using a premiss that has already been accepted. Ferretti argues that Rawls's idea of consensus is conservative, given the fact that the reasons currently accepted do not provide the resources to address some of the ongoing irregularities. This suggests that we need new perspectives in public debates. Also, in the Rawlsian model, despite public justification being based on shared reasons, it is possible to support injustice against minorities, because members of minorities often offer reasons that are not commonly accepted, but also reasons that are not 'shared' in the normative sense indicated by Rawls. Likewise, in Rawls's view, a number of negotiations and vetoes on rule proposals can be dismissed as unreasonable and decisions that are challenged are declared to be justified, despite the challenges. This represents a case of undesirable exclusion of minorities.

Gaus considers it wrong to select reasons that may enter the public justification of a law and to establish that only some reasons are appropriate in public justification, in the way Rawls does this. Instead, the role of public debate is to articulate the values and reasons that a wide variety of people support despite their diverging worldviews. People should be able to express the reasons for their support, both in public debate and when voting on political issues. Justification is obtained when the variety of reasons employed by people with diverging worldviews converge on the same public decisions. When there is no convergence, a proposal is defeated and deemed unjustified in the process of public justification, by virtue of the opposition of some people. However, Ferretti claims that, in contrast to Gaus, she presents his model with idealized people, referred to as Members of the Public, rather than real-life people. Thus, we face the problem of what to do with real opinions that real people with all sorts of dubious, or, even flawed, epistemic engagements (rather than idealized members of the public), express for or against certain proposals. In a public debate, people sometimes cast doubt on very well-established concepts, for example by casting doubt on widely accepted scientific knowledge. We therefore seem to need some normative guidance to know when these objections have a place in public justification. In opposition to Gaus's thesis that it is sufficient for a law to be justified from all perspectives, and that the common result counts as public justification, Ferretti emphasizes that the public justification of a law implies both epistemic and motivational reasons.

Thus, Ferretti dismisses both models, saying that Rawls's concept of shared reasons is conservative and internally exclusive and, although the joint agreement reached by Gaus seeks to be more inclusive, it separates public reasoning from public justification. In both models, Ferretti argues, the critical role of public reason is threatened in certain key ways.

The theory of public justification offered by Ferretti is based on reconnecting public reasons with the actual beliefs of people about the reasons that they (and others) have, and the arguments that citizens exchange with each other to ensure that agreement is not static or passively accepted but open to the scrutiny of alternative evaluative perspectives. She aims to show how the reasons that we have, and the reasons that we give others, are interconnected and influence each other by exploring the ways in which agreement and disagreement are both vital for a liberal society, and how the reasons that we have and the reasons that we give to one another are interconnected and exercise mutual influence (30). Thus, different notions of good can be a source of disagreement, but despite such disagreement, we recognize the fundamental importance of treating others as free moral agents, which, according to Ferretti, requires justice.

In chapter 3, she argues that a moderate interpretation of foundationalism is shown to be appropriate for a theory of public justification. She introduces Locke's ethics of belief, or belief governance, by stating that a well-grounded belief needs not to be indefeasible. The concept of the ethics of belief assumes that we may be held responsible for what we believe, which requires that we exercise deliberate control over our beliefs. This includes gathering information and deciding whether to accept or reject it. What individuals can be held responsible for are such actions in the process of belief formation. The focus of the discussion is on the rules that we use to convey evidence and weigh probability, rather than on the beliefs themselves.

In a morally pluralistic society, Ferretti argues for a rational examination of beliefs in which belief reformation and governance should be at the core of a project for public ethics. In her view, Locke's theory of beliefs and the idea of alethic obligation represent a valid approach to these ends. Locke asserts that each of us has an obligation to believe what is true, and thus presents the first rigorous formulation of what has come to be known as alethic obligation (from the Greek aletheia, truth) (44). Alethic obligation applies indirectly as a requirement to resist doxastic practices that do not have truth (or high probability) as a main criterion of inquiry. Through this, Ferretti provides a novel answer by combining moral and epistemic factors in a way that allows us to bear responsibility for our views. We must assert that the reasons we have are true. Citizens should be responsible believers and defer to experts, according to Ferretti, who are able to match their beliefs with those held by the scientific community.

Such a viewpoint, in my opinion, has a flaw. The issue that I want to highlight here is that Ferretti's theory does not respond to the demand that she has established for a theory of public justification. This is the requirement that those who are subject to a government can freely assent to its decisions. In fact, Ferretti does not specify what we should do about the problem of lay people not understanding the reasons of experts due to their lack of scientific terminology or because they have no political knowledge, which is why they turn out to be irresponsible and irrational. It appears that the value of the public's perspective and the justified judgment of experts is limited to those who have previously done their homework on the subject. But most people do not have the ability to question experts (even when they are wrong or when there is no consensus in the scientific community). They cannot recognize experts or when someone is just pretending to be one, and then they turn to untrustworthy and easier to understand sources. An example of this could be the many conspiracy theories and video essays on the global pandemic currently going on. Thus, Ferretti's Lockean proposal does not satisfy the requirement that government activities must be justified to all citizens for public purposes.

Ferretti builds on Locke in chapter 4, pointing out that citizens have an alethic obligation to employ the method of probability when they want to convey their reasons to others. She starts with the concept of the ethics of belief, which assumes that we may be held responsible for what we believe, which requires that we exercise deliberate control over our beliefs. According to Locke's theory, the nature of beliefs contains an essential ambiguity, which provides answers to the question of how to approach different perspectives. This includes gathering information and deciding whether to accept or reject it. What individuals can be held responsible for are their actions in the process of belief formation. As a result, the focus of the discussion is on the rules that we use to convey evidence and weigh probability, rather than on the beliefs themselves.

She opposes the method of probability to a subjective approach, and she shows how conflict can develop if we understand the alethic obligation in a subjective way, using an example of Galileo's beliefs that did not derive from the probability method. This method selects the kind of beliefs that are properly employed in public justification. On the one hand, there are non-givable reasons based on intimate experience, and reasons that are contingently or necessarily un-givable, that are not properly employed in public justification. On the other hand, there are reasons that are considered in public justification. Such are beliefs that correspond to the shared ethics of beliefs.

Ferretti returns to this topic in the next chapter, Facing Disagreement, stating that it is difficult to decide which proposals or positive rules should be endorsed from a public perspective and how much personal freedom should be granted. Namely, justification, as defined by public reasoning, can resolve a wide range of issues, but it also has significant drawbacks. Thus, justified public laws and choices should be upheld strongly, but with a fallibilist mindset that permits us to perceive them as perpetually revisable and changeable.

In chapter 6, Ferretti advocates a view of freedom that is consistent with the idea that different people have different ideas about what is good. In such a conception, freedom is equal for all, which implies, too, that it is limited for each person. This, she believes, ensures or defends a certain degree of independence from outside pressures.

She continues with chapter 7, by stating that cultural claims can be described as claims to freedom in cases where others, or the government, claim interference with cultural practices (152). In this context, Ferretti argues that respect for people as free requires (i) respect for the fact that people exercise their freedom in groups and (ii) the limits of public justification when the matter is constituted by deciding what people in groups should be free to do.

She responds to Brian Barry's remarks in this section. He says that the liberal commitment to equality requires similar treatment for all people, irrespective of their sex, race, or culture, with no space for a 'politics of difference'. Ferretti agrees with Barry that some public norms must universally apply to cultural groups, irrespective of their differences, but she deems his expectations to be excessively strict. The reason for her view is based on the fallibilism and limitations of public justification, she argues in the previous chapters. Thus, when the matter is represented by cultural claims, Ferretti believes that a liberal conception of public life must not neglect the fact that people disagree about public issues, often in extreme ways. Consequently, the prevailing culture of society should not be used to justify broad norms that overlook such disagreements. Instead, the goal is represented by the harmonious coexistence of freedom and equality of citizens, which implies some restrictions on interference in inside group relations. As a result, she argues that the reasons for multicultural policies are grounded in an idea: (i) of respect for people as free, which requires respect for the fact that people exercise their freedom in groups and (ii) on the limit of public justification in relation to decisions that concern whether people should be free to exercise their cultural practices inside their communities.

I find (ii) problematic. Ferretti's strategy on the question of multicultural respect for communities seems objectionable to me. In particular, I think that in her view there is a hardly sustainable distinction between the private and the public sphere. This is problematic, on the one hand, because some multicultural claims are explicitly directed to the public domain. An example is the recent Vatican protests against a newly proposed law, called the Zan Law, that would punish discrimination and incitement to violence against the LGBT community, women, and people with disabilities. The Vatican claims the law will legally restrict the religious freedoms guaranteed by the treaty between the Vatican and the Republic of Italy. According to the Vatican, with the protection of these groups, Catholics could also face legal action for expressing opinions on LGBT issues. However, the Prime Minister of Italy, Mario Draghi, rejected Vatican's complaint in the name of the secularity of the state. Here we have an illustration of the problematic distinction between the domains that are defined as public and those that are not. Thus, it is not sufficiently clear which space needs to be excluded from the interference through public justification and universal norms.

Some internal cultural practices, on the other hand, are completely unacceptable in terms of universal justice and universal rights. Here, we see the dubious sustainability of some cases of protection of the non-public sphere, because too important universal norms and values are at stake. Let me return to Barry's assertion that the liberal commitment to equality requires similar treatment for all people, irrespective of their sex, race or culture, with no space for a 'politics of difference'. Such politics include exemptions of parents from some forms of care of their children, like health care, based on cultural or religious reasons. A good illustration of this is a case of denial of treatment that happened in 2016 in Rijeka (Croatia). A nine-year-old was diagnosed with lymph node cancer, and when he arrived at the hospital, his neck was visibly swollen. But after a day in the hospital, his parents pulled him out of the hospital, despite the doctor's insistence that he should receive chemotherapy. They signed the outing explaining that they wanted a second opinion and subjected him to alternative methods of treatment because the child's father claimed that chemotherapy was 'war poison'. A further example is represented by the illustration Barry gives of the Jewish and Muslim traditions to slaughter animals in conformity with particularly cruel practices.

Ferretti criticizes Barry's argument as a harsh expression that prevents tolerance and the freedom of people, who, according to her, must have the opportunity to live according to the reasons that, in their views, justify practices. However, as expressed in the examples, allowing religious and other cultural reasons to justify practices in the public domain makes it difficult to establish a boundary of the legitimacy of these reasons and the practices that they justify. The question is important, because, by allowing free choices to members of a group with certain customs and principles, others are deprived of their freedom of choice, or other basic rights. Thus, even Ferretti's solution of multicultural policies and reasons does not meet the condition of a liberal state that all citizens be treated as equal and free.

To conclude, Ferretti's proposal has relevant merits. She has made a vital and creative addition to the debate on public justification with The Public

Perspective. She succeeds in reminding liberal theory of some of its foundations, with the original contribution, in the contemporary context, of revitalizing the Lockean probability method and the ethics of belief in such a way that they can be utilized as guidelines for contemporary liberal theories of democracy. Thus, her book offers an original proposal that inserts, in an interesting way, epistemological considerations into a public justification theory respectful of pluralism.

BOOK REVIEW

Marcin Będkowski, Anna Brożek, Alicja Chybińska, Stepan Ivanyk, and Dominik Traczykowski (Eds.)

FORMAL AND INFORMAL METHODS IN PHILOSOPHY

Brill | Rodopi, 2020, pp. vii + 320
ISBN-10: 9004420495
ISBN-13: 978-90-04-42050-2
Hardback, €149.00 / $179.00

IVAN RESTOVIĆ
Institute of Philosophy, Zagreb

This book is Volume 113 of the *Poznań Studies in the Philosophy of the Sciences and the Humanities Series*, being the 12th book of the subseries *Polish Analytical Philosophy*. It consists of 16 chapters authored by different scholars, most of which address the Polish tradition of analytic philosophy, especially the Lvov-Warsaw School—one of the most notable movements in this country's intellectual history, founded by Kazimierz Twardowski (1866–1938). As it is the case with many other movements, the tenets of the School's members are not uniform. However, a distinguishing feature of the Lvov-Warsaw School is an analytic approach to philosophy characterized by the use of particular formal and less formal methods developed by its members. More often than not, there is a strong emphasis on logic. How philosophy was done in this school can be inferred from the book itself since, as the Editors note in the Introduction, "the majority of the authors of the presented volume are genetically connected with the Lvov-Warsaw School, namely being indirect students and followers of members of the first generations of this formation" (6).

In what follows, I provide an overview of each of the chapters. Not in all cases do I also offer my opinion on the texts or the theses provided therein. But before that, let me give a few general remarks about the collection as a whole. First of all, I believe this volume would be a valuable addition to the library of anyone interested in the history of analytic philosophy and the methodology of philosophy. The Lvov-Warsaw School is not widely mentioned in philosophy curricula, and upon reading this book I strongly believe this omission should be rectified. I do, however, have a few critiques of the book. One considers the title. The Editors claim that it

Correspondence: irestovic@ifzg.hr

"refers to the tension between formal and informal elements in the way of practicing analytical philosophy" (2). However, this tension is not explicitly explored in the present volume, exception being Chapters 2, 4, 6 and 9, the only texts that mention the word "informal" in the relevant sense. Most of the texts *do* talk about, propose and use various (historical) formal and informal methods, but the two opposing accounts are seldom contrasted. My second concern is about the cover of the book. It features two photographs depicting two scholars, but nowhere in the book does it say who these people are. I have some ideas about which members of the Lvov-Warsaw School they might be, but I'm not as sure as to be comfortable enough to share my hypotheses. I think this information should have been made available also to the readers not (that) familiar with the School. Lastly, some of the chapters may have benefited from a closer proofreading.

1) Mieszko Tałasiewicz: "Metareflection: A Method for Philosophy" (pp. 9–40)

This chapter offers "a phenomenological description of a way in which one can practice philosophy" (12), in opposition to a (more popular) stance according to which the method of philosophy is the conceptual analysis of data given to us by philosophical intuition. The author stresses the importance and indispensability of the first-person view in philosophy—calling this method "metareflection"—where intuition is not wholly dispensed with, but is understood as being "made on the basis of explicit reasoning" (24) and "is subject to calibration and correction" (27). Especially interesting is the term "conceptual synthesis", which "involves having to introduce new technical terms or attaching a new technical sense to previous everyday expressions" (15). We should, however, be extra careful when dealing with concepts referring to vital social practices, such as justice and responsibility, where the usual understanding of the terms arguably needs to be preserved as much as possible.

The author doesn't accept philosophical exceptionalism, arguing that scientists themselves surely can and often do engage in philosophy, but that in philosophy there is a difference of degree to which the first-person analysis is (supposed to be) used. Nor can, he continues, philosophers just "spout nonsense" (28) about things empirically verifiable. He emphasizes the importance of philosophical training, especially of the distinctions introduced in the philosophical tradition, to name a few: Brentano's intentional vs. unintentional states, act vs. content vs. subject of presentation in Twardowski, and Donellan's referential vs. attributive use. The view proposed in this chapter also incorporates a stance towards

thought experiments, which are not understood as merely a "cheap substitute for a real-life experiment" (34).
This paper, the longest in the book, offers an engaging and thought-provoking introduction to the volume. (But, on the other hand, it does not explicitly concern the philosophy of the Lvov-Warsaw School, so those who came for an introduction to this particular brand of philosophy may perhaps skip to the second chapter.) As the author himself admits, further elaborations of some claims made in the text "would require a book, not a paper" (26). It is certainly something to look forward to.

2) Jacek Jadacki: "Semi-Formal Analysis of the Formality-Informality Opposition in the Spirit of the Lvov-Warsaw School" (pp. 41–55)

The main thesis of this chapter is that opposing formal to informal theories—especially in the case of logic—"has no rational basis" (48). The author claims that there is no such thing as an informal theory—a theory can only be more or less formal. But he also claims that "there is no formula that would be fully formal" (50), i.e. 'contentless', since variables always have a range, i.e. a domain. He develops his argument by first meticulously specifying and distinguishing all the transformations one can do on sentences, namely: enlargement, generalization, extrapolation, variabilization, standardization, schematization, and clarification. All of them are needed to eliminate the unwanted features of (the arguments put forward in) the natural language, such as ellipticity, amphibolicity, polysemia, occasionality, and approximation. Following the philosophical tenets of Łukasiewicz, Ajdukiewicz, Bocheński and Twardowski, he concludes that "[i]n practice, what is practiced under the banner of 'informal logic' is sometimes the result of operations that have been called 'clarification' here, or [sometimes] such an extension of classical logic that would be [a] more adequate theory of argumentation" (53).

In my understanding of the author's point, all that informal logic purports to do can be done formally, in the spirit of the Lvov-Warsaw School. Also, the very analysis that the author provides, which is according to his theory (merely) semi-formal, can itself be done more formally, but such an analysis is "waiting for its creator" (54). Personally, although I find the arguments proposed in this text compelling, I find that the author does not engage enough with the literature from the field (mis)identified as informal logic. The author quotes only a passage from the editorial introduction to the first issue of journal *Informal Logic* from 1978 where it is clearly stated that the informal logic means different things to different people, as well as the entry on informal logic from the *Stanford Encyclopedia of Philosophy*, where it is, admittedly, stated that "the goals of informal logic have been pursued in the Polish tradition of 'pragmatic logic'" (53, n. 10).

But there surely have to be *some* (methodological) differences, especially given that there's a lack of an agreed-upon definition, demarcation and goals of the field the author criticizes.

3) Marcin Będkowski, Anna Brożek, Alicja Chybińska, Stepan Ivanyk and Dominik Traczykowski: "Analysis – Paraphrase – Axiomatization: Philosophical Methods in the Lvov-Warsaw School" (pp. 56–74)

This chapter offers a reconstruction of three methods of doing philosophy used by the members of the Lvov-Warsaw School: analysis of concepts, semantic paraphrase, and axiomatization. It starts with a short yet informative description of the philosophical program of the School, pointing to some differences in approaches among its key members. In keeping with the tradition, the authors take a clear stance towards the notion of method in philosophy: "We share the view of the members of the LWS that philosophy is a science in a broad sense, and that various methods are used in it" (58). Their definition of method is—not altogether unobjectionably—tied to the aim of the research: "[T]he most useful definition of 'method' is one relativized to the aim" (58). The authors offer an evaluation of methods with respect to reliability, providing a distinction between reliable and infallible methods, as well as between local and global methods. Preceding the reconstruction, the four basic ingredients are outlined needed in order to characterize a given philosophical method, one of them being a clear indication of the applied conceptual or technological tools.

In the main part of the paper, the authors provide the successive stages of each of the three philosophical methods. They reconstruct them from the methodological remarks of the members of the School, as well as from the way they deal with specific philosophical problems. They draw from Łukasiewicz, Czeżowski, Twardowski, Ajdukiewicz, Kotarbiński and Leśniewski, and give (reconstructions of) examples from their works. To the reader, the preferred way of dealing with philosophical problems in the Lvov-Warsaw School is clear from the outset, and can be summarized by this sentence from the concluding section of the paper: "[It] is easy to notice the linguistic approach to problems and the trust in the instruments of logic (broadly understood)" (72). In the said section, we also find a brief comparison of the Lvov-Warsaw school and other similar movements in early twentieth-century analytic philosophy.

As a not-fully-initiated reader, I left with one question still lingering. The sophisticated formal methodology and the utmost clarity of the concepts used by the Lvov-Warsaw School notwithstanding, there is still one term that escapes definition: "[I]n each of them [i.e. methods of the School] an important role—at some stage—is played by intuition" (71). Not that intuition cannot be defined or accepted as a kind of insight, but—in my own opinion—it may forever remain a nebulous term.

4) Friedrich Stadler: "From Methodenstreit to the 'Science Wars' – an Overview on Methodological Disputes between the Natural, Social, and Cultural Sciences" (pp. 77–100)

This chapter presents several historical variants of the dispute about the unity versus the plurality of the scientific method. The author wants to show that the same/similar debate arose many times throughout history in many different guises and under many different names, covering related problems like, for example, unity vs. plurality of sciences, adherence to vs. rejection of different "cultures" of the sciences (i.e. of humanities, social and natural sciences), identification of vs. differentiation between understanding and explaining, and opposing views on the context of discovery and the context of justification.

In the introductory part, the author provides a brief overview of the variants of the Methodenstreit, providing more than a dozen of its historical iterations. Some of them are discussed in more detail in subsequent sections of the chapter: disputes in economics between the Austrian School and the German Historical School beginning in 1883, the Methodenstreit in the historical sciences lasting from 1891 to 1899, the debate around Hempel-Oppenheim's calls for methodological unification, different views about and around the Vienna Circle, competing interpretation of Weber's stance on methodology, and, finally, the "science wars" related to the Sokal hoax. All these disputes, the author suggests, can be investigated both from meta-theoretical and contextual points of view. This is done in the present text as well: We are given a plethora of influential names and works, where the influences are in each variant of the dispute meticulously traced back. But all the different positions are also summarized and classified according to their underlying philosophical assumptions, although not all of them arose in the field of philosophy *per se*.

The author—as far as I understood—doesn't take an explicit stance towards the issues discussed in this text, but his position on the role of historical analysis—a topic widely discussed across various iterations of the debate on method—can perhaps be inferred from the following, instructive, quote: "[T]he history of the Methodenstreit facilitates a better understanding and provides good arguments for both sides, in addition to helping to prevent a mere repetition of the good old debates" (97). This chapter is characterized by an abundance of references, which I'm sure makes it hard to grasp entirely for a reader not (that) familiar with the field of history of science and philosophy. At the same time, however, it offers such a reader a great starting point for future research in the field, at the same time preventing them from re-inventing the wheel.

5) Krzysztof Brzechczyn: "Periodization as a Disguised Conceptualization of Historical Development: A Case Study of a Theory of the Historical Process Developed in the Poznań School of Methodology" (pp. 101–125)

This chapter provides an outline of the philosophy of history of the Poznań School of Methodology, developed by Leszek Nowak and his colleagues. The text starts with an argument for distinguishing between periodization and chronology, where the former is a kind of division that should be more informed by theory. Unfortunately, as the author reports, this is rarely done by historians: they are often not explicit about their underlying theoretical assumptions when dividing time into periods. To this, discussions held in the Poznań School of Methodology are rare exceptions.

The chapter describes two distinct approaches to the philosophy of history taken by the members of the School: the adaptive interpretation of historical materialism and non-Marxian historical materialism. The adaptive interpretation, to which a substantial part of the chapter is devoted, was developed to solve the "well-known interpretive difficulties" of Marxism: It was not always clear how to interpret the cause-and-effect relationships between "global productive forces and relations of production, a social base and a legal and political superstructure, social and economic conditions and particular states of social consciousness" (103). The author provides Nowak's solution, the adaptive understanding, and describes its three varieties: "[t]he mechanism of the adaptation of systems of production to the level of productive forces", "adaptive dependency between the superstructure and the economic base" and the adaptive "dependency of social consciousness on social being" (104-5). We also find a diagrammatical representation of the structure of class formation according to Nowak, one of many such representations in this text.

Other members of the Poznań School working within the framework of the adaptive interpretation are also presented. There is description of the periodization of the pre-class epoch in the works of Burbelka, as well as of different conceptions of transitions between "formations", i.e. sub-periods, in the class epoch offered by Łastowski and Buczkowki. Following the description of the adaptive interpretation, the author presents some problems for it, including the place for, significance and status of "social momentum" (114) in relation to the economic one.

This part of the text leads into the portrayal of the "non-Marxian historical materialism", developed by Nowak after the application of the adaptive interpretation to "the construction of a theory of socialism appeared to be unconvincing" (118). Here we can find a rather interesting differentiation between the means of production, coercion and indoctrination, but also the author's critiques of this variety of historical materialism. One of them is about the division of societies into oriental and occidental, which he argues "is too rough to grasp the developmental diversity of non-European societies" (119).

6) Ryszard Kleszcz: "Władysław Tatarkiewicz: Metaphilosophical Notes" (pp. 126–149)

This chapter offers a (partial) reconstruction of Tatarkiewicz's stance on philosophical method and of his metaphilosophy, the fields he is less famous for than for his work in history of philosophy, aesthetics, and art history. The reconstruction is done based on his numerous works and letters, with ample representative quotations. The paper starts with a description of Tatarkiewicz's lasting philosophical influences, including Aristotle, Twardowski and British analytic philosophy. We are then given a depiction of Tatarkiewicz's stance towards analytic philosophy—an approach he opted for, following the postulates of common sense, (conceptual) clarity and precision. The author, however, points out that there are limits set for this kind of philosophizing: "Tatarkiewicz did not overestimate the possibility of using philosophical tools in the domain of religion" (134). The author then goes on to discuss a closely related question of the role of logic in (meta)philosophy, contrasting Tatarkiewicz's position with that of Łukasiewicz. Tatarkiewicz's "affinity for analytical thought" (134) notwithstanding, he was closer to an "informal attitude" (135) about logic.

Next, we find a detailed description of the three types of knowledge/perspectives according to Tatarkiewicz: natural, scientific, and philosophical. Natural perspective can be found in every individual and it "does not require any particular education or professional preparation" (139). It gives opinions about the world as a whole. Scientific perspective is, as one can expect, more rigorous, but "does not aspire to gain knowledge about every realm of reality" (140). Both perspectives are, however, "deformed to some extent" (142), each in its own way. But when building a worldview, a choice has to be made, and "[i]f such a choice is to be made from an external, somehow neutral point of view, the role of arbitrator must be entrusted to philosophy" (143). For Tatarkiewicz, the author reports, philosophy is a science in a broad sense, a "discipline with the widest scope and one that uses the most general concepts" (143). It applies scientific methods but goes beyond them. However, we are not given a definite answer as to what these methods are. Instead, what we find is an appreciation of different views and approaches: “[T]he object of philosophy is not constant but changes depending on the era” (145).

In the last section, the author provides a synthesis/summary of Tatarkiewicz’s metaphilosophical and methodological tenets. At the very end, he states that it is “not possible to fully and systematically determine [Tatarkiewicz’s] position” (146) and hence the justification for the wording of the title—metaphilosophical “notes”. Given the flexibility and permissiveness of Tatarkiewicz’s (meta)philosophy, I did not find that to be a disappointment. What I would personally like to have seen, however, is a more detailed comparison between philosophical and natural knowledge, especially given that (if I understood correctly) they both strive to encompass the whole of the world.

7) Tadeusz Szubka: “Casimir Lewy and the Lvov-Warsaw School” (pp. 150–160)

This chapter discusses the reasons why Kazimierz (Casimir) Lewy, a student and later a lecturer at the University of Cambridge, was “rather resistant” (150) to the philosophy of the Lvov-Warsaw School, even though he started his philosophical development in Warsaw and was moved to philosophy by Kotarbiński, a member of the School. In the first section, clearing up first an ambiguity found in the literature about whether it was a paper by or on Kotarbiński that inspired Lewy—opting for the first option—the author describes four episodes of Lewy’s involvement with Polish analytic philosophy. He helped Zbigniwe Jordan publish “a general sketch of the pre-war achievements of the Lvov-Warsaw School” (153), and on three occasions he wrote critical reviews of two logic textbooks by Tarski and one by Czeżowski. Concerning the textbooks, Lewy praised the logic therein, but was highly skeptical about their philosophical assumptions.

Initially, while reading the section about these four encounters, I developed an expectation about where the chapter would go next, which ultimately turned out to be wrong. From the tone and wording of the section, I thought the author would make the claim that Lewy’s encounters were partial and unrepresentative, and that he wouldn’t have been as critical had he got more acquainted with the philosophy of the other members of the Lvov-Warsaw School. Instead, the chapter goes on to describe three main reasons for the critical attitude Lewy expressed towards the School. All of them are “diverging philosophical perspectives” (156). Firstly, what distinguished Lewy from the ontologically conservative Lvov-Warsaw

School was the fact that he was "unrepentant in his affirmation of the existence of abstract objects, including concepts and propositions, and of modalities" (156). Secondly, Lewy's attitude towards logic was "more flexible" (157)—he was open to using other logics that the classical extensional logic to deal with philosophical problems. Lastly, there is "Lewy's reluctance to weaken the relationship holding between analysandum and analysans in correct analysis" (158), unlike the approach taken by Carnap, which can be said, the author tentatively suggests, to be similar to the approach taken by the Lvov-Warsaw School.

Admittedly the anti-climactic nature of the second section may have been all on me. So, the section about Lewy's encounters with the School should be read as episodes that provided him with an understanding of what the philosophy of the Lvov-Warsaw School generally was.

8) Srećko Kovač: "Remarks on the Origin and Foundations of Formalisation" (pp. 163–179)

This chapter rehabilitates and argues for a mechanistic view of formal reasoning. The text starts by describing "modern standards of the certainty and exactness of knowledge" set by the founders of modern logic, standards according to which "one cannot be fully satisfied with a given theory until it is formalised, that is, presented in a shape of a formal system" (163). Especially highlighted is Łukasiewicz's axiomatic approach (to philosophy). Following the works of Łukasiewicz and Bocheński, the author makes the case for the claim that the said standards go back to Aristotle, who not only established formal logic, but also a general theory of axiomatics (albeit, seen from the viewpoint of modern standards, with "some shortcomings in [...] presentation and wording" (165)). As the author explains, "Aristotle's approach resembled the requirements for a formal system as formulated by Frege" (166).

Next, considering, among others, Frege's, Hilbert's, Kant's and Łukasiewicz's remarks on formal systems, the author explores the relation between the "sensible giveness" (166) of concrete, written, signs used in concrete proofs and the necessary, i.e. presupposed, "abstract and 'ideal' or 'conceptual' pre-understanding of expressions" (167). Following the logical and philosophical work of Tarski, Gödel and Turing, the author establishes and defends his central claim that "[t]he concept of a formal system can be rendered precise in its 'abstract' ('absolute') sense independently of any formalism" (168) and, if envisaged as a Turing machine, can be "reduced to mechanical (and thus causal) terms and rendered objective" (169). Such a view can be attributed to Aristotle, whose understanding of syllogism, the author suggests, "was basically dependent on causal terms (e.g., premises as causes of a syllogism)" (169). This is tied to Wittgenstein's reflection on machines, according to whom a machine or a picture of it "can be used as a symbol for a certain way of operation" and thus his "symbolic machine shares its abstractness with a Turing machine" (170). The author investigates some possible influences on Wittgenstein, among which there may be Croatian philosopher Faust Vrančić, whose book *Machinae novae* was a part of Wittgenstein's private library.

Following the conclusions drawn about formal reasoning as a mechanical (causal) procedure, the author provides a formal account of this procedure, which "should possess general features of determinacy" (171). As a starting point, he uses Minari's modal reformulation of Łukasiewicz's three-valued logic, adapting its axiomatization and adding to the language the tools of justification logic in order to allow for expressing more specified causal justifications. For the proposed axiomatic system, he proves soundness and completeness.

9) Krzysztof Wójtowicz: "The Status of Mathematical Proofs and the Enhanced Indispensability Argument" (pp. 180–194)

This chapter identifies a tension between the two ways of choosing ontological commitments regarding mathematical objects, seen from the perspective of the two versions of the indispensability argument proposed by mathematical realists. The author starts by describing and contrasting the original indispensability argument as first proposed by Quine, and the enhanced indispensability argument advocated by Baker. The former regards as indispensable only those mathematical entities that are logically necessary in scientific explanations, while the latter focuses on those mathematical entities that carry explanatory power.

The central question the chapter raises is the following: Does the explanatory power come from mathematical theorems themselves, or does it (at least partially) come from the proofs of theorems? The author sides with the latter option but notes, however, that it is then important to consider the two different visions about the nature of mathematical proofs. According to the first, "[a] mathematical proof is an intellectual activity which is not constrained by purely formal conditions", it is "an operation on concepts, and semantic aspects have a non-reducible character" (188-9). On the second view, "[a] mathematical proof is a formal construct whose semantic aspects are insignificant—only compliance with formal rules counts" (189). Even though he recognizes that mathematics is not usually practiced in line with the second view, "[p]roofs from everyday mathematical practice […] being a mixture of natural and symbolic

languages" (188), the author notes that, from the perspective of ontological commitment, the latter view needs to be taken into account.
Here, the author suggests, the field of reverse mathematics may provide valuable insights, because it establishes the "strength of assumptions necessary to prove theorems […] [a]nd in terms of ontological commitments—it provides a tool for identifying them" (190-1). In line with the second, stricter, view of mathematical proofs, analyses in terms of reverse mathematics include translating proofs into the language of second-order arithmetic. This, however, "from the point of view of everyday mathematics is a very artificial procedure" (191) and, consequently (and importantly), is likely to have a negative impact on the "explanatory virtues" (191) of the proof. As the author warns, there may appear two versions of a proof, one using weak assumptions but lacking in explanatory power, i.e. "leaving a feeling of cognitive insufficiency" (188), and the other which explains, but uses stronger assumptions. From the perspective of ontological commitment, the author concludes, the enhanced indispensability argument faces a drawback when compared to the original indispensability argument: The use of reverse mathematics helps us to see that more explanatory power may lead to a more baroque (mathematical) ontology.

10) Kordula Świętorzecka: "A Case of Metalogical Explanation of Logical Normativity" (pp. 195–205)

This chapter proposes a view that normativity of logic can be explained in terms of metalogical properties of the inference relation. The author takes inspiration from various philosophical understandings of Kant's, Frege's and Carnap's views on normativity, warning us that they "fluctuate between contradictory interpretations" (195). For instance, there are in the literature opposing answers on whether Kant saw logic as normative. MacFarlane, Hanna and Lu-Adler claim that he did. Alternatively, Tolley "suggests a plausible interpretation of the concept of normativity according to which Kant is not a normativist at all" (196). Situations like these, the author suggests, prompt us to inquire about a precise and non-ambiguous definition of normativity of logic that is in accordance with the standards of modern logic. Her approach thus starts from the "conviction" that "if philosophical creativity is to concern matters in the close vicinity of scientific considerations, then it should consider as much as possible the subjects and the methods of the latter" (198).

The concept of normativity of logic presented in this chapter is restricted to situations where logic is applied to "somehow distinguished non-logical reasonings" (198). These are not reasonings that have nothing to do with logic, they are non-logical only because they are not put forward in a language of symbolic logic. A further restriction is that the given approach, for the sake of simplicity, considers only non-logical reasonings expressed in a language which is "morphologically similar" (198) to that of propositional logics. To build her case, the author provides preliminary notions from the contemporary methodology of deductive systems. Among other things, she offers precise definitions of structural consequence operation, valid inference, logic, and well-defined logic. She then presents a morphologically similar language to express simple reasonings, as well as a way to formalize it in the language of propositional logic. It is in this sense, the author suggests, that we can understand normativity: "To phrase the description of […] reasonings in normative terms, we can say that they respect norms of a given logic, or that the logic is normative with respect to them" (202). On this approach, "the question of the normativity of any reasoning is reduced to the problem of the existence of a formalization that translates a reasoning into a generally verifiable inference" (202).

The author recognizes some pragmatic limitations of the proposed view, the most serious probably being that simple reasonings are in her approach expressed in a language designed to be similar to that of propositional logic. This is, however, not the language used in philosophical reasoning, the formalization of which may prove significantly more difficult. She leaves this concern for another occasion, but notes that rephrasing philosophical talk to fit the language of logic may also be considered a normative task. It remains to be seen if less formalistically-minded philosophers will find this approach to normativity understandable and/or convincing.

11) Sébastien Richard: "Leśniewski's Intuitive Formalism" (pp. 206–228)

This chapter describes the philosophical position of Stanisław Leśniewski, which Tarski calls "intuitionistic formalism". As Leśniewski never fully explained how this position was to be understood, the author sets out to explain/reconstruct what it is and how Leśniewski applied it in his work. The text starts with an explanation of the name of the Polish logician's philosophical stance: his view of formalism and of intuition.

The author claims there are two parts to Leśniewski's philosophy: critical and constructive. The first "concerns some formal systems built by other logicians" (207), where these systems are criticized on account of their meaning. In the construction of a formal system, we should at every point know what its constituting expressions are about: "The formalism […] comes after the intuition in order 'to encode and communicate' it in a more precise way" (208). This is in opposition to Hilbertian formalist stance in philosophy of mathematics, where statements and symbols have meaning only relative to the role they play in a theory, although Leśniewski, like Hilbert, takes an axiomatic approach. Intuitionistic formalism also cannot be subsumed under Brouwerian intuitionism since, as the author notes, Leśniewski accepts the principle of excluded middle. Recognizing there is a tension between intuitionism and formalism, the author opts for another name given to Leśniewski's philosophy—"intuitive formalism".

These considerations are followed by a reconstruction of the meaning of the term "intuition", which for the Polish logician is both about the language and the world, concerning "how to speak about the way the world is" (211). In the description of the critical part of Leśniewski's philosophy, we are also given some concrete examples—his position on Russell and Whitehead's Principia: the critique of their use and explanation of the assertion-sign and the critique of their equivocation of the two readings of the negation-sign.

Regarding the constructive part of intuitive formalism, the author describes how this philosophy is used by Leśniewski in construction of his three formal systems: Protothetic, Ontology and Mereology, the motivation for which is to find a more "intuitive" solution for the Russellian paradox of classes which was, as the author states on multiple occasions, discovered independently also by Leśniewski himself. The text concludes with a clear description of Mereology, the system based on Protothetic and Ontology, where its philosophical assumptions are made explicit and distinguished from those of other systems proposed to solve the antimony of classes.

Having read this chapter, I can indeed say that Leśniewski's solution to me seems to be superior to Russell's—a case in point being the identification of "every unary collective class with its unique element" (225)—and I would recommend this text to anyone who decides to grapple with (the solution to) "Russell's" paradox.

12) Zuzana Rybaříková: "The Case of Logic: Łukasiewicz-Prior's Discussion on Logic" (pp. 229–238)

This chapter concerns the philosophical differences between Łukasiewicz and Prior that lead them to use opposing systems of logic when approaching philosophical problems. Even though the title announces that what will be addressed is the "discussion" between these two logicians, the reader should rather expect a *contrast* between their views, featuring a lot more of Prior's comments on Łukasiewicz's work than *vice versa*. However, this "asymmetry of discussion" may well be the result of historical facts rather than a flaw of the chapter: Łukasiewicz may just have not engaged that deeply with Prior's work, but the text, especially given its, in my opinion, misleading title, leaves this mystery unresolved.

The opening section outlines and explains the possible origins of some similarities between the views of the two logicians (and philosophers). The two remaining sections are dedicated to the logic/philosophy of Łukasiewicz and Prior, respectively. Concerning the former, we find a description of his view of the philosophical method, which he considered to be wanting in comparison to the precise methods of natural sciences, leading him to an analysis of philosophical problems by means of (developing) mathematical logic. The author discusses the philosophical topics considered by Łukasiewicz, most notably his analysis and rejection of determinism and his view of causality, influenced particularly by Łukasiewicz's "passion for human freedom" (231, n. 1), which ultimately led him to reject "the meta-logical law of bivalence" (233). We also find an informative description of different many-valued logics developed by Łukasiewicz, but also remarks on his anti-psychologist stance, his preference towards extensional logic and his possible Platonism.

Regarding Prior, the author provides an outline of his philosophical development, followed by a depiction of the influence the Polish logician had on him. We find out that Prior at first adopted Łukasiewicz's system of logic, but later "discovered several controversial aspects" (235) therein. Prior criticized Łukasiewicz's systems on account of, among other things, allowing the law of contradiction and the law of excluded middle not to hold universally, and not being genuinely indeterministic. Prior also, unlike his Polish fellow logician, preferred intensional logic and was a nominalist. In the concluding paragraph, the author states that "[i]t was primarily the philosophical convictions of both authors that gave rise to the differences in their views on logic" (236), ending the text with a thought-provoking question: [D]oes it still mean that mathematical logic is a precise tool in philosophy, if the choice of the system of logic is affected by the philosophical preferences of each philosopher?" (237).

13) Aleksandra Horecka: "The Semiotic Method in Art Theory and Aesthetics in the Lvov-Warsaw School" (pp. 241–256)

This chapter is about the various semiotic theories developed by the members of the Lvov-Warsaw School and the proposed applications of these theories to analysis and classification of works of art. It focuses mostly on Wallis's account, but considers in detail also the views of Twardowski, Pelc, Blaustein, Witwicki and Tatarkiewicz. The text starts with the necessary philosophical preliminaries for the application of semiotics to aesthetics and to the theory of art, where the latter is not—

unlike the other two—considered a "philosophical field" (242). (In other parts of the text, however, aesthetics and theory of art are not further distinguished and are considered together.)

In order to successfully undertake this application, the author states, the objects of aesthetics/theory of art have to be understood as/in terms of signs. She describes the two different approaches regarding the ontology of signs: the monocategorical vs. the polycategorical view, suggesting that art is better analyzed in terms of the latter, according to which there are different kinds of signs, and which most members of the Lvov-Warsaw School themselves ascribed to. She then goes on to consider and compare competing definitions and classifications of signs proposed by the members of the School. Special attention is given to the explanation of and the interplay between the three domains of semiotics: semantics, pragmatics and syntax, particularly to different accounts of the latter domain, about which the author says: "In the case of applying the semiotic method to the theory of art, it becomes necessary to develop a specific theory of the structure of semiotic objects and the theory of the combination of multiple parts into a unified harmonious whole" (246).

The part of the chapter concerning the theory of art and aesthetics provides some definitions of (form and content of) a work of art given by the members of the Lvov-Warsaw School, as well as their different accounts on whether there can be a (part of a) work of art that is not a sign. This text, however, is not only theoretical: The author provides photographs in color of Romanesque columns located in the Cistercian monastery in Wąchock in Poland, which she analyzes according to some elements of Wallis's semiotic syntax. We find out, among other things, why demons are located at the bottom, and flower at the top. A strong conceptual apparatus proposed in the first part of the chapter enables us also to make sense of the claim that "[b]ecause the column as a whole is part of the house of God, it must be entirely good" (249).

14) Marcin Będkowski: "From Concepts and Contents to Connotations: Łukasiewicz's Theory of Conceptual Analysis and Its Further Evolution" (pp. 257–277)

This chapter offers a reconstruction of Łukasiewicz's theory of conceptual analysis, i.e. of the methodological remarks present in his philosophy. These remarks were put forward mostly as preliminaries to his analysis of the concept of cause, but, as the author suggests, some scholars consider them "even more valuable than the solution of the main issue" (259). However, the author stresses the fact that "Łukasiewicz's conception has unfortunately not provoked many comments or studies" (257). Wanting to ameliorate this situation, this chapter describes Łukasiewicz's understanding of concepts, his view of conceptual and logical analyses (with an emphasis on the use of inductive and deductive method), as well as, importantly, his underlying philosophical assumptions—all of which are guided by "the ideal of accuracy offered by the deductive sciences" (258). But it does not stop there.

Having provided a recapitulation of Łukasiewicz's methodological tenets, the author recognizes some "minor deficiencies", but also some "more serious errors" (265) therein. Among the former is a lack of explanation of the difference between a concept and objects that fall under it; among the latter is simultaneous acceptance of conceptual realism and the claim that concepts are constructed. The author admits he would not set out to give the problems "the attention they undoubtedly deserve" (266). He does, however, offer an amendment to Łukasiewicz's philosophy which makes clearer the relations between concepts, names of concepts, meaning of names, designata of names and connotations of names.

The chapter also provides the views on conceptual analysis of some other members of the Lvov-Warsaw School, considering the influences by and on Łukasiewicz. For instance, we find out that it was probably Łukasiewicz who made Twardowski, the founder of the School, change his position from psychologism to moderate antipsychologism. We also find an interesting analysis of Łukasiewicz's and the Committee's opinions on his habilitation dissertation, with which he was ultimately not satisfied with, and which the Committee accepted not on account of the positions expressed, but on account of analytic rigor and clarity. Following is a description of the School member's diverging (but also fluctuating) positions on the relations between meaning, content, connotation and concept, on which there are two opposing tendencies: to identify—as Łukasiewicz does—or to differentiate—as done by, among others, Ajdukiewicz and Kotarbiński. The text ends with a (invitation to a further) comparison between Łukasiewicz and Moore, who "can be regarded as the pioneers of the 20th century philosophical analysis" (274), but among which the former is undeservingly less popular.

15) Alicja Chybińska: "Kotarbiński's Methodological Reism: Framework and Inspirations" (pp. 278–296)

This chapter offers a reconstruction of an unrecognized aspect of Kotarbiński's reism. As the author reports, it is widely assumed that the position of this Polish philosopher had two aspects: the ontological and the semantic reism. However, she shows that this can be called into question, also recognizing a place for Kotarbiński's reism regarding methodology.

Along with the said reconstruction, this chapter gives an analysis of the influence of Twardowski, the founder of the Lvov-Warsaw School, on the philosophy of Kotarbiński, his student and thesis supervisor.
Regarding methodological reism, the author starts her argument by distinguishing between the "ontological thesis" and the "semantical thesis" (279) of reism. According to the former, the only objects that exist are concrete objects. According to the latter, every meaningful sentence contains only names of concrete objects or names that can be paraphrased in terms of such names. Unlike Kotarbiński, the author claims that these theses are independent. Tied to, but different from, the thesis about semantics is that about the method according to which one is to formulate their philosophical language and thought. The "semantical thesis" of reism is about clarity of expression and, as the author aptly recognizes, "clarity is a methodological concept characteristic of normative methodology" (282). She formulates four theses expressing different relations between clarity and lack of "apparent names", i.e. names that do not refer to concrete entities, identifying among them the position held by Kotarbiński. In connection to these theses, she also proposes three postulates of methodological reism, from the weakest to the strongest.

The part of the chapter concerning influence offers ample representative quotations from Kotarbiński and Twardowski in order to prove the (dis)similarities between the positions of the two, as well as to trace the effect the latter had on the former. The author distinguishes between "positive" and "negative" influence Twardowski had on Kotarbiński. Positive influence, i.e. the positions Kotarbiński accepted from his teacher, concern, for example, the view on the connectedness between "the vices of speaking and the vices of thinking" and "respecting the principle of clarity and embodying it both in teaching and in scientific work" (290). What Kotarbiński didn't accept are his teacher's pluralistic ontological commitments, which are described in detail. However, the author makes the claim that Kotarbiński's reism, "an original Polish conception" (294), would probably have not existed had there not been for the differences between him and Twardowski: Having faced his teacher's position, particularly expressed in his dissertation, Kotarbiński was inspired to develop his own philosophy. On the other hand, it was the fact that Twardowski "neither promoted his ideas over others' nor forced his own philosophical solutions on his students" (293) that gave rise to an atmosphere in which Kotarbiński could develop his standpoint.

16) Anna Brożek: "Interdisciplinarity: Analysis of the Concept and Some Examplifications in the Lvov-Warsaw School" (pp. 297–313)

This chapter offers what could be called a philosophy of interdisciplinarity. The chief aim of the text is to distinguish between the essential and merely apparent senses and uses of the term, which is, the author states, presently "accompanied by great conceptual chaos" (298). She starts her conceptual analysis by distinguishing between the five different aspects of a scientific discipline, out of which she gives the most attention to domain or the set of objects, methods and language: Interdisciplinarity will be grounded in differences between the aspects of two or more disciplines. Regarding domains of disciplines, an important and illuminative distinction is made between "material" and "formal object" of investigation. For instance, "[a] man as an individual or man as a species is the material object of many disciplines which approach it from different perspectives, that is they have different formal objects" (300).

This leads to an analogous distinction between two kinds of interdisciplinarity: material vs. formal. The former is exemplified in the above quote. The latter occurs when two or more disciplines study different material objects but use the same tools. An instance of this would be "game theory—invented in the context of gambling and then successfully used in economics, sociology, computer science, biology and ethics" (303). The author stresses, however, that the similarity/sameness of material/formal objects is not sufficient for interdisciplinarity. What is also needed is "a suitable integrating language" (302). Interdisciplinary language, a language of a genuinely interdisciplinary field, should differ from the languages of disciplines it concerns.

Having defined interdisciplinarity in the real sense(s), the author offers a critique of the ways this term is often used, talking about its several "overuses". Notably, she relates the proposed theory to the real world of scientific practice, observing and questioning the role of institutions and grant providers on various understandings of interdisciplinarity, as well as on the very division of sciences into disciplines. If I understood correctly, according to the theory proposed in this chapter, interdisciplinarity is seen as something temporary: It leads either to an emergence of a new discipline or to a unification of disciplines. This is a claim that, in my opinion, may be disputed while still accepting the overall analysis of interdisciplinarity provided in this chapter.

In the second, shorter, part of the text, the author offers an analysis of the philosophy of Twardowski, the founder of the Lvov-Warsaw School, and his students Witwicki and Łukasiewicz, establishing that the former's work was interdisciplinary in the material sense, while that of the rest was intradisciplinary, albeit with some "interdisciplinary stamps" (311) that they inherited from their teacher.

BOOK REVIEW

Rafe McGregor
A CRIMINOLOGY OF NARRATIVE FICTION,
Bristol University Press, 2021, pp. 176
ISBN: 978-1529208054
Hardback, €74.70 / $66.09

IRIS VIDMAR JOVANOVIĆ
University of Rijeka, Faculty of Humanities and Social Sciences

According to Rafe McGregor, fictional narrative representations can explain the causes of crime and social harm, which is why they should be employed to direct public policy and the practice of criminal justice professionals. More to the point, McGregor argues that those fictional works dealing with crime, crime-related practice and harm have the potential to expose the causes of that harm, and thus to contribute to reducing it.

Underneath this precise and straightforward idea is a rather complex theoretical framework stretching from literary aesthetics to various branches of criminology. McGregor's primary interest is to establish his account as a contribution to criminological studies, supported by philosophical theories on the cognitive value of fiction, which would recognize that criminological fiction should not be reduced to criminological imagination, but should instead be recognized for the concrete benefits it can induce.

McGregor positions his theory (in the first three chapters, and with a summary in the conclusion) with respect to narrative criminological framework, cultural criminological and critical realist framework. This part of the book may seem the most technical and demanding for those coming to it outside of criminology, and the most thought-provoking and challenging to the criminologists. While much under the influence of Lois Presser (whose work he identifies as the leading voice in the narrative criminology), McGregor is careful to highlight the differences among them, two of which are the most relevant. Unlike Presser, McGregor is concerned with fictional, rather than real life stories; and he is not engaged with exploring the ethical aspect of stories (having already done so in his Narrative Justice). With respect to fictional criminologies, McGregor is

Correspondence: ividmar@ffri.uniri.hr

(chapter 3) very detailed in comparing and contrasting his work to the one done within the cultural criminological framework by Jon Frauley, Nicole Rafter and Vincenzo Ruggiero.

McGregor analyses three epistemic roles for narrative fiction in criminological inquiry. The semiotic one refers to narrative fiction providing knowledge of the production and reception of representations of crime and its control; fiction's pedagogical role is to facilitate, augment, or enhance the communication of criminological knowledge and its etiological role relates to providing knowledge of the causes of crime or social harm in virtue of providing phenomenological, counterfactual and mimetic knowledge (more on this below). For the most part, McGregor is interested in fiction's etiological role, claiming that only those works which are imbued with such value can contribute to crime reduction. The crucial issue then is to explain which works in fact have such a value.

The other theoretical line in McGregor's theory is his presupposing (rather than arguing in favor of) the doctrine of literary cognitivism, the view that narrative literary fiction is a source of knowledge. McGregor has already established himself as a fervent advocator of this theory and in this book he applies his bent of the theory to particular case studies: an array of works he takes to exemplify his take on the narrative fiction's contribution to criminology. As McGregor sees it, there are three types of knowledge available in works of fiction, transferable to three types of values. First, phenomenological value is the value of the representation of the subjective experience of offenders derived from the capacity of literary works to convey the phenomenological knowledge of what is like to be a perpetrator of a certain crime. Second, literary fictions have counterfactual value defined as the extent to which a given work "provides knowledge of reality by means of exploring alternatives to that reality" (91). Third, mimetic value relates to the representation's capacity to provide knowledge of the everyday reality, primarily, as McGregor argues, the type of knowledge that is not available "to nonfictional representations for reasons of access, ethics or law" (113).

On McGregor's view, there are three types of crime that can be exposed through fiction: state, ordinary, and organized. To this is also related a three-partite division of modes of representations: literary, cinematic and hybrid. Although there is no necessary relation among the criminological values and modes of representation, in the sense that all values are available through all forms of representations, there is a stronger relation between cinematic mode of representation and mimetic knowledge. Elaborating on that claim is the topic of the penultimate chapter, where

McGregor engages with Berys Gaut and Greg Currie's theories on film and the types of realism film can advance.

Chapters five to seven are dedicated to exemplifying McGregor's theoretical claims regarding criminological fiction by extensive, informative and thought-provoking analysis of the case studies, all of which are taken from the popular culture rather than high art domain—examples include novels (e.g. Martin Amis' *The Zone of Interest*), films (*Miami Vice, No Country for Old Men*) and series (ITV's *Broadchurch*). The relevance of popular culture is in particular emphasized in the chapter dealing with cinema, as McGregor invokes (echoing Noel Carroll's arguments) the accessibility of popular art. On his view, part of what makes fiction, primarily cinema, such a powerful tool for the criminological investigation and for the communication of knowledge is its immense popularity, itself a result of its availability with the masses.

The emphasis on works from popular culture is further relevant for McGregor's concerns with fiction (rather than with art in the evaluative sense) or narratives (a topic he already addressed in his *Narrative Justice*, where he argued that the fiction/non-fiction distinction is of lesser importance for the narrative's capacity to deliver phenomenological knowledge). One of the main aims he sets out for himself is to provide a space within criminology for taking fiction seriously, that is, for showing that "fiction can provide actual data that complements the data provided by traditional academic and documentary sources" (3). Such fiction's capacity is related to its giving knowledge of what certain experiences are like, in giving knowledge about the non-existent situations and detailed and accurate knowledge of everyday reality.

McGregor is aware that his arguments are "counterintuitive and (…) highly unpopular" (3) with the criminologists, and is more concerned with proving them wrong than with converting the skeptics of the cognitive benefits of fiction. Unlike some scholars who recognize similar power with crime fiction and are concerned with tracking the mimetic elements in popular works dealing with crime in order to establish their potency with respect to providing understanding of crime,[1] McGregor does not seem to be too concerned with the traditional notion of fiction as breaking the fidelity to the world/life constraint. This isn't necessarily a fault in the book, since many philosophers have argued that fiction is not divorced from the truth, from how things are, from the state of the world. McGregor may be right in simply building upon that foundation, pointing only in the

[1] A good example is Peter Swirski's *American Crime Fiction*: *A Cultural History of Nobrow Literature as Art* (2016, Palgrave Macmillan).

penultimate chapter to the fact that some works (his example is *Beverly Hills Cop*) may have pedagogical but not etiological value, having sacrificed such value for the wider accessibility of the film. However, it is not quite obvious that this example suffices to provide means of distinguishing reliable from unreliable works (i.e. works with etiological value from those lacking it), particularly given his endorsement of the accessibility condition for the works' overall success in reaching the wider audience—a condition so crucial to his argument. In other words, it may be interesting to press McGregor on developing a more clear-cut criterion that helps differentiate between those works which transfer criminological knowledge from those which do not. For those who share McGregor's intuition, the examples he offers may be enough, particularly when supported with such masterful analysis as his account of the *Broadchurch*, where he tackles the legal issues related to rape, public prejudice related to the victims and perpetrators, and the like.

For those however who are on the fence, the book may not be sufficiently convincing, despite McGregor's insightful analysis of the representational devices employed by the works to convey knowledge he attributes to his examples. For example, I share his conviction that the fictional description of the lived experiences of the perpetrators of the crime can explain the causes of the crime,[2] but I am reluctant in accepting McGregor's further claim, according to which such a link (from lived experience to understanding the causes of the crime) can indeed contribute to its reduction. Not all criminological fiction is a cautionary tale and some descriptions may simply be deficient in some way, even if the work seems to have aetiological value. In addition, one may feel that McGregor is too quick to take the experience of one (fictional) character as representative of a class of people who are in some sense similar to that character, as he does in suggesting the analogy among fictional undercover cop and undercover agents in the real world. Furthermore, I wonder why McGregor does not consider the perspective of a victim of a crime as in any way potent with phenomenological (what is like) and mimetic (how it is) values. Given his take on the rape case in *Broadchurch*, one would expect him to make a case for the perspective of a victim.

[2] In arguing this, McGregor is restating some of his arguments from his *Narrative Justice* (2018, Rowman & Littlefield*)*, primarily the concept of the „standard mode of engagement" originally developed by Greg Currie. I criticized such approach to cognitive value of fiction in Vidmar Jovanović (2020), "Becoming Sensible: Thoughts on Rafe McGregor's Narrative Justice", *The Journal of Aesthetic Education*. I will not restate my arguments here for reasons of space, though I think they apply with even greater force, given McGregor's focus on fiction.

Leaving such worries for the conferences, let me end by recommending this book to those interested in literary cognitivism, in fiction and in the link between fiction and our social reality. While occasionally hard to read due to McGregor's adherence to the analytic style, the author offers sufficient repetitions and concluding statements to allow for comprehension. Given the popularity of mass art nowadays, his book is a much needed account of why it should not be dismissed as light, trivial or lacking in cultural and educational values. Furthermore, crime genre has always had a special place in our culture and within the humanities. McGregor's book is an immensely insightful contribution to exploring, reaffirming and honoring its status and value.[3]

[3] This work has been supported in part by Croatian Science Foundation, grant UIP-2020-02-1309.

ABSTRACTS (SAŽECI)

TRUE GRIT AND THE POSITIVITY OF FAITH

Finlay Malcolm
University of Hertfordshire

Michael Scott
University of Manchester

ABSTRACT

Most contemporary accounts of the nature of faith explicitly defend what we call 'the positivity theory of faith' – the theory that faith must be accompanied by a favourable evaluative belief, or a desire towards the object of faith. This paper examines the different varieties of the positivity theory and the arguments used to support it. Whilst initially plausible, we find that the theory faces numerous problematic counterexamples, and show that weaker versions of the positivity theory are ultimately implausible. We discuss a distinct property of faith that we call 'true grit', such that faith requires one to be resilient toward the evidential, practical, and psychological challenges that it faces. We show how true grit is necessary for faith, and provides a simpler and less problematic explanation of the evidence used to support the positivity theory.

Keywords: Propositional faith; objectual faith; desire; evaluative belief; positive attitude

ISTINSKA NEPOKOLEBLJIVOST I POZITIVNOST VJERE

Finlay Malcolm
University of Hertfordshire

Michael Scott
University of Manchester

SAŽETAK

Većina suvremenih teorija o prirodi vjere izričito brani ono što nazivamo 'teorijom pozitivnosti vjere' - teoriju da vjera mora biti popraćena povoljnim evaluativnim vjerovanjem ili željom za predmetom vjere. Ovaj rad istražuje različite varijante teorije pozitivnosti i argumente kojima se podržava. Iako incijalno izgleda uvjerljiva, smatramo da se teorija suočava s brojnim problematičnim protuprimjerima i pokazujemo da su slabije verzije teorije pozitivnosti u konačnici neuvjerljive. Raspravljamo o određenom svojstvu vjere koje nazivamo ‘istinska nepokolebljivost’, u smislu da vjera traži od osobe da bude otporna na evidencijske, praktične i psihološke izazove s kojima se suočava. Pokazujemo kako je istinska nepokolebljivost neophodna za vjeru i pruža jednostavnije i manje problematično objašnjenje dokazne građe koja se koristi u svrhu opravdanja teorije pozitivnosti.

Ključne riječi: Propozicijska vjera; predmetna vjera; želja; evaluativno vjerovanje; pozitivni stav

PURE POWERS ARE NOT POWERFUL QUALITIES

Joaquim Giannotti
University of Birmingham

ABSTRACT

There is no consensus on the most adequate conception of the fundamental properties of our world. The pure powers view and the identity theory of powerful qualities claim to be promising alternatives to categoricalism, the view that all fundamental properties essentially contribute to the qualitative make-up of things that have them. The pure powers view holds that fundamental properties essentially empower things that have them with a distinctive causal profile. On the identity theory, fundamental properties are dispositional as well as qualitative, or powerful qualities. Despite the manifest difference, Taylor (2018) argues that pure powers and powerful qualities collapse into the same ontology. If this collapse objection were sound, the debate between the pure powers view and the identity theory of powerful qualities would be illusory: these views could claim the same advantages and would suffer the same problems. Here I defend an ontologically robust distinction between pure powers and powerful qualities. To accomplish this aim, I show that the collapse between pure powers and powerful qualities can be resisted. I conclude by drawing some positive implications of this result.

Keywords: Pure powers; powerful qualities; dispositionalism; collapse objection; dispositional essentialism

ČISTE MOĆI NISU MOĆNE KVALITETE

Joaquim Giannotti
University of Birmingham

SAŽETAK

Ne postoji konsenzus o najadekvatnijoj koncepciji temeljnih svojstava našega svijeta. Gledište čiste moći i teorija identiteta moćnih kvaliteta tvrde da su obećavajuće alternative kategorikalizmu, gledištu prema kojemu sva temeljna svojstva u osnovi doprinose kvalitativnom sastavu stvari koje ih imaju. Prema gledište čistih moći, temeljna svojstva u osnovi daju karakterističan uzročni profil stvarima koje ih imaju. Prema teoriji identiteta, temeljna svojstva su dispozicijska, ali također su i kvalitativna ili moćna svojstva. Unatoč očiglednoj razlici, Taylor (2018) argumentira da se čiste moći i moćne kvalitete urušavaju u istu ontologiju. Kada bi ovaj prigovor kolabiranja bio dobar, rasprava između gledišta čistih moći i teorije identiteta moćnih kvaliteta bi bila iluzorna: ta gledišta bi imala iste prednosti i iste probleme. Ovdje branim ontološki robustnu razliku između čistih moći i moćnih kvaliteta. Kako bih postigao taj cilj, pokazujem da se može odoljeti kolapsu između čistih moći i moćnih kvaliteta. Zaključujem povlačenjem nekih pozitivnih implikacija ovog rezultata.

Ključne riječi: Čiste moći; moćne kvalitete; dispozicionalizam; prigovor kolapsa; dispozicijski esencijalizam

ACTS THAT KILL AND ACTS THAT DO NOT — A PHILOSOPHICAL ANALYSIS OF THE DEAD DONOR RULE

Cheng-Chih Tsai
Center for Holistic Education, MacKay Medical College

ABSTRACT

In response to recent debates on the need to abandon the Dead Donor Rule (DDR) to facilitate vital-organ transplantation, I claim that, through a detailed philosophical analysis of the Uniform Determination of Death Act (UDDA) and the DDR, some acts that seem to violate DDR in fact do not, thus DDR can be upheld. The paper consists of two parts. First, standard apparatuses of the philosophy of language, such as sense, referent, truth condition, and definite description are employed to show that there exists an internally consistent and coherent interpretation of UDDA which resolves the Reduction Problem and the Ambiguity Problem that allegedly threaten the UDDA framework, and as a corollary, the practice of Donation after the Circulatory Determination of Death (DCDD) does not violate DDR. Second, an interpretation of the DDR, termed 'No Hastening Death Rule' (NHDR), is formulated so that, given that autonomy and non-maleficence principles are observed, the waiting time for organ procurement can be further shortened without DDR being violated.

Keywords: DDR; UDDA; DCDD; NHDR; vital organ; causation

DJELA KOJA UBIJAJU I DJELA KOJA NE UBIJAJU — FILOZOFSKA ANALIZA PRAVILA MRTVOG DONORA

Cheng-Chih Tsai
Center for Holistic Education, MacKay Medical College

SAŽETAK

Kao odgovor na nedavne rasprave o potrebi napuštanja pravila mrtvog donora (DDR) radi olakšavanja transplantacije vitalnih organa, oslanjajući se na detaljnu filozofsku analizu Jedinstvenog zakona o utvrđivanju smrti (UDDA) i DDR-a, tvrdim da za neka djela za koja se čini da krše DDR, zapravo ga ne krše, stoga se DDR može podržati. Rad se sastoji od dva dijela. Prvo, koriste se standardni alati filozofije jezika, poput smisla, referencije, istinosnih uvjeta i određenog opisa kako bi se pokazalo da postoji interno konzistentna i koherentna interpretacija UDDA koja rješava problem redukcije i problem dvosmislenosti koji navodno prijete UDDA okviru, te kao posljedica toga, praksa darivanja nakon cirkulacijskog utvrđivanja smrti (DCDD) ne krši DDR. Drugo, tumačenje DDR-a, nazvano 'Pravilo brze smrti' (NHDR), formulirano je tako da se, s obzirom na poštivanje načela autonomije i ne-zlonamjernosti, vrijeme čekanja na nabavu organa može dodatno skratiti bez kršenja DDR-a.

Ključne riječi: DDR; UDDA; DCDD; NHDR; vitalni organ; uzročnost

IS THERE CHANGE ON THE B-THEORY OF TIME?

Luca Banfi
University College Dublin

ABSTRACT

The purpose of this paper is to explore the connection between change and the B-theory of time, sometimes also called the Scientific view of time, according to which reality is a four-dimensional spacetime manifold, where past, present and future things equally exist, and the present time and non-present times are metaphysically the same. I argue in favour of a novel response to the much-vexed question of whether there is change on the B-theory or not. In fact, B-theorists are often said to hold a 'static' view of time. But this far from being innocent label: if the B-theory of time presents a model of temporal reality that is static, then there is no change on the B-theory. From this, one can reasonably think as follows: of course, there is change, so the B-theory must be false. What I plan to do in this paper is to argue that in some sense there is change on the B-theory, but in some other sense, there is no change on the B-theory. To do so, I present three instances of change: Existential Change, namely the view that things change with respect to their existence over time; Qualitative Change, the view that things change with respect to how they are over time; Propositional Change, namely the view that things (i.e. propositions) change with respect to truth value over time. I argue that while there is a reading of these three instances of change that is true on the B-theory, and so there is change on the B-theory in this sense, there is a B-theoretical reading of each of them that is not true on the B-theory, and therefore there is no change on the B-theory in this other sense.

Keywords: Change; B-theory of time; existence; properties; propositions

POSTOJI LI PROMJENA PREMA B-TEORIJI VREMENA?

Luca Banfi
University College Dublin

SAŽETAK

Svrha ovog rada je istražiti vezu između promjene i B-teorije vremena, koja se ponekad naziva i Znanstvenim pogledom na vrijeme, prema kojem je stvarnost četverodimenzionalni prostor-vremenski manifold, u kojem prošlost, sadašnjost i budućnost jednako postoje, a sadašnje vrijeme i ne-sadašnje vrijeme su metafizički isto. Argumentiram u prilog novom odgovoru na kompleksno pitanje postoji li promjena prema B-teoriji. Zapravo, često se kaže da B-teoretičari imaju ‘statični’ pogled na vrijeme. Međutim, ovo je daleko od nevine semantičke razlike: ako B-teorija vremena predstavlja model vremenske stvarnosti koji je statičan, tada prema B-teoriji nema promjena. Na temelju ovoga se razumno može smatrati sljedeće: naravno, promjena postoji, dakle B-teorija mora biti lažna. U ovom radu tvrdim da u jednom smislu postoji promjena prema B-teoriji, međutim u drugom smislu nema promjene prema B-teoriji. U tu svrhu predstavljam tri slučaja promjene: Egzistencijalna promjena, naime gledište da se stvari mijenjaju s obzirom na njihovo postojanje tijekom vremena; Kvalitativna promjena, gledište da se stvari mijenjaju s obzirom na njihovo postojanje tijekom vremena; Propozicijska promjena, naime gledište da se stvari (tj. propozicije) s vremenom mijenjaju u odnosu na njihovu istinosnu vrijednost. Tvrdim da, iako postoji interpretacija ove tri instance promjene koja je istinita prema B-teoriji, pa tako i promjena prema B-teoriji u tom smislu, postoji B-teorijska interpretacija svakog od njih koja nije istinita prema B-teoriji, stoga prema tome nema promjene prema B-teoriji u ovom drugom smislu.

Ključne riječi: Promjena; B-teorija vremena; postojanje; svojstva; propozicije

AGAINST PHENOMENAL BONDING

S Siddharth
National Institute of Advanced Studies (A recognized research centre of University of Mysore)

ABSTRACT

Panpsychism, the view that phenomenal consciousness is possessed by all fundamental physical entities, faces an important challenge in the form of the combination problem: how do experiences of microphysical entities combine or give rise to the experiences of macrophysical entities such as human beings? An especially troubling aspect of the combination problem is the subject-summing argument, according to which the combination of subjects is not possible. In response to this argument, Goff (2016) and Miller (2017) have proposed the phenomenal bonding relation, using which they seek to explain the composition of subjects. In this paper, I discuss the merits of the phenomenal bonding solution and argue that it fails to respond satisfactorily to the subject-summing argument.

Keywords: Panpsychism; combination problem; subject-summing; phenomenal bonding; constitutive panpsychism

PROTIV FENOMENALNOG POVEZIVANJA

S Siddharth
National Institute of Advanced Studies (A recognized research centre of University of Mysore)

SAŽETAK

Panpsihizam, gledište prema kojemu svi temeljni fizički entiteti posjeduju fenomenalnu svijest, suočava se s važnim izazovom u obliku problema kombinacije: kako se iskustva mikrofizičkih entiteta kombiniraju ili dovode do iskustva makrofizičkih entiteta kao što su ljudska bića? Posebno zabrinjavajući aspekt problema kombinacije je argument sumiranja subjekata prema kojemu kombinacija subjekata nije moguća. Kao odgovor na ovaj argument, Goff (2016) i Miller (2017) sugeriraju da postoji fenomenalni odnos povezivanja, pomoću kojeg nastoje objasniti kompoziciju subjekata. U ovom radu raspravljam o uvjerljivosti rješenja koje se temelji na relaciji fenomenalnog povezivanja i argumentiram da ono ne uspijeva na zadovoljavajući način odgovoriti na argument sumiranja subjekata.

Ključne riječi: Panpsihizam; problem kombinacije; sumiranje subjekata; fenomenalno povezivanje; konstitutivni panpsihizam

MOTIVATIONAL INTERNALISM AND THE SECOND-ORDER DESIRE EXPLANATION

Xiao Zhang
Birmingham

ABSTRACT

Both motivational internalism and externalism need to explain why sometimes moral judgments tend to motivate us. In this paper, I argue that Dreier' second-order desire model cannot be a plausible externalist alternative to explain the connection between moral judgments and motivation. I explain that the relevant second-order desire is merely a constitutive requirement of rationality because that desire makes a set of desires more unified and coherent. As a rational agent with the relevant second-order desire is disposed towards coherence, she will have some motivation to act in accordance with her moral judgments. Dreier's second-order desire model thus collapses into a form of internalism and cannot be a plausible externalist option to explain the connection between moral judgments and motivation.

Keywords: Motivational internalism; externalism; second-order desire; practical rationality

MOTIVACIJSKI INTERNALIZAM I OBJAŠNJENJE ŽELJE DRUGOG REDA

Xiao Zhang
Birmingham

SAŽETAK

Motivacijski internalizam i eksternalizam moraju objasniti zašto nas ponekad moralni sudovi motiviraju. U ovom radu argumentiram da Dreierov model želje drugog reda ne može biti uvjerljiva eksternalistička alternativa za objašnjenje veze između moralnih sudova i motivacije. Objašnjavam da je relevantna želja drugog reda samo konstitutivni zahtjev racionalnosti jer ta želja čini skup želja jedinstvenijim i koherentnijim. Budući da je racionalni djelatnik s relevantnom željom drugog reda sklon koherentnosti, imat će određenu motivaciju da djeluje u skladu sa svojim moralnim sudovima. Dreierov model želje drugog reda tako se urušava u oblik internalizma i ne može biti uvjerljiva eksternalna opcija za objašnjenje veze između moralnih sudova i motivacije.

Ključne riječi: Motivacijski internalizam; eksternalizam; želja drugog reda; praktična racionalnost

BOOK REVIEW Maria Paola Feretti, THE PUBLIC PERSPECTIVE: PUBLIC JUSTIFICATION AND THE ETHICS OF BELIEF, Rowman & Littlefield, 2018

Iva Martinić
University of Rijeka, Faculty of Humanities and Social Sciences

ABSTRACT

BOOK REVIEW: Maria Paola Feretti, THE PUBLIC PERSPECTIVE: PUBLIC JUSTIFICATION AND THE ETHICS OF BELIEF, Rowman &

Littlefield, 2018, ISBN-10 1786608723, ISBN-13: 978-1786608727, Hardback, $126.00, e-Book, $38.00

RECENZIJA KNJIGE Maria Paola Feretti, THE PUBLIC PERSPECTIVE: PUBLIC JUSTIFICATION AND THE ETHICS OF BELIEF, Rowman & Littlefield, 2018

Iva Martinić
Sveučilište u Rijeci, Filozofski fakultet

SAŽETAK

RECENZIJA KNJIGE: Maria Paola Feretti, THE PUBLIC PERSPECTIVE: PUBLIC JUSTIFICATION AND THE ETHICS OF BELIEF, Rowman & Littlefield, 2018, ISBN-10 1786608723, ISBN-13: 978-1786608727, Tvrdi uvez, $126.00, e-knjiga, $38.00

BOOK REVIEW Marcin Będkowski, Anna Brożek, Alicja Chybińska, Stepan Ivanyk, and Dominik Traczykowski (Eds.) FORMAL AND INFORMAL METHODS IN PHILOSOPHY, Brill | Rodopi, 2020

Ivan Restović
Institute for Philosophy, Zagreb

ABSTRACT

BOOK REVIEW: Marcin Będkowski, Anna Brożek, Alicja Chybińska, Stepan Ivanyk, and Dominik Traczykowski (Eds.), FORMAL AND INFORMAL METHODS IN PHILOSOPHY, Brill | Rodopi, 2020, pp. vii + 320, ISBN-10: 9004420495, ISBN-13: 978-90-04-42050-2, Tvrdi uvez, €149.00 / $179.00

RECENZIJA KNJIGE Marcin Będkowski, Anna Brożek, Alicja Chybińska, Stepan Ivanyk, and Dominik Traczykowski (ur.) FORMAL AND INFORMAL METHODS IN PHILOSOPHY, Brill | Rodopi, 2020

Ivan Restović
Institute for Philosophy, Zagreb

SAŽETAK

RECENZIJA KNJIGE: Marcin Będkowski, Anna Brożek, Alicja Chybińska, Stepan Ivanyk, and Dominik Traczykowski (Eds.) FORMAL AND INFORMAL METHODS IN PHILOSOPHY, Brill | Rodopi, 2020, str. vii + 320, ISBN-10: 9004420495, ISBN-13: 978-90-04-42050-2, Tvrdi uvez, €149.00 / $179.00

BOOK REVIEW Rafe McGregor, A CRIMINOLOGY OF NARRATIVE FICTION, Bristol University Press, 2021, pp. 176, ISBN: 978-1529208054, Hardback, €74.70 / $66.09

Iris Vidmar Jovanović
University of Rijeka, Faculty of Humanities and Social Sciences

ABSTRACT

BOOK REVIEW: Rafe McGregor, A CRIMINOLOGY OF NARRATIVE FICTION, Bristol University Press, 2021, pp. 176, ISBN: 978-1529208054, Hardback, €74.70 / $66.09

RECENZIJA KNJIGE Rafe McGregor, A CRIMINOLOGY OF NARRATIVE FICTION, Bristol University Press, 2021, str. 176, ISBN: 978-1529208054, Tvrdi uvez, €74.70 / $66.09

Iris Vidmar Jovanović
Sveučilište u Rijeci, Filozofski fakultet

SAŽETAK

RECENZIJA KNJIGE: Rafe McGregor, A CRIMINOLOGY OF NARRATIVE FICTION, Bristol University Press, 2021, str. 176, ISBN: 978-1529208054, Tvrdi uvez, €74.70 / $66.09

Translated by Marko Jurjako (Rijeka)

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;
- be sent exclusively electronically to the Editors (eujap@ffri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- be in American or British English;
- be no longer than 9000 words, including references (for Original and Review Articles).
- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (https://eujap.uniri.hr/publication-ethics/). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (https://eujap.uniri.hr/instructions-for-reviewers/).

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Book with multiple authors

T: (Hirstein, Sifferd, and Fagan 2018, 100)

R: Hirstein, William, Katrina Sifferd, and Tyler Fagan. 2018. *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, Massachusetts: The MIT Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, Richard. 2006. 'The Epistemic Role of Testimony: Internalist and Externalist Perspectives'. In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 77–91. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780199276011.003.0004.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. 'Normative requirements'. *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, John. 2010. 'Sentimentalism: Its Scope and Limits'. Ethical Theory and Moral Practice 13 (2): 125–36.
https://doi.org/10.1007/s10677-009-9210-6.

Article with multiple authors in a journal
T: (Churchland and Sejnowski 1990)
R: Churchland, Patricia S., and Terrence J. Sejnowski. 1990. 'Neural Representation and Neural Computation'. *Philosophical Perspectives 4*. https://doi.org/10.2307/2214198

T: (Dardashti, Thébault, and Eric Winsberg 2017)
R: Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. 2017. 'Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity'. *The British Journal for the Philosophy of Science* 68 (1): 55–89.
https://doi.org/10.1093/bjps/axv010

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing ‘forthcoming’ instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

The journal allows the author(s) to hold the copyright without restrictions. In the reprints, the original publication of the text in EuJAP must be acknowledged by mentioning the name of the journal, the year of the publication, the volume and the issue numbers and the article pages.

EuJAP subscribes to Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Users can freely copy and redistribute the material in any medium or format, remix, transform, and build upon the material for any purpose. Users must give appropriate credit, provide a link to the license, and indicate if changes were made. Users may do so in any reasonable manner, but not in any way that suggests the licensor endorses them or their use. Nonetheless, users must distribute their contributions under the same license as the original.

Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link to the journal's web pages and HRČAK.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:
International:
individuals € 30
institutions € 50
Croatia:
individuals 100 kn
institutions 375 kn

Bank: Zagrebačka banka d.d. Zagreb
SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455
Only for subscribers from Croatia,
please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed and abstracted in SCOPUS, The Philosopher's Index, European Reference Index for the Humanities (ERIH PLUS), Directory of Open Access Journals (DOAJ), PhilPapers, and Portal of Scientific Journals of Croatia (HRČAK), ANVUR (Italy), Sherpa Romeo